

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
17 August 2006 (17.08.2006)

PCT

(10) International Publication Number
WO 2006/086357 A2(51) International Patent Classification:
C12N 15/861 (2006.01)(21) International Application Number:
PCT/US2006/004213

(22) International Filing Date: 6 February 2006 (06.02.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/651,843 10 February 2005 (10.02.2005) US(71) Applicant (for all designated States except US): **MERCK & CO., INC.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BETT, Andrew, J.** [CA/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US). **CASIMIRO, Danilo, R.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US). **SHIVER, John, W.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US). **EMINI, Emilio, A.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US). **CHASTAIN, Michael** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US). **KASLOW, David, C.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).(74) Common Representative: **MERCK & CO., INC.**; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).

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

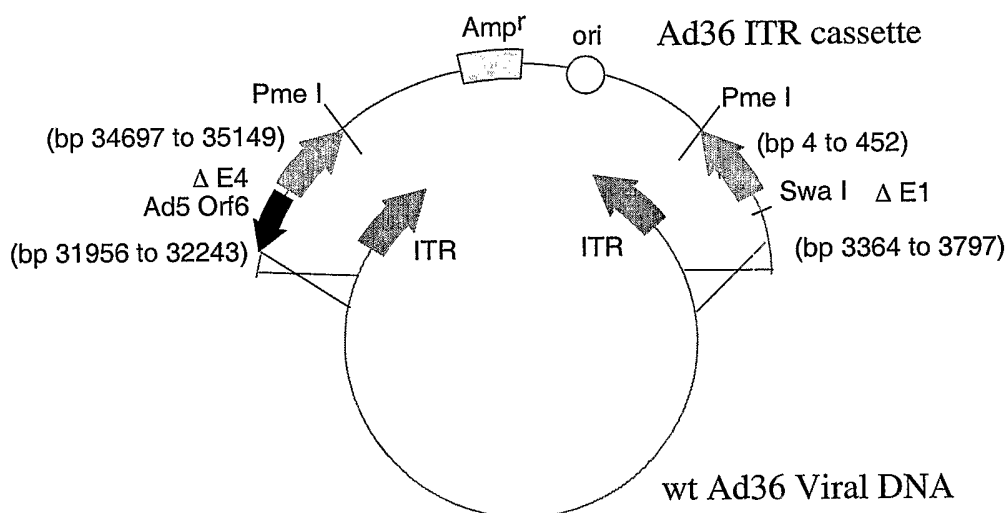
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT,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

Published:

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

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

(54) Title: ADENOVIRUS SEROTYPE 36 VECTORS, NUCLEIC ACID AND VIRUSES PRODUCED THEREBY

(57) **Abstract:** Adenoviral serotypes differ in their natural tropism. The various serotypes of adenovirus have been found to differ in at least their capsid proteins (*e.g.*, penton-base and hexon proteins), proteins responsible for cell binding (*e.g.*, fiber proteins), and proteins involved in adenovirus replication. This difference in tropism and capsid proteins among serotypes has led to many research efforts aimed at redirecting the adenovirus tropism by modification of the capsid proteins. The present invention bypasses such requirement for capsid protein modification as it presents a recombinant, replication-defective adenovirus of serotype 36, a rare adenoviral serotype, and methods for generating the alternative, recombinant adenovirus. Additionally, means of employing the recombinant adenovirus for delivery and expression of heterologous genes are provided.

WO 2006/086357 A2

TITLE OF THE INVENTION

ADENOVIRUS SEROTYPE 36 VECTORS, NUCLEIC ACID AND VIRUSES PRODUCED
THEREBY

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefits of U.S. provisional application serial no. 60/651,843, filed February 10, 2005.

BACKGROUND OF THE INVENTION

10 Adenoviruses are nonenveloped, icosahedral viruses that have been identified in several avian and mammalian hosts; Horne *et al.*, 1959 *J. Mol. Biol.* 1:84-86; Horwitz, 1990 In *Virology*, eds. B.N. Fields and D.M. Knipe, pps. 1679-1721. The first human adenoviruses (Ads) were isolated over four decades ago. Since then, over 100 distinct adenoviral serotypes have been isolated which infect various mammalian species, 51 of which are of human origin; Straus, 1984, In *The Adenoviruses*, ed. H. Ginsberg, pps. 451-498, New York:Plenus Press; Hierholzer *et al.*, 1988 *J. Infect. Dis.* 158:804-813; Schnurr and Dondero, 1993, *Intervirology*; 36:79-83; De Jong *et al.*, 1999 *J Clin. Microbiol.*, 37:3940-5. The human serotypes have been categorized into six subgenera (A-F) based on a number of biological, chemical, immunological and structural criteria which include hemagglutination properties of rat and rhesus monkey erythrocytes, DNA homology, restriction enzyme cleavage patterns, percentage G+C
15 content and oncogenicity; Straus, *supra*; Horwitz, *supra*.

Adenoviruses are attractive targets for the delivery and expression of heterologous genes. Adenoviruses are able to infect a wide variety of cells (dividing and non-dividing), and are very efficient in introducing their DNA into infected host cells. Adenoviruses have not been found to be associated with severe human pathology in immuno-competent individuals. The viruses can be produced at high
25 virus titers in large quantities. The adenovirus genome is very well characterized, consisting of a linear double-stranded DNA molecule of approximately 30,000-45,000 base pairs. Furthermore, despite the existence of several distinct serotypes, there is some general conservation found amongst the various serotypes.

Safety in utilizing adenoviruses as gene delivery vehicles can be enhanced by rendering
30 the viruses replication-defective through deletion/modification of the essential early-region 1 ("E1") of the viral genomes, rendering the viruses devoid (or essentially devoid) of E1 activity and, thus, incapable of replication in the intended host/vaccinee; *see, e.g.*, Brody et al, 1994 *Ann N Y Acad Sci.*, 716:90-101. Deletion of adenoviral genes other than E1 (*e.g.*, in E2, E3, and/or E4), furthermore, creates adenoviral vectors with greater capacity for heterologous gene inclusion. Presently, two well-characterized

adenovirus serotypes of subgroup C, serotypes 5 ("Ad5") and 2 ("Ad2") form the basis of the most widely used gene delivery vectors.

One concern surrounding the use of adenovectors relates to the potential for cellular and humoral immune responses elicited by and directed towards the viruses themselves (Chirmule *et al.*, 1999 *Gene Ther.* 6:1574-1583). Although an immune response associated with the initial administration of a vector may be advantageous (Zhang *et al.*, 2001 *Mol. Ther.* 3:697-707), the generation of systemic levels of adenovirus-specific neutralizing antibodies may cause poor transduction when the vectors are readministered (booster immunizations; Kass-Eisler *et al.*, 1996 *Gene Ther.* 3:154-162; Chirmule *et al.*, 1999 *J. Immunol.* 163:448-455). The scientific literature and data from our own epidemiological studies suggest that most North Americans have anti-Ad5 neutralizing antibody titers, and about one third have relatively high titers (>200). Other parts of the world typically exhibit higher frequencies and levels of anti-Ad5 antibodies. Serospecific antibodies to these and other adenoviral serotypes resulting from such natural adenovirus infections in humans may affect the extent of response to the administration of heterologous polypeptides by adenovectors; Chirmule *et al.*, 1999 *Gene Ther.* 6:1574-1583.

Accordingly, there is a need to develop adenoviral vectors based on alternate adenovirus serotypes as gene transfer vectors, particularly where the serotypes are less prevalent than adenovirus serotypes 2 and 5.

Adenovirus serotype 36, a subgroup D adenovirus, was originally isolated in 1980 (R. Wigand *et al.*, 1980 *Arch. Virol.* 64:225-233). Its antigenic relationship to 46 other human adenoviruses, determined in reference horse antisera, has been discussed; J.C. Hierholzer *et al.*, 1991 *Arch. Virol.* 121:179-197. Partial sequence for the Ad36 E1 region (1100 bp) is disclosed by N.V. Dhurandhar *et al.*, GenBank (Accession No. AY368049); *see also* Accession Nos. AAQ72378 and AAQ72377. The sequence of the virus associated RNA region for Ad36 and partial sequence for the pre-terminal protein and 52/55K proteins (521 bp) is disclosed by Ma & Matthews, 1996 *J. Virol.*, 70:5083-5099, and GenBank (Accession Nos. U52555, AAB00276 and AAB00277). Sequence is disclosed for the hexon protein of Ad36; *see, e.g.*, GenBank Accession Nos. AB099367 and BAD15141.

The fields of vaccines and gene therapy would greatly benefit from additional knowledge concerning alternative adenoviral serotypes, particularly those serotypes such as Ad36 which are not well represented in the human population. Of particular interest are recombinant adenoviral vectors based on alternative adenoviral serotypes, and means of obtaining such recombinant adenoviral vectors. This need in the art is met with the disclosure of the present application related to recombinant adenoviral vectors based on adenoviral serotype 36.

SUMMARY OF THE INVENTION

The present invention relates to recombinant, replication-defective adenovirus vectors of serotype 36, a rare adenoviral serotype, and methods for generating recombinant adenoviral vectors based on the alternative serotype. Additionally, means of employing said recombinant adenoviral vectors in the delivery and expression of heterologous genes are provided. Recombinant, replication-defective adenoviral vectors of serotype 36 wherein the vectors comprise one or more transgenes (heterologous genes) operatively linked to regulatory sequences are, furthermore, disclosed herein. Host administration of such recombinant adenovirus serotype 36 vectors, whether administered alone or in a combined modality and/or prime boost regimen, results in the efficient expression of the incorporated transgene and effectively induces an immune response capable of specifically recognizing the particular antigen administered. Recombinant viruses in accordance with this description have an innate ability to evade pre-existing immunity directed towards adenovirus serotypes which tend to be more prevalent in the human population (*e.g.*, Ad5 and Ad2). Use of such recombinant adenoviruses, therefore, offers an enhanced means for expressing a particular heterologous nucleic acid of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1 through 1A-11 illustrate a nucleic acid sequence for adenovirus serotype 36 (SEQ ID NO: 1). An available ATCC product number for Ad36 is VR-913.

Figure 2 illustrates a homologous recombination scheme utilized to recover pAd36ΔE1ΔE4Ad5Orf6.

Figures 3A-1 through 3A-10 illustrate a nucleic acid sequence for pAd36ΔE1ΔE4Ad5Orf6 (SEQ ID NO: 2).

Figure 4 illustrates a homologous recombination scheme utilized to recover pAd36ΔE1ΔE3ΔE4Ad5Orf6.

Figure 5 illustrates, in tabular format, levels of Gag-specific T cell responses induced using Ad36ΔE1gagΔE4Ad5Orf6 and/or MRKAd5gag vectors in a prime-boost vaccination protocol. Values reflect the mock-subtracted numbers of IFN-γ secreting cells per million PBMC; wk, week.

Figure 6 illustrates, in tabular format, the numbers of CD4+ and CD8+ Gag-specific T cells per million lymphocytes in Ad36-immunized macaques at week 8.

Figure 7 illustrates the nucleic acid sequence (SEQ ID NO: 3) of an optimized human HIV-1 gag open reading frame.

Figure 8 illustrates the nucleic acid sequence encoding the gag expression cassette (SEQ ID NO: 4). The various regions of the figure are as follows: (1) a first underlined segment of nucleic acid sequence encoding the immediate early gene promoter region from human cytomegalovirus; (2) a first segment of lowercase letters which is not underlined, which segment of DNA contains a convenient

restriction enzyme site; (3) a region in caps which contains the coding sequence of HIV-1 gag; (4) a second segment of lowercase letters which is not underlined, which segment of DNA contains a convenient restriction enzyme site; and (5) a second underlined segment, this segment containing nucleic acid sequence encoding a bovine growth hormone polyadenylation signal sequence.

Figures 9A-1 through 9A-2 illustrate a codon optimized wt-pol sequence, wherein sequences encoding protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity (SEQ ID NO: 6). The open reading frame starts at an initiating Met residue at nucleotides 10-12 and ends at a termination codon at nucleotides 2560-2562.

Figures 10A-1 through 10A-2 illustrate the open reading frame (SEQ ID NO: 7) of the wild type pol construct disclosed as SEQ ID NO: 6.

Figures 11A-1 through 11A-3 illustrate the nucleotide (SEQ ID NO: 8) and amino acid sequence (SEQ ID NO: 9) of IA-Pol. Underlined codons and amino acids denote mutations, as listed in Table 1 herein.

Figure 12 illustrates a codon optimized version of HIV-1 jrfl nef (SEQ ID NO: 10). The open reading frame starts at an initiating methionine residue at nucleotides 12-14 and ends at a "TAA" stop codon at nucleotides 660-662.

Figure 13 illustrates the open reading frame (SEQ ID NO: 11) of codon optimized HIV jrfl Nef.

Figures 14A-1 through 14A-2 illustrate a nucleotide sequence comparison between wild type nef (jrfl) and codon-optimized nef. The wild type nef gene from the jrfl isolate consists of 648 nucleotides capable of encoding a 216 amino acid polypeptide. WT, wild type sequence (SEQ ID NO: 12); opt, codon-optimized sequence (contained within SEQ ID NO: 10). The Nef amino acid sequence is shown in one-letter code (SEQ ID NO: 11).

Figure 15 illustrates nucleic acid (herein, "opt nef (G2A, LLAA)"; SEQ ID NO: 13) which encodes optimized HIV-1 Nef wherein the open reading frame encodes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2) and substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175. The open reading frame starts at an initiating methionine residue at nucleotides 12-14 and ends at a "TAA" stop codon at nucleotides 660-662.

Figure 16 illustrates the open reading frame (SEQ ID NO: 14) of opt nef (G2A, LLAA).

Figure 17 illustrates nucleic acid (herein, "opt nef (G2A)"; SEQ ID NO: 15) which encodes optimized HIV-1 Nef wherein the open reading frame encodes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2). The open reading frame starts at an initiating methionine residue at nucleotides 12-14 and ends at a "TAA" stop codon at nucleotides 660-662.

Figure 18 illustrates the open reading frame (SEQ ID NO: 16) of opt nef (G2A).

Figure 19 illustrates a schematic presentation of nef and nef derivatives. Amino acid residues involved in Nef derivatives are presented. Glycine 2 and Leucine 174 and 175 are the sites involved in myristylation and dileucine motif, respectively.

Figure 20 illustrates the nucleic acid sequence encoding the SEAP expression cassette (SEQ ID NO: 17). The various regions of the figure are as follows: (1) a first underlined segment of nucleic acid sequence encoding the immediate early gene promoter region from human cytomegalovirus; (2) a first segment of lowercase letters which is not underlined, which segment of DNA contains a convenient restriction enzyme site; (3) a region in caps which contains the coding sequence of the human placental SEAP gene; (4) a second segment of lowercase letters which is not underlined, which segment of DNA contains a convenient restriction enzyme site; and (5) a second underlined segment, this segment containing nucleic acid sequence encoding a bovine growth hormone polyadenylation signal sequence.

DETAILED DESCRIPTION OF THE INVENTION

Rare adenoviral serotypes possess an inherent advantage over the more commonly exploited adenoviral serotypes such as adenoviral serotypes 2 and 5 primarily because preexisting immunity is unlikely to limit their efficient delivery to, and expression of heterologous genes at, their target site. Different adenoviral serotypes also exhibit distinct tropisms by reason of their varying capsid structure and, thus, present the potential for targeting different tissues and possibly leading to the elicitation of superior immune responses when used for vaccine or gene therapy purposes. These rare serotypes when rendered replication-defective, however, have typically been difficult to propagate and rescue and, therefore, have not been fully characterized to date.

Applicants have recently managed to successfully rescue and propagate one such rare, replication-defective alternative serotype, adenovirus serotype 36, a subgroup D adenovirus, and herein demonstrate the effective functioning of these adenoviruses in the delivery and expression of heterologous transgenes.

Accordingly, the present invention relates to recombinant adenoviral vectors of serotype 36 suitable for use in gene therapy or vaccination protocols. A nucleic acid sequence disclosed herein for adenovirus serotype 36 (SEQ ID NO: 1) is illustrated in Figures 1A-1 to 1A-11, although any functional homologue or different strain of adenovirus serotype 36 forms an embodiment hereof, and can be utilized in accordance with the vectors, methods and compositions of the present invention; as one of ordinary skill in the art will appreciate. Sequence variation within Ad36 sequence has been noted by Applicants. Accordingly, Ad36 vectors possessing sequence variation are encompassed as embodiments hereof. One such sequence variation is as follows: (1) with reference to SEQ ID NO: 1, base pair 27820 exists also as a "T" rather than the "A" as reflected in SEQ ID NO: 1.

One of skill in the art, provided with the sequence information disclosed herein, can furthermore identify other variants of adenovirus serotype 36 sequences. Serotype classification is well understood in the art. Adenovirus serotypes have been distinguished in the art via a number of art-appreciated biological, chemical, immunological and structural criteria including but not limited to hemagglutination properties of rat and rhesus monkey erythrocytes, DNA homology, restriction enzyme cleavage patterns, percentage G+C content and oncogenicity; Straus, *supra*; Horwitz, *supra*. A given serotype can be identified by any number of methods including, but not limited to, restriction mapping of viral DNA; analyzing mobility of viral DNA; analyzing mobility of virion polypeptides on SDS-polyacrylamide gels following electrophoresis; comparison of sequence information to known sequences particularly from capsid genes (*e.g.*, hexons) which contain sequences that define a serotype; and comparison of sequence information with reference sera for a particular serotype available from the ATCC. Classification of adenovirus serotypes by SDS-PAGE has been discussed in Wadell *et al.*, 1980 *Ann. N.Y. Acad. Sci.* 354:16-42. Classification of adenovirus serotypes by restriction mapping has been discussed in Wadell *et al.*, *Current Topics in Microbiology and Immunology* 110:191-220.

Adenovirus serotype 36 vectors in accordance with the present invention are at least partially deleted/mutated in E1 such that any resultant viruses are devoid (or essentially devoid) of E1 activity, rendering the vectors incapable of replication in the intended host. Preferably, the E1 region is completely deleted or inactivated. The adenoviruses may contain additional deletions in E3, and other early regions, albeit in situations where E2 and/or E4 is deleted, E2 and/or E4 complementing cell lines may be required to generate recombinant, replication-defective adenoviral vectors.

Adenoviral vectors of use in the methods of the present invention can be constructed using well known techniques, such as those reviewed in Graham & Prevec, 1991 *In Methods in Molecular Biology: Gene Transfer and Expression Protocols*, (Ed. Murray, E.J.), p. 109; and Hitt *et al.*, 1997 "Human Adenovirus Vectors for Gene Transfer into Mammalian Cells" *Advances in Pharmacology* 40:137-206.

E1-complementing cell lines used for the propagation and rescue of recombinant adenoviruses as described herein should provide elements essential for the viruses to replicate, whether the elements are encoded in the cell's genetic material or provided *in trans*. It is, furthermore, preferable that the E1-complementing cell lines and the vectors not contain overlapping elements which could enable homologous recombination between the nucleic acid of the vector and the nucleic acid of the cell line, potentially leading to replication competent virus (or replication competent adenovirus "RCA"). Often, propagation cells are human cells derived from the retina or kidney, although any cell line capable of expressing the appropriate E1 and any other critical deleted region(s) can be utilized to generate adenovirus suitable for use in the methods of the present invention. Embryonal cells such as amniocytes have been shown to be particularly suited for the generation of E1 complementing cell lines. Several cell

lines are available and include but are not limited to the known cell lines PER.C6® (ECACC deposit number 96022940), 911, 293, and E1 A549. PER.C6® cell lines are described in WO 97/00326 (published January 3, 1997) and issued U.S. Patent No. 6,033,908. PER.C6® is a primary human retinoblast cell line transduced with an E1 gene segment that complements the production of replication defective (FG) adenovirus, but is designed to prevent generation of replication competent adenovirus by homologous recombination. 293 cells are described in Graham *et al.*, 1977 *J. Gen. Virol.* 36:59-72. For the propagation and rescue of non-group C adenoviral vectors, a cell line expressing an E1 region which is complementary to the E1 region deleted in the viruses being propagated can be utilized. For example, a specific example of cells suitable for the propagation of recombinant Ad36 E1-deleted vectors express the early region 1 (E1) of adenovirus 36 or another group D serotype. Alternatively, a cell line expressing regions of E1 and E4 derived from the same serotype can be employed; *see, e.g.*, U.S. Patent No. 6,270,996. Another alternative would be to propagate non-group C adenovirus in available E1-expressing cell lines (*e.g.*, PER.C6®, A549 or 293). This latter method involves the incorporation of a critical E4 region into the adenovirus to be propagated. The critical E4 region is native to a virus of the same or highly similar serotype as that of the E1 gene product(s) (particularly the E1B 55K region) of the complementing cell line, and comprises typically, at a minimum, E4 open reading frame 6 ("ORF6"); *see* PCT/US2003/026145, published March 4, 2004. One of skill in the art can readily appreciate and carry out numerous other methods suitable for the production of recombinant, replication-defective adenoviruses of use in the methods of the present invention. Following viral production in whatever means employed, viruses may be purified, formulated and stored prior to host administration.

Such methods for producing recombinant, replication-defective adenoviruses of serotype 36 are considered part of the present invention. Particularly, such methods comprising the acts of (1) introducing recombinant, replication-defective adenoviral vectors of serotype 36 into appropriate adenoviral E1-complementing cells and (2) allowing for the production of viral particles. Viral particles so produced and host cells comprising recombinant, replication-defective adenoviral serotype 36 vectors of the present invention form additional aspects of the present invention. "Isolated host cells" are defined herein as a population of cells not including a transgenic human being.

Adenoviral vectors in accordance with the present invention are very well suited to effectuate the expression of heterologous polypeptides, especially in situations where an individual's immune response effectively prevents administration or readministration via the more commonly employed adenoviral serotypes. Accordingly, specific embodiments of the present invention are recombinant, replication-defective adenoviral vectors of serotype 36 which comprise a heterologous nucleic acid encoding a polypeptide(s) of interest. The expressed nucleic acid can be DNA and/or RNA, and can be double or single stranded. The nucleic acid can be inserted in an E1 parallel (transcribed 5' to 3' relative to the vector backbone) or anti-parallel (transcribed 3' to 5' relative to the vector backbone)

orientation. The nucleic acid can be codon-optimized for expression in the desired host (*e.g.*, a mammalian host). The heterologous nucleic acid can be in the form of an expression cassette. A gene expression cassette can contain (a) nucleic acid encoding a protein or antigen of interest; (b) a heterologous promoter operatively linked to the nucleic acid encoding the protein/antigen; and (c) a transcription termination signal.

In specific embodiments, the heterologous promoter is recognized by a eukaryotic RNA polymerase. One example of a promoter suitable for use in the present invention is the immediate early human cytomegalovirus promoter (Chapman *et al.*, 1991 *Nucl. Acids Res.* 19:3979-3986). Further examples of promoters that can be used in the present invention are the strong immunoglobulin promoter, the EF1 alpha promoter, the murine CMV promoter, the Rous Sarcoma Virus promoter, the SV40 early/late promoters and the beta actin promoter, albeit those of skill in the art can appreciate that any promoter capable of effecting expression of the heterologous nucleic acid in the intended host can be used in accordance with the methods of the present invention. The promoter may comprise a regulatable sequence such as the Tet operator sequence. Sequences such as these that offer the potential for regulation of transcription and expression are useful in circumstances where repression/modulation of gene transcription is sought. The adenoviral gene expression cassette may comprise a transcription termination sequence; specific embodiments of which are the bovine growth hormone termination/polyadenylation signal (bGHpA) or the short synthetic polyA signal (SPA) of 49 nucleotides in length defined as follows: AATAAAAGATCTTTATTTTCATTAGATCTGTGTGTT-
GGTTTTTTGTGTG (SEQ ID NO: 5). A leader or signal peptide may also be incorporated into the transgene. In specific embodiments, the leader is derived from the tissue-specific plasminogen activator protein, tPA.

Heterologous nucleic acids of interest typically encode immunogenic and/or therapeutic proteins. Preferred therapeutic proteins are those which elicit some measurable therapeutic benefit in the individual host upon administration. Preferred immunogenic proteins are those proteins which are capable of eliciting a protective and/or beneficial immune response in an individual. A specific embodiment of the present invention, illustrated herein, is the delivery of nucleic acid encoding representative immunogenic proteins (HIV Gag, Nef and/or Pol) by the vectors, methods and compositions disclosed, albeit any gene encoding a therapeutic or immunogenic protein can be used in accordance with the methods disclosed herein and form important embodiments hereof. The vectors, methods and compositions of the present invention can be used to effectuate the delivery of any polypeptide whose presence/function brings about a desired effect in a given host, particularly a therapeutic/immunogenic effect useful in the treatment/alteration/modification of various conditions associated with, caused by, effected by (positively or negatively), exacerbated by, or modified by the presence or absence of a particular nucleic acid, protein, antigen, fragment, or activity associated with

any of the foregoing. Adenovirus serotype 36 vectors were found to induce significant levels of gag-specific T cells; Figure 5. Moreover, the results indicated that immunization with the disclosed vectors was able to elicit both HIV-specific CD4+ and CD8+ T cells; Figure 6. These results were particularly manifested when Ad36 was used in a prime-boost protocol with an adenoviral vector of serotype 5.

One aspect of the present invention, therefore, relates to adenovirus serotype 36-based vectors carrying heterologous nucleic acid encoding an HIV antigen(s)/protein(s), vector compositions, and methods of using same. Human Immunodeficiency Virus ("HIV") is the etiological agent of acquired human immune deficiency syndrome (AIDS) and related disorders. HIV is an RNA virus of the Retroviridae family and exhibits the 5'LTR-gag-pol-env-LTR 3' organization of all retroviruses. The integrated form of HIV, known as the provirus, is approximately 9.8 Kb in length. Each end of the viral genome contains flanking sequences known as long terminal repeats (LTRs).

Heterologous nucleic acid encoding an HIV antigen/protein may be derived from any HIV strain including, but not limited to, HIV-1 and HIV-2, strains A, B, C, D, E, F, G, H, I, O, IIIB, LAV, SF2, CM235, and US4; *see, e.g.,* Myers *et al., eds.* "Human Retroviruses and AIDS: 1995 (Los Alamos National Laboratory, Los Alamos NM 97545). Another HIV strain suitable for use in the methods disclosed herein is HIV-1 strain CAM-1; Myers *et al., eds.* "Human Retroviruses and AIDS": 1995, IIA3-IIA19. This gene closely resembles the consensus amino acid sequence for the clade B (North American/European) sequence. HIV gene sequence(s) may be based on various clades of HIV-1; specific examples of which are Clades A, B, and C. Sequences for genes of many HIV strains are publicly available from GenBank and primary, field isolates of HIV are available from the National Institute of Allergy and Infectious Diseases (NIAID) which has contracted with Quality Biological (Gaithersburg, MD) to make these strains available. Strains are also available from the World Health Organization (WHO), Geneva Switzerland.

HIV genes encode at least nine proteins and are divided into three classes; the major structural proteins (Gag, Pol, and Env), the regulatory proteins (Tat and Rev); and the accessory proteins (Vpu, Vpr, Vif and Nef). The *gag* gene encodes a 55-kilodalton (kDa) precursor protein (p55) which is expressed from the unspliced viral mRNA and is proteolytically processed by the HIV protease, a product of the *pol* gene. The mature p55 protein products are p17 (matrix), p24 (capsid), p9 (nucleocapsid) and p6. The *pol* gene encodes proteins necessary for virus replication - protease (Pro, P10), reverse transcriptase (RT, P50), integrase (IN, p31) and RNase H (RNase, p15) activities. These viral proteins are expressed as a Gag or Gag-Pol fusion protein which is generated by a ribosomal frame shift. The 55 kDa gag and 160 kDa gagpol precursor proteins are then proteolytically processed by the virally encoded protease into their mature products. The *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity. The *env* gene encodes the viral

envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and then cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41). Gp120 and gp41 remain associated and are displayed on the viral particles and the surface of HIV-infected cells. The *tat* gene encodes a long form and a short form of the Tat protein, a RNA binding protein which is a transcriptional transactivator essential for HIV replication. The *rev* gene encodes the 13 kDa Rev protein, a RNA binding protein. The Rev protein binds to a region of the viral RNA termed the Rev response element (RRE). The Rev protein promotes transfer of unspliced viral RNA from the nucleus to the cytoplasm. The Rev protein is required for HIV late gene expression and in turn, HIV replication.

Nucleic acid encoding any HIV antigen may be utilized in the methods and compositions of the present invention (specific examples of which include but are not limited to the aforementioned genes, nucleic acid encoding active and/or immunogenic fragments thereof, and/or modifications/derivatives of any of the foregoing). The present invention contemplates as well the various codon-optimized forms of nucleic acid encoding HIV antigens, including codon-optimized HIV gag (including but by no means limited to p55 versions of codon-optimized full length ("FL") Gag and tPA-Gag fusion proteins), HIV pol, HIV nef, HIV env, HIV tat, HIV rev, and modifications/derivatives of immunological relevance. Embodiments exemplified herein employ nucleic acid encoding codon-optimized Nef antigens; codon-optimized p55 Gag antigens; and codon-optimized Pol antigens. Codon-optimized HIV-1 gag genes are disclosed in PCT International Application PCT/US00/18332, published January 11, 2001 (WO 01/02607). Codon-optimized HIV-1 env genes are disclosed in PCT International Applications PCT/US97/02294 and PCT/US97/10517, published August 28, 1997 (WO 97/31115) and December 24, 1997 (WO 97/48370), respectively. Codon-optimized HIV-1 pol genes are disclosed in U.S. Application Serial No. 09/745,221, filed December 21, 2000 and PCT International Application PCT/US00/34724, also filed December 21, 2000. Codon-optimized HIV-1 nef genes are disclosed in U.S. Application Serial No. 09/738,782, filed December 15, 2000 and PCT International Application PCT/US00/34162, also filed December 15, 2000. It is well within the purview of the skilled artisan to choose an appropriate nucleotide sequence including but not limited to those cited above which encodes a specific HIV antigen, or immunologically relevant portion or modification/derivative thereof. "Immunologically relevant", "immunogenic" or "antigenic" as defined herein means (1) with regard to a viral antigen, that the protein is capable, upon administration, of eliciting a measurable immune response within an individual sufficient to retard the propagation and/or spread of the virus and/or to reduce/contain viral load within the individual; or (2) with regards to a nucleotide sequence, that the sequence is capable of encoding for a protein capable of the above. One of skill in the art can, furthermore, appreciate that any nucleic acid encoding for a protein, antigen, derivative or fragment

capable of effectuating a desired result (sequences that may or may not be codon-optimized) is of use in the methods and compositions of the instant invention.

An example of a codon-optimized gag gene that can be utilized in the vectors, methods and compositions of the present invention is that disclosed in PCT/US00/18332, published January 11, 2001 (*see* Figure 7; SEQ ID NO: 3). The sequence is derived from HIV-1 strain CAM-1 and encodes full-length p55 gag. The gag gene of HIV-1 strain CAM-1 was selected as it closely resembles the consensus amino acid sequence for the clade B (North American/European) sequence (Los Alamos HIV database). The sequence was designed to incorporate human preferred ("humanized") codons in order to maximize *in vivo* mammalian expression (Lathe, 1985, *J. Mol. Biol.* 183:1-12).

Open reading frames for various synthetic *pol* genes comprising coding sequences for reverse transcriptase (or RT which consists of a polymerase and RNase H activity) and integrase (IN) are disclosed in PCT/US00/34724. The protein sequences therein are based on Hxb2r, a clonal isolate of IIIB. This sequence has been shown to be closest to the consensus clade B sequence with only 16 nonidentical residues out of 848 (Korber, et al., 1998, Human retroviruses and AIDS, Los Alamos National Laboratory, Los Alamos, New Mexico).

A particular embodiment of this portion of the invention relates to vectors comprising codon optimized nucleotide sequences which encode wt-pol constructs (herein, "wt-pol" or "wt-pol (codon optimized)") wherein sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO: 6 (Figures 9A-1 to 9A-2), the open reading frame being contained from an initiating Met residue at nucleotides 10-12 to a termination codon from nucleotides 2560-2562. The open reading frame of the wild type pol construct (SEQ ID NO: 7; Figures 10A-1 to 10A-2) contains 850 amino acids.

Alternative specific embodiments relate to vectors comprising codon optimized HIV-1 pol wherein, in addition to deletion of the portion of the wild type sequence encoding the protease activity, a combination of active site residue mutations are introduced which are deleterious to HIV-1 pol (RT-RH-IN) activity of the expressed protein. Therefore, the present invention relates to adenoviral vectors comprising HIV-1 pol devoid of sequences encoding PR activity and containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant which is part and parcel of an adenoviral vector construct of use in the methods and compositions disclosed herein may include but is not limited to a mutated nucleic acid molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of HIV-1 Pol. In a specific embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation (or

mutations) within the Pol coding region which effectively abolishes RT, RNase H and IN activity. A specific HIV-1 pol-containing construct contains at least one point mutation which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, specific embodiments relate to adenoviral vectors comprising HIV-1 pol wherein the encoding nucleic acid comprises nine codon substitution mutations which result in an inactivated Pol protein (IA Pol: SEQ ID NO: 8, Figures 11A-1 to 11A-3) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Therefore, one exemplification contemplated employs an adenoviral vector construct which comprises, in an appropriate fashion, a nucleic acid molecule which encodes IA-Pol, which contains all nine mutations as shown below in Table 1. An additional amino acid residue for substitution is Asp551, localized within the RNase domain of Pol. Any combination of the mutations disclosed herein may be suitable and therefore may be utilized in the vectors, methods and compositions of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the preferred mutation is a point mutation resulting in a substitution of the wild type amino acid with an alternative amino acid residue.

Table 1

	<u>wt aa</u>	<u>aa residue</u>	<u>mutant aa</u>	<u>enzyme function</u>
20	Asp	112	Ala	RT
	Asp	187	Ala	RT
	Asp	188	Ala	RT
	Asp	445	Ala	RNase H
	Glu	480	Ala	RNase H
25	Asp	500	Ala	RNase H
	Asp	626	Ala	IN
	Asp	678	Ala	IN
	Glu	714	Ala	IN

30 It is preferred that point mutations be incorporated into the IApol mutant adenoviral vector constructs of the present invention so as to lessen the possibility of altering epitopes in and around the active site(s) of HIV-1 Pol. To this end, SEQ ID NO: 8 (Figures 11A-1 to 11A-3) discloses the nucleotide sequence which codes for a codon optimized pol in addition to the nine mutations shown in Table 1 and referred to herein as "IApol".

In one specific example of an adenoviral vector comprising *pol* (see Table 1), all residues that comprise the catalytic triad of the polymerase, namely Asp112, Asp187, and Asp188, can be substituted with alanine (Ala) residues (Larder, et al., *Nature* 1987, 327: 716-717; Larder, et al., 1989, *Proc. Natl. Acad. Sci.* 1989, 86: 4803-4807). Three additional mutations can be introduced at Asp445, Glu480 and Asp500 to abolish RNase H activity, with each residue being substituted for an Ala residue, respectively (Davies, et al., 1991, *Science* 252:, 88-95; Schatz, et al., 1989, *FEBS Lett.* 257: 311-314; Mizrahi, et al., 1990, *Nucl. Acids. Res.* 18: pp. 5359-5353). HIV *pol* integrase function can be abolished through three mutations at Asp626, Asp678 and Glu714. Again, each of these residues can be substituted with an Ala residue (Wiskerchen, et al., 1995, *J. Virol.* 69: 376-386; Leavitt, et al., 1993, *J. Biol. Chem.* 268: 2113-2119). Amino acid residue Pro3 of SEQ ID NO: 8 marks the start of the RT gene. The complete amino acid sequence of IA-Pol is disclosed herein as SEQ ID NO: 8 and shown in Figures 11A-1 to 11A-3.

It will be understood that any combination of mutations may be suitable and therefore utilized in the adenoviral HIV constructs, methods and compositions disclosed herein either administered alone, with other heterologous genes, in a combined modality regime and/or as part of a prime-boost regimen. For example, it may be possible to mutate only 2 of the 3 residues within the respective reverse transcriptase, RNase H, and integrase coding regions while still abolishing these enzymatic activities.

Another feature of the present invention are methods, vectors and compositions employing adenoviral vector constructs comprising codon optimized HIV-1 Pol comprising a eukaryotic trafficking signal peptide or a leader peptide such as that found in highly expressed mammalian proteins such as immunoglobulin leader peptides. Any functional leader peptide may be tested for efficacy. The respective DNA may be modified by known recombinant DNA methodology. In the alternative, as noted above, a nucleotide sequence which encodes a leader/signal peptide may be inserted into a DNA vector housing the open reading frame for the Pol protein of interest. Regardless of the cloning strategy, the end result is a vector construct which comprises vector components for effective gene expression in conjunction with nucleotide sequences which encode a modified HIV-1 Pol protein of interest, including but not limited to a HIV-1 Pol protein which contains a leader peptide.

The design of gene sequences disclosed herein incorporates the use of human preferred ("humanized") codons for each amino acid residue in the sequence in order to maximize *in vivo* mammalian expression (Lathe, 1985, *J. Mol. Biol.* 183:1-12). As can be discerned by inspecting the codon usage in SEQ ID NOs: 6 and 8, the following codon usage for mammalian optimization is preferred: Met (ATG), Gly (GGC), Lys (AAG), Trp (TGG), Ser (TCC), Arg (AGG), Val (GTG), Pro (CCC), Thr (ACC), Glu (GAG); Leu (CTG), His (CAC), Ile (ATC), Asn (AAC), Cys (TGC), Ala (GCC), Gln (CAG), Phe (TTC) and Tyr (TAC). For an additional discussion relating to mammalian (human) codon optimization, see WO 97/31115 (PCT/US97/02294). It is intended that the skilled artisan may use

alternative versions of codon optimization or may omit this step when generating HIV vaccine constructs within the scope of the present invention. Therefore, the present invention also relates to vectors, methods and compositions comprising/utilizing non-codon optimized or partially codon optimized versions of nucleic acid molecules and associated recombinant adenoviral HIV constructs which encode
5 the various wild type and modified forms of the HIV proteins. However, codon optimization of these constructs constitutes a preferred embodiment of this invention.

Codon optimized versions of HIV-1 nef and HIV-1 nef modifications of use in specific embodiments herein can be found in U.S. Application Serial No. 09/738,782, filed December 15, 2000 and PCT International Application PCT/US00/34162, also filed December 15, 2000. Specific codon
10 optimized nef and nef modifications relate to nucleic acid encoding HIV-1 Nef from the HIV-1 jrfl isolate wherein the codons are optimized for expression in a mammalian system such as a human. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO: 10 (Figure 12), while the expressed open reading frame is disclosed herein as SEQ ID NO: 11. Figures 14A-1 to 14A-2 illustrate a comparison of wild type vs. codon optimized nucleotides comprising the open reading frame of HIV-nef.
15 The open reading frame for SEQ ID NO: 10 comprises an initiating methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides 660-662. The open reading frame of SEQ ID NO: 10 provides a 216 amino acid HIV-1 Nef protein expressed through utilization of a codon optimized DNA vaccine vector. The 216 amino acid HIV-1 Nef (jrfl) protein is disclosed herein as SEQ ID NO: 11; Figure 13. A modified nef optimized coding region forming an additional embodiment herein relates to a
20 nucleic acid molecule encoding optimized HIV-1 Nef wherein the open reading frame codes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2) and substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175, herein described as opt nef (G2A, LLAA). A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO: 13, while the expressed open reading frame is disclosed herein as SEQ ID NO: 14. Yet another modified nef optimized coding region
25 forming an embodiment hereof relates to a nucleic acid molecule encoding optimized HIV-1 Nef wherein the open reading frame codes for modifications at the amino terminal myristylation site (Gly-2 to Ala-2), herein described as opt nef (G2A). A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO: 15, while the expressed open reading frame is disclosed herein as SEQ ID NO: 16.

HIV-1 Nef is a 216 amino acid cytosolic protein which associates with the inner surface
30 of the host cell plasma membrane through myristylation of Gly-2 (Franchini et al., 1986, *Virology* 155: 593-599). While not all possible Nef functions have been elucidated, it has become clear that correct trafficking of Nef to the inner plasma membrane promotes viral replication by altering the host intracellular environment to facilitate the early phase of the HIV-1 life cycle and by increasing the infectivity of progeny viral particles. In one aspect of the invention, the methods, vectors and
35 compositions of the present invention employ an adenoviral vector(s) comprising codon-optimized nef

sequence modified to contain a nucleotide sequence which encodes a heterologous leader peptide such that the amino terminal region of the expressed protein will contain the leader peptide. The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their biosynthetic pathways since their final destination, the cellular location at which they perform their function, becomes their permanent residence. Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct destinations. Defined sequence motifs exist in proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins. An effective induction of CTL responses often requires sustained, high level endogenous expression of an antigen. As membrane-association via myristylation is an essential requirement for most of Nef's function, mutants lacking myristylation, by glycine-to-alanine change, change of the dileucine motif and/or by substitution with a leader sequence, will be functionally defective, and therefore will have improved safety profile compared to wild-type Nef for use as an HIV-1 vaccine component.

In specific embodiments, therefore, the nucleotide sequence is modified to include a leader or signal peptide of interest. This may be accomplished by known recombinant DNA methodology. In the alternative, as noted above, insertion of a nucleotide sequence may be inserted into a DNA vector housing the open reading frame for the Nef protein of interest.

It has been shown that myristylation of Gly-2 in conjunction with a dileucine motif in the carboxy region of the protein is essential for Nef-induced down regulation of CD4 (Aiken et al., 1994, *Cell* 76: 853-864) via endocytosis. It has also been shown that Nef expression promotes down regulation of MHCI (Schwartz et al., 1996, *Nature Medicine* 2(3): 338-342) via endocytosis. The present invention contemplates adenoviral vectors which comprise sequence encoding a modified Nef protein altered in trafficking and/or functional properties and the use thereof in the methods and compositions of the present invention. The modifications introduced into the adenoviral vector HIV constructs of the present invention include but are not limited to additions, deletions or substitutions to the nef open reading frame which results in the expression of a modified Nef protein which includes an amino terminal leader peptide, modification or deletion of the amino terminal myristylation site, and modification or deletion of the dileucine motif within the Nef protein and which alter function within the infected host cell.

A recombinant adenoviral construct of use in accordance with the methods and compositions disclosed herein can comprise sequence encoding optimized HIV-1 Nef with modifications

at the amino terminal myristylation site (Gly-2 to Ala-2) and substitution of the Leu-174-Leu-175 dileucine motif to Ala-174-Ala-175. This open reading frame is herein described as opt nef (G2A,LLAA) and is disclosed as SEQ ID NO: 13, which comprises an initiating methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides 660-662. The nucleotide sequence of this
5 codon optimized version of HIV-1 jrfl nef gene with the above mentioned modifications is disclosed herein as SEQ ID NO: 13; Figure 15. The open reading frame of SEQ ID NO: 13 encodes Nef (G2A,LLAA), disclosed herein as SEQ ID NO: 14; Figure 16.

Another recombinant adenoviral construct of use in accordance with the methods and compositions disclosed herein can comprise sequence encoding optimized HIV-1 Nef with modifications
10 at the amino terminal myristylation site (Gly-2 to Ala-2). This open reading frame is herein described as opt nef (G2A) and is disclosed as SEQ ID NO: 16, which comprises an initiating methionine residue at nucleotides 12-14 and a "TAA" stop codon from nucleotides 660-662. The nucleotide sequence of this codon optimized version of HIV-1 jrfl nef gene with the above mentioned modification is disclosed herein as SEQ ID NO: 15; Figure 17. The open reading frame of SEQ ID NO: 15 encodes Nef (G2A),
15 disclosed herein as SEQ ID NO: 16; Figure 18.

Figure 19 shows a schematic presentation of nef and nef derivatives. Amino acid residues involved in Nef derivatives are presented. Glycine 2 and Leucine 174 and 175 are the sites involved in myristylation and dileucine motif, respectively.

Adenoviral vectors of use in the methods and compositions of the present invention may
20 comprise one or more HIV genes/encoding nucleic acid. The administration of at least one recombinant adenoviral vector(s) comprising two or more HIV genes, their derivatives, or modifications are anticipated as well as exemplified herein. Two or more HIV genes can be expressed on at least one recombinant adenoviral vector construct and/or two or more HIV genes can be expressed across two or more constructs. Therefore, the present invention offers the possibility of using the methods and
25 compositions of the present invention to evade/bypass host immunity and effectuate a multi-valent HIV gene administration, specific examples, but not limitations of which, include the administration of adenoviral vectors comprising nucleic acid sequence encoding (1) Gag and Nef polypeptides, (2) Gag and Pol polypeptides, (3) Pol and Nef polypeptides, and (4) Gag, Pol and Nef polypeptides.

Multiple genes/encoding nucleic acid may be ligated into a proper shuttle plasmid for
30 generation of a pre-adenoviral plasmid comprising multiple open reading frames. Open reading frames for the multiple genes/encoding nucleic acid can be operatively linked to distinct promoters and transcription termination sequences. In other embodiments, the open reading frames may be operatively linked to a single promoter, with the open reading frames operatively linked by an internal ribosome entry sequence (IRES; as disclosed in WO 95/24485), or suitable alternative allowing for transcription of
35 the multiple open reading frames to run off of a single promoter. In certain embodiments, the open

reading frames may be fused together by stepwise PCR or suitable alternative methodology for fusing together two open reading frames. Various combined modality administration regimens suitable for use in the present invention are disclosed in PCT/US01/28861, published March 21, 2002.

Multi-valent vectors of this description form an important aspect of the present invention as are methods of using same in eliciting cellular-mediated immune responses specific for the HIV antigens contained therein. It is well within the purview of one of skill in the art to arrive at and effectively utilize various fusion/multi-valent constructs.

The present invention encompasses methods for (1) effectuating a therapeutic response in an individual and (2) inducing an immune response (including a cellular-mediated immune response) comprising administering to an individual a recombinant adenovirus serotype 36 vector in accordance with the present invention. One aspect of the present invention are methods for generating an enhanced immune response against one or more antigens (bacterial, viral (*e.g.*, HIV) or other (*e.g.*, cancer)) which comprise the administration of a recombinant adenovirus serotype 36 vehicle expressing the antigen of interest. Administration of recombinant Ad36 vectors in this manner provides for improved cellular-mediated immune response, particularly where there is pre-existing immunity in a given host to the more well-represented adenovirus serotypes (*e.g.*, Ad2 and Ad5). An effect of the improved vaccine administration methods should be a lower transmission rate to (or occurrence rate in) previously uninfected individuals (*i.e.*, prophylactic applications) and/or a reduction in the levels of virus/bacteria/foreign agent within an infected individual (*i.e.*, therapeutic applications). As relates to HIV indications, an effect of the improved vaccine administration methods should be a lower transmission rate to previously uninfected individuals (*i.e.*, prophylactic applications) and/or a reduction in the levels of viral loads within an infected individual (*i.e.*, therapeutic applications) so as to prolong the asymptomatic phase of HIV infection. Administration, intracellular delivery and expression of the recombinant Ad36 vectors elicits a host CTL and Th response.

Accordingly, the present invention relates to methodology regarding administration of the recombinant Ad36 viral vectors (or immunogenic compositions thereof, herein termed vaccines) to provide effective immunoprophylaxis, to prevent establishment of an infection following exposure to the viral (for instance, HIV), bacterial or other agent, or as a post-infection therapeutic vaccine to mitigate infection to result in the establishment of a lower virus/bacteria/other load with beneficial long term consequences.

The recombinant adenovirus serotype 36 vectors disclosed herein may form the subject of a single administration or be part of a broader prime/boost-type administration regimen. Prime-boost regimens can employ different viruses (including but not limited to different viral serotypes and viruses of different origin), viral vector/protein combinations, and combinations of viral and polynucleotide administrations. In this type of scenario, an individual is first administered a priming dose of a

protein/antigen/derivative/modification utilizing a certain vehicle (be that a viral vehicle, purified and/or recombinant protein, or encoding nucleic acid). Multiple primings, typically 1-4, are usually employed, although more may be used. The priming dose(s) effectively primes the immune response so that, upon subsequent identification of the protein/antigen(s) in the circulating immune system, the immune response is capable of immediately recognizing and responding to the protein/antigen(s) within the host. Following some period of time, the individual is administered a boosting dose of at least one of the previously delivered protein(s)/antigen(s), derivatives or modifications thereof (administered by viral vehicle/protein/nucleic acid). The length of time between priming and boost may typically vary from about four months to a year, albeit other time frames may be used as one of ordinary skill in the art will appreciate. The follow-up or boosting administration may also be repeated at selected time intervals. A mixed modality prime and boost inoculation scheme should result in an enhanced immune response, specifically where there is pre-existing anti-vector immunity.

Selection of the alternate administration vehicle (be it viral/nucleic acid/protein) to be employed in conjunction with the vectors disclosed herein in a prime-boost administration regimen is not critical to the successful practice hereof. Any vehicle capable of delivering the antigen (or effectuating expression of the antigen) to sufficient levels such that a cellular and/or humoral-mediated response is elicited should be sufficient to prime or boost the presently disclosed administration. Suitable viral vehicles include but are not limited to distinct serotypes of adenovirus, including but not limited to adenovirus serotypes 5, 6, 24, 34 and 35 (*see, e.g.*, PCT/US00/18332, published Jan. 11, 2001 (Ad5); PCT/US01/28861, published Mar. 21, 2002 (Ad5); PCT/US02/32512, published April 17, 2003 (Ad6); PCT/US2003/026145, published March 4, 2004 (Ad24, Ad34); PCT/NL00/00325, published November 23, 2000 (Ad35)). Alternatively, the adenoviral administration can be followed or preceded by a viral vehicle of diverse origin. Examples of different viral vehicles include but are not limited to adeno-associated virus ("AAV"; *see, e.g.*, Samulski *et al.*, 1987 *J. Virol.* 61:3096-3101; Samulski *et al.*, 1989 *J. Virol.* 63:3822-3828); retrovirus (*see, e.g.*, Miller, 1990 *Human Gene Ther.* 1:5-14; Ausubel *et al.*, *Current Protocols in Molecular Biology*); pox virus (including but not limited to replication-impaired NYVAC, ALVAC, TROVAC and MVA vectors, *see, e.g.*, Panicali & Paoletti, 1982 *Proc. Natl. Acad. Sci. USA* 79:4927-31; Nakano *et al.* 1982 *Proc. Natl. Acad. Sci. USA* 79: 1593-1596; Piccini *et al.*, In *Methods in Enzymology* 153:545-63 (Wu & Grossman, eds., Academic Press, San Diego); Sutter *et al.*, 1994 *Vaccine* 12:1032-40; Wyatt *et al.*, 1996 *Vaccine* 15:1451-8; and U.S. Patent Nos. 4,603,112; 4,769,330; 4,722,848; 4,603,112; 5,110,587; 5,174,993; and 5,185,146); and alpha virus (*see, e.g.*, WO 92/10578; WO 94/21792; WO 95/07994; and U.S. Patent Nos. 5,091,309 and 5,217,879). Prime-boost protocols exploiting adenoviral and pox viral vectors for delivery of HIV antigens are discussed in International Application No. PCT/US03/07511, published September 18, 2003. An alternative to the above immunization schemes would be to employ polynucleotide administrations (including but not

limited to “naked DNA” or facilitated polynucleotide delivery) in conjunction with an adenoviral prime and/or boost; *see, e.g.*, Wolff *et al.*, 1990 *Science* 247:1465, and the following patent publications: U.S. Patent Nos. 5,580,859; 5,589,466; 5,739,118; 5,736,524; 5,679,647; WO 90/11092 and WO 98/04720. Another alternative would be to employ purified/recombinant protein administration in a prime-boost scheme along with adenovirus.

Potential hosts/vaccinees/individuals that can be administered the recombinant adenoviral vectors of the present invention include, but are not limited to, primates and especially humans and non-human primates, and include any non-human mammal of commercial or domestic veterinary importance.

Compositions of adenoviral vectors including, but not limited to, vaccine compositions, administered in accordance with the methods and compositions of the present invention may be administered alone or in combination with other viral- or non-viral-based DNA/protein vaccines. They also may be administered as part of a broader treatment regimen. The present invention, thus, encompasses those situations where the disclosed adenovirus constructs are administered in conjunction with other therapies; including but not limited to other antimicrobial (*e.g.*, antiviral, antibacterial) agent treatment therapies. Any specific antimicrobial agent(s) is not critical to successful practice of the methods disclosed herein. The antimicrobial agent can, for example, be based on/derived from an antibody, a polynucleotide, a polypeptide, a peptide, or a small molecule. Any antimicrobial agent that effectively reduces microbial replication/spread/load within an individual is sufficient for the uses described herein.

Antiviral agents antagonize the functioning/life cycle of a virus, and target a protein/function essential to the proper life cycle of the virus; an effect that can be readily determined by an *in vivo* or *in vitro* assay. Some representative antiviral agents which target specific viral proteins are protease inhibitors, reverse transcriptase inhibitors (including nucleoside analogs; non-nucleoside reverse transcriptase inhibitors; and nucleotide analogs), and integrase inhibitors. Protease inhibitors include, for example, indinavir/CRIXIVAN®; ritonavir/NORVIR®; saquinavir/FORTOVASE®; nelfinavir/VIRACEPT®; amprenavir/AGENERASE®; lopinavir and ritonavir/KALETRA®. Reverse transcriptase inhibitors include, for example, (1) nucleoside analogs, *e.g.*, zidovudine/RETROVIR® (AZT); didanosine/VIDEX® (ddI); zalcitabine/HIVID® (ddC); stavudine/ZERIT® (d4T); lamivudine/EPIVIR® (3TC); abacavir/ZIAGEN® (ABC); (2) non-nucleoside reverse transcriptase inhibitors, *e.g.*, nevirapine/VIRAMUNE® (NVP); delavirdine/RESCRIPTOR® (DLV); efavirenz/SUSTIVA® (EFV); and (3) nucleotide analogs, *e.g.*, tenofovir DF/VIREAD® (TDF). Integrase inhibitors include, for example, the molecules disclosed in U.S. Application Publication No. US2003/0055071, published March 20, 2003; and International Application WO 03/035077. The antiviral agents, as indicated, can target as well a function of the virus/viral proteins, such as, for instance

the interaction of regulatory proteins tat or rev with the trans-activation response region ("TAR") or the rev-responsive element ("RRE"), respectively. An antiviral agent is, preferably, selected from the class of compounds consisting of: a protease inhibitor, an inhibitor of reverse transcriptase, and an integrase inhibitor. Preferably, the antiviral agent administered to an individual is some combination of effective
5 antiviral therapeutics such as that present in highly active anti-retroviral therapy ("HAART"), a term generally used in the art to refer to a cocktail of inhibitors of viral protease and reverse transcriptase.

One of skill in the art can appreciate that the present invention can be employed in conjunction with any pharmaceutical composition useful for the treatment of microbial infections. Antimicrobial agents are typically administered in their conventional dosage ranges and regimens as
10 reported in the art, including the dosages described in the Physicians' Desk Reference, 54th edition, Medical Economics Company, 2000.

Compositions comprising the recombinant viral vectors may contain physiologically acceptable components including, but not limited to, buffer, normal saline or phosphate buffered saline, sucrose, other salts and polysorbate. In specific embodiments the viral particles are formulated in A195
15 formulation buffer. In certain embodiments, the formulation has: 2.5-10 mM TRIS buffer, preferably about 5 mM TRIS buffer; 25-100 mM NaCl, preferably about 75 mM NaCl; 2.5-10% sucrose, preferably about 5% sucrose; 0.01 -2 mM MgCl₂; and 0.001%-0.01% polysorbate 80 (plant derived). The pH should range from about 7.0-9.0, preferably about 8.0. One skilled in the art will appreciate that other conventional vaccine excipients may also be used in the formulation. In specific embodiments, the
20 formulation contains 5mM TRIS, 75 mM NaCl, 5% sucrose, 1mM MgCl₂, 0.005% polysorbate 80 at pH 8.0. This has a pH and divalent cation composition which is near the optimum for virus stability and minimizes the potential for adsorption of virus to glass surface. It does not cause tissue irritation upon intramuscular injection. It is preferably frozen until use.

The amount of viral particles in the vaccine composition(s) to be introduced into a
25 vaccine recipient will depend on the strength of the transcriptional and translational promoters used and on the immunogenicity of the expressed gene product(s). In general, an immunologically or prophylactically effective dose of 1×10^7 to 1×10^{12} particles and preferably about 1×10^{10} to 1×10^{11} particles per adenoviral vector is administered directly into muscle tissue. Subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as
30 intraperitoneal, intravenous, or inhalation delivery are also contemplated.

Administration of additional agents able to potentiate or broaden the immune response (*e.g.*, the various cytokines, interleukins), concurrently with or subsequent to parenteral introduction of the viral vectors of this invention, is appreciated herein as well and can be advantageous.

The benefits of administration as described herein should be (1) a comparable or broader
35 population of individuals successfully immunized/treated with recombinant adenoviral vectors, and (2) in

situations of immunization, a lower transmission rate to (or occurrence rate in) previously uninfected individuals (*i.e.*, prophylactic applications) and/or a reduction in/control of the levels of virus/bacteria/foreign agent within an infected individual (*i.e.*, therapeutic applications).

5 The following non-limiting Examples are presented to better illustrate the workings of the invention.

Example 1

Construction of pAd36ΔE1ΔE4Ad5Orf6

10 To construct pAd36ΔE1ΔE4Ad5Orf6 (An Ad36 pre-Ad plasmid containing an E1 deletion and an E4 deletion substituted with Ad5 Orf6 in order to enable efficient propagation in existing group C/Ad5 E1 complementing cell lines), an Ad36 ITR cassette was constructed containing sequences from the right (bp 31956 to 32243 and bp 34697 to 35149) and left (bp 4 to 452 and bp 3364 to 3797) end of the Ad36 genome (*see* Figures 1A-1 to 1A-11) separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. The four segments were generated by
15 PCR and cloned sequentially into pNEB193, generating pNEBAd36-4a. Next the Ad5 Orf6 open reading frame was generated by PCR and cloned between Ad36 bp 32243 and 34697 generating pNEBAd36-4aAd5Orf6 (the ITR cassette). PNEB193 is a commonly used commercially available cloning plasmid (New England Biolabs cat# N3051S) containing a bacterial origin of replication, ampicillin resistance
20 gene and a multiple cloning site into which the PCR products were introduced. The ITR cassette contains a deletion of E1 sequences from Ad36 bp 453 to 3363 with a unique *Swa* I restriction site located in the deletion and an E4 deletion from Ad36 bp 32244 to 34696 into which Ad5 Orf6 was introduced in an E4 parallel orientation. In this construct Ad5Orf6 expression is driven by the Ad36 E4 promoter. The Ad36 sequences (bp 31956 to 32243 and bp 3364 to 3797) in the ITR cassette provided
25 regions of homology with the purified Ad36 viral DNA in which bacterial recombination could occur following cotransformation into BJ 5183 bacteria (Figure 2). The ITR cassette was also designed to contain unique restriction enzyme sites (*Pme*I) located at the end of the viral ITR's so that digestion would release the recombinant Ad36 genome from the plasmid sequences. Potential clones were screened by restriction analysis and one clone was selected as pAd36ΔE1ΔE4Ad5Orf6. Pre-Adenovirus
30 plasmid pAd36ΔE1ΔE4Ad5Orf6 contains Ad36 sequences from bp 4 to 452; bp 3364 to bp 32243 and bp 34697 to bp 35149 with Ad5 Orf6 cloned between bp 32243 and bp 34697. This plasmid was completely sequenced (*see* Figures 3A-1 to 3A-10). The bp numbering in the above description refers to the wt sequence for both Ad36 and Ad5.

*Example 2*Insertion of HIV-1 gag and SEAP transgenes into pAd36ΔE1ΔE4Ad5Orf6

In order to introduce a gag or SEAP expression cassette (*see* Figures 8 and 20, respectively) into the E1 region of pAd36ΔE1ΔE4Ad5Orf6, bacterial recombination was again used. An HIV-1 gag expression cassette consisting of the following: 1) the immediate early gene promoter from human cytomegalovirus, 2) the coding sequence of the human immunodeficiency virus type 1 (HIV-1) gag (strain CAM-1; 1526 bp) gene, and 3) the bovine growth hormone polyadenylation signal sequence, was cloned into the E1 deletion in Ad36 shuttle plasmid, pNEBAd36-2 (a precursor to the Ad36 ITR cassette described above), generating pNEBAd36CMVgagBGHpA. pNEBAd36-2 contains Ad36 sequences from the left end of the genome (bp 4 to 452 and bp 3364 to 3797) that define the E1 deletion. The gag expression cassette was obtained from a previously constructed plasmid and cloned into the E1 deletion between bp 452 and 3364 in the E1 parallel orientation. The shuttle vector containing the gag transgene was digested to generate a DNA fragment consisting of the gag expression cassette flanked by Ad36 bp 4 to 452 and bp 3364 to 3797 and the fragment was purified after electrophoresis on an agarose gel. Cotransformation of BJ 5183 bacteria with the shuttle vector fragment and pAd36ΔE1ΔE4Ad5Orf6, linearized in the E1 region by digestion with *Swa* I, resulted in the generation of the Ad36 gag-containing pre-Adenovirus plasmid pAd36ΔE1gagΔE4Ad5Orf6 by homologous recombination. Potential clones were screened by restriction analysis.

A similar strategy was used to generate Ad36 pre-Ad plasmids containing a SEAP expression cassette. In this case a SEAP expression cassette consisting of: 1) the immediate early gene promoter from human cytomegalovirus, 2) the coding sequence of the human placental SEAP gene, and 3) the bovine growth hormone polyadenylation signal sequence was cloned into the E1 deletion in the Ad36 shuttle plasmid, pNEBAd36-2, generating pNEBAd36CMVSEAPBGHpA. The transgene was then recombined into pAd36ΔE1ΔE4Ad5Orf6 as described above for the gag transgene.

*Example 3*Rescue of pAd36ΔE1ΔE4Ad5Orf6, pAd36ΔE1gagΔE4Ad5Orf6 and pAd36ΔE1SEAPΔE4Ad5Orf6 into Virus

In order to rescue pre-adenovirus plasmids pAd36ΔE1ΔE4Ad5Orf6, pAd36ΔE1gagΔE4Ad5Orf6 and pAd36ΔE1SEAPΔE4Ad5Orf6 into virus, the plasmids were each digested with *Pme* I and transfected into T-25 flasks of PER.C6® cells using the calcium phosphate co-precipitation technique (Cell Pfect Transfection Kit, Amersham Pharmacia Biotech Inc). *Pme*I digestion releases the viral genome from plasmid sequences allowing viral replication to occur after cell entry. When cytopathic effect (CPE) was complete, approximately 7-10 days post transfection, the infected cells and media were harvested, freeze/thawed three times and the cell debris pelleted by centrifugation.

The cell lysate was then used to infect T-225 flasks of PER.C6® cells at 80-90% confluence. Once CPE was reached, infected cells and media were harvested, freeze/thawed three times and the cell debris pelleted by centrifugation. Clarified cell lysates were then used to infect 2-layer NUNC cell factories of PER.C6® cells. Following complete CPE, the virus was purified by ultracentrifugation on CsCl density gradients. In order to verify the genetic structure of the rescued viruses, viral DNA was extracted using pronase treatment followed by phenol chloroform extraction and ethanol precipitation. Viral DNA was then digested with *HindIII* and treated with Klenow fragment to end-label the restriction fragments with P33-dATP. The end-labeled restriction fragments were then size-fractionated by gel electrophoresis and visualized by autoradiography. The digestion products were compared with the digestion products of the corresponding pre-Adenovirus plasmid (that had been digested with *PmeI/HindIII* prior to labeling) from which they were derived. The expected sizes were observed, indicating that the viruses had been successfully rescued.

Example 4

Construction of pMRKAd36ΔE1ΔE3ΔE4Ad5Orf6

To increase the cloning capacity of our Ad36-based vectors, early region 3 ("E3") was deleted, generating pre-Ad plasmid pAd36ΔE1ΔE3ΔE4Ad5Orf6. To construct pAd36ΔE1ΔE3ΔE4Ad5Orf6 (an Ad36 pre-Ad plasmid containing an E1 deletion, an E3 deletion and an E4 deletion substituted with Ad5 Orf6), an Ad36 E3 shuttle cassette was constructed containing sequences that defined the desired E3 deletion (bp 26077 to 26525 and bp 30334 to 30773). The two segments were generated by PCR and cloned sequentially into pNEB193, generating pNEBAd36E3-2. The E3 shuttle cassette contains a deletion of E3 sequences from Ad36 bp 26526 and bp 30333 with a unique *Pac I* restriction site located in the deletion. The E3 shuttle cassette was digested to liberate the Ad36 DNA defining the E3 deletion and the fragment was purified after electrophoresis on an agarose gel. Cotransformation of BJ 5183 bacteria with the shuttle vector fragment and pAd36ΔE1ΔE4Ad5Orf6 (linearized in the E3 region by digestion with *Pac I*) resulted in the generation of Ad36 pre-Adenovirus plasmid pAd36ΔE1ΔE3ΔE4Ad5Orf6 by homologous recombination (Figure 4). Potential clones were screened by restriction analysis and one clone was selected as pAd36ΔE1ΔE3ΔE4Ad5Orf6. Pre-Adenovirus plasmid pAd36ΔE1ΔE3ΔE4Ad5Orf6 contains Ad36 sequences from bp 4 to 452; bp 3364 to bp 26525; bp 30334 to 32243; and bp 34697 to bp 35149 with Ad5 Orf6 cloned between bp 32243 and bp 34697. The bp numbering in the above description refers to the wt sequence for both Ad36 and Ad5.

*Example 5**In Vivo Studies*

A. Immunization

Cohorts of three rhesus macaques were immunized with either 10^8 vp or 10^{10} vp of Ad36 Δ E1gag Δ E4Ad5Orf6 (at weeks 0, 4, 27). A third cohort received priming immunizations with 10^8 vp MRKAd5 gag (disclosed in PCT/US01/28861, published March 21, 2002) followed by a 10^{10} vp booster of Ad36 Δ E1gag Δ E4Ad5Orf6. Rhesus macaques were between 3-10 kg in weight. In all cases, the total dose of each vaccine was suspended in 1 mL of buffer. The macaques were anesthetized (ketamine/xylazine) and the vaccines were delivered intramuscularly ("i.m.") in 0.5-mL aliquots into both deltoid muscles using tuberculin syringes (Becton-Dickinson, Franklin Lakes, NJ). Plasma and peripheral blood mononuclear cells (PBMC) samples were collected following standard protocols.

B. ELISPOT and ICS Assays

Ninety-six-well flat-bottomed plates (Millipore, Immobilon-P membrane) were coated with 1 μ g/well of anti-gamma interferon (IFN- γ) mAb MD-1 (U-Cytech-BV) overnight at 4°C. The plates were then washed three times with PBS and were blocked with R10 medium (RPMI, 0.05 mM 2-mercaptoethanol, 1 mM sodium pyruvate, 2 mM L-glutamate, 10 mM HEPES, 10% fetal bovine serum) for 2 hours at 37°C. The medium was discarded from the plates, and freshly isolated PBMCs were added at $1-4 \times 10^5$ cells/well. The cells were stimulated in the absence (mock) or presence of a gag peptide pool (4 μ g/mL per peptide). The pool, which consisted of 15-aa peptides shifted by 4 aas (Synpep, CA), was constructed from the entire HIV-1 CAM1 gag sequence. The cells were then incubated for 20-24 hours at 37°C in 5% CO₂. Plates were washed six times with PBST (PBS, 0.05% Tween 20) and 100 μ L/well of 1:400 dilution of anti-IFN- γ polyclonal biotinylated detector antibody solution (U-Cytech-BV) was added, and the plates were incubated overnight at 37°C. The plates were washed six times with PBST. Color was developed by incubating in NBT/BCP (Pierce) for 10 minutes. Spots were counted under a dissecting microscope and normalized to 1×10^6 PBMC.

Intracellular staining for IFN- γ production (ICS) was conducted following a previously established protocol; Casimiro *et al.*, 2003 *J. Virol.* 77:6305-6313.

Figure 5 lists, in tabular format, the mock-corrected levels of gag-specific T cells as measured by the IFN- γ ELIspot assay.

C. Results

The Ad36-based vaccine vector was able to effectively induce gag-specific T cells at a dose level of 10^{10} vp. MRKAd5gag-primed animals exhibited responses of greater than 100 SFC/ 10^6 PBMC after the first dosing. A homologous Ad36 boost at wk 27 improved the levels of gag-specific T cells to values that were comparable to the peak levels established prior to the boost. These results indicate that the Ad36 is a potent vector for inducing HIV-specific immunity. A significant utility for the

Ad36 vector is to boost HIV-specific immune responses elicited by priming using the MRKAd5gag vector (Figure 5, group 3). In this case, 2 of 3 animals had peak post-boost levels of gag-specific T cells exceeding their respective peak post-prime values.

Analyses of the T cell subsets of the gag-specific immunity by intracellular cytokine staining (ICS) revealed that the Ad36 vector is capable of eliciting both cytotoxic and helper responses.

WHAT IS CLAIMED IS:

1. A recombinant adenoviral vector of serotype 36 which is at least partially deleted in E1 and devoid of E1 activity.

5

2. A recombinant adenoviral vector in accordance with claim 1 which comprises heterologous nucleic acid.

3. A recombinant adenoviral vector in accordance with claim 1 which comprises an E4 gene or a segment of an E4 gene comprising open reading frame 6 ("ORF6") of an alternative serotype.

10

4. A recombinant adenoviral vector in accordance with claim 3 wherein the alternative serotype is adenovirus serotype 5.

15

5. A recombinant adenoviral vector in accordance with claim 2 comprising a gene expression cassette, which comprises:

- a) nucleic acid encoding a protein or antigen of interest;
- b) a heterologous promoter operatively linked to the nucleic acid of a); and
- c) a transcription termination sequence.

20

6. A recombinant adenoviral vector in accordance with claim 5 wherein the heterologous nucleic acid comprises codons optimized for expression in a human host.

25

7. A recombinant adenoviral vector in accordance with claim 2 wherein the heterologous nucleic acid encodes for an HIV-1 antigen.

8. A recombinant adenoviral vector in accordance with claim 7 wherein the heterologous nucleic acid encodes for at least one antigen selected from the group consisting of: HIV-1 Gag, Nef, or Pol.

30

9. A population of cells comprising the recombinant adenoviral vector of claim 2.

10. A population of cells comprising the recombinant adenoviral vector of claim 3.

35

11. A method for producing recombinant, replication-defective adenovirus particles comprising:

- a) transfecting a recombinant adenoviral vector of claim 2 into a population of cells; and
- b) harvesting the resultant recombinant, replication-defective adenovirus.

12. A method for producing recombinant, replication-defective adenovirus particles comprising:

- a) transfecting a recombinant adenoviral vector of claim 3 into a population of cells; and
- b) harvesting the resultant recombinant, replication-defective adenovirus.

13. Purified recombinant, replication-defective adenovirus particles harvested in accordance with the method of claim 11.

14. Purified recombinant, replication-defective adenovirus particles harvested in accordance with the method of claim 12.

15. A composition comprising purified recombinant adenovirus particles in accordance with claim 13.

16. A composition comprising purified recombinant adenovirus particles in accordance with claim 14.

17. A composition in accordance with claim 15 which comprises a physiologically acceptable carrier.

18. A composition in accordance with claim 16 which comprises a physiologically acceptable carrier.

19. A method for delivery and expression of heterologous nucleic acid encoding a protein or antigen of interest, which comprises administering the composition of claim 15 to an individual.

20. A method in accordance with claim 19 wherein administration is preceded or followed by administration of heterologous nucleic acid encoding a protein or antigen of interest with a different vector.

21. A method in accordance with claim 20 wherein the different vector is an adenovirus of a distinct serotype.

5 22. A method for generating an immune response against an antigen in an individual, which comprises: administering to the individual a composition in accordance with claim 17 wherein the heterologous nucleic acid comprises nucleic acid encoding for said antigen.

10 23. A method in accordance with claim 22 wherein the heterologous nucleic acid encodes for at least one antigen selected from the group consisting of: HIV-1 Gag, Nef, or Pol.

1/44

```

1 catcatcaat aatatacccc acaaagtaaa caaaagttaa tatgcaaattg agctttttgaa
61 tttggggcgt ggccgacgct gattggacga gaagtgggtga tgcaaattgac gtcacgacgc
121 acggccgacg gtcgccgcgg aggcgtggcc tagtccggaa gcaagtcgcg gggctgatga
181 cgtataaaaa agcggacttt agaccggaa acggccgatt ttcccgcggg caccgccgga
241 tatgaggtaa ttctgggcgg atgcaagtga aattaggcca ttttggcgcg aaaactgaat
301 gaggaagtga aaagtgaaaa ataccgggce cgcccagggc ggaatattta ccgagggccg
361 agagactttg accgattacg tgggggtttc gattgcggtg tttttttcgc gaatttccgc
421 gtccgtgtca aagtcgggtg tttatgtcac agatcagctg atccacaggg tatttaaaac
481 agtcgagccc gtcaagaggg cactcttgag tgccagcgag tagagatttc tctgagctcc
541 gctcccagag accgagaaaa atgagacacc tgcgccttct accttcaact gtgcccggcg
601 acctggctgt gattatgctg gaggactttg tgaatacagt tctggaggac gaactgcac
661 cagagccatt tgagctggga cctacacttc aggacctcta tgatctggag gtagatgcc
721 atgatgacga ccctaacgaa gaggtgtgta atttaattat tccagaatct atgattcttc
781 aggctgacat agccagtga gcatagtta ctctctaca tactccact ctgcctccca
841 tacctgaatt ggaggaggat gaagaaatag acctccggtg ctacaggaa ggttttctc
901 ccagcgattc agaggacgaa caggggtgagc agcagatggc tctaattctc gatttagctt
961 gtgtgattgt ggaggaacaa gttgtgattg aaaaatctac cgagccagta caaggctgta
1021 ggaactgcca gtatcaccgg gataagtcgc gagaccgaa cgcttcctgc cttctgtgtt
1081 acatgaaatc tactttcagc tttatttaca gtaagtggag tgaatgtgag agaggctgag
1141 tgcttaacac ataactgtaa tgcttgaaca gctgtgctaa gtgtgggtta tttttgttac
1201 taggtccggt gtcagaggat gagtcacac cctcagaaga agaccaccg tctccccctg
1261 agctgtcagg cgaaacgccc ctgcaagtgc acagaccac cccagtcaga gccagtggcg
1321 agaggcgagc agctgtagaa aaaattgagg acttgttaca tgacatgggt ggggatgaac
1381 ctttgacact gagcttgaaa cgcccagga actaggcgca gctgcgctta gtcattgtga
1441 aataaagttg tacaataaaa gtatatgtga cgcatgcaag gtgtgggtta tgactcatgg
1501 gcggggctta gtcctatata agtggcaaca cctgggcact tgggcacaga ctttcaggga
1561 gttcctgatg gatgtgtgga ctatccttgc agactttagc aagacacgcc ggcttgtaga
1621 ggaatgttca gacgggtgct ccgggttctg gagacactgg tttggaactc ctctatctcg
1681 cctggtgtat acagtttaaga aggattataa agaggaattt gaaaatcttt ttgctgactg
1741 ctctggtctg ctgattctc tgaatcttgg ccaccagtcc cttttccagg aaagggactt
1801 ccacagcctt gatttttcca gccaggggcg cactacagcc ggggttgctt ttgtggtttt
1861 tctggttgac aaatggagcc aggacacca actgagcagg ggctacatcc tggacttcgc
1921 agccatgcac ctgtggaggg cctggatcag gcagcgggga cagagaatct tgaactactg
1981 gcttctacag ccagcagctc cgggtcttct tctgtctacac agacaaacat ccatgttgga
2041 ggaagaaatg aggcaggcca tggacgagaa cccgaggagc ggcctggacc ctccgtcgga
2101 agaggagctg gattgaatca ggtatccagc ctgtaccagc agcttagcaa ggtgctgaca
2161 tccatggcca ggggagtga gagggagagg agcagtgagg gtaataccgg gtatgtgacc
2221 gagctgactg ccagctctgat gaatcggaag cgcccagagc gccttacctg gtacgagcta
2281 cagcaggagt gcagggatga gataggcctg atgcaggata aatatggcct ggagcagata
2341 aaaacccatt ggttgaaccc agatgaggat tgggaggagg ctattaagaa gtatgccaa
2401 atagccctgc gccagattg caagtacata gtgaccaaga ccgtgaatat cagacatgcc
2461 tgctacatct cggggaacgg ggcagaggtg gtcacgata ccctggacaa ggccgccttc
2521 aggtgttgca tgatgggaat gagagcagga gtgatgaata tgaattccat gatcttcatg
2581 aacattaagt tcaatggaga gaagtttaat ggggtgctgt tcatggccaa cagccacatg
2641 accctgcatg gctgcagctt cttcggtttc aacaacatgt gcgccgaggt ctggggagct
2701 gctaagatca ggggatgtaa gttttatggc tgctggatgg gcgtggtcgg aagacccaa
2761 agcgagatgt ctgtgaagca gtgtgtgttt gagaaatgct acctgggagt ctctaccgag
2821 ggcaatgcta gagtgagaca ctgctcttcc atggagacgg gctgcttctg cctgggtgaag
2881 ggcacggcct ctctgaagca taatatggtg aagggtgca cggatgagcg catgtacaac
2941 atgctgacct gcgattcggg ggtctgccat atcctgaaga acatccatgt tcaagtgcc
3001 cccagaaaga agtggccagt gtttgagaat aacctgctga ttagccagac caagctgctg
3061 ggtgccagaa ggggcacctt ccagccgtac cagtgcact ttagccagac caagctgctg
3121 ttggagaacg atgccttctc caggggtgaac ctgaacggca tctttgacat ggatgtctcg
3181 gtgtacaaga tcctgagata cgatgagacc aagtccaggg tgccgcgttg cgagtggggg
3241 ggcagacaca ccaggatgca accagtggcc ctggatgtga cagaggagct gagaccagac
3301 cacctggtga tggcctgtac cgggaccgag ttcagctcca gtggggagga cacagattag
3361 aggtaggttt tgagtagtgg gcgtggctaa ggtgagtata aaggcgtgtg cttacagagg
3421 tctttttgct tttctgcaga catcatgaac gggaccggcg gggccttcga aggggggctt

```

FIG. 1A-1

```

3481 ttttagccctt atttgacaac ccgcctgccg ggatggggccg gagttcgtca gaatgtgatg
3541 ggatctacgg tggatggggc cccagtgcct ccagcaaatt cctcgaccat gacctacgcg
3601 accgtgggga gctcgtcgtc cgacagcacc gccgcagccg cggcagccgc agccgccatg
3661 acagcgacga gactggcctc gagctacatg cccagcagca gcagtagccc ctctgtgccc
3721 agttccatca tcgccgagga gaaactgctg gccctgctgg cagagctgga agccctgagc
3781 cgccagctgg ccgccctgac ccagcaggtg tccgagctcc gcgagcaaca gcagcagcaa
3841 aataaatgat tcaataaaca cagattctga ttcaaacagc aaagcatctt tattatttat
3901 tttttcgcgc gcggtaggcc ctgggtccacc tctcccgatc attgagagtg cggtggaattt
3961 tttccaggac ccggtagagg tgggattgga tgttgaggta catgggcatg agcccgctcc
4021 gggggtggag gtagcaccac tgcattggcct cgtgctctgg ggtcgtgttg tagatgatcc
4081 agtcatagca ggggcgctgg gcgtggtgct ggatgatgtc cttaaggagg agactgatgg
4141 ccacggggag ccccttggtg taggtgttgg cgaagcgggt gagctgggag ggatgcatgc
4201 ggggggagat gatgtgcagt ttggcctgga tcttgagggt ggcaatgttg ccgccagat
4261 ccgcctggg gttcatgttg tgcaggacca ccaggacggg gtagcccggt cacttgggga
4321 acttatcatg caacttgga gggaatgctg ggaagaattt ggagacgccc ttgtgcccgc
4381 ccaggttttc catgcaactc tccatgatga tggcgatggg cccgtgggct gcggctttgg
4441 caaagacgtt tctgggggtc gagacatcgt aattatgctc ctgggtgaga tcatcataag
4501 acattttaat gaatttgagg cggagggtgc cagattgggg gacgatagtt ccctcgggc
4561 ccggggcaaa gtccccctc cagatctgca tctcccaggc ttcatctcgc gaggggggga
4621 tcatgtccac ctgcggggcg atgaaaaaaa cggtttcccg ggcgggggtg atgagctgcg
4681 aggagagcag gtttctcaac agctgggact tgccgcaccc ggtcggggcg tagatgacct
4741 cgatgacggg ttgcagggtg tagttcaagg acatgcagct gccgtcgtcc cggaggaggg
4801 gggccacctc gttgagcatg tctctgactt ggaggttttc ccggacgagc tcgccgagga
4861 ggcggtcccc gcccagcgag aggagctctt gcagggaagc aaagtttttc aggggcttga
4921 gtccgtcggc catgggcatc ttggcgaggg tctgcgagag gagctccagg cggctccaga
4981 gctcggtgac gtgctctacg gcactctgat ccagcagact tctcgtttc gggggttggg
5041 acgactgcga ctgtagggca cgagacgatg ggcgtccagc gcggccagcg tcatgtcctt
5101 ccagggtctc agggctccgc tgagtgtggt cctcgtcacg gtgaagggtt gggccccggg
5161 ctggggcctt gcaagggtgc gcttgagact catcctgctg gtgctgaaac gggcacggtc
5221 ttcgccctgc gcgtcggcga gatagcagtt gacctgagc tcgtagttag gggcctcggc
5281 ggcgtggccc ttggcgcgga gcttgccctt ggaagagcgc ccgcaggcgg gacagagaag
5341 ggattgcagg gcgtagagct tgggtgcgag aaaaacggac tcgggggcga aggcgtccgc
5401 tccgcagtgg gcgcagacgg tctcgcactc gacgagccag gtgagctcgg gctgctcggg
5461 gtcaaaaacc agttttcccc cgttcttttt gatgcgcttc ttacctcgcg tctccatgag
5521 tctgtgtccg cgctcgggtg caaacaggct gtcgggtgtc ccgtagacgg acttgatggg
5581 cctgtcctgc agggacgtcc cgcggctctc ctcgtagaga aactcggacc actctgagac
5641 aaaggcgcg gcacacgcca agacaaagga ggcacagtg gcgggtagc ggtcgttgtc
5701 caccaggggg tccacctttt ccaccgtgtg cagacacatg tccccctcct ccgcatccaa
5761 gaagggtgatt ggcttgtagg tgtaggccac gtgaccgggg gtccccgacg ggggggtata
5821 aaagggggcg ggtctgtgct cgtcctcact ctcttccgcg tcgctgtcca cgagcgccag
5881 ctggttggggt aggtattccc tctctagagc gggcatgacc tcggcactca ggttgtcagt
5941 ttctagaaac gaggaggatt tgatgttggc ctgcccgtgc gcgatgcttt ttaggagact
6001 ttcattccatc tggtcagaaa agacaatttt tttattgtca agcttgggtg caaaggagcc
6061 atagaggggc ttggatagaa gcttggcgat ggatctcatg gtctgatttt tgtcacggtc
6121 ggcgcgctcc ttggccgcga tgttttagct gacatactcg cgcgcgacgc acttccattc
6181 ggggaagacg gtggtgcgct cgtcgggcac gatcctgacg cgcagccgc ggttatgcag
6241 ggtgaccagg tccacgctgg ttggccacct gccgcgagg ggctcgttgg tccagcagag
6301 tctgccgccc ttgcgcgagc tgaacgtggg cagcacatca agcagatgct cgtcaggggg
6361 gtccgcacgc atggtgaaga tgcccgga ca gaggttcctt tcaaaataat cgatttttga
6421 ggatgcatca tccaaggcca tctgccactc gcgggcggcc agcgctcgct cgtaggggtt
6481 gaggggcgga ccccagggca tgggatgcgt gagggcgagg gcgtacatgc cgcagatgtc
6541 atacacatag atgggctccg agaggatgcc gatgtagggt ggataacagc gccccccgcg
6601 gatgctggcg cgcacgtagt catacaactc gtgcgagggg gccaaagaagg cggggccgag
6661 attggtgcgc tggggctgct cggcgcgga gacgatctgg cgaaagatgg cgtgcgagtt
6721 tgaggagatg gtgggcccgtt ggaagatgtt aaagtggggc tgaggcaggc ggaccgagtc
6781 gcggatgaag tgccgcgtagg agtcttgtag cttggcgacg agctcggcgg tgacgaggac
6841 gtccatggca cagtagtcca gcgtttcgcg gatgatgtca taaccgcctt ctctttctt
6901 ctcccacagc tcgcggttga gggcgctact ctcgtcatcc ttccagtact cccggagcgg

```

FIG. 1A-2

```

6961 gaatcctcga tcgtccgcac ggtaagagcc cagcatgtag aatgggttca cggccttgta
7021 gggacagcag cccttctcca cggggagggc gtaagcttga gcggccttgc ggagcgaggt
7081 gtgcgtcagg gcgaaggtgt ctctgaccat gactttcaag aactggtact tgaaatccga
7141 gtcgtcgcag ccgccgtgct cccagagctc gaaatcgggt cgcttcttcg agaggggggtt
7201 aggcagagcg aaagtgcagt cattgaagag aatcttgctt gcccgcgcca tgaaattgcg
7261 ggtgatgcgg aaagggcccg ggacggaggc tcggttggtg atgacctggg cggcgagcac
7321 gatctcgtcg aagccgttga tgttgtgccc gacgatgtag agttccatga atcgcgggcg
7381 gcctttgatg tgcggcagct ttttgagctc ctcgtaggtg aggtcctcgg gtcattgcag
7441 tccgtgctgc tcgagcgcgc actcctggag atgtgggttg gcttgcataa aggaagccca
7501 gagctcgcgg gccatgaggg tctggagctc gtcgcgaaag agacggaact gctggccccc
7561 ggccatcttt tcgggtgtga cgcagtagaa ggtgaggggg tcccgcctcc agcgatccca
7621 gcgtaagcgc acggcgagat cgcgagcgag ggcgaccagc tcgggggtccc cggagaattt
7681 catgaccagc atgaagggga cgagctgctt gccgaaggac cccatccagg ttaggtttc
7741 tacatcgtag gtgacaaaga gccgctccgt gcgaggatga gagccgattg ggaagaactg
7801 gatttcctgc caccagttgg acgagtggct gttgatgtga tgaaagtaga aatcccgcg
7861 gcgaaccgag cactcgtgct gatgcttgta aaagcgtccg cagtactcgc agcgtgcac
7921 gggctgtacc tcatccacga gatacacagc gcgtcccctg aggaggaact tcaggagtgg
7981 cggccctggc tggtgggttt catgttcgcc tgcgtgggac tcacctggg gctcctcgag
8041 gacggagagg ctgacgagcc cgcgcgggag ccaggtccag atctcggcgc ggccggggcg
8101 gagagcgaag acgagggcgc gcagttggga gctgtccatg gtgtcgcgga gatccaggtc
8161 cgggggcagg gttctgaggt tgacctcgta gaagcgggtg agggcggtgc tgagatgcag
8221 atggtacttg atttctacgg gtgagttggg ggccgtgtcc acgcattgca tgagcccgt
8281 gctgcgcggg gccacgaccg tgccgcgggt cgcttttaga agcgggtgtc cggacgcgct
8341 cccggcggca gcggcggttc cggcccccg cgagggggcg gcagaggcac gtcggcggtg
8401 cgctcgggca ggtcccgggt ctgcgccttg agagcgctgg cgtgcgcgac gacgcggcgg
8461 ttgacatcct ggatctgccc cctctgcgtg aagaccacgg gcccctgac tttgaacctg
8521 aaagacagtt caacagaatc aatctcggcg ctattgacgg cggcctgacg caggatctct
8581 tgcacgtgct ccgagttgtc ctggtaggcg atctcggaca tgaactgctc gatctcctcc
8641 tcctggagat gcgcgcggcc cgcgcgctcc acggtggcgg cgaggtcatt cgagatgcga
8701 cccattagct gcgagaaggg gcccaggccg ctctcgttcc agacgcggct gtagaccag
8761 tccccgtcgg cgtcgcgcgc gcgcatgacc acctgcgcga ggttgagctc cacgtgccgc
8821 gtgaagacgg cgtagttagc caggcgctgg aagaggtagt tgaggggtgt ggcgatgtgc
8881 tcggtgacga agaagtacat gatccagcgg cgcaggggca tctcgtgat gtcgccgatg
8941 gcctccagcc tttccatggc ctctagaaaa tccacggcga agttgaaaaa ctgggcgttg
9001 cgggcccaga ccgtgagctc gtcttccagg agccggatga gctcggcgat ggtggcgcgc
9061 acctcgcgct cgaaatcccc ggggacctcc tctcttccat cttcttccat ctcgacctct
9121 tcttctatct ctctctctgg gggcggtggt tggggcgggg cccgacgacg acggcgacgc
9181 accgggaggt ggtcgacgaa gcgctcgatc atctccccgc ggccgacgac catggtttcg
9241 gtgacggcgc gaccccgctt gcgaggacgc agcgtgaaga cgccgcgggt catctcccgg
9301 taatggggcg ggtccccggt gggcagcgat agggcgctga cgatgcattc tatcaattgc
9361 ggtgtagggg acgtgagcgc gtcgagatcg accggatcgg agaattcttc gaggaaagcg
9421 tctagccaat cgcagtcgca aggtaaagctc aaacacgtag cagccctgtg gacgctgtta
9481 gaattgcggt tgctgatgat gtaattgaag taggcgtttt taaggcgggc gatggtggcg
9541 aggaggacca ggtccttggg tcccgtttgc tggatgcgga gccgctcggc catgccccag
9601 gcctggccct gacaccggct caggttcttg tagtagtcat gcatgagcct ttcaatgtca
9661 tcactggcgg aggcggagtc ttccatgcgg gtgaccccga cgccctgag cggctgcacg
9721 agcgccaggt cggcgacgac gcgctcggcg aggatggcct gttgcacgcg ggtgaggggtg
9781 tcctggaagt catccatgtc gacgaagcgg tggtaggccc cggtgttgat ggtgtaggtg
9841 cagttggcca tgagcgacca gttgacggtc tgcaggccgg gttgcacgac ctcgagtagc
9901 ctgagccgcg agaaggcgcg cgagtcgaag acgtagtcgt tgcaggtgcg cacgaggtac
9961 tggtagccga ctaggaagtg cggcgggcgg tggcggtaga gcggccagcg ctgggtggcc
10021 ggcgcgcccg gggccaggct ctcgagcatg aggcggtggt agccgtagag gtagcgggac
10081 atccaggtga tgccgcggcg ggtggtggag gcgcgcggga actcgcggac gcggttccag
10141 atggttgcga gcggcaggaa ataatccatg gtcggcacgg tctggccggt gagacgcgcg
10201 cagtcattga cgctctagag gcaaaaacga aagcggttga gcgggctctt cctccgtagc
10261 ctggcggaac gcaaacgggt taggccgcgt gtgtacccc gttcgaagtc ctcgaatca
10321 ggctggagcc gcgactaacg tggtagttgc actcccgtct cgaccgagc ccgatagcgc
10381 ccaggatagc gcggagagcc ctttttgccc accgagtggt gtcgctagac ttgaaagcgg

```

FIG. 1A-3

```

10441 ccgaaaaccc cgccgggtag tggctcgcgc ccgtagtctg gagaagcatc gccaggggtg
10501 agtcgcggca gaacccgggt cgcgacggc cgcggcgagc gggacttggt caccgcccg
10561 atttaaagac ccacagccag ccgacttctc cagttacggg agcgagcccc cttttttctt
10621 tttgccagat gcatcccgtc ctgcgccaaa tgcgtcccac cccccggcg accaccgca
10681 ccgcgccgct aacaggcgcc ggcgctagcc agccacagac agagatggac ttggaagagg
10741 gcgaagggtt ggcgagactg ggggcgcgct ccccgagagc acaccccgcc gtgcagctgc
10801 agaaggacgt gcgcccggcg tacgtgcctg cgcagaacct gttcagggag cgcagcgggg
10861 aggagccccg ggagatgcgc gactgccggt ttcgggcggg cagggagctg cgcgagggcc
10921 tggaccgcca gcgctgctg cgcgacgagg atttcgagcc gaacgagcag acggggatca
10981 gccccgcgcg cgcgcacgtg ggcgcggcca acctggtgac ggcctacgag cagacgggtg
11041 agcaggagcg caacttccaa aagagtttca acaaccacgt gcgcacgctg atagcgcgcg
11101 aggaggtggc cctgggcctg atgcacctgt gggacctggc ggaggccatc gtgcagaacc
11161 cggacagcaa gcctctgacg gcgcagctgt tcctggtggt gcagcacagc agggacaacg
11221 aggcgttcag ggaggcgctg ctgaacatcg ccgagccccg gggtcgctgg ctgctggagc
11281 tgatcaacat cttgcagagc atcgtagtgc aggagcgagc cctgagcctg gccgagaagg
11341 tggcggcgat caactactcg gtgctgagcc tgggcaagtt ttacggcgcg aagatttaca
11401 agacgccgta cgtgcccata gacaaggagg tgaagataga cagcttttac atgcgcctg
11461 cgctcaaggt gctgacgctg agcgacgacc tgggcgtgta tcgcaacgac cgcattccaca
11521 aggcggtgag cagagcccg cggcgcgagc tgagcgaccg cgagctgatg ctgagcctgc
11581 gccgggcgct ggtagggggc gccgcggcg gcgaggagtc ctacttcgac atggggcgcg
11641 acctgcattg gcagccgagc cggcgcgctt tggaggccgc ctacggtcca gaggacttgg
11701 atgaggatga ggaagaggag gaggatgcac ccgttgcggg gtactgacgc ctccgtgatg
11761 tgtttttaga tgtcccagca agccccggac cccgccataa gggcgcgct gcaaagccag
11821 ccgtccggtc tagcatcgga cgactgggag gccgcgatgc aacgcatcat ggccctgacg
11881 acccgcaacc ccgagtcctt tagacaacag ccgcaggcca acagactctc ggccattctg
11941 gaggcgggtg tccccctctg gaccaacccc acgcacgaga aggtgctggc gatcgtgaac
12001 gcgctggcgg agaacaaggc catccgtccc gacgaggccg ggtgggtgta caacgccctg
12061 ctggagcgcg tgggcgcgta caacagcagc aacgtgcagt ccaacctgga ccggctggtg
12121 acggagcgtg gcgaggccgt ggcgcagcgc gagcggttca agaacgaggg cctgggctcg
12181 ctggtggcgc tgaacgcctt cctggcgacg cagccggcga acgtgccgcg agggcaggac
12241 gattacacca actttatcag cgcgctgcgg ctgatggtga ccgaggttcc ccagagcgag
12301 gtgtaccagt cgggcccggg ctactttttc cagacaagcc ggcagggcct gcagacggtg
12361 aacctgagtc aggttttcaa gaacctgcgc gggctgtggg gcgtgcaggc gcccggtggg
12421 gaccggtcga cggtgagcag cttgctgacg cccaactcgc ggetgctgct gctgctgatc
12481 gcgccttcca ccgacagcgg cagcgtgaac cgcaactcgt acctgggcca cctgctgacg
12541 ctgtaccgcg aggccatagg ccaggcgagc gtggacgagc agacctcca ggcagatcac
12601 agcgtgagcc gcgcgctggg gcagaacgac accgacagtc tgagggccac cctgaacttt
12661 ttgctgacaa atagacagca gaagatccc ggcgagtagc cactgtcggc cgaggaggaa
12721 aggatcctga gatattgtga gcagagcgta gggctgttcc tgatgcagga ggggtgccacc
12781 ccagcgcccg cgctggacat gaccgcgcgc aacatggaac ctacgatgta cgccgccaac
12841 cggccgttca tcaataagct gatggactac ctgcaccgcg cggcgcccat gaacacggac
12901 tacttttcaa acgccatatt gaacctgcgc tggcttccgc cgccgggggt ctacacgggc
12961 gagtacgaca tgcccagacc caacgacggg ttcctgtggg acgacgtgga cagcgcgggtg
13021 ttctcgccga cctttcaaaa gcgccaggag gcgccgccga gcgagggcgc ggtggggagg
13081 agcccccttc ctagcttagg gagtttgcac agcttgccgg gctcgtgtaa cagcggcagg
13141 gtgagccggc cgcgcttgct gggcgaggac gagtacctga acgactcgct gctgcagccg
13201 ccgcgggtca agaaccgat ggccaataac gggatagaga gtctggtgga caaactgaac
13261 cgctggaaga cctacgctca ggaccatagg gaggctgcgc ccgcgcccg gcgacagcgc
13321 cacgaccggc agcggggcct ggtgtgggac gacgaggact cggccgacga tagcagcgtg
13381 ttggacttgg gcgggagcgg tggggccaac ccgttcgcgc atctgcagcc cagactgggg
13441 cggcggatgt tttgaatgca aaataaaact caccaaggcc atagcgtgcg ttctcttctt
13501 tgtagagat gaggcgtgcg gtggtgtctt cctctcctcc tccctcgtac gagagcgtga
13561 tggcgcaggc gacctggag gttccgtttg tgccctccgc gtatatgggt cctacggagg
13621 gcagaaacag cattcgttac tcggagctgg ctccgcagta cgacaccact cgcgtgtact
13681 tgggtggaaa caagtcggcg gacatcgctt ccctgaacta ccaaaacgac cacagcaact
13741 tcctgaccac ggtggtgcag aacaacgatt tcaccccgcc cgaggccagc acgacagca
13801 taaattttga cgagcggtcg cggtggggcg gtgatctgaa gaccattctg cacaccaaca
13861 tgcccaatgt gaacgagtag atgttcacca gcaagtttaa ggcgcggggt atggtggcta

```

FIG. 1A-4

```

13921 gaaagcatcc caaagatgta gatgccagtg atttaagcaa ggatatctta gagtataagt
13981 ggttttgagtt taccctgccc gagggcaact tttccgagac catgaccata gacctgatga
14041 acaacgccat cttggaaaac tacttgcaag tggggcggca gaatggcgtg ctggagagcg
14101 atatcgaggt caagtttgac agcaggaatt tcagactggg ctgggacccg gtgaccaagc
14161 tggatgatgcc aggggtctac acctacgagg ccttccaccc ggacgtgggtg ctactgccgg
14221 gctgcgggggt ggacttcacc gagagccgcc tgagcaacct cctgggcatt cgcaagaagc
14281 aaccttttca agagggcttc agaatcatgt atgaggatct agaaggggggt aacatccccg
14341 ctctcctgga taccaaaaaa tatctggata gcaagaagga acttgaggat gctgccaaagg
14401 aagctgcaaa gcaacaggga gatggtgctg tcactagagg cgataccacc ctcactgtag
14461 ctcaagaaaa agcagctgaa aaggagctag tgatcgtacc aattgaaaag gatgagagca
14521 acagaagtta caacctgata aaggacaccc atgacacccct gtaccgaagc tggtagctgt
14581 cctataccta cggggacccc gagaaggggg tgcagtcgtg gacgtgtctc accaccccgg
14641 acgtcacctg cggcgcgagg caagtctact ggtcgtgtgc ggacctcatg caagaccccg
14701 tcaccttccg ctctaccacg caagtcagca actaccccgt ggtcggcgcc gagctcatgc
14761 ccttccgcgc caagagcttt tacaacgacc tcgccgteta ctcccagctc atccgcagct
14821 acacttccct caccacgctc ttcaaccgct tccccgacaa ccagatcctc tgcgccccgc
14881 ccgcgccccac catcaccacc gtcagtgaac acgtgcctgc tctcacagat cacgggacgc
14941 ttccgctgcg cagcagtatc cgcggagtcc agcagtgac cgtcactgac gccctcgcc
15001 gcacctgtcc ctacgtctac aaggccctgg ccatagtcgc gccgcgctg ctctccagtc
15061 gcaccttcta aaaaatgtct attctcatct cgcccagcaa taacaccggc tgggggtctta
15121 ctaggcccag caccatgtac ggaggagcca agaagcgctc ccagcagcac cccgtccgcy
15181 tccgcggtca cttccgcgct ccctggggag cttacaagcg ggggcgcaact gccaccgcy
15241 ccgcgctgcy caccaccgct gacgacgtca tcgactcggg ggtcgcgcac gcgcgcaact
15301 acacccccgc cccctccacc gtggacgcgg tcatcgacag cgtggtggcc gacgcgcgcg
15361 actatgccag acgcaagagc cggcgcgac ggatcgccag gcgccaccgg agcacgcccg
15421 ccatgcgcgc cgcccgggct ctgctgcgcc gcgccagacg caggggcccgc cgggccatga
15481 tgcgagccgc gcgcgcgcgc gccactgac ccccgagag caggactcgc agacgagcgg
15541 ccgcgcgcgc tgccgcgcgc atctctagca tgaccagacc caggcgcgga aacgtgtact
15601 ggggtgcgca ctcgctcacg ggcgtgcgcg tgcccgtgcy caccgctct cctcgtccct
15661 gatctaagtc ttgtgtctc cccgcgaagc gacgatgtca aagcgcaaaa tcaaggagga
15721 gatgctccag gtcgtgcgcc cggagattta cggacccccg gaccagaaac cccgcaaaat
15781 caagcgggtt aaaaaaaagg atgaggtgga cgagggggca gtagagtbtg tgcgcgagtt
15841 cgctccgcgg cggcgcgtaa attggaaggg gcgcaggggt cagcgcggtg tgcggcccgg
15901 cacggcggtg gtgttcacgc ccggcgagcy gtcctcggtc aggagcaagc gtagctatga
15961 cgaggtgtac ggcgacgacg acatcctgga ccaggcgcg gcgaggttcgc gcgagttcgc
16021 ctacgggaag cggctgcgcg aagaggagct gatctcgtg ccgctggacg aaagcaacc
16081 cagcccgagc ctgaagcccg tgaccctgca caggtgtctg cccagcgcg tgcgtgctc
16141 gagccgcggg gtcagcgcg agggcgagag catgtacccg accatgcaga tcatggtgcc
16201 caagcgccgg cgcgtggagg acgtgctgga caccgtgaaa atggatgtgg agcccaggtt
16261 caaggtgcgc cccatcaagc aggtggcgcc gggcctgggc gtgcagaccg tggacattca
16321 gatccccacc gacatggatg tcgacaaaaa accctcgacc agcatcgagg tgcagaccga
16381 cccctggctc ccagcctcca ccgtaccgt ctccactttt acccgcccca cggctaccga
16441 gcctccagag aggcgaagat ggggcgcgcg cagccggctg atgcccaact acgtgttgca
16501 tccttccatc atcccagcgc cgggctaccg cggcacccgg tactacgcca gccgcaggcg
16561 cccagccgcc aaacgcgcgc gccgcaactgc caccgcgcgc cgtctggccc cgcaccacc
16621 gcgcgcgcta accacgcgcc ggggcgcgtc gctcgttctg cccaccgtgc gctaccacc
16681 cagcatcctt taatccgtgt gctgtgatac tgttgagag agatggctct cacttgccgc
16741 ctgcgcaccc ccgtcccga ttaccagga agatcccgc gcaggagagg catggcaggc
16801 agcggcctga accgcgcgcg gcggcgggcc atgcgcaggc gcctgagtgg cgggtttctg
16861 ccgcgcgtca tcccataat cgcgcgcggc atcggcacga tcccgggcat agcttccgtt
16921 gcgctgcagg cgtcgcagcg ccgttgatgt gcgaataaag cctctttaga ctctgacaca
16981 cctggctcctg tataattttt gaatggaaga catcaatttt gcgtccctgg ctccgcggca
17041 cggcacgcgg ccgttcatgg gcacctggaa cgagatcggc accagccagc tgaacggggg
17101 cgcttcaat tggagcagtg tctggagcgg gcttaaaaaa ttcggctcga cgctccggac
17161 ctatgggaac aaggcctgga atagtagcac ggggcagttg ttaagggaag agctcaaaga
17221 ccagaacttc cagcagaagg tggtagcagg gctggcctcg ggcattaacg ggggtgggga
17281 catcgcaaac caggccgtgc agcgcgagat aaacagccgc ctggacccgc ggccgccac
17341 ggtggtggag atggaagatg caactcttcc gccgccccaa ggcgagaagc ggccgcggcc

```

FIG. 1A-5

6/44

```

17401 cgacgcggag gagacgatcc tgcaggtgga cgagccgccc tcgtacgagg aggccgtcaa
17461 ggccggcatg cccaccacgc gcatcatcgc gccgctggcc acgggtgtaa tgaaaccgcg
17521 cacccttgac ctgcctccac caccacgcc cgctccaccg aaggcagctc cggttgtaga
17581 ggcccctccg gtggcgaccg ccgtgcgcgc cgtccccgcc cgccgccagg cccagaactg
17641 gcagagcacg ctgcacagta tcgtgggcct gggagtgaag agtctgaagc gccgccgatg
17701 ctattgagag aggaaagagg acactaaagg gagagcttaa cttgtatgtg ccttaccgcc
17761 agagaacgcg cgaagatggc caccctctcg atgatgccgc agtgggcgta catgcacatc
17821 gccgggcagg acgcctcgga gtacctgagc ccgggtctgg tgcagtttgc ccgcgccacc
17881 gacacgtact tcagcctggg caacaagtth aggaacccca cggtggcccc gaccatgat
17941 gtgaccacgg accggtccca gcgtctgacg ctgcgcttcg tgcccgtgga tcgcgaggac
18001 accacgtact cgtacaaggc gcgcttcaact ctggccgtgg gcgacaaccg ggtgctagac
18061 atggccagca cgtactttga catccgcggc gtcctggacc gcggtcccag cttcaaacc
18121 tactcgggca cggcttaca cagtttggcc ccaaggggcg ccccaactc cagtcagtgg
18181 actgacaaag aacggcaaaa tgggtggaca ccaccacta caaaagatgt taaaaaaca
18241 ttcggagtag cagccagggg agggcttcat attactgata aaggactaca aataggagaa
18301 gatgaaaata acgaggatgg tgaagaagag atatatgcag aaaaacttt ccagccagaa
18361 cctcaagtag gagaggaaaa ctggcaagat actgatgttt tctatggcgg cagagcgctt
18421 aaaaaggaaa ccaaattgaa accatgctat ggctcttttg ccagacctac caatgaaaaa
18481 ggaggtcaag cttaaatttt tctttgatct taaacaaaat gacactggaa ctactcaaaa
18541 acattagctt tctttgatct taaacaaaat gacactggaa ctactcaaaa ccagccagat
18601 gttgtcatgt acactgaaaa tgtgtatctg gaaaccccag acacccatgt ggtgtacaaa
18661 cctggcaagg aagatacaag ctccgctgct aaccttacac aacagtccat gcccaacagg
18721 cccaactaca ttggtttcag ggacaacttt gtggggctca tgtattacaa cagcactggc
18781 aacatgggtg tgctggctgg tcaggcctct cagttgaatg ctgtggttga cttgcaagac
18841 agaaacaccg agctgtctta tcagctcttg ctagattctc tgggtgacag aaccagatac
18901 tttagcatgt ggaattctgc ggtggacagc tatgatccag atgtcaggat cattgagaat
18961 cacggtgttg aagatgagct tccaaattat tgcttcccac tggatggatc tggcagcaat
19021 accgcataatc aaggtgttaa atatgaaaac ggagctggca atggaagctg gaaagtagat
19081 ggcgaaagtg cttctcagaa tcagatcgcc aagggtaatc tgtatgccat ggagataaac
19141 cttcaggcca acctgtggaa gagttttctg tactcgaacg tggcgctgta tctaccagac
19201 tcctacaagt acacgccggc caacatcacg ctgcccacca acaccaacac ctacgagtag
19261 atgaacggcc gcgtggtggc accctcgctg gtggatgcct atgtcaacat cgggtgcccg
19321 tggctcgctg accccatgga caacgtcaac cccttcaacc accaccgcaa cgcggtctg
19381 cgctaccgct ccatgcttct gggcaacggc cgctacgtgc ccttccacat ccaagtggcc
19441 caaaagttct ttgccatcaa gaacctgctc ctgcttcccg gttcctacac ctacgagtgg
19501 aacttccgca aggatgtcaa catgatcctg caaagttccc tcggcaacga cctgcgcgtc
19561 gacggcgctt ccgtccgctt cgacagcgct aacctctatg ccaccttctt cccatggcg
19621 cacaacaccg cctccaccct tgaagccatg ctgcgcaacg acaccaacga ccagtccttc
19681 aacgatacc tctcgccgcg caacatgctc tacccaatcc cggccaaggc caccaacgtg
19741 cccatctcca tcccctcgcg caactggggc gccttccgcg gctggagtth cacccggtc
19801 aagaccaagg aaactccctc cctcggtctg ggttttcgac cctactttgt ctactcgggc
19861 tccattccct acctcgacgg aaccttctac ctcaaccaca ccttcaagaa ggtctccatc
19921 atgttcgact cctcggtcag ctggcccggc aacgaccggc tgctcacgcc gaacgagttc
19981 gagatcaagc gcagcgtcga cggggagggg tacaacgtgg cccaatgcaa catgactaag
20041 gactgggttc tcgtccagat gctctctcat tacaacattg gctaccaggg cttctacgtg
20101 cctgagggtt acaaggaccg catgtactcc ttcttccgca acttccagcc catgagcagg
20161 cagggtggtc atgagatcaa ctacaaggac tacaaggccg tcacctgccc cttccagcac
20221 aacaactcgg gcttaccgg ctacctcgca cccaccatgc gtcaggggca gccatacccc
20281 gccaacttcc cctaccgct catcggccag acagccgtgc cctccgtcac ccagaaaaag
20341 ttcctctgcg acagggctcat gtggcgcatc ccttcttcca gcaacttcat gtccatgggc
20401 gccctcaccg acctgggtca gaacatgctc tacgccaact cggcccacgc gctcgacatg
20461 accttcgagg tggaccccat ggatgagccc accctcctct atcttctctt cgaagtttct
20521 gacgtggtca gagtgacca gccgcaccgc ggcgtcatcg aggcgtcta cctgcgcacg
20581 cccttctccg ccggaaacgc caccacataa gcatgagcgg ctccagcgaa cgagagctcg
20641 cgcccatcgt gcgcgacctg ggctgcgggc cctacttttt gggaaacccac gacaagcgct
20701 tccctggctt cctcgccggc gacaagctgg cctgcgccat cgtcaacacg gccggccgcg
20761 agaccggagg cgtgcaactg ctgccttcg gctggaaccc gcgtcgcgc cctgtctaca
20821 tgttcgacct ctttggggtc tcggaccgcc ggctcaagca gatttacagc ttcgagtacg

```

FIG. 1A-6

```

20881 aggccatgct gcgcgcgcagc gccctggcct cctcgcccga ccgctgtctc agcctcgagc
20941 agtccacca gaccgtgcag gggcccgaact ccgcccgcctg cggacttttc tgttgcattg
21001 tcttgcatgc cttcgtgcac tggcccgaact gaccatgga cgggaacccc accatgaact
21061 tgctgacggg ggtgcccac ggcattgctac aatcgccaca ggtgctgccc accctccggc
21121 gcaaccagga ggagctctac cgcttcctcg cgcgcactc cccttacttt cgatcccacc
21181 gcgcgcgcac cgaacacgcc accgcttttg acaaaatgaa acaactgcgt gtatctcaat
21241 aaacagcact tttattttac atgactgga gtatatgcaa gttatttaaa agtcgaaggg
21301 gttctcgcg cgtcgttgt gcgcgcgcgt ggggagggcc acgttgcggt actggtactt
21361 gggaagccac ttgaactcgg gcatcaccag tttgggaacc ggaatctcgg ggaaggtctc
21421 gctccacatg cgcggctca tctgcagggc gccagcatg tccggggcgg agatcttgaa
21481 atcgcatgtt ggaccggtgc tctgcgcgcg cgagttgcgg tacacggggt tgcagcactg
21541 gaacaccatc agactggggt acttcacact ggccagcacg ctcttgctcg tgatctgac
21601 cttgtcaagg tcctcggcgt tgctcaggcc aaacggggtc atcttgcaac gctggcggcc
21661 caggaagggc acgctctgag gcttgtggtt aactcgcag tgaatgggca ttagcatcat
21721 cccgcgcgcg cgctgcatat tcgggtagag ggccttgaca aaggccgaga tctgtttgaa
21781 agcttgcgtg gccttggtc cctcgctgaa aaacagccc cagctcttcc cgtgaactg
21841 gttattcccg caaccggcat cctgcacgca gcagcgcgcg tcattggttg tcagttgcac
21901 cagctcgcgt cccagcggg tctgggtcac cttggccttg ctgggttgct ccttcagcgc
21961 gcgctgcgcg ttctcgtcgt tcacatccat ctcaccacg tggctcttgt ggatcatcac
22021 cgctccatgc agacacttga gctggccttc cacctcgggt cagccgtgat cccacagggc
22081 gcatccggtg cactcccagt tcttatgcgc gatcccgcgt tggctgaaga tgtaaccttg
22141 caacaggcga cccatgacgg tgctaaaggc tttctgggtg gtgaagggtc gttgcagacc
22201 gcgggcctcc tcgttcatcc aggtctggca catcttctg aagatctcgg tctgctctg
22261 catgagcttg taagcatcgc gcaggccgct gtcgacgcgg tagcgttcca tcagcacgtt
22321 catggtatcc atgcccttct cccaggacga gaccagaggc agactcaggg ggttgccgac
22381 gttcaggaca ccggggggtc cggtctcgac gatgcgtttt ccgtccttgc ctctctcaa
22441 cagaaccgga ggttggtga atcccactcc cagatcacg gcgtcttcc ggggcattct
22501 ttcgtctggg tctaccttgg tcacatgctt ggtcttctg gcttgcctt tttttggagg
22561 gctgtccacg gggaccacgt cctcctccga agaccgcgg agaccaccgt gatactttcg
22621 gcgcttggtg ggcagaggag gtggcggcga ggggctcct tcctgctccg gcggatagcg
22681 gcgcgaccg tggccccggg gcggagtggc ctctcgctcc atgaaccggc gcacgtcctg
22741 actgccgcg gccattgttt cctaggggaa gatggaggag cagccgcgta agcaggagca
22801 ggaggaggac ttaaccaccc acgagcaacc caaaatcgag caggacctg gcttcgaaga
22861 gccggctcgt ctagaacccc cacaggatga acaggagcac gagcaagacg caggccagga
22921 ggagaccgac gctgggctcg agcatggcta cctgggagga gaggaggatg tgctgctgaa
22981 acacctgcag cgccagtccc tcacctcctg ggacgccttg gccgaccgga gccgaacccc
23041 cctcagcgtc gaggagctgt gtcgggccta cgagctcaac ctcttctcgc cgcgcgtgcc
23101 ccccaaacgc cagcccaacg gcacctgcga gcccacccg cgtctcaact tctatcccgt
23161 ctttgcggtc cccgaggccc ttgccacctc tcacatcttt ttcaagaacc aaaagatccc
23221 cgtctcctgc cgcgcacaacc gcaccgcgc cgacgcgctc ctcgctctgg ggcccggcgc
23281 gcgcatacct gatatcgctt ccctggaaga ggtgcccaag atcttcgaag ggctcggtcg
23341 ggacgagacg cgcgcggcga acgctctgaa agaaacagca gaggaagagg gtcacactag
23401 gcgcctggtg gagttggaag gcgacaacgc caggctgggt gtgctcaage gcagcgtcga
23461 gctcaccac ttcgcctacc ccgcggttaa cctccgcgcc aaggtcatgc gtcgcatcat
23521 ggatcagctt atcatgcccc acatcgaggc catcgatgag acccaagagc agcgcgccga
23581 ggacgcccgg ccctggttca gcgacgagat gctgcgcgcg tggctcggga ccgcgaccc
23641 ccaggctttg gaacagcggc gcaagctgat gctggccgta gtcctggtca cctcgagct
23701 cgaatgcatg cgccgcttct tctcgacacc cgagaccctg cgcaaggctc agagaccct
23761 gcactacact ttcagacacg gtttcgtcag gcaagcctgc aagatctcca acgtggagct
23821 gaccaacctg gtctcctgcc tggggatcct gcatgagaac cgcctggggc agacagtgt
23881 cactctacc ctgaaggcg aggcacggcg ggactatgtc cgcgactgcg tcttctctt
23941 tctatgccac acatggcaag cagccatggg cgtgtggcag cagtgtctcg aggacgaaa
24001 cctgaaggag ctggacaagc ttcttgctag aaaccttaaa aagttgtgga cgggcttcga
24061 cgagcgcacc gtcgcctcgg acctggccga gatcgtttcc cccgagcgc tgaggcatac
24121 gctgaaaggc gggctgcccg acttcatgag ccagagcatg ttgcaaaact accgcacttt
24181 cattctcgag cgtcgggta tcctgcccgc cacctgcaac gccttccct cgcactttgt
24241 cccgctgagc taccgcgagt gtccccgcgc gctgtggagc cactgtacc tcttcagct
24301 ggctaactac atctcctacc actcggacgt gatcgaggac gtgagcggcg aggggctgct

```

FIG. 1A-7

```

24361 cgagtgccac tgccgctgca acctgtgtct cccgcaccgc tccctgggtct gcaacccccca
24421 gctcctgagc gagacccagg tcatcggtac cttcgagctg caaggtccgg agaagtccac
24481 cgctccgctg aaactcacgc cgggggtgtg gacttccgcg tacctgcgca aatttgtacc
24541 cgaagactac cacgcccacg agataaagtt cttcgaggac caatcgctc cgcagcacgc
24601 ggatctcacg gcctgcgtca tcacccaggg cgcgatcctc gcccaattgc atgccatcca
24661 aaaatccccg caagagtttc ttctgaaaaa gggtagaggg gtctacctgg acccccagac
24721 gggcgaggtg ctcaaccggg gtctccccc gcatgccgag gaagaagcag gagccgctag
24781 tggaggagat ggaagaagaa tgggacagcc aggagagga ggacgaatgg gaggaggaga
24841 cagaggagga agaattggaa gaggtggaag aggagcaggc aacagagcag cccgtcgccg
24901 caccatccgc gccgcagcc cggcggtca cggatacaac ctccgctccg gtcaagcctc
24961 ctcgtagatg ggatcgagtg aagggtgacg gtaagcacga gcggcagggc taccgatcat
25021 ggagggccca caaagccgcg atcatcgctt gcttgcaaga ctgcgggggg aacatcgctt
25081 tcgcccgcgg ctacctgtct ttcaccgcg ggggtgaacat cccccgcaac gtgttgcat
25141 actaccgtca cttcacagc taagaaaaag caagtaagag gagtccggg aggaggcctg
25201 aggatcgcg cgaacgagcc ctcgaccacc agggagctga ggaaccggat cttcccact
25261 ctttatgcca tttttcagca gagtcgaggt cagcagcaag agctcaaagt aaaaaatcgg
25321 tctctgcgct cgctcaccgg cagttgtctg taccacaaa acgaagatca gtgcagcgc
25381 actctcgaag acgcccaggc tctgttccac aagtactcg cgctcactct taaagactaa
25441 ggcgcccca cccgaaaaa aggcgggaat tacctcatcg ccaccatgag caaggagatt
25501 cccaccctt acatgtggag ctatcagccc cagatggggc tggccgcggg cgctcccag
25561 gactactcca cccgatgaa ctggctcagt gccggccctc cgatgatctc acgggtcaac
25621 ggggtccgta accatcgaaa ccagatatgt ttggagcagg cggcggtcac atccacgccc
25681 agggcaaagc tcaaccgcg taattggccc tccaccctgg tgtatcagga aatccccggg
25741 ccgactaccg tactacttcc gcgtgacgca ctggccgaag tccgatgac taactcaggt
25801 gtccagctgg ccggcggcgc ttcccgggtg ccgctccgcc cacaatcggg tataaaaacc
25861 ctgatgatcc gaggcagagg cacacagctc aacgacgagt tggtagctc ttcatcggt
25921 ctgcgaccgg acggagtgtt ccaactagcc ggagccggga gatcatctt cactccaac
25981 caggcctacc tgaccttgca gagcagctct tcggagcctc gctccgagg actcggaacc
26041 ctccagtctc tggaggagt ttgtgccctg gtctacttca accccttctc gggatcgcca
26101 ggcctctacc cggacgagtt cataccgaac ttcgacgcag tgagagaagc ggtggacggc
26161 tacgactgaa tgtcccatgg tgactcggct gagctcgctc ggttgaggca tctggaccac
26221 tgccgcccgc tgcgtgtctt cgcccgggag agctgcggac tcatctactt tgagctgccc
26281 gaggagcacc ccaacggccc tgcacacgga gtacggatca ccgtagaggg caccgccgag
26341 tctcacctgg tcaggttctt caccagcaa cccttccctg tccgagcggga cccgggccc
26401 accacctaca ccgtctactg catctgtcct accccaaagt tgcattgagaa tttttgctgt
26461 actctttgtg gtgagtttaa taaaagctga actaagaacc tactttggaa tcccttgctg
26521 tcatcaaatc cacaagacca tcaacttacc ttctgaggaa caggtgaact ttacctgcaa
26581 gccacacaag aagtacgtca cctggtttta ccagaacact actctagcag tagccaacac
26641 ctgtctgaac gacggtgttc ttcttccaaa caatctcacc agtggactaa ctttctcagt
26701 gaaaagggca aagctaattc ttcatcgccc tattgtagaa ggaacttacc agtgtcagag
26761 cggaccttgc ttccacagtt tcacttttgt atactaacac tccataaacc ggcagcagca
26821 agaaactaac cttctttctg atactaacac tccataaacc ggaggtgagc tctgggttcc
26881 ctctctgaca gaggggggta gtcataattg agcggtcggg tatttgattt taggggtggt
26941 cctgggtggg tgcatagcgg tgctatatta ccttcccttg tgggtcgaaa tcaggggtatt
27001 tatctgctgg gtcagacatt gtggggagga accatgaagg ggctcttgct gattatcctt
27061 tccctggttg ggggtttact ggcctgccac gaacagccac gatgtaacat caccacagge
27121 aatgagagga acgactgtct tgtagtgate aaatgcgagc accagtgtcc tctcaacatt
27181 acattcaaga ataagaccat gggaaatgta tgggtgggat tctggcaacc aggagatgag
27241 cagaactaca cggtcactat ccatggtagc gatggaaatc acactttcgg tttcaaattc
27301 atttttgaag tcatgtgtga tatcacactg catgtggcta gacttcatgg cttgtggccc
27361 cctaccaagg agaacatggt tgggttttct ttggcttttg tgatcatggc ctgtgcaatg
27421 tcaggtctgc tggtaggggc tctagtgtgg ttctgaagc gcaagcccag gtacggaaat
27481 gaggagaagg aaaaattgct ataaatcttt ttctcttcgc agaaccatga atactttgac
27541 cagtgtcgtg ctgctctctc ttttagttat taatgtggaa tgtgccgatc ctattctagt
27601 tagtgtagat tggggaaaaa atcttacatt agagggctct aaagaaacac cagttgaatg
27661 gtgggggtga agaaacatac aacaactgtg catagggaat caaaccaaac ataaagagct
27721 aagtcacaga tgtaatgtcc agaacataac tttactgttt gtaaatacta gttttaatgg
27781 agactacttt gggtttaaaa atgataacag cggtatgaaa cattataaag tcacagttat

```

FIG. 1A-8

9/44

27841 accccctaata cctccactc ggaaacctct ttctcctcca cactatgtaa acgcaactat
27901 ggggcaaaac ctaacattag tggggcctgc aaacattcca gttacttggc ttagtgaata
27961 tggcaggttg tgtgagggca aaaaaatttt gcacaaagaa ttaaatcaca cctgtaacga
28021 acagaacctc acgttactgt ttgttaatat gacacacaac ggaccatatt ttggccttga
28081 caaatacaac attgatagag agcagtatga ggtttctatt attagtttgt ttaaagttgg
28141 cgctggacag aagaaaattg ggaaaggaca gaaaaaggag gaaaagacaa aaccaaactc
28201 tagtgatttg ggacaaagac aatccagacc aaagaaaaaa gatattgttg aagagggtcca
28261 aatcaaaaca ggagaaaatc gaacccttgt tgggtccacct ggaaaagttg attggattaa
28321 actttccagt ggaaacaata atgttcttaa gttgtgtaat ggcgacaagt atattaaaca
28381 cacatgtgat ggtcaaaatt taacattaat taatgtgact agaatttatg acggaactta
28441 ttatggttct agcaatgatg gctcaagtca ttacaaagtt accatctatg aattacacaa
28501 agttaataaa actaaatcta tgcttaagcc atacactaca aaaagaacta cagtgaatgc
28561 aacagatgac agtgctcaca aaattgcttt gcagcaggaa aataatgggc aaacagaaaa
28621 tgatcaagaa tcaaaaattc catctgctac tgtggcaatc gtgggtgggag tgattgcggg
28681 cttcataact ataatcattg tcattctgtg ctacatctgc tgccgcaagc gtcccagggc
28741 atacaataat atggtagacc cactactcag cttctcttac tgagactcag tcactttcat
28801 ttcagaacca tgaaggcttt cacagcttgc gttctgatta gcataattac acttagttta
28861 gcagcaccta aaccagaagt atatacaca gttaatgtca ctaggggtgg gaatgctaca
28921 ctatgtggac catttaacaa taacacatgg acaagatata atgatgatgg gagaaaaaac
28981 ggatggatga atatttgtaa atggtcagac ccatcatata catgtcatag taatggaagc
29041 cttagtattt ttgctttcaa cttagtcca ggtaaatata aagttcaaag ttacactaac
29101 agttataatg gattagatgg ttatgaaaaa cttgaagtta aaatgtttta tctaacagta
29161 attgagcctc caaccactag agcaccacc acagttagga caactaagga aacaacacag
29221 cctaccactg taccactac acatccaacc accacagtca gtacaactat tgagaccact
29281 actcatacta cacagctaga cacaacagtg cagaatacta ctttactgat tgaattttta
29341 ctaagaggga atgaaagtac tactgatcag acagaggcta cctcaagtgc ctcagcagt
29401 actgcaaatt taacttcgct tgcttggact aatgaaaccg gagtatcatt gatgcattgg
29461 cagccttact caggtttgga tattcaaatt acttttctgg ttgtctgtgg gatctttatt
29521 cttgtggttc ttctgtactt tgtctgctgc aaagccagag agaaatctag tagggccatc
29581 tacaggccag taatcgggga acctcagcct ctccaagtgg aagggggtct aaggaatctt
29641 ctcttctctt ttctcagtatg gtgatcagcc atgattccta ggttcttctt atttaacatc
29701 ctcttctgtc tcttcaacat ctgcgctgcc ttgcggccg tctcgccacg ctcgcccagc
29761 tgtctcgggc ccttcccac ctacctctc tttgcctgct cactctgcac ctgcgtctgc
29821 agcattgtct gcctggctgt caccttctct cagctcatcg actgggtgctg cgcgcgttac
29881 aattatctcc accacagtcc cgaatacagg gacaagaacg tagccagaat ctttaaggctc
29941 atctgaccat gcagactctg ctgatactgc tatccctcct ctcccctgct cttgctgact
30001 gtaaaatttg ggacatatgg aatttcttag actgttatca agagaaaatg gatatgcctt
30061 cctattactt ggtgattgtg ggtgtagtca tgggtctgct ctgcactttc tttgctatca
30121 tgatctatcc ctgttttgat ctgcgctgga actctgttga ggcattcaca tacacactag
30181 aaagcagttc actagcctcc acgcgcgcc ccacaccgcc tccccgcaga aatcagttcc
30241 ccctgattca gtacttagaa gagccccctc cccggccccc ttccactggt agctactttc
30301 acataaccgg cggcgatgac tgaccaccac ctggacctcg agatggacgg ccaggcctcc
30361 gagcagcgca tcctgcaact gcgcgtccgt cagcagcagg agcggggccg caaggagctc
30421 ctcgatgcca tcaacatcca ccagtgaag aagggcatct tctgcctggt caaacaggca
30481 aagatcacct acgagctcgt gtccaacggc aaacagcatc gcctcaccta tgagatgcc
30541 cagcagaagc agaagttcac ctgcatggtg ggcgtcaacc ccatagtcac caccagcag
30601 tcgggcgaga ccagcggctg catccactgc tccctgcgaaa gccccagtg catctactcc
30661 ctcccaaga ccccttgctg acttcgcgac ctccctccca tgaactgatt gattaaagcc
30721 cagaaaccaa tcaaaccccc tccccatca ccccaaataa caatcattgg aaataatcat
30781 tcaataaaga tcaattactt gaaatctgaa agtatgtctc tgggtgtagt ttgcagcagc
30841 acctcggtac cctcctccca gctctggtac tccagtcccc ggcggggcggc gaacttcctc
30901 cacaccttga aagggatgtc aaattcctgg tccacaattt tcattgtctt ccctctcaga
30961 tgtcaaagag gctccgggtg gaagatgact tcaaccccgt ctacccttat ggctacgcgc
31021 ggaatcagaa tatccccttc ctcaactccc cctttgtctc ctccgatgga ttccaaaact
31081 tcccccttg ggtcctgtca ctcaaaactgg ctgatccaat cgccatcgcc aatgggaatg
31141 tctcactcaa ggtgggaggg ggactcactg tagaacaaca gtctggaata ctgagtgtgg
31201 atactaaggc acccttgcaa gttgcaaatg acaacaaatt ggagctatct tatgatgatc
31261 catttaaggt agagaataac aaacttgga ttaaagctgg ccatgggtta gcagttgtaa

FIG. 1A-9

10/44

```

31321 ctaaagaaaa cacaagtctt cctagtctag ttggaacact tgtagtttta actggaaaag
31381 gaataggtag tggatcaagt gcacatggag gaactattga tgtaagactt ggtgaaggag
31441 gtgggttatc atttgatgaa aaaggagact tagtagcttg ggacaaaaaa aatgatacac
31501 gcaccctttg gacaacacct gatccttctc caaatgcaa agttgaaaca gcaagagact
31561 caaagctaac cttagcactt acaaaatgtg gtagtcaa atttggccact gtatctttac
31621 ttgttgttac gggcaaatat gctattataa gtgacacagt caacccaaag cagttctcta
31681 ttaagttact gtttaatgac aagggtgttt tgtaaagtga ctcaaactct gatgggacat
31741 attggaacta tagaagcaac aataacaaca taggcactcc ttataaagag gctgttggtt
31801 ttatgccaag cacaacagct tatcctaagc caaccaacaa caccagcaca tgcgaaaaca
31861 aaaaagttag tcaaggtaaa aataaaattg taagcaatat ttatcttggg ggagaggtat
31921 atcaaccagg atttattgtt gttaaattta atcaggaaac tgatgccaat tgtgcatact
31981 ctattacatt tgattttgga tggggtaagg tgtataagga tcctatacca tatgatacct
32041 cttctttttac tttctcatat atcgctcaag aatgaaagac caataaacgt gtttttcatt
32101 gaaaattttc atgtatcttt attgattttt acaccagcac gggtagtcag tctcccacca
32161 ccagcccatt tcacagtgtg aacaattctc tcagcacggg tggccttaaa taggggaatg
32221 ttctgattag tgcgggaact ggatttagtg tctataatcc acacagtttc ctggcgagcc
32281 aaacggggat cggtgattga gatgaagccg tcctctgaaa agtcttccaa gcgggcctca
32341 cagtcceaag tcacagtctg gtggaatgag aagaacgcac agattcatal tcggaanaaa
32401 ggatgggttc gtgcctctcc atcagcgccc tcaacagtct ctgccgcggg ggctcgggtg
32461 ggctgctgca gatgggatcg ggatcgcaag tctctctgac tatgatcccc acagccttca
32521 gcatcagtct cctggtgcgt cgggcacagc acctcatcct gatctcgtc atgttctcac
32581 agtaagtgca gcacataatc accatgttat tcagcagccc ataattcagg gtgctccagc
32641 caaagctcat gttggggatg atggaacca cgtgaccatc ataccaaatg cggcagtata
32701 tcaggtgcct gccctcatg aacacactgc ccatatacat gatctctttg ggcattgttc
32761 tgttcacaat ctggcggtac caggggaagc gctggttgaa catgcacccg taaatgactc
32821 tcctgaacca cacggccagc aggggtgcct ccgcccggca ctgcaggag cccggggagc
32881 aacagtggca atgcaggatc cagcgctcgt acccgctcac catctgagct ctcaccaagt
32941 ccagggtagc ggggcacagg cacactgaca tacatctttt taaaattttt atttctctg
33001 gggtaggatc catatcccag gggactggaa actcttgagg cagggtaaa ccagcagcac
33061 atggtaatcc acggacagaa cttacattat gatattcagc atgatacaaa tcgggcagca
33121 gggggtgttg ttcagtcagt gaggccttgg tctcctcctc agatcgtggg aaacgggccc
33181 tgcggtatgg atgatggcgg agcgaggtcg attgttctct ggtgctcatt gtagtgacc
33241 ctcttgcgta ccttgctgta cttctgccag cagaaagtgg cccgggaaca gcagataccc
33301 ctctcctgct cgtcctttcg ctgctgccgc tcagtcatcc aactgaagta catccattcc
33361 cgaaggttct ggagaagttc ctctgcatct gatgaaacaa aaagcccgtc catgcgaatt
33421 cccctcatca catcagccag gactctgtag gccatcccc tccagttaat gctgcttctg
33481 ctatcattca gagggggcgg tggcaggatt ggaagaacca ttattttttt actccaaacg
33541 gtcgcgcagc aatataaaat tcaggtcacg gaggtggcac ctctctcctc cactgttttg
33601 gtggaacag acagccaaat caaaaattat gcgattctca aggtgctcta ctgtggcttc
33661 cagcagaggc tccacacgta catccagaaa caacagcaca ttaaaagcgg gcccgccatc
33721 ctgctcctca atcatcatat tacagtctct aacctcccc aggtaatttt cgtttttcca
33781 gccttgaatt atcgctacaa actgttggtg taagtccact ccacacatgt tgaaaagctc
33841 ccacagcgcc cctccacct tcataagcag gcagaccttc ataatagcaa cagatctggc
33901 tgctccacaa cctgcagcgt gttcaaaaaca aagtcactca agtcctcggc cactacagat
33961 ctgagctcgc gtctcagtg tagctgtaaa aagtcactca agtcctcggc cactacagat
34021 gccaaagtcag agccagggct aagcgttgga ctggcaagcg tgaggagtag tcttagagct
34081 ccataattag ccccaaaaaa ctgcatgctg gaataagctc tctttgtgtc accggtgatt
34141 ccttccaaga ggtgagtgat aaagcgaggt aggtgctctt taatcatagc agtaatggaa
34201 aagtcctcta aataagtcac tagggcccca gggaccacaa tgtggtagct gacaccgcgc
34261 cgctcaagca tgggttagtag agatgagagt ctgaaaaaca gaaagcatgc actaaaccag
34321 ggtggctatt ctactgaag gaaaaatcac tctctccagc agcagggttc ccactgggtg
34381 gccctcgcgg acatacaaaa atcggtccgt gtgggttaaag agcagcacag agagatcatg
34441 tcttctcccg gcaaatatca catcggactg ggttagtatg cccctggcat ggtagtcatt
34501 caaggccata aatctgccct gatatccagt aggaaccagc acgctcactc tgaggtgaag
34561 caatatcacc ccattgcggg gaatgtggaa agattcgggg caaaagaaga tatactatt
34621 gctagcccct tcctggacgg gagcgatccc tccagggcta tctatgaaag catacagaga
34681 ttcagccata gctcagccc cttaccagta gacagagagc acagcgaac cactgacgta
34741 agcagcgact gactaccac tgaccagct ccctatttaa aggcacctta cactgacgta

```

FIG. 1A-10

11/44

```
34801 atgaccaaag gtctaaaaac cccgccaaaa aaaaacacac acgccctggg tgtttttcgc
34861 gaaaacactt ccgcgttgtc acttcctcgt atcgatttcg tgactcaact tccgggttcc
34921 cacgttacgt cacttctgcc cttacatgtg gctccgccgt agggcgccat cttgcccacg
34981 tccaaaatgg cttccatgtc agaccacgcc tccgcgccga ccgttgacgc cgtgcgtcgt
35041 gacgtcagta acggctcttc caacagccaa tcagcgtcgg ccccgcccca aattcaaaag
35101 ctcatttgca tattaacttt tgtttacttt gtgggggtata ttattgatga tg
SEQ ID NO: 1
```

FIG. 1A-11

12/44

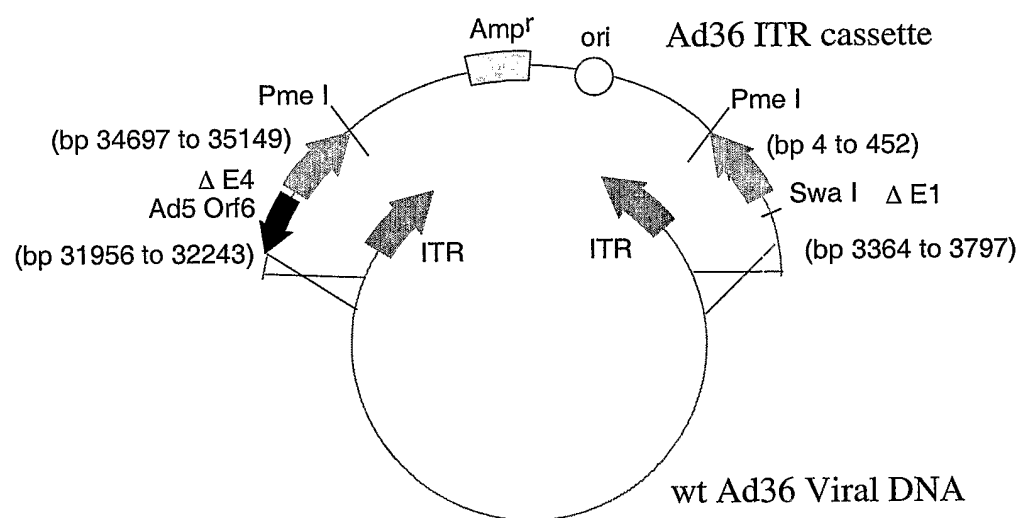


FIG. 2

13/44

```

1 aaaccatcaa taatatatccc cacaaagtaa acaaaagtta atatgcaa at gagcttttga
61 atttggggcg tggccgacgc tgattggacg agaagtgggtg atgcaa atga cgtcacgacg
121 cacggccgac ggtcgccgcg gaggcgtggc ctagtccgga agcaagtcgc ggggctgatg
181 acgtataaaa aagcggactt tagaccgga aacggccgat tttcccgcg tccacgcccgg
241 atatgaggta attctgggcg gatgcaagtg aaattaggcc attttggcgc gaaaactgaa
301 tgagggaagtg aaaagtga aaataccgggc ccgcccaggg cggaatattt accgagggcc
361 gagagacttt gaccgattac gtggggggtt cgattgcggt gtttttttcg cgaatttccg
421 cgtccgtgtc aaagtccggt gtttatgtca cagcgccgc atttaa atgg cgcgcctagg
481 ttttgagtag tgggcgtggc taaggtgagt ataaaggcgg tgtcttacga ggtcctttt
541 gcttttctgc agacatcatg aacgggaccg gcggggcctt cgaagggggg ctttttagcc
601 cttatttgac aaccgcctg ccgggatggg ccggagtctg tcagaatgtg atgggatcta
661 cgggtggatg gcgcccagtg cttccagcaa attcctcgac catgacctac gcgaccgtgg
721 ggagctcgtc gctcgacagc accgcccag ccgcggcagc cgcagccgcc atgacagcga
781 cgagactggc ctcgagctac atgcccagca gcagcagtag cccctctgtg ccagttcca
841 tcatcgccga ggagaaactg ctggccctgc tggcagagct ggaagccctg agccgccagc
901 tggccgcctt gaccagcag gtgtccgagc tccgcgagca acagcagcag caaaataaat
961 gattcaataa acacagattc tgattcaa acgcaaagcat ctttattatt ttttttttcg
1021 cgcgcggtag gccctgggtc acctctccc atcattgaga gtgcggtgga ttttttccag
1081 gaccggtag aggtgggatt ggatgttgag gtacatgggc atgagcccg cccgggggtg
1141 gaggtagcag cactgcatgg cctcgtgtc tggggtcgtg ttgtagatga tccagtcata
1201 gcaggggcgc tgggcgtggt gctggatgat gtccttaagg aggagactga tggccacggg
1261 gagccccttg gtgtaggtgt tggcgaagcg gttgagctgg gagggatgca tgcgggggga
1321 gatgatgtgc agtttggcct ggatccttag gttggcaatg ttgccgccca gatcccgcct
1381 ggggttcatg ttgtgcagga ccaccaggac ggtgtagccc gtgcacttgg ggaacttatc
1441 atgcaacttg gaagggaatg cgtggaagaa tttggagacg cccttgtgcc cgcgccaggt
1501 ttccatgcac tcatccatga tgatggcgat gggcccgtgg gctgcggctt tggcaaagac
1561 gtttctgggg tcagagacat cgtaattatg ctccctgggtg agatcatcat aagacatttt
1621 aatgaatttg gggcggaggg tgcagattg ggggacgata gttccctcgg gcccggggc
1681 aaagtcccc tcacagatct gcatctccca ggctttcatc tcggaggggg ggatcatgtc
1741 cacctgcggg gcgatgaaaa aaacggtttt cggggcgggg gtgatgagct gcgaggagag
1801 caggtttctc aacagctggg acttgccgca cccggtcggg ccgtagatga ccccgatgac
1861 gggttgcagg tggtagttca aggacatgca gctgccgtcg tcccggagga ggggggcccac
1921 ctcgttgagc atgtctctga cttggagggt tcccggagc agctcgccga ggaggcggtc
1981 cccgccagc gagaggagct cttgcaggga agcaaagttt ttcaggggct tgagtccgtc
2041 ggccatgggc atcttggcga gggctctgca gaggagctcc aggcggtccc agagctcgg
2101 gacgtgctct acggcatctc gatccagcag acttccctcg ttcgggggtt gggagcactg
2161 cgactgtagg gcacgagacg atgggcgtcc agcgcggcca gcgtcatgtc cttccagggt
2221 ctcagggtcc gcgtgagtg ggtctccgtc acggtgaaag ggtggggccc gggctgggcg
2281 cttgcaagg tgcgttgag actcatcctg ctggtgctga aacgggcacg gtcttcgccc
2341 tgcgcgtcgg cgagatagca gttgaccatg agctcgtagt tgagggcctc ggcggcgtgg
2401 cccttggcgc ggagcttgcc cttggaagag cggccgagc cgggacagag aagggattgc
2461 agggcgtaga gcttgggtgc gagaaaaacg gactcggggg cgaaggcgtc cgctccgcag
2521 tgggcgcaga cggctctgca ctcgacgagc caggtagact cgggctgctc ggggtcaaaa
2581 accagttttc cccgcttctt tttgatgcgc ttcttacctc ggcgtctccat gagtctgtgt
2641 ccgcgctcgg tgacaaacag gctgtcgggt tccccgtaga cggacttgat gggcctgtcc
2701 tgcagggacg tcccgcggtc ctccctcgt agaaactcgg accactctga gacaaaggcg
2761 cgcgtccacg ccaagacaaa ggaggccacg tgcgaggggt agcgtcgtt gtccaccagg
2821 ggggtccact tttccaccgt gtgcagacac atgtccccct cctccgcac caagaagggt
2881 attgcaattg aggtgtaggc cacgtgaccg ggggtccccg acgggggggt ataaaagggg
2941 ggggtctgt gctcgtctc actctcttcc gcgtcgtgt ccacgagcgc cagctgttgg
3001 ggtaggtatt cctctctag agcgggcatg acctcggcac tcaggttgtc agtttctaga
3061 aacgaggagg atttgatgtt ggcctgcct gccgcgatg tttttaggag actttcatcc
3121 atctggtcag aaaagacaat ttttttattg tcaagcttg atggtctgat tttgtcacg
3181 gcgttgata gaagcttggc gatggatctc atggtctgat tttgtcacg cgcacttcca
3241 tccttggccg cgatgtttag ctggacatac tgcgcgcga cgcacttcca ttcggggaag
3301 acggtggtgc gctcgtcggg cacgatcctg acgcgccagc cgcggttatg caggtgacc
3361 aggtccacgc tgggtggccac ctcgccgcgc aggggctcgt tggccagca gacttgccg
3421 cccttgcgcg agcagaacgg gggcagcaca tcaagcagat gctcgtcagg ggggtccgca

```

FIG. 3A-1

14/44

```

3481 tcgatgggtga agatgcccg agatgcccgg acagagttcc ttgtcaaaat aatcgatttt tgaggatgca
3541 tcatccaagg ccatctgcca ctgcggggcg gccagcgctc gctcgtaggg gttgaggggc
3601 ggaccccagg gcatgggatg cgtgagggcg gaggcgtaca tgccgcagat gtcatacaca
3661 tagatgggct ccgagaggat gccgatgtag gtgggataac agcgccccc gcggatgctg
3721 gcgcgcacgt agtcatacaa ctcgtgcgag ggggccaaaga aggcggggcc gagattgggtg
3781 cgctggggct gctcggcgcg gaagacgatc tggcgaaaga tggcgtgcga gtttgaggag
3841 atgggtggg cc gttggaagat gttaaagtgg gcgtgaggca ggcggaccga gtcgcggatg
3901 aagtgcgcgt aggagtcttg cagcttggcg acgagctcgg cggtagcagag gacgtccatg
3961 gcacagtagt ccagcgtttc gcgatgatg tcataaccg cctctccttt cttctccac
4021 agctcgcggt tgagggcgta ctctcgtca tcttccagt actccggag cggaatcct
4081 cgatcgccg cacgtaaga gccagcatg tagaaatgg tccaggcctt gtagggacag
4141 cagcccttct ccacggggag ggcgtaagct tgagcggcct tgccgagcga ggtgtgcgtc
4201 agggcggaag tgtctctgac catgactttc aagaactgg acttgaaatc cgagtcgtcg
4261 cagccgccgt gctcccagag ctcgaaatcg gtgcgcttct tcgagagggg gttaggcaga
4321 gcgaaagtga cgtcattgaa gagaatcttg cctgcccgcg gcatgaaatt gcgggtgatg
4381 cggaaaggg ccgggacgga ggctcggttg ttgatgacct ggcggcgag gacgatctcg
4441 tcgaagccgt tgatgtttg cccgacgat tagagttcca tgaatcgcg gcggcctttg
4501 atgtgcggga gctttttgag ctctcgtag gtgaggtcct cggggcattg cagtcctgctg
4561 tgctcgagcg ccactcctg gagatgtggg ttggcttgca tgaaggaagc ccagagctcg
4621 cggggccatga gggctctggag ctctcgcga aagagacgga actgctggcc caccggccatc
4681 ttttcgggtg tgacgcagta gaaggtgagg gggctccgct cccagcgatc ccagcgtaag
4741 cgcacggcga gatcgcgagc gagggcgacc agctcgggg ccccgagaa aggtgtaggt ttctacatcg
4801 agcatgaagg ggacgagctg cttgccgaag gacccatcc aggtgtaggt ttctacatcg
4861 taggtgacaa agagccgctc cgtgcgagga tgagagccga ttgggaagaa ctggatttcc
4921 tgccaccagt tggacgagtg gctgttgatg tgatgaaagt agaaatcccg ccggcgaacc
4981 gagcactcgt gctgatgctt gtaaaagcgt ccgcagtact cgcagcgctg cagggctctg
5041 acctcatcca cgagatacac agcgcgtccc ttgaggagga acttcaggag tggcgccctt
5101 ggctggtggt tttcatgttc gcctgcgtgg gactcacct ggggctcctc gaggacggag
5161 aggctgacga gccgcgcgg gagccaggtc cagatctcgg cgcggcgggg gcggagagcg
5221 aagacgaggg cgcgcagttg ggagctgtcc atggtgtcgc ggagatccag gtccgggggc
5281 agggttctga ggttgacctc gtagaagcgg gtgagggcgt gcttgagatg cagatggtag
5341 ttgattttct cgggtgagtt ggtggccgtg tccacgcatt gcatgagccc gtagctgcgc
5401 gggggccacga ccgtgcgcgc gtgcgctttt agaagcggtg tcgcgagcgc gctcccggcg
5461 gcagcgcgcg ttcgggcccc gcgggcaggg cgggcagagg cagtcggcg tggcgctcgg
5521 gcaggtcccc gtgctgcgcc ctgagagcgc tggcgtgcgc gacgacgcgg cggttgacat
5581 cctggatctg ccgcctctgc gtgaagacca cgggccccgt gactttgaac ctgaaagaca
5641 gttcaacaga atcaatctcg gcgtcattga cggcgccctg acgcaggatc tcttgacgt
5701 cgcccgagtt gtcctggtag gcatctcgg acatgaactg ctcgatctcc tctcctgga
5761 gatcgcgcg gccgcgcgc tccacgggtg cggcgaggtc attcgagatg cgaccatta
5821 gctgcgagaa ggcgcccagg ccgctctcgt tccagacgcg gctgtagacc acgtccccgt
5881 cggcgctcgc cgcgcgcatg accacctgcg cgaggttag ctccacgtgc cgcgtgaaga
5941 cggcgtagtt gcgcaggcgc tgggaagagg agttgagggt ggtggcgatg tgctcggtag
6001 cgaagaagta catgatccag cggcgaggg gcatctcgct gatgtcgccg atggcctcca
6061 gcctttccat ggcctcgtag aaatccacgg cgaagttgaa aaactgggcg ttgcgggccg
6121 agaccgtgag ctgctcttcc aggagccgga tgagctcggc gatggtggcg cgcacctcgc
6181 gctcgaaatc cccggggacc tctctcttct cctcttcttc catgacgacc tcttcttcta
6241 tttcttcttc tgggggcggg ggtggtggcg gggcccgacg acgacggcga cgcaccggga
6301 gacggtcgac gaagcgctcg atcatctccc gcggcgggcg acgcatgggt tgggtgacgg
6361 cgcgaccccg ttcgcgagga cgcagcgtga agacgcggcc ggtcatctcc cggtaatggg
6421 gcgggtcccc gttgggcagc gatagggcgc tgacgatgca tcttatcaat tgcggttag
6481 gggacgtgag cgcgtcgaga tcgaccggat cggagaatct ttcgaggaat gcgcttagcc
6541 aatcgagtc gcaaggtaag ctcaaacacg tagcagccct gtggacgctg ttagaattgc
6601 ggttgctgat gatgtaattg aagtaggcgt ttttaaggcg gcggtgggtg gcgaggagga
6661 ccaggtcctt gggctccgct tgctggatgc ggagccgctc ggccatgccc caggcctggc
6721 cctgacaccg gctcaggttc ttgtagtagt catgcatgag cctttcaatg tcatcactgg
6781 cggaggcgga gtcttccatg cgggtgacct cgacgcccct gagcggtgag acgagccca
6841 ggtcggcgac gacgcgctcg gcgaggatgg cctgtgcac gcggtgaggt gtgacctgga
6901 agtcatccat gtcgacgaag cgggtgtagg ccccggtgtt gatggttag gtgcagttgg

```

FIG. 3A-2

```

6961 ccatgagcga ccagttgacg gtctgcaggg cgggttgccac gacctcggag tacctgagcc
7021 gcgagaaggg gcgcgagtcg aagacgtagt cgttgccaggt gcgcacgagg tactggttagc
7081 cgactaggaa gtgcggcggc ggctggcggg agagcggcca gcgctgggtg gccggcgcg
7141 ccggggccag gtcctcgagc atgaggcggg ggtagccgta gaggtagcgg gacatccagg
7201 tgatgccggc ggcgggtggtg gaggcgcgcg ggaactcgcg gacgcgggtt cagatgttgc
7261 gcagcggcag gaaataatcc atggctcggc cggctctggc ggtgagacgc gcgcagtcac
7321 tgacgctcta gaggcaaaaa cgaaagcggg tgagcgggct cttcctccgt agcctggcgg
7381 aacgcaaacg ggtagggccg cgtgtgtacc ccggttcgag tcccctcgaa tcaggctgga
7441 gccgcgacta acgtggtatt ggcactcccg tctcgacccg agcccgatag ccgccaggat
7501 acggcggaga gccctttttg ccgaccgagt ggggtcgcga gacttgaaag cggccgaaaa
7561 cccgcgcggg tagtggctcg cgcccgtagt ctggagaagc atcgccaggg ttgagtcgcg
7621 gcagaacccg gttcgcggac ggccgcggcg agcgggactt ggtcaccgcc ccgatttaaa
7681 gaccacagc cagccgactt ctccagttac gggagcgcgc cccctttttt ctttttgcca
7741 gatgcatccc gtcctgcgcc aaatgcgtcc ccccccccg gcgaccaccg cgaccgcggc
7801 cgtaacaggg gccggcgcta gccagccaca gacagagatg gacttggaag agggcgagg
7861 gctggcgaga ctggggggcg cgtccccgga gcgacacccc cgcgtgcagc tgcagaagga
7921 cgtgcgcccg gcgtacgtgc ctgcgcagaa cctgttcagg gaccgcagcg gggaggagcc
7981 cgaggagatg cgcgactgcc ggtttcggc gggcagggag ctgcgcagcg gcctggaccg
8041 ccagcgcgtg ctgcgcgacg aggatttcga gccgaacgag cagacgggga tcagccccgc
8101 gcgcgcgcac gtggcggcgg ccaacctggt gacggcctac gagcagacgg tgaagcagga
8161 gcgcaacttc caaaagagtt tcaacaacca cgtgcgcacg ctgatagcgc gcgaggaggt
8221 ggccctgggc ctgatgcacc tgtgggacct ggcggaggcc atcgtgcaga acccggacag
8281 caagcctctg acggcgcagc tgttccctgt ggtgcagcac agcagggaca acgaggcgtt
8341 cagggaggcg ctgctgaaca tcgccgagcc cgaggggtcg tggtgctgg agctgatcaa
8401 catcttgtag agcatcgtag tgcaggagcg cagcctgagc ctggccgaga aggtggcggc
8461 gatcaactac tcggtgctga gcctgggcaa gttttacgcg cgcaagattt acaagacgcc
8521 gtacgtgccc atagacaagg aggtgaagat agacagcttt tacatgcga gaccgcatcc
8581 ggtgtgtagc ctgagcgacg acctgggcgt gtatcgcaac gaccgcatcc acaaggccgt
8641 gagcacgagc cggcggcgcg agctgagcga ccgcgagctg atgctgagcc tgcgccgggc
8701 gctggtaggg ggcgcgcgcg gcggcgagga gtcctacttc gacatggggg cggacctgca
8761 ttggcagccg agccggcgcg ccttgagggc cgcctacggt ccagaggact tggatgagga
8821 tgaggaagag gaggaggatg cacccgttgc ggggtactga cgcctccgtg atgtgtttt
8881 agatgtccca gcaagccccg gaccccgcca taaggcgggc gctgcaaagc cagccgtccg
8941 gtctagcatc ggacgactgg gagggcgcg gcaaacgcat catggccctg acgacccgca
9001 accccgagtc ctttagacaa cagccgcagg ccaacagact ctcgccattt ctggaggcgg
9061 tgggtccctc tcggaccaac cccacgcagc agaagggtgt ggcgatcgtg aacgcgtgg
9121 cgggagaaca ggccatccgt cccgacgagg ccgggctggg gtacaacgcc ctctggagc
9181 gcgtgggccc ctacaacagc acgaacgtgc agtccaacct ggaccggctg gtgacggacg
9241 tgcgcgaggg cgtggcgtag acgcagccgg cgaacgtgcc ggcgagggcag gacgattaca
9301 cgctgaacgc cttcctggcg acgcagccgg tgaccgaggt tcccagagc gaggtgtacc
9361 ccaactttat cagcgcgctg cggctgatgg taccagagaa gccggcaggg cctgcagacg
9421 agtcggggcc ggactacttt ttccagacaa ggggcgtgtg ggggcctgca ggcgcccgtg
9481 gtcaggcttt caagaacctg acgcccactt cgcggctgtg gctgctgctg atcgcgccct
9541 cgacggttag cagcttgctg acgcccactt cgtacctggg ccacctgctg acgctgtacc
9601 tcaccgacag cggcagcgtg caggtggacg agcagacctt ccaggagatc acgagcgtga
9661 gcgaggccat agggcaggcg ggggcagaac gcacccgaca gtctgagggc caccctgaac
9721 gccgcgcgct ggggcagaac cccgcgagat cccgcgaggt ggccgaggag gaaaggatcc
9781 caaatagaca gcagaagatc cccgcgaggt gtagggctgt tcctgatgca ggagggtgcc
9841 tgagatatgt gcagcagagc gtagggctgt cgcaacatgg aacctagcat gtacgccgcc
9901 ccgcgctgga catgaccgcg gctgatggac tacctgcacc gcgcggcgcc catgaacacg
9961 tcatcaataa attgaacccg cactggcttc cgccgcgggg gttctacacg ggcgagtagc
10021 caaacgccat ccccaacgac ggggttcctg gggacgagct ggacagcgcg gtgttctcgc
10081 acatgcccga ccccaacgac gaggcgcgcg catagcttgc cgggctcggg gaacagcgcc
10141 cgacctttca aaagcgccag agggagtttg gacgagtagc tgaacgactc gctgtcgcag
10201 ttcctagctt agggagtttg gctgggagag gacgagtagc aacgggtagg ggacaaactg
10261 ggccgcgctt catggccaat aacgggtagg agagtctggt aaccgctgga aaccgctgga
10321 tcaagaacgc catggccaat aacgggtagg agagtctggt aaccgctgga aaccgctgga
10381 agacctacgc tcaggacccat agggagcctg cgccgcgcgc gcggcgacag cgccacgacc

```

FIG. 3A-3

16/44

```

10441 ggcagcgggg cctggtgtgg gacgacgagg actcggccga cgatagcagc gtgttggact
10501 tgggcgggag cgggtggggcc aaccgcgttc cgcactctga gccagactg gggcgccgga
10561 tgttttgaat gcaaaataaa actcaccaag gccatagcgt gcgttctctt ccttgttaga
10621 gatgaggcgt gcggtggtgt cttcctctcc tcctccctcg tacgagagcg tgatggcgca
10681 ggcgaccctg gaggttccgt ttgtgcctcc gcggtatatg gtccttacgg agggcagaaa
10741 cagcattcgt tactcggagc tggctccgca gtacgacacc actcgcgtgt acttggtgga
10801 caacaagtcg gcggacatcg cttccctgaa ctacaaaac gaccacagca acttccctgac
10861 cacggtggtg cagaacaacg atttcacccc cgccgaggcc agcacgcaga cgataaattt
10921 tgacgagcgg tcgcggtggg gcggtgatct gaagaccatt ctgcacacca acatgcccaa
10981 tgtgaacgag tccatgttca ccagcaagtt taaggcgcgg gtgatggtgg ctagaaagca
11041 tcccaaagat gtagatgcca gtgatttaag caaggatata ttagagtata agtggtttga
11101 gtttaccctg cccgagggca acttttccga gaccatgacc atagacctga tgaacaacgc
11161 catcttgga aactacttgc aagtggggcg gcagaatggc gtgctggaga gcgatatcgg
11221 agtcaagttt gacagcagga atttcagact gggctgggac ccggtgacca agctggtgat
11281 gccaggggtc tacacctacg aggccttcca cccgacgtg gtgctactgc cgggctgcgg
11341 ggtggacttc accgagagcc gcctgagcaa cctcctgggc attcgcaaga agcaaccttt
11401 tcaagagggc ttcagaatca tgtatgagga tctagaaggg ggtaacatcc ccgtctcctt
11461 ggataccaaa aaatatctgg atagcaagaa ggaacttgag gatgctgcca aggaagctgc
11521 aaagcaacag ggagatggtg ctgtcactag aggcgatacc cacctcactg tagctcaaga
11581 aaaagcagct gaaaaggagc tagtgatcgt accaattgaa aaggatgaga gcaacagaag
11641 ttacaacctg atcaaggaca cccatgacac cctgtaccga agctggtacc tgcctatac
11701 ctacggggac cccgagaagg ggggtgcagtc gtggacgctg ctcaccaccc cggacgtcac
11761 ctgcgggcgc gagcaagtct actggtcgct gccggacctc atgcaagacc ccgtcacctt
11821 ccgctctacc cagcaagtca gcaactaccc cgtggtcggc gccgagctca tgccttccg
11881 cgccaagagc ttttacaacg acctcgccgt ctactcccag ctcacccgca gctacacctc
11941 cctcaccac gtcttcaacc gcttccccga caaccagatc ctctgcgcgc cgccgcgcgc
12001 caccatcacc accgtcagtg aaaacgtgcc tgctctcaca gatcacggga cgcttccgct
12061 gcgcagcagt atccgcggag tccagcgagt gaccgtcact gacgcccgtc gccgcacctg
12121 tccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtgctctcca gtcgcacctt
12181 ctaaaaaatg tctattctca tctcgcccag caataacacc ggctggggtc ttactaggcc
12241 cagcaccatg tacggaggag ccaagaagcg ctcccagcag caccctgccc gcgtccgcgg
12301 tcacttccgc gctccctggg gagcttacia gcgggggcgc actgccaccg ccgcgcgcgt
12361 gcgcaccacc gtcgacgacg tcatcgactc ggtggtcgcc gacgcgcgca actacacccc
12421 cgccccctcc accgtggacg cggtcacatga cagcgtggtg gccgacgcgc gcgactatgc
12481 cagacgcaag agccggcggc gacggatcgc caggcgccac cggagcacgc ccgccatgcy
12541 cgccgcccgc gctctgctgc gccgcgccag acgcacgggc cgccgggcca tgatgcgagc
12601 cgcgcgccgc gccgccactg caccctccgc aggcaggact cgacagcag cgccgcgcgc
12661 cgctgcgcgc gccatctcta acccaggcgc ggaaacgtgt actgggtgcy
12721 cgactccgtc acgggcgtgc gcgtgccctg gcgcacctgt cctcctcgtc cctgatctaa
12781 tgcttgtgtc ctcccccgca agcgacgatg tcaaagcgca aaatcaagga ggagatgctc
12841 caggtcgctc ccccgagat ttacggaccc cgggaccaga aaccctcgaa aatcaagcgg
12901 gttaaaaaaa aggatgaggt ggacgagggg gcagtagagt ttgtgcgcga gttcgctccg
12961 cggcggcgcy taaattggaa ggggcgcagg gtgcagcgcy gtttgcgccc cggcacggcy
13021 gtggtgttca cgcccggcga gcggtcctcg gtcaggagca agcgtagcta tgacgaggtg
13081 tacggcgacg acgacatcct ggaccaggcy gcggagcggy cgggcgagtt cgcctacggg
13141 aagcggtcgc gcgaagagga gctgatctcg ctgcccaggy acgaaagcaa cccacgcgcg
13201 agcctgaagc ccgtgacctt gcagcaggtg ctgcccaggy cgggtgctgt cccgagccgc
13261 ggggtcaagc gcgagggcga gagcatgtac ccgacctgcy agatcatggt gcccaagcgc
13321 cggcgcgtgy aggcgtgctt ggacacctgt aaaatggatg tggagcccga ggtcaaggtg
13381 cgccccatca agcaggtggc gccgggcttg ggcgtgcaga ccgtggacat tcagatcccc
13441 accgacatgy atgtcgacaa aaaacctcg accagcatcy aggtgcagac cgacccttgy
13501 ctccagcct ccaccgctac cgtctccact ttaccgccc ccacggctac cgagcctccc
13561 aggagggcga gatggggcgc cgccagccgy ctgatgccc actacgtgtt gcaccttccc
13621 atcatcccga cgccgggcta ccgcggcacc cgggtactac ccagccgcag gcgcccagcc
13681 gccaaacgcy gccggcgcac tgccacctgc cgccgtctgy ccccgcccg cgtgcgcgcg
13741 gtaaccacgc gccggggcgc ctgcctcgtt ctgcccacgy tgcgtatac cccagcatc
13801 ctttaatccg tgtgctgtga tactgttgca gagagatggc tctcacttgc ccctgcgca
13861 tccccgtccc gaattaccga ggaagatccc gccgcaggag aggcattggca ggcagcgccc

```

FIG. 3A-4

17/44

```

13921 tgaaccgcgcg ccggcggcgcg gccatgcgca ggcgcctgag tggcggggttt ctgcccgcgc
13981 tcatcccat aatcgccgcg gccatcggca cgatcccggg catagcttcc gttgcgctgc
14041 aggcgtcgca gcgcggttga tgtgcgaata aagcctcttt agactctgac acacctggtc
14101 ctgtatatatt ttagaatgga agacatcaat tttgcgtccc tggctccgcg gcacggcacg
14161 cggccgttca tgggcacctg gaacgagatc ggcaccagcc agctgaacgg gggcgcttcc
14221 aattggagca gtgtctggag cgggcttaaa aatttcgggt cgacgctccg gacctatggg
14281 aacaaggcct ggaatagtag cacggggcag ttgttaaggg aaaagctcaa agaccagaac
14341 ttccagcaga aggtggtgga cgggctggcc tcgggcatta acggggtggg ggacatcgcg
14401 aaccaggccg tgcagcgca gataaacagc cgcttgacc cgcgccgcc cacggtggtg
14461 gagatggaag atgcaactct tccgcgcgcc aaaggcgaga agcgccgcg gccgcagcg
14521 gagagacga tctgaggt ggacgagcgc cctcgtacg aggaggcgt caaggccggc
14581 atgcccacca cgcgcacat cgcgcgcgtg gccacgggtg taatgaaacc cgccacctt
14641 gacctgcctc caccaccac gccgcctcca ccgaaggcag ctccggttgt gcaggccct
14701 ccggtggcga ccgcgtgcg ccgcgtcccc gccgcgcgcc aggccagaa ctggcagagc
14761 acgctgcaca gtatcgtggg cctgggagtg aaaagtctga agcgccgcg atgctattga
14821 gagaggaaag aggacactaa agggagagct taacttgtat gtgccttacc gccagagaac
14881 gcgcgaagat ggccaccccc tcgatgatgc cgcagtgggc gtacatgcac atcgccgggc
14941 aggacgcctc ggagtacctg agcccggttc tggtgcagtt tgcccgccc accgacacgt
15001 acttcagcct gggcaacaag tttaggaaac ccacggtggc cccgacccat gatgtgacca
15061 cggaccggtc ccagcgtctg acgctgcgct tcgtgcccg ggatcgcgag gacaccacgt
15121 actcgtacaa ggcgcgcttc actctggccg tgggcgacaa ccgggtgcta gacatggcca
15181 gcacgtactt tgacatccgc ggcgtcctgg acccggttcc cagcttcaaa ccctactcgg
15241 gcacggctta caacagtttg gcccccaagg gcgcccccaa ctccagtcag tggactgaca
15301 aagaacggca aaatggtgga caaccacca ctacaaaaga tgttacaaaa acattcggag
15361 tagcagccag gggagggctt catattactg ataaaggact acaaatagga gaagatgaaa
15421 ataacgagga tgggtgaagaa gagatatatg cagacaaaac tttccagcca gaacctcaag
15481 taggagagga aaactggcaa gatactgatg ttttctatgg cggcagagcg cttaaaaagg
15541 aaacaaaaat gaaacatgc tatggctctt ttgcccagacc taccaatgaa aaaggagtc
15601 aagctaaatt tttaaatggc gaaaacggtc aaccttctaa agatcaagat attacattag
15661 ctttctttga tcttaaacaa aatgacactg gaactactca aaaccagcca gatgttgtca
15721 tgtacactga aaatgtgtat ctggaaaccc cagacacca tgtggtgtac aaacctggca
15781 aggaagatac aagctccgct gctaacctta cacaacagtc catgcccaac agggccaact
15841 acattggttt cagggacaac tttgtggggc tcagtgtatta caacagcact ggcaacatgg
15901 gtgtgctggc tggtcaggcc tctcagttga atgctgtggt tgacttgcaa gacagaaaca
15961 ccgagctgtc ttatcagctc ttgctagatt ctctgggtga cagaaccaga tacttttagc
16021 tgtggaattc tgcggtggac agctatgatc cagatgtcag gatcattgag aatcacggtg
16081 ttgaagatga gcttccaaat tattgcttcc cactggatgg atctggcagc aataccgcat
16141 atcaagggtg taaatatgaa aacggagctg gcaatggaag ctggaaagta gatggcgaag
16201 ttgcttctca gaatcagatc gccaaaggta atctgtatgc catggagata aaccttcagg
16261 ccaacctgtg gaagagtttt ctgtactcga acgtggcgct gtatctacca gactcctaca
16321 agtacacgcc ggccaacatc acgctgcccc ccaacaccaa cacctacgag tacatgaacg
16381 gccgcgtggg ggcaccctcg ctggtggatg cctatgtcaa catcggtgcc cgctggtcgc
16441 tggaccccat ggacaacgct aaccccttca accaccaccg caacgcgggt ctgcgctacc
16501 gctccatgct tctgggcaac ggccgctacg tgcccttcca catccaagtg ccccaaaagt
16561 tctttgccat caagaacctg ctctgcttc ccggttccca cacctacgag tggaaacttc
16621 gcaaggatgt caacatgatc ctgcaaagtt cctctggcaa cgacctgcgc gtcgacggcg
16681 cctccgtccg cttcgacagc gtcaacctct atgccacctt cttcccatg ggcacaaca
16741 ccgcctccac ccttgaagcc atgctgcgca acgacaccaa cgaccagtcc ttcaacgact
16801 acctctcggc cgccaacatg ctctacccaa tcccggccaa ggccaccaac gtgcccactt
16861 ccatccctc gcgcaactgg gccgccttcc ggggctggag tttcaccggg ctcaagacca
16921 aggaaactcc ctccctcggc tcgggtttcg acccctactt tgtctactcg ggctccattc
16981 cctacctcga cggaaccttc tactcaacc acaccttcaa gaaggtctcc atcatgttcg
17041 actcctcggc cagctggccc ggcaacgacc ggetgctcac gccgaacgag ttcgagatca
17101 agcgcagcgt cgacggggag ggctacaacg tggcccaatg caacatgact aaggactggg
17161 tcctcgtcca gatgctctct cattacaaca ttggctacca gggcttctac gtgcctgagg
17221 gttacaagga ccgcatgtac tccttcttcc gcaacttcca gcccatgagc aggcaggtgg
17281 tcgatgagat caactacaag gactacaagg ccgtcaccct gcccttccag cacaacaact
17341 cgggcttcac cggctacctc gcaccacca tgcgtcaggg gcagccatac cccgccaact

```

FIG. 3A-5

18/44

```

17401 tccccctaccc gctcatcggc cagacagccg tgcctctcgt caccagaaa aagttcctct
17461 gcgacagggg catgtggcgc atccccctt ccaagcaact catgtccatg ggcgcctca
17521 ccgacctggg tcagaacatg ctctacgcca actcggccca cgcgctcgac atgaccttcg
17581 aggtggaccc catggatgag cccaccctcc tctatcttct ctcgaagtt ttcgacgtgg
17641 tcagagtgc cagccgcac cgcggcgta tcgaggccgt ctacctgcgc acgccttct
17701 ccgccgaaa cgcaccaca taagcatgag cggctccagc gaacgagagc tcgcgcccat
17761 cgtgcgcgac ctgggctgcg ggccctactt tttgggaacc cagcacaagc gcttccctgg
17821 ctctctcgcc ggcgacaagc tggcctgcgc catcgtcaac acggccggcc gcgagaccgg
17881 aggcgtgcac tggctcgctt tcggtggaa cccgctcg cgcacctgct acatgttcga
17941 cccctttggg ttctcgacc gccggtcaa gcgatttac agcttcgagt acgaggccat
18001 gctgcgcgc agcgccctgg cctcctcgcc cgaccgtgt ctcagcctcg agcagtcac
18061 ccagaccgtg caggggcccg actccgcgc ctgaggactt ttctgttgca tgttcttgca
18121 tgccttcgtg cactggcccg accgacctat ggacgggaac cccacctga acttgctgac
18181 ggggggtgcc aacggcatgc tacaatcgcc acaggtgctg cccaccctcc ggcgcaacca
18241 ggaggagctc taccgcttcc tcgcgcgcca ctcccttac ttcgatccc accgcgcgc
18301 catcgaacac gccaccgctt ttgacaaaat gaaacaactg cgtgtatctc aataaacagc
18361 acttttattt tacatgcaact ggagtatatg caagttattt aaaagtcgaa ggggttctcg
18421 cgctcgtcgt tgtgcgcgc gctggggagg gccacgttgc ggtactggtta cttgggaagc
18481 cacttgaact cggggatcac cagtttggga accggaatct cggggaaggt ctcgctccac
18541 atgcgcggc tcactcgag gcgcgcagc atgtccgggg cggagatctt gaaatcgag
18601 ttgggaccgg tgctctgcgc gcgcgagttg cgttacacgg ggttgacgca ctggaacacc
18661 atcagactgg ggtacttcac actggccagc acgctcttgt cgggtgatctg atccttgtca
18721 aggtcctcgg cgttgctcag gccaaacggg gtcactctgc acagctggcg gccaggaag
18781 ggcacgctct gaggttgtg gttacactcg cagtgaatgg gcattagcat catccccgcg
18841 ccgcgctgca tattcgggta gagggccttg acaaaggccg agatctgttt gaaagcttgc
18901 tgggccttgg ctccctcgct gaaaaacagc ccgcagctct tcccgtgaa ctggttattc
18961 ccgcaaccgg catcctgcac gcagcagcgc gcgtcatggc tggtcagttg caccacgctc
19021 cgtccccagc ggttctgggt cactttggcc ttgctgggt gctccttcag cgcgcgtgt
19081 ccgttctcgc tggtcacatg catctccacc acgtggtctt tgtggatcat caccgttcca
19141 tgcagacact tgagctggcc ttccacctcg gtgcagccgt gatcccacag ggcgcattccg
19201 gtgcaactcc agttcttatg cgcgatcccg ctgtggctga agatgtaacc ttgcaacagg
19261 cgacccatga cgggtgctaaa ggctttcttg gtggtgaagg tcagttgcag accgcgggcc
19321 tcctcgttca tccaggtctg gcacatcttc tggaagatct cggctctgctc tggcatgagc
19381 ttgtaagcat cgcgcaggcc gctgtcgacg cggtagcggt ccatcagcac gttcatggta
19441 tccatgccct tctcccagga cgagaccaga ggcagactca ggggggttgcg cacgttcagg
19501 acaccggggg tcgcgggctc gacgatgcgt tttccgtcct tgccttctt caacagaacc
19561 ggaggctggc tgaatcccac tcccacgatc acggcgtctt cctggggcat ccttctgtct
19621 gggctctac tttgtcacatg cttggtcttt ctggcttctt tcttttttgg agggctgtcc
19681 acggggacca cgtcctcctc cgaagaccgg gagccacccc gctgatactt tggcgcttg
19741 gtgggcagag gaggtggcgg cgaggggctc ctctcctgct ccggcggata gcgcgcgac
19801 ccgtggcccc ggggcggagt ggctctcgc tccatgaacc ggcgcacgct ctgactgccg
19861 ccggccattg tttcctaggg gaagatggag gagcagccgc gtaagcagga gcaggaggag
19921 gacttaacca cccacgagca acccaaaatc gagcaggacc tgggcttcga agagccggct
19981 cgtctagaac cccacagga tgaacaggag cacgagcaag acgcaggcca ggaggagacc
20041 gacgctgggc tcgagcatgg ctacctggga ggagaggagg atgtgctgct gaaacacctg
20101 cagegccagt cctcatcct ccgggacgcc ctggccgacc ggagcgaac cccctcagc
20161 gtcgaggagc tgtgtcgggc ctacgagctc aacctcttct cgcgcgctg gcccccaca
20221 cgccagccca accgcacctg cgagcccaac ccgctctca acttctatcc cgtctttgcg
20281 gtccccgagg cccttgccac ctatcacatc tttttcaaga accaaaagat ccccgctctc
20341 tgccgcgcca accgcaccgg cgccgacgcg ctctcctgctc tggggcccg ggcgcgcata
20401 cctgatatcg cttccctgga agaggtgccc aagatcttcg aagggtcag tccggacgag
20461 acgcgcgcgg cgaacgctct gaaagaaaca gcagaggaag agggtcacac tagcgcctg
20521 gtagagtgg aaggcgacaa cgccaggctg gtcgtgctca agcgcagcgt cgagctcacc
20581 cacttcgcct accccgcggt taacctccc cccaagggtc tgctgcgcat catggatcag
20641 cttatcatgc cccacatcga ggccatcgat gagaccaag agcagcgcgc cgaggacgcc
20701 cggcccgtgg tcagcgacga gatgctcgcg cgctggctcg ggaccgcga ccccgaggt
20761 ttggaacagc ggcgcaagct gatgctggcc gtagtcctgg tcacctcga gctcgaatgc
20821 atgcgcgct tcttctgcga ccccgagacc ctgcgcaagg tcgaggagac cctgcactac

```

FIG. 3A-6

20881 acttttcagac acgggtttcgt caggcaagcc tgcaagatct ccaacgtgga gctgaccaac
20941 ctgggtctcct gcctgggggat cctgcatgag aaccgcctgg ggcagacagt gctccactct
21001 accctgaagg gcgaggcacg gcgggactat gtccgcgact gcgtctttct ctttctatgc
21061 cacacatggc aagcagccat gggcggtgtg cagcagtgct tgcaggacga gaacctgaag
21121 gagctggaca agcttcttgc tagaaacctt aaaaagtgtg ggacgggctt cgacgagcgc
21181 accgtcgcct cggacctggc cgagatcggt tccccgagc gcctgaggca tacgtgaaa
21241 ggccgggctgc ccgacttcat gagccagagc atgttgcaaa actaccgcac tttcattctc
21301 gagcgctcgg gtatcctgcc cgccacctgc aacgccttcc cctccgactt tgtcccgtg
21361 agctaccgcg agtgtcccc gccgctgtgg agccactgct acctcttgca gctggctaac
21421 tacatctcct accactcggg cgtgatcgag gacgtgagcg gcgaggggct gctcgagtgc
21481 cactgccgct gcaacctgtg cccccgcac cgctccctgg tctgcaacct ccagctcctg
21541 agcgagaccc aggtcatcgg taccttcgag ctgcaaggct cggagaagtc caccgctccg
21601 ctgaaactca cgccgggggt gtggacttcc gcgtacctgc gcaaatttgt acccgaagac
21661 taccacgccc atgagataaa gttcttcgag gaccaatcgc gtccgcagca cgcggatctc
21721 acggcctgcg tcatcaccca gggcgcgatc ctgcaccaat tgcatgccat ccaaaaatcc
21781 cgccaagagt ttcttctgaa aaagggtaga ggggtctacc tggaccccc gacgggagag
21841 gtgctcaacc cgggtctccc ccagcatgcc gaggaagaag caggagcgcg tagtggagga
21901 gatggaagaa gaatgggaca gccaggcaga ggaggacgaa tgggaggagg agacagagga
21961 ggaagaattg gaagagggtg aagaggagca ggcaacagag cagcccgctg ccgcaccatc
22021 cgcgccggca gccccggcgg tcacggatac aacctccgct ccggtcaagc ctctctgtag
22081 atgggatcga gtgaagggtg acggtaagca cgagcggcag ggctaccgat catggagggc
22141 ccacaaagcc gcgatcatcg cctgcttgca agactgcggg gggaacatcg ctttcgcccc
22201 ccgctacctg ctcttccacc gcgggggtgaa catccccgc aacgtgttgc attactaccg
22261 tcaccttcac agctaagaaa aagcaagtaa gaggagtgcg cggaggaggc ctgaggatcg
22321 cgcgcaacga gccctcgacc accagggagc tgaggaaacc gatcttcccc actctttatg
22381 ccatttttca gcagagtcca ggtcagcagc aagagctcaa agtaaaaaat cggctctctg
22441 gctcgtcac ccgcagttgc ttgtaccaca aaaacgaaga tcagctgcag cgcactctcg
22501 aagacgccga ggctctgttc cacaagtact gcgcgctcac tcttaaagac taaggcgcg
22561 ccaccggaa aaaaggcggg aattacctca tgcaccat gagcaaggag attcccaccc
22621 cttacatgtg gagctatcag cccagatgg gcctggcgcg gggcgctcc caggactact
22681 ccaccgcgat gaactggctc agtgccggcc cctcgatgat ctacgggtc aacgggggtc
22741 gtaaccatcg aaaccagata ttgttgagc aggcggcggt cacatccacg cccagggcaa
22801 agctcaaccc gcgtaattgg ccctccaccc tgggtgatca ggaaatcccc gggccgacta
22861 ccgtactact tccgcgtgac gactggccg tgcgcgtcc gccacaatc ggggtataaaa accctgatga
22921 tggccggcgg cgcttccccg tgcgcgtcc gccacaatc ggggtataaaa accctgatga
22981 tccgaggcag aggcacacag ctcaacgacg agttgggtgag ctcttcgatc ggtctctgac
23041 cggacggagt gttccaacta gccggagcgc ggagatcatc cttactccc aaccaggcct
23101 acctgacctt gcagagcagc tcttcgggac ctgcgtccgg aggcacgga accctccagt
23161 tcgtggagga gtttgtgccc tgggtctact tcaaccctt ctccgggatc ccaggcctct
23221 acccgagcga gttcataccg aacttcgacg cagttagaga agcgggtggac ggctacgact
23281 gaatgtccca tgggtgactc gctgagctcg ctccggttag gcatctggac cactgccgcc
23341 gcctgcgctg cttcgcccgg gagagctgcg gactcatcta ctttgagctg cccgaggagc
23401 accccaacgg ccctgcacac ggagtacgga tcaccgtaga gggcaccgcc gagtctcacc
23461 tggtcagggt cttcacccag caacccttcc tggtcgagcg ggaccggggc gccaccactt
23521 acacggtcta ctgcatctgt cctaccccaa agttgcatga gaatttttgc tgtactcttt
23581 gtgggtgagtt taataaaagc tgaactaaga acctactttg gaatcccttg tcgtcatcaa
23641 atccacaaga ccatcaactt cacctttgag gaacaggtga actttacctg caagccacac
23701 aagaagtacg tcacctgggt ttaccagaac actactctag cagtagccaa cacctgctcg
23761 aacgacgggt ttcttcttcc aaacaatctc accagtggac taactttctc agtgaaaagg
23821 gcaaagctaa ttcttcatcg ccctattgta gaaggaaact accagtgtca gagcggacct
23881 tgcttccaca gtttctactt ggtgaacgtt accggcagca gcacagtcgc tccagaaact
23941 aaccttcttt ctgatactaa cactcctaaa accggagggt agctctgggt tccctctctg
24001 acagaggggg gtagtcatat tgaagcggtc ggggtatttg ttttaggggt ggtcctgggt
24061 ggggtgcatag cgggtgctata ttaccttctt tgctgggtcg aaatcagggt atttatctgc
24121 tgggtcagac attgtgggga ggaaccatga aggggctctt gctgattatc ctttccctgg
24181 ttgggggttt actggcctgc cacgaacagc cacgatgtaa catcaccaca ggcaatgaga
24241 ggaacgactg ctctgtagtg atcaaatgcg agcaccagtg tctctcaaac attacattca
24301 agaataagac catgggaaat gtatgggtgg gattctggca accaggagat gagcagaact

FIG. 3A-7

```

24361 acacgggtcac tatccatggt agcgatggaa atcacacttt cggtttcaaa ttcattttttg
24421 aagtcattgtg tgatatcaca ctgcatgtgg ctagacttca tggccttggtg cccctacca
24481 aggagaacat gggttggttt tctttggctt ttgtgatcat ggcctgtgca atgtcaggtc
24541 tgctggtagg ggctctagtg tggttcctga agcgcaagcc cagggtacgga aatgaggaga
24601 aggaaaaatt gctataaatc tttttctctt cgcagaacca tgaatacttt gaccagtgtc
24661 gtgctgctct ctcttttagt tattaatgtg gaatgtgccg atcctattct agttagtgtg
24721 gattgggggaa aaaatcttac attagagggg cctaaagaaa caccagttaga atgggtgggt
24781 ggaagaaaca tacaacaact gtgcataggg aatcaaacca aacataaaga gctaagtcac
24841 agatgtaatg tccagaacat aactttactg ttgttaaata ctagttttta tggagactac
24901 tttgggttta aaaatgataa cagcggtagt aaacattata aagtcacagt tataccccct
24961 aaacctcca ctcggaaacc tctttctcct ccacactatg taaacgcaac tatggggcaa
25021 aacctaacat tagtggggcc tgcaaacatt ccagttactt ggcttagtga atatggcacg
25081 ttgtgtgagg gcaaaaaaat tttgcacaaa gaattaaatc acacctgtaa cgaacagaac
25141 ctcacgttac tgtttgttaa tatgacacac aacggaccat attttggtt tgacaaatac
25201 aacattgata gagagcagta tgaggtttct attattagtt tgtttaaagt tggcgtgga
25261 cagaagaaaa ttgggaaagg acagaaaaag gaggaaaaga caaaaccaa ctctagtgtg
25321 ttgggacaaa gacaatccag accaaagaaa aaagatattg ttgaagaggt ccaaatacaa
25381 acaggagaaa atcgaacctt tgttggtcca cctggaaaag ttgattggat taaactttcc
25441 agtggaaaca ataattgtct taagtgtgtt aatggcgaca agtatattaa acacacatgt
25501 gatggtcaaa atttaacatt aattaatgtg actagaattt atgacggaac ttattatggt
25561 tctagcaatg atggctcaag tcattacaaa gttaccatct atgaattaca caaagttaat
25621 aaaactaaat ctatgcttaa gccatacact acaaaaagaa ctacagtga tgcaacagat
25681 gacagtgtct acaaaattgc tttgcagcag gaaaataatg ggcaaacaga aatgatcaa
25741 gaatcaaaaa ttccatctgc tactgtggca atcgtgggtg gagtgattgc gggcttcata
25801 actataatca ttgtcattct gtgctacatc tgctgccgca agcgtcccag ggcatacaat
25861 aatatggtag acccactact cagcttctct tactgagact cagtcacttt catttcagaa
25921 ccatgaaggc ttccacagct tgcgttctga tttagcataa tacacttagt tttagcagac
25981 ctaaaccaga agtatataca caagttaatg tcactagggg tgggaatgct acactagatg
26041 gaccatttaa caataacaca tggacaagat atcatgatga tgggagaaaa aacggatgga
26101 tgaatatattg taaatggtca gacccatcat acacatgtca tagtaatgga agccttagta
26161 tttttgcttt caacattagt tcaggtaaatt ataaagttca aagttacact aacagttata
26221 atggattaga tggttatgaa aaacttgaag ttaaaatggt taatctaaca gtaattgagc
26281 ctccaaccac tagagcacc accacagtta ggacaactaa ggaacaaca cagcctacca
26341 ctgtaccac tacacatcca accaccacag tcagtacaac tattgagacc actactcata
26401 ctacacagct agacacaaca gtgcagaata ctactttact gattgaattt ttactaagag
26461 ggaatgaaag tactactgat cagacagagg ctacctcaag tgccttcagc agtactgcaa
26521 atttaacttc gcttgcttgg actaatgaaa cgggagtatc attgatgag gccagcctt
26581 actcaggttt ggatattcaa ttacttttct tggttgtctg tgggatcttt attcttgtg
26641 ttcttctgta ctttgtctgc tgcaaagcca gagagaaatc tagtaggccc atctacaggc
26701 cagtaatcgg ggaacctcag cctctccaag tgggaagggg tctaaggaat cttctcttct
26761 ctttttcagt atggtgatca gccatgattc ctaggttctt cctatttaac atcctcttct
26821 gtctcttcaa catctgcgct gccttcgcgg cgtctcgcga cgcctcgcgc gactgtctcg
26881 ggcccttccc cacctacctc ctctttgccc tgctcacctg cacctgcgtc tgcagcattg
26941 tctgcctggt cgtcaccttc ctgcagctca tcgactgggt ctgcgcgctg tacaattatc
27001 tccaccacag tcccgaaata agggacaaga acgtagccag aatcttaagg ctcatctgac
27061 catgcagact ctgctgatac tgctatccct cctctcccct gcccttgctg actgtaaatt
27121 tgcggacata tggaaatttct tagactgtta tcaagagaaa atggatatgc cttcctatta
27181 cttggtgatt gtgggtgtag tcatgtgtct ctcctgcact ttctttgcta tcatgatcta
27241 cccctgtttt gatctcggct ggaactctgt tgaggcattc acatacacac tagaaagcag
27301 ttcactagcc tccacgccgc caccacacc gcctccccgc agaaatcagt tccccctgat
27361 tcagtactta gaagagcccc ctccccggcc cccttccact gttagctact ttacataaac
27421 cggcggcgat gactgaccac cacctggacc tcgagatgga cggccaggcc tccgagcagc
27481 gcatcctgca actgcgcgtc cgtcagcagc aggagcgggc cgccaaggag ctctcgtatg
27541 ccatcaacat ccaccagtgc aagaagggca tcttctgcct ggtcaaacag gcaaagatca
27601 cctacgagct cgtgtccaac ggcaaacagc atcgcctcac ctatgagatg cccagcaga
27661 agcagaagtt cacctgcatg gtgggcgtca acccatagt catcaccagc cagtcgggcg
27721 agaccagcgg ctgcatccac tgctcctgcg aaagccccga gtgcacttac tccctcctca
27781 agaccctttg cggacttcgc gacctcctcc ccatgaactg attgattaaa gcccagaaac

```

FIG. 3A-8

21/44

```

27841 caatcaaacc cccttcccca tcaccccaaa taacaatcat tggaaataat cattcaataa
27901 agatcactta cttgaaatct gaaagtatgt ctctgggtgta gttgttcagc agcacctcgg
27961 taccctcctc ccagctctgg tactccagtc cccggcgggc ggcgaacttc ctccacacct
28021 tgaaagggat gtcaaattcc tgggtccaaa ttttcattgt cttccctctc agatgtcaaa
28081 gaggctccgg gtggaagatg acttcaacc cgtctacccc tatgggtacg cgcggaatca
28141 gaatatcccc ttctctactc ccccttttgt ctctccgat ggattccaaa acttcccccc
28201 tggggtcctg tcaactcaaac tgggtgatcc aatcgccatc gccaatggga atgtctcact
28261 caagggtgga gggggactca ctgtagaaca acagtctgga aaactgagtg tggatactaa
28321 ggcacccttg caagttgcaa atgacaacaa attggagcta tcttatgatg atccatttaa
28381 ggtagagaat aacaaacttg gaattaaagc tggccatggg ttagcagttg taactaaaga
28441 aaacacaagt ctctctagtc tagttggaac acttgtagtt ttaactggaa aaggaatagg
28501 tactggatca agtgcacatg gaggaactat tgatgtaaga cttggtgaag gaggtgggtt
28561 atcatttgat gaaaaaggag acttagtagc ttgggacaaa aaaaatgata cacgcaccct
28621 ttggacaaca cctgatcctt ctccaaattg caaagttgaa acagcaagag actcaaagct
28681 aaccttagca cttacaaaat gtggtagtca aattttggcc actgtatctt tacttgttgt
28741 tacgggcaaa tatgctatta taagtgcac agtcaaccca aagcagttct ctattaagtt
28801 actgtttaat gacaagggtg ttttgtaaag tgactcaaat cttgatggga catattggaa
28861 ctatagaagc aacaataaca acataggcac tccttataaa gaggctgttg gttttatgcc
28921 aagcacaaca gcttatccta agccaaccaa caaccacagc acagatccgg ataaaaaggt
28981 gagtcaaggt aaaaaataaaa ttgtaagcaa tatttatctt ggaggagagg tatatcaacc
29041 aggatttatt gttgttaaat ttaatcagga aactgatgcc aattgtgcat actctattac
29101 atttgatttt ggatggggta aggtgtataa ggatcctata ccatatgata cctcttcttt
29161 tactttctca tatatcgctc aagaatgaaa gaccaataaa cgtgtttttc attgaaaatt
29221 ttcatgtatc tttattgatt ttacaccag cacgggtagt cagtctccca ccaccagccc
29281 atttcacagt gtaaacaatt ctctcagcac ggtgggctt aaatagggga atgttctgat
29341 tagtgcgga actggagtcg acctacatgg ggttagagtc ataatcgtgc atcaggatag
29401 ggcggtggtg ctgcagcagc gcgcgaataa actgctgccc cgcgcgtccc gtcctgcagg
29461 aatacaacat ggcagtggtc tcctcagcga tgattcgcac cgcccgagc ataaggcgcc
29521 ttgtcctccg ggcacagcag cgcaccctga tctacttaa atcagcacag taactgcagc
29581 acagcaccac aatatgttgc aaaatcccac agtgcaaggc gctgtatcca aagctcatgg
29641 cggggaccac agaaccacg tggccatcat accacaagcg caggtagatt aagtggcgac
29701 cctcataaaa cacgctggac ataaacatta cctcttttgg catgttgtaa ttcaccacct
29761 cccggtacca tataaacctc tgattaaaca tggcgccatc caccaccatc ctaaaccagc
29821 tggccaaaac ctgcccgcgc gctatacact gcagggaacc gggactggaa caatgacagt
29881 ggagagccca ggactcgtaa ccatggatca tcatgctcgt catgatatca atgttggcac
29941 aacacaggca cacgtgcata cacttccctc ggattacaag ctccctcccgc gttagaacca
30001 tatcccaggg aacaacccat tcctgaatca gcgtaaatcc cacactgcag ggaagacctc
30061 gcaggtaaat cacgttgtgc attgtcaaa tgttacattc gggcagcagc ggatgatcct
30121 ccagtatggt agcgcgggtt tctgtctcaa aaggaggtag acgatcccta ctgtacggag
30181 tgccgcgaga caaccgagat cgtgttggtc gtagtgtcat gccaaatgga acgcccggagc
30241 tagtcatagc tagccccgct taccagtaga cagagagcac agcagtacaa gcgccaacag
30301 cagcgactga ctaccactg acccagctcc ctatttaaag gcaccttaca ctgacgtaat
30361 gaccaaaggt ctaaaaaccc cgccaaaaaa cgatttcgtg actcaacttc cgggttccca
30421 aaacacttcc gcgttgtcac ttctcgtat tccgcgtag ggcgccatct tgcccacgtc
30481 cgttacgtca cttctgccct tacatgtggc cgcgcgcgacc gttgacgcgc tgcgctgtga
30541 caaaatggct tccatgtcag accacgcctc agcgtcggcc cgcgccaaa ttcaaaagct
30601 cgtcagtaac ggctcttcca acagccaatc tttactttgt ggggtatatt attgatgttt
30661 catttgcata ttaacttttg ttactttgt gtgaaattgt tatccgctca caattccaca
30721 ggcgtaatca tggtcatagc tggttcctgt taaagtgtaa agcctggggg gcctaagtag
30781 caacatacga gccggaagca cactgcccgc tttccagtcg ggaaacctgt cgtgccagct
30841 cacattaatt gcgttgcgct ggcgggggag aggcggtttg cgtattgggc gctcttccgc
30901 gcattaatga atcgccaac tgcgctcggt cgttcggctg cggcgagcgg tatcagctca
30961 ttctcgtctc actgactcgc tatccacaga atcaggggat aacgcaggaa agaacatgtg
31021 ctcaaaggcg gtaatacggg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca
31081 agcaaaaggc cccctgacg agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa
31141 taggctccgc cccctgacg accaggcgtt tccccctgga agctccctcg tgcgctctcc
31201 cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc
31261 tgttccgacc ctgcccgtta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc

```

FIG. 3A-9

22/44

```
31321 gcttttctcat agctcacgct gtaggtatct cagttcgggtg taggtcggtc gctccaagct
31381 gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg
31441 tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag
31501 gattagcaga gcgaggtagt taggcgggtgc tacagagtgc ttgaagtggg ggcctaacta
31561 cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg
31621 aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtgggtttttt
31681 tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt
31741 ttctacgggg tctgacgctc agtggaaacga aaactcacgt taagggattt tggtcatgag
31801 attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaataaat
31861 ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc
31921 tatctcagcg atctgtctat ttcggtctatc catagttgcc tgactccccg tcgtgtagat
31981 aactacgata cgggagggct taccatctgg cccagtgct gcaatgatac cgcgagaccc
32041 acgctcacgc gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag
32101 aagtggctct gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag
32161 agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt
32221 ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg
32281 agttacatga tccccatgt tgtgcaaaaa agcggtttagc tccttcgggc ctccgatcgt
32341 tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc
32401 tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc
32461 attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtaa tacgggataa
32521 taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacggt cttcggggcg
32581 aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca ctcgtgcacc
32641 caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag
32701 gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tggtgaatac tcatactctt
32761 cctttttcaa tattattgaa gcatttatca gggttattgt ctcagtgcgc gatacatatt
32821 tgaatgtatt tagaaaaata aacaaatagg ggttcgcgc acatttcccc gaaaagtgcc
32881 acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata ggcgtatcac
32941 gaggcccttt cgtctcgcgc gtttcgggtga tgacggtgaa aacctctgac acatgcagct
33001 cccggagacg gtcacagctt gtctgtaagc ggatgccggg agcagacaag cccgtcaggg
33061 cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac tatgcggcat cagagcagat
33121 tgtactgaga gtgcaccata tgcgggtgtga aataccgcac agatgcgtaa ggagaaaata
33181 ccgcatcagg cgccattcgc cattcaggct gcgcaactgt tgggaagggc gatcggtgcg
33241 ggcctcttcg ctattacgcc agctggcgaa agggggatgt gctgcaaggc gattaagttg
33301 ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagtg aattcggtt
```

SEQ ID NO: 2

FIG. 3A-10

23/44

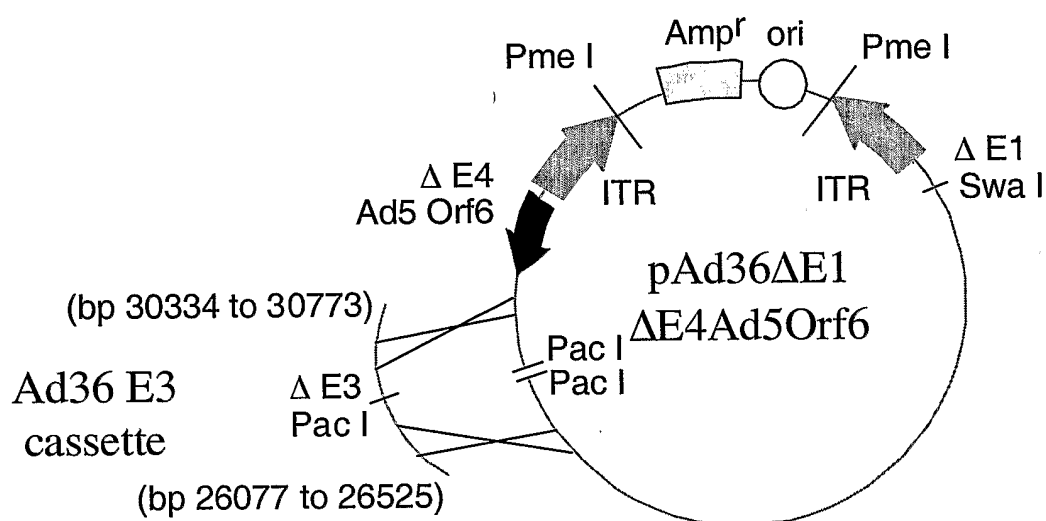


FIG. 4

24/44

Grp #	Prime (Week 0, 4)	Boost (Week 27)	Monk #	Pre	Week 4	Week 8	Week 27	Week 31	Week 35
1	Ad36ΔE1gagΔE4Ad5Orf6 10 [^] 10 vp	Ad36ΔE1gagΔE4Ad5Orf6 10 [^] 10 vp	01D151	0	621	389	65	746	564
			01D170	0	775	359	173	734	588
			01D223	0	113	70	11	38	60
2	Ad36ΔE1gagΔE4Ad5Orf6 10 [^] 8 vp	Ad36ΔE1gagΔE4Ad5Orf6 10 [^] 8 vp	01D239	14	26	39	19	15	10
			01D245	3	0	0	1	0	0
			01D319	0	0	9	30	108	91
3	MRKAd5gag 10 [^] 8 vp	Ad36ΔE1gagΔE4Ad5Orf6 10 [^] 10 vp	01D289	2	1383	1032	444	1289	805
			01D291	9	120	78	18	286	163
			01D296	0	566	587	138	1480	1115

FIG. 5

25/44

Vaccine (Wk 0, 4)	Monkey ID	CD3+CD4+		CD3+CD8+	
		Mock	Gag	Mock	Gag
Ad36ΔE1gagΔE4Ad5Orf6 10 ¹⁰ vp	01D151	114	485	145	712
	01D170	21	273	63	736
	01D223	63	261	125	241

FIG. 6

26/44

atgggtgctagggcttctgtgctgtctgggtgggtgagctggacaagtgggagaagatcaggctgaggcctgggtggc
aagaagaagtacaagctaagcacattgtgtgggcctccaggagctggagaggttctgtgaaccctggcc
tgctggagacctctgaggggtgcaggcagatcctgggccagctccagccctccctgcaaacaggctctgagga
gctgaggctccctgtacaacacagtggctaccctgtactgtgtgcaccagaagattgatgtgaaggacaccaagg
aggccctggagaagattgaggaggagcagaacaagtccaagaagaaggcccagcaggctgctgctggcac
aggcaactccagccagggtgtccagaactacccattgtgcagaacctccagggccagatggtgcaccaggc
catctccccccggaccctgaatgcctgggtgaagggtggaggagaaggccttctccctgaggatgaccca
tgttctctgcccgtgtctgaggggtgccacccccaggacctgaacaccatgctgaacacagtggggggccatcag
gtgccatgcagatgctgaaggagaccatcaatgaggaggctgctgagtgggacaggctgcatcctgtgcacg
ctggccccattgccccggccagatgaggagcccaggggctctgacattgctggcaccacctccacctcca
ggagcagattggctggatgaccaacaaccccccatccctgtgggggaaatctacaagagggtgatcctctg
ggcctgaacaagattgtgaggatgtactccccacctccatcctggacatcaggcaggggcccaaggagccct
tcagggactatgtggacagggtctacaagacctgagggctgagcaggcctcccaggagggtgaagaactggat
gacagagaccctgctggtgcagaatgccaacctgactgcaagaccatcctgaaggccctgggcccctgctgcc
accttgaggagatgatgacagcctgccaggggtggggggccctggtcacaaggccagggtgctggctga
ggcatgtcccagggtgaccaactccgccaccatcatgatgcagaggggcaacttcaggaaccagaggaaga
cagtgaagtgttcaactgtggcaagggtgggccacattgccaagaactgtagggccccaggaagaagggt
gctggaagtgtggcaaggaggggccaccagatgaaggactgcaatgagaggcaggccaacttcctgggcaa
aatctggccctccacaagggcaggcctggcaacttcctccagtccaggcctgagcccacagccccctccgag
gagtccttcagggttggggaggagaagaccacccccagccagaagcaggagcccattgacaaggagctgta
ccccctggcctccctgaggctcctgtttggcaacgacccctcctcccagtaaaataaagcccgggcagat
(SEQ ID NO: 3)

FIG. 7

27/44

1 ccattgcata cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaca
61 ttaccgccat gttgacattg attattgact agttattaat agtaatcaat tacgggggtca
121 ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgccct
181 ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta
241 acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac
301 ttggcagtac atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacgggt
361 aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag
421 tacatctacg tattagtcac cgctattacc atggtgatgc ggttttggca gtacatcaat
481 gggcgtggat agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat
541 gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc
601 ccattgacgc aaatgggccc taggcgtgta cgggtggagg tctatataag cagagctcgt
661 ttagtgaacc gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga
721 caccgggacc gatccagcct ccgcgccggg gaacgggtgca ttggaacgcg gattccccgt
781 gccaagagtg agatctacca TGGGTGCTAG GGCTTCTGTG CTGTCTGGTG GTGAGCTGGA
841 CAAGTGGGAG AAGATCAGGC TGAGGCCTGG TGGCAAGAAG AAGTACAAGC TAAAGCACAT
901 TGTGTGGGCC TCCAGGGAGC TGGAGAGGTT TGCTGTGAAC CCTGGCCTGC TGGAGACCTC
961 TGAGGGGTGC AGGCAGATCC TGGGCCAGCT CCAGCCCTCC CTGCAACAG GCTCTGAGGA
1021 GCTGAGGTCC CTGTACAACA CAGTGGCTAC CCTGTACTGT GTGCACCAGA AGATTGATGT
1081 GAAGGACACC AAGGAGGCCC TGGAGAAGAT TGAGGAGGAG CAGAACAAGT CCAAGAAGAA
1141 GGCCCAGCAG GCTGCTGCTG GCACAGGCAA CTCCAGCCAG GTGTCCCAGA ACTACCCCAT
1201 TGTGCAGAAC CTCCAGGGCC AGATGGTGCA CCAGGCCATC TCCCCCGGA CCCTGAATGC
1261 CTGGGTGAAG GTGGTGGAGG AGAAGGCCTT CTCCCCTGAG GTGATCCCCA TGTTCTCTGC
1321 CCTGTCTGAG GGTGCCACCC CCCAGGACCT GAACACCATG CTGAACACAG TGGGGGGCCA
1381 TCAGGCTGCC ATGCAGATGC TGAAGGAGAC CATCAATGAG GAGGCTGCTG AGTGGGACAG
1441 GCTGCATCCT GTGCACGCTG GCCCCATTGC CCCCAGCCAG ATGAGGGAGC CCAGGGGCTC
1501 TGACATTGCT GGCACCACCT CCACCCTCCA GGAGCAGATT GGCTGGATGA CCAACAACCC
1561 CCCCATCCCT GTGGGGGAAA TCTACAAGAG GTGGATCATC CTGGGCCTGA ACAAGATTGT
1621 GAGGATGTAC TCCCCACCT CCATCCTGGA CATCAGGCAG GGCCCCAAGG AGCCCTTCAG
1681 GGAATATGTG GACAGGTTCT ACAAGACCTT GAGGGCTGAG CAGGCCTCCC AGGAGGTGAA
1741 GAACTGGATG ACAGAGACCC TGCTGGTGCA GAATGCCAAC CCTGACTGCA AGACCATCCT
1801 GAAGGCCCTG GGCCCTGCTG CCACCCTGGA GGAGATGATG ACAGCCTGCC AGGGGGTGGG
1861 GGGCCCTGGT CACAAGGCCA GGGTGCTGGC TGAGGCCATG TCCCAGGTGA CCAACTCCGC
1921 CACCATCATG ATGCAGAGGG GCAACTTCAG GAACCAGAGG AAGACAGTGA AGTGCTTCAA
1981 CTGTGGCAAG GTGGGCCACA TTGCCAAGAA CTGTAGGGCC CCCAGGAAGA AGGGCTGCTG
2041 GAAGTGTGGC AAGGAGGGCC ACCAGATGAA GGAATGCAAT GAGAGGCAGG CCAACTTCCT
2101 GGGCAAAATC TGGCCCTCCC ACAAGGGCAG GCCTGGCAAC TTCTCCAGT CCAGGCCTGA
2161 GCCCACAGCC CCTCCCGAGG AGTCCTTCAG GTTTGGGGAG GAGAAGACCA CCCCAGCCA
2221 GAAGCAGGAG CCCATTGACA AGGAGCTGTA CCCCCTGGCC TCCCTGAGGT CCCTGTTTGG
2281 CAACGACCCC TCCTCCCAGT AAaataaagc ccgggcagat ctgatctgct gtgccttcta
2341 gttgccagcc atctgttgtt tgccccctccc ccgtgccttc cttgacctg gaagggtgcca
2401 ctcccactgt cctttcctaa taaaatgagg aaattgcatc gcattgtctg agtaggtgtc
2461 attctattct ggggggtggg gtggggcagc acagcaaggg ggaggattgg gaagacaata
2521 gcaggcatgc tggggatgcg gtgggctcta

SEQ ID NO: 4

FIG. 8

28/44

AGATCTACCA TGGCCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC
ATGGATGGCC CCAAGGTGAA GCAGTGGCCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG
GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC
TACAACACCC CTGTGTTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG
GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC
CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGGATGTGGG GGATGCCTAC
TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC
AACAAATGAGA CCCCTGGCAT CAGGTACCAG TACAATGTGC TGCCCCAGGG CTGGAAGGGC
TCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC
CCTGACATTG TGATCTACCA GTACATGGAT GACCTGTATG TGGGCTCTGA CCTGGAGATT
GGGCAGCACA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC
ACCCCTGACA AGAAGCACCA GAAGGAGCCC CCCTTCCTGT GGATGGGCTA TGAGCTGCAC
CCCGACAAGT GGACTGTGCA GCCCATTTGT CTGCCTGAGA AGGACTCCTG GACTGTGAAT
GACATCCAGA AGCTGGTGGG CAAGCTGAAC TGGGCCTCCC AAATCTACCC TGGCATCAAG
GTGAGGCAGC TGTGCAAGCT GCTGAGGGGC ACCAAGGCCC TGACTGAGGT GATCCCCCTG
ACTGAGGAGG CTGAGCTGGA GCTGGCTGAG AACAGGGAGA TCCTGAAGGA GCCTGTGCAT
GGGGTGTACT ATGACCCCTC CAAGGACCTG ATTGCTGAGA TCCAGAAGCA GGGCCAGGGC
CAGTGGACCT ACCAAATCTA CCAGGAGCCC TTCAAGAACC TGAAGACTGG CAAGTATGCC
AGGATGAGGG GGGCCACAC CAATGATGTG AAGCAGCTGA CTGAGGCTGT GCAGAAGATC
ACCACTGAGT CCATTGTGAT CTGGGGCAAG ACCCCCAAGT TCAAGCTGCC CATCCAGAAG
GAGACCTGGG AGACCTGGTG GACTGAGTAC TGGCAGGCCA CCTGGATCCC TGAGTGGGAG
TTTGTGAACA CCCCCCCCCT GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATTTGTG
GGGGCTGAGA CTTTCTATGT GGATGGGGCT GCCAACAGGG AGACCAAGCT GGGCAAGGCT
GGCTATGTGA CCAACAGGGG CAGGCAGAAG GTGGTGACCC TGACTGACAC CACCAACCAG
AAGACTGAGC TCCAGGCCAT CTACCTGGCC CTCCAGGACT CTGGCCTGGA GGTGAACATT
GTGACTGACT CCCAGTATGC CCTGGGCATC ATCCAGGCCC AGCCTGATCA GTCTGAGTCT
GAGCTGGTGA ACCAGATCAT TGAGCAGCTG ATCAAGAAGG AGAAGGTGTA CCTGGCCTGG
GTGCCTGCCC ACAAGGGCAT TGGGGGCAAT GAGCAGGTGG ACAAGCTGGT GTCTGCTGGC
ATCAGGAAGG TGCTGTTTCT GGATGGCATT GACAAGGCCC AGGATGAGCA TGAGAAGTAC
CACTCCAACCT GGAGGGCTAT GGCCTCTGAC TTCAACCTGC CCCCTGTGGT GGCTAAGGAG
ATTGTGGCCT CCTGTGACAA GTGCCAGCTG AAGGGGGAGG CCATGCATGG GCAGGTGGAC
TGCTCCCCCTG GCATCTGGCA GCTGGACTGC ACCCACCTGG AGGGCAAGGT GATCCTGGTG
GCTGTGCATG TGGCCTCCGG CTACATTGAG GCTGAGGTGA TCCCTGCTGA GACAGGCCAG
GAGACTGCCT ACTTCTGCT GAAGCTGGCT GGCAGGTGGC CTGTGAAGAC CATCCACACT

FIG. 9A-1

29/44

GACAATGGCT CCAACTTCAC TGGGGCCACA GTGAGGGCTG CCTGCTGGTG GGCTGGCATC
AAGCAGGAGT TTGGCATCCC CTACAACCCC CAGTCCCAGG GGGTGGTGGA GTCCATGAAC
AAGGAGCTGA AGAAGATCAT TGGGCAGGTG AGGGACCAGG CTGAGCACCT GAAGACAGCT
GTGCAGATGG CTGTGTTCAT CCACAAC TTC AAGAGGAAGG GGGGCATCGG GGGCTACTCC
GCTGGGGAGA GGATTGTGGA CATCATTGCC ACAGACATCC AGACCAAGGA GCTCCAGAAG
CAGATCACCA AGATCCAGAA CTCAGGGTG TACTACAGGG ACTCCAGGAA CCCCCTGTGG
AAGGGCCCTG CCAAGCTGCT GTGGAAGGGG GAGGGGGCTG TGGTGATCCA GGACAACTCT
GACATCAAGG TGGTGCCAG GAGGAAGGCC AAGATCATCA GGGACTATGG CAAGCAGATG
GCTGGGGATG ACTGTGTGGC CTCCAGGCAG GATGAGGACT AAAGCCCCGGG CAGATCT
(SEQ ID NO:6)

FIG. 9A-2

30/44

Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly
Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr
Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala
Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu
Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys

FIG. 10A-1

31/44

Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
Glu Asp (SEQ ID NO:7)

FIG. 10A-2

32/44

AGATCTACCATGGCCCCCATCTCCCCATTGAGACTGTGCCTGTGAAGCTGAAGCCTGGCATGGATGGCCCCAAGGTGAA
BgIII MetAlaProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLy
1 10 20

GCAGTGGCCCCCTGACTGAGGAGAAGATCAAGGCCCTGGTGGAAATCTGCACCTGAGATGGAGAAGGAGGGCAAAATCTCCA
sGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
30 40 50

AGATTGGCCCCGAGAACCCCTACAACACCCCTGTGTTTGCCATCAAGAAGAAGGACTCCACCAAGTGGAGGAAGCTGGTG
ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal
60 70

GACTTCAGGGAGCTGAACAAGAGGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCTGGCCTGAAGAA
AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLy
80 90 100

GAAGAAGTCTGTGACTGTGCTGGCTGTGGGGATGCCTACTTCTCTGTGCCCTGGATGAGGACTTCAGGAAGTACACTG
sLysLysSerValThrValLeuAlaValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrA
110 120 130

CCTTCACCATCCCCCTCCATCAACAATGAGACCCCTGGCATCAGGTACCAGTACAATGTGCTGCCCCAGGGCTGGAAGGGC
laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGly
140 150

TCCCCTGCCATCTTCCAGTCTCCATGACCAAGATCCTGGAGCCCTTCAGGAAGCAGAACCCTGACATTGTGATCTACCA
SerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGl
160 170 180

GTACATGGCTGCCCTGTATGTGGGCTCTGACCTGGAGATTGGGCAGCACAGGACCAAGATTGAGGAGCTGAGGCAGCACC
nTyrMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
190 200 210

TGCTGAGGTGGGGCCTGACCACCCCTGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGTGGATGGGCTATGAGCTGCAC
euLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis
220 230

CCCGACAAGTGGACTGTGCAGCCCATTTGTGCTGCCTGAGAAGGACTCCTGGACTGTGAATGACATCCAGAAGCTGGTGGG
ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGl
240 250 260

CAAGCTGAACTGGGCCTCCCAAATCTACCCTGGCATCAAGGTGAGGCAGCTGTGCAAGCTGCTGAGGGGCACCAAGGCCC
yLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
270 280 290

TGACTGAGGTGATCCCCCTGACTGAGGAGGCTGAGCTGGAGCTGGCTGAGAACAGGGAGATCCTGAAGGAGCCTGTGCAT
EuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis
300 310

GGGGTGACTATGACCCCTCCAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAAATCTA
GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTy
320 330 340

CCAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGGCCACACCAATGATGTGAAGCAGCTGA
rGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
350 360 370

CTGAGGCTGTGCAGAAGATCACCCTGAGTCCATTGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAG
hrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys
380 390

FIG. 11A-1

33/44

GAGACCTGGGAGACCTGGTGGACTGAGTACTGGCAGGCCACCTGGATCCCTGAGTGGGAGTTGTGAACACCCCCCCCCT
GluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLe
400 410 420

GGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCCATTGTGGGGGCTGAGACCTTCTATGTGGCTGGGGCTGCCAACAGGG
uValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAlaGlyAlaAlaAsnArgG
430 440 450

AGACCAAGCTGGGCAAGGCTGGCTATGTGACCAACAGGGGCAGGCAGAAGGTGGTGACCCTGACTGACACCACCAACCAG
luThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGln
460 470

AAGACTGCCCTCCAGGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAACATTGTGACTGCCTCCCAGTATGC
LysThrAlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAlaSerGlnTyrAl
480 490 500

CCTGGGCATCATCCAGGCCAGCCTGATCAGTCTGAGTCTGAGCTGGTGAACCAGATCATTGAGCAGCTGATCAAGAAGG
aLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysG
510 520 530

AGAAGGTGTACCTGGCCTGGGTGCCTGCCCACAAGGGCATTGGGGGCAATGAGCAGGTGGACAAGCTGGTGTCTGCTGGC
luLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly
540 550

ATCAGGAAGGTGCTGTTCCTGGATGGCATTGACAAGGCCAGGATGAGCATGAGAAGTACCCTCCAAGTGGAGGGCTAT
IleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnAspGluHisGluLysTyrHisSerAsnTrpArgAlaMe
560 570 580

GGCCTCTGACTTCAACCTGCCCCCTGTGGTGGCTAAGGAGATTGTGGCCTCCTGTGACAAGTGCCAGCTGAAGGGGGAGG
tAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluA
590 600 610

CCATGCATGGGCAGGTGGACTGCTCCCCCTGGCATCTGGCAGCTGGCCTGCACCCACCTGGAGGGCAAGGTGATCCTGGTG
laMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAlaCysThrHisLeuGluGlyLysValIleLeuVal
620 630

GCTGTGCATGTGGCCTCCGGCTACATTGAGGCTGAGGTGATCCCTGCTGAGACAGGCCAGGAGACTGCCTACTTCTGCT
AlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLe
640 650 660

GAAGCTGGCTGGCAGGTGGCCTGTGAAGACCATCCACACTGCCAATGGCTCCAAGTTCCTGAGGGCCACAGTGAGGGCTG
uLysLeuAlaGlyArgTrpProValLysThrIleHisThrAlaAsnGlySerAsnPheThrGlyAlaThrValArgAlaA
670 680 690

CCTGCTGGTGGGCTGGCATCAAGCAGGAGTTTGGCATCCCCTACAACCCCCAGTCCCAGGGGGTGGTGGCCTCCATGAAC
laCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValAlaSerMetAsn
700 710

AAGGAGCTGAAGAAGATCATTGGGCAGGTGAGGGACCAGGCTGAGCACCTGAAGACAGCTGTGCAGATGGCTGTGTTTCAT
LysGluLeuLysLysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPheIl
720 730 740

CCACAACCTCAAGAGGAAGGGGGGCATCGGGGGCTACTCCGCTGGGGAGAGGATTGTGGACATCATTGCCACAGACATCC
eHisAsnPheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleG
750 760 770

AGACCAAGGAGCTCCAGAAGCAGATCACCAAGATCCAGAACTTCAGGGTGTACTACAGGGACTCCAGGAACCCCTGTGG
lnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArgAspSerArgAsnProLeuTrp
780 790

FIG. 11A-2

34/44

AAGGGCCCTGCCAAGCTGCTGTGGAAGGGGGAGGGGGCTGTGGTGATCCAGGACAACCTCTGACATCAAGGTGGTGCCCAG
LysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysValValProAr
800 810 820

GAGGAAGGCCAAGATCATCAGGGACTATGGCAAGCAGATGGCTGGGGATGACTGTGTGGCCTCCAGGCAGGATGAGGACT
gArgLysAlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAspx
830 840 850

AAAGCCCGGGCAGATCT (SEQ ID NO: 8)
Xx BglIII (SEQ ID NO: 9)

FIG. 11A-3

35/44

GATCTGCCAC CATGGGCGGC AAGTGGTCCA AGAGGTCCGT GCGCGGCTGG TCCACCGTGA
GGGAGAGGAT GAGGAGGGCC GAGCCCGCCG CCGACAGGGT GAGGAGGACC GAGCCCGCCG
CCGTGGGCGT GGGCGCCGTG TCCAGGGACC TGGAGAAGCA CGGCGCCATC ACCTCCTCCA
ACACCGCCGC CACCAACGCC GACTGCGCCT GGCTGGAGGC CCAGGAGGAC GAGGAGGTGG
GCTTCCCCGT GAGGCCCCAG GTGCCCCTGA GGCCCATGAC CTACAAGGGC GCCGTGGACC
TGTCCCACTT CCTGAAGGAG AAGGGCGGCC TGGAGGGCCT GATCCACTCC CAGAAGAGGC
AGGACATCCT GGACCTGTGG GTGTACCACA CCCAGGGCTA CTTCCCCGAC TGGCAGAACT
ACACCCCCGG CCCCGGCATC AGGTTCCCC TGACCTTCGG CTGGTGCTTC AAGCTGGTGC
CCGTGGAGCC CGAGAAGGTG GAGGAGGCCA ACGAGGGCGA GAACAACCTGC CTGCTGCACC
CCATGTCCCA GCACGGCATC GAGGACCCCG AGAAGGAGGT GCTGGAGTGG AGGTTGACT
CCAAGCTGGC CTTCCACCAC GTGGCCAGGG AGCTGCACCC CGAGTACTAC AAGGACTGCT
AAAGCCCGG C (SEQ ID NO: 10)

FIG. 12

36/44

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO: 11)

FIG. 13

37/44

WT	- ATG GGT GGC AAG TGG TCA AAA CGT AGT GTG CCT GGA TGG TCT	-42
OPT	- ATG GGC GGC AAG TGG TCC AAG AGG TCC GTG CCC GGC TGG TCC	-14
	M G G K W S K R S V P G W S	
WT	- ACT GTA AGG GAA AGA ATG AGA CGA GCT GAG CCA GCA GCA GAT	-84
OPT	- ACC GTG AGG GAG AGG ATG AGG AGG GCC GAG CCC GCC GCC GAC	-28
	T V R E R M R R A E P A A D	
WT	- AGG GTG AGA CGA ACT GAG CCA GCA GCA GTA GGG GTG GGA GCA	-126
OPT	- AGG GTG AGG AGG ACC GAG CCC GCC GCC GTG GGC GTG GGC GCC	-42
	R V R R T E P A A V G V G A	
WT	- GTA TCT CGA GAC CTG GAA AAA CAT GGA GCA ATC ACA AGT AGC	-168
OPT	- GTG TCC AGG GAC CTG GAG AAG CAC GGC GCC ATC ACC TCC TCC	-56
	V S R D L E K H G A I T S S	
WT	- AAT ACA GCA GCT ACC AAT GCT GAT TGT GCC TGG CTA GAA GCA	-210
OPT	- AAC ACC GCC GCC ACC AAC GCC GAC TGC GCC TGG CTG GAG GCC	-70
	N T A A T N A D C A W L E A	
WT	- CAA GAG GAT GAG GAA GTG GGT TTT CCA GTC AGA CCT CAG GTA	-252
OPT	- CAG GAG GAC GAG GAG GTG GGC TTC CCC GTG AGG CCC CAG GTG	-84
	Q E D E E V G F P V R P Q V	
WT	- CCT TTA AGA CCA ATG ACT TAC AAG GGA GCT GTA GAT CTT AGC	-294
OPT	- CCC CTG AGG CCC ATG ACC TAC AAG GGC GCC GTG GAC CTG TCC	-98
	P L R P M T Y K G A V D L S	
WT	- CAC TTT TTA AAA GAA AAG GGG GGA CTG GAA GGG CTA ATT CAC	-336
OPT	- CAC TTC CTG AAG GAG AAG GGC GGC CTG GAG GGC CTG ATC CAC	-112
	H F L K E K G G L E G L I H	
WT	- TCA CAG AAA AGA CAA GAT ATC CTT GAT CTG TGG GTC TAC CAC	-378
OPT	- TCC CAG AAG AGG CAG GAC ATC CTG GAC CTG TGG GTG TAC CAC	-126
	S Q K R Q D I L D L W V Y H	
WT	- ACA CAA GGC TAC TTC CCT GAT TGG CAG AAC TAC ACA CCA GGG	-420
OPT	- ACC CAG GGC TAC TTC CCC GAC TGG CAG AAC TAC ACC CCC GGC	-140
	T Q G Y F P D W Q N Y T P G	
WT	- CCA GGA ATC AGA TTT CCA TTG ACC TTT GGA TGG TGC TTC AAG	-462
OPT	- CCC GGC ATC AGG TTC CCC CTG ACC TTC GGC TGG TGC TTC AAG	-154
	P G I R F P L T F G W C F K	

FIG. 14A-1

38/44

WT	- CTA GTA CCA GTT GAG CCA GAA AAG GTA GAA GAG GCC AAT GAA	-504
OPT	- CTG GTG CCC GTG GAG CCC GAG AAG GTG GAG GAG GCC AAC GAG	-168
	L V P V E P E K V E E A N E	
WT	- GGA GAG AAC AAC TGC TTG TTA CAC CCT ATG AGC CAG CAT GGG	-546
OPT	- GGC GAG AAC AAC TGC CTG CTG CAC CCC ATG TCC CAG CAC GGC	-182
	G E N N C L L H P M S Q H G	
WT	- ATA GAG GAC CCG GAG AAG GAA GTG TTA GAG TGG AGG TTT GAC	-588
OPT	- ATC GAG GAC CCC GAG AAG GAG GTG CTG GAG TGG AGG TTC GAC	-196
	I E D P E K E V L E W R F D	
WT	- AGC AAG CTA GCA TTT CAT CAC GTG GCC CGA GAG CTG CAT CCG	-630
OPT	- TCC AAG CTG GCC TTC CAC CAC GTG GCC AGG GAG CTG CAC CCC	-210
	S K L A F H H V A R E L H P	
WT	- GAG TAC TAC AAG GAC TGC TGA (SEQ ID NO: 12)	-651
OPT	- GAG TAC TAC AAG GAC TGC TAA (contained within SEQ ID NO: 10)	
	E Y Y K D C (SEQ ID NO: 11)	-216

FIG. 14A-2

39/44

GATCTGCCAC CATGGCCGGC AAGTGGTCCA AGAGGTCCGT GCCCGGCTGG TCCACCGTGA
GGGAGAGGAT GAGGAGGGCC GAGCCCGCCG CCGACAGGGT GAGGAGGACC GAGCCCGCCG
CCGTGGGCGT GGGCGCCGTG TCCAGGGACC TGGAGAAGCA CGGCGCCATC ACCTCCTCCA
ACACCGCCGC CACCAACGCC GACTGCGCCT GGCTGGAGGC CCAGGAGGAC GAGGAGGTGG
GCTTCCCCGT GAGGCCCCAG GTGCCCCTGA GGCCCATGAC CTACAAGGGC GCCGTGGACC
TGTCCCCTT CCTGAAGGAG AAGGGCGGCC TGGAGGGCCT GATCCACTCC CAGAAGAGGC
AGGACATCCT GGACCTGTGG GTGTACCACA CCCAGGGCTA CTTCCCCGAC TGGCAGAACT
ACACCCCCGG CCCCAGCATC AGGTTCCCCC TGACCTTCGG CTGGTGCTTC AAGCTGGTGC
CCGTGGAGCC CGAGAAGGTG GAGGAGGCCA ACGAGGGCGA GAACAACATGC GCCGCCCACC
CCATGTCCCA GCACGGCATC GAGGACCCCG AGAAGGAGGT GCTGGAGTGG AGGTTCGACT
CCAAGCTGGC CTTCCACCAC GTGGCCAGGG AGCTGCACCC CGAGTACTAC AAGGACTGCT
AAAGCCCGGG C (SEQ ID NO: 13)

FIG. 15

40/44

Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Ala Ala His
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
His Pro Glu Tyr Tyr Lys Asp Cys (SEQ ID NO: 14)

FIG. 16

41/44

```
1 gatctgccac catggccggc aagtgggtcca agagggtccgt gcccggtctgg tccaccgtga
61 gggagaggat gaggagggcc gagcccgcgc cgcacagggt gaggaggacc gagcccgcgc
121 cagtgggctg gggcgccgtg tccagggacc tggagaagca cggcgccatc acctcctcca
181 acaccgccc caccaacgcc gactgcgccct ggctggaggc ccaggaggac gaggaggtgg
241 gcttccccgt gaggccccag gtgcccctga ggcccatgac ctacaagggc gccgtggacc
301 tgtcccactt cctgaaggag aaggcgccgc tggagggcct gatccactcc cagaagaggc
361 aggacatcct ggacctgtgg gtgtaccaca ccaggggcta cttccccgac tggcagaact
421 acacccccgg ccccgccatc aggttcccc tgaccttcgg ctggtgcttc aagctggtgc
481 ccgtggagcc cgagaagggt gaggaggcca acgagggcga gaacaactgc ctgctgcacc
541 ccatgtccca gcacggcatc gaggaccccg agaaggaggt gctggagtgg aggttcgact
601 ccaagctggc cttccaccac gtggccaggg agctgcaccc cgagtactac aaggactgct
661 aaagcccggg c (SEQ ID NO: 15)
```

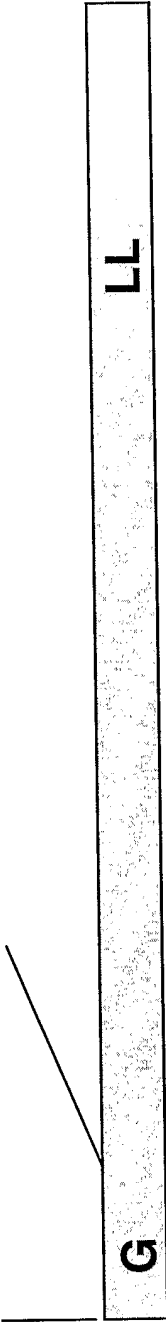
FIG. 17

42/44

1 MAGKWSKRSV PGWSTVRERM RRAEPAADRV RTEPAAVGV GAVSRDLEKH GAITSSNTAA
61 TNADCAWLEA QEDEEVGFPV RPQVPLRPMT YKGAVDLSHF LKEKGGLEGL IHSQKRQDIL
121 DLWVYHTQGY FPDWQNYTPG PGIRFPLTFG WCFKLVPVEP EKVEEANEGE NNCLLHPMSQ
181 HGIEDPEKEV LEWRFDKLA FHHVARELHP EYYKDC (SEQ ID NO: 16)

FIG. 18

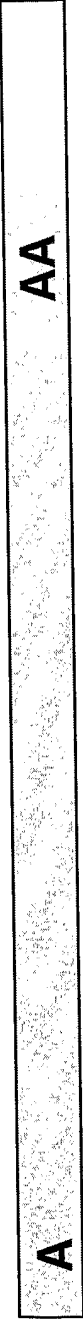
MGGKWSKRSV (within SEQ ID NO: 11)



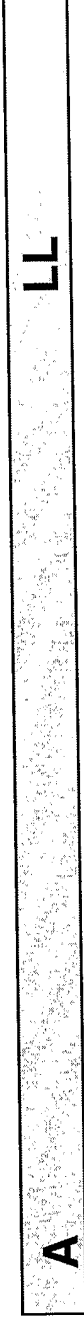
Nef

2 **174,175**

FIG. 19 Nef G2A, LLAA



2 **174,175**



Nef G2A

2 **174,175**

44/44

1 ccattgcata cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaca
 61 ttaccgccat gttgacattg attattgact agttattaat agtaatcaat tacgggggtca
 121 ttagttcata gccatataat ggagttccgc gttacataac ttacggtaaa tggcccgcct
 181 ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta
 241 acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac
 301 ttggcagtag atcaagtgt tcatatgcc agtacgcccc ctattgacgt caatgacggt
 361 aaatggcccg cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag
 421 tacatctacg tattagtcac cgctattacc atgggtgatgc ggttttggca gtacatcaat
 481 gggcgtggat agcggtttga ctacggggga tttccaagtc tccaccccat tgacgtcaat
 541 gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc
 601 ccattgacgc aaatgggccc taggcgtgta cgggtggagg tctatataag cagagctcgt
 661 ttagtgaacc gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga
 721 caccgggacc gatccagcct ccgcccgcgg gaacgggtgca ttggaacgcg gattccccgt
 781 gccaaagagt agatctaagt aagcttcctg CATGCTGCTG CTGCTGCTGC TGCTGGGCCT
 841 GAGGCTACAG CTCTCCCTGG GCATCATCCC AGTTGAGGAG GAGAACCCGG ACTTCTGGAA
 901 CCGCAGAGCA GCCGAGGCCC TGGGTGCCGC CAAGAAGCTG CAGCCTGCAC AGACAGCCGC
 961 CAAGAACCTC ATCATCTTCC TGGGCGATGG GATGGGGGTG TCTACGGTGA CAGCTGCCAG
 1021 GATCCTAAAA GGGCAGAAGA AGGACAAACT GGGGCCCTGAG ATACCCCTGG CCATGGACCG
 1081 CTTCCCATAT GTGGCTCTGT CCAAGACATA CAATGTAGAC AAACATGTGC CAGACAGTGG
 1141 AGCCACAGCC ACGGCCCTACC TGTGCGGGGT CAAGGGCAAC TTCCAGACCA TTGGCTTGAG
 1201 TGCAGCCGCC CGCTTTAACC AGTGCAACAC GACACGCGGC AACGAGGTCA TCTCCGTGAT
 1261 GAATCGGGCC AAGAAAGCAG GGAAGTCAGT GGGAGTGGTA ACCACCACAC GAGTGCAGCA
 1321 CGCCTCGCCA GCCGGCACCT ACGCCACAC GGTGAACCGC AACTGGTACT CGGACGCCGA
 1381 CGTGCCTGCC TCCGCCCGCC AGGAGGGGTG CCAGGACATC GCTACGCAGC TCATCTCCAA
 1441 CATGGACATT GACGTGATCC TAGGTGGAGG CCGAAAGTAC ATGTTTCGCA TGGGAACCCC
 1501 AGACCCCTGAG TACCCAGATG ACTACAGCCA AGGTGGGACC AGGCTGGACG GGAAGAATCT
 1561 GGTGCAGGAA TGGCTGGCGA AGCGCCAGGG TGCCCGGTAT GTGTGGAACC GCACTGAGCT
 1621 CATGCAGGCT TCCCTGGACC CGTCTGTGAC CCATCTCATG GGTCTCTTTG AGCCTGGAGA
 1681 CATGAAATAC GAGATCCACC GAGACTCCAC ACTGGACCCC TCCCTGATGG AGATGACAGA
 1741 GGCTGCCCTG CGCCTGCTGA GCAGGAACCC CCGCGGCTTC TTCTCTTTG TGGAGGGTGG
 1801 TCGCATCGAC CATGGTCATC ATGAAAGCAG GGCTTACCGG GCACTGACTG AGACGATCAT
 1861 GTTCGACGAC GCCATTGAGA GGGCGGGCCA GCTCACCAGC GAGGAGGACA CGCTGAGCCT
 1921 CGTCACTGCC GACCACTCCC ACGTCTTCTC CTTCGGAGGC TACCCCTGCG GAGGGAGCTC
 1981 CATCTTCGGG CTGGCCCCCTG GCAAGGCCCG GGACAGGAAG GCCTACACGG TCCTCCTATA
 2041 CGGAAACGGT CCAGGCTATG TGCTCAAGGA CCGCGCCCGG CCGGATGTTA CCGAGAGCGA
 2101 GAGCGGGAGC CCCGAGTATC GGCAGCAGTC AGCAGTGCCC CTGGACGAAG AGACCCACGC
 2161 AGGCGAGGAC GTGGCGGTGT TCGCGCGCGG CCCGCAGGCG CACCTGGTTC ACGGCGTGCA
 2221 GGAGCAGACC TTCATAGCGC ACGTCATGGC CTTCGCCGCC TGCTTGAGC CTTACACCGC
 2281 CTGCGACCTG GCGCCCCCG CCGGCACCAC CGACGCCCGC CACCCGGGTT AAcccgtggt
 2341 ccccgcggtg ctctctctgc tggccgggac atcaggtggc ccccgctgaa ttggaatcga
 2401 tcagaattca gtcgacgata tctgatcacg atctgatctg ctgtgccttc tagttgccag
 2461 ccactctgtt tttgcccctc ccccgctgct tccttgacct tgggaagggtg cactcccact
 2521 gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt
 2581 ctgggggggtg ggggtggggca gcacagcaag ggggaggatt gggaagacaa tagcaggcat
 2641 gctgggggatg cgggtgggctc ta (SEQ ID NO: 17)

FIG. 20

SEQUENCE LISTING

<110> Merck & Co, Inc.

<120> ADENOVIRUS SEROTYPE 36 VECTORS, NUCLEIC
ACIDS AND VIRUS PRODUCED THEREBY

<130> 21798 PCT

<150> 60/651,843

<151> 2005-02-10

<160> 17

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 35152

<212> DNA

<213> adenovirus serotype 36

<400> 1

```

catcatcaat aatatacccc acaaagtaaa caaaagttaa tatgcaaatag agctttttgaa 60
tttggggcgt ggccgacgct gattggacga gaagtgggtga tgcaaatagac gtcacgacgc 120
acggccgacg gtcgcccgcg aggcgtggcc tagtccggaa gcaagtcgcg gggctgatga 180
cgtataaaaa agcggacttt agaccggaa acggccgatt ttcccgcggt cagccccgga 240
tatgaggtaa ttctgggagg atgcaagtga aattaggcca ttttggcgcg aaaactgaat 300
gaggaagtga aaagtgaaaa ataccggggc cgcccagggc ggaatatatta ccgagggccg 360
agagactttg accgattacg tgggggtttc gattgcggtg tttttttcgc gaattttccgc 420
gtccgtgtca aagtcgggtg tttatgtcac agatcagctg atccacaggg tatttaaacc 480
agtcgagccc gtcaagaggc cactcttgag tgccagcgag tagagatttc tctgagctcc 540
gctcccagag accgagaaaa atgagacacc tgcgccttct accttcaact gtgcccggcg 600
acctggctgt gattatgctg gaggactttg tgaatacagt tctggaggac gaactgcac 660
cagagccatt tgagctggga cctacacttc aggacctcta tgatctggag gtagatgccc 720
atgatgacga ccctaacgaa gaggtgtgta atttaataatt tccagaatct atgattcttc 780
aggctgacat agccagtga gccatagtta ctctcttaca tactcccact ctgcctccca 840
tacctgaatt ggaggaggat gaagaaatag acctccggtg ctacgaggaa ggttttccctc 900
ccagcgattc agaggacgaa cagggtgagc agcagatggc tctaattctc gatttagctt 960
gtgtgattgt ggaggaacaa gttgtgattg aaaaatctac cgagccagta caaggctgta 1020
ggaactgcca gtatcacccg gataagtcgg gagaccgaa cgcttcctgc gctctgtgtt 1080
acatgaaatc tacttttcagc tttattttaca gtaagtggag tgaatgtgag agaggctgag 1140
tgcttaacac ataactgtaa tgcttgaaca gctgtgctaa gtgtgggttta tttttgttac 1200
taggtccggt gtcagaggat gagtcatcac cctcagaaga agaccacccg tctccccctg 1260
agctgtcagg cgaaacgccc ctgcaagtgc acagaccac cccagtcaga gccagtggcg 1320
agaggcgagc agctgtagaa aaaattgagg acttggttaca tgacatgggt ggggatgaac 1380
ctttggacct gagcttgaaa cgccccagga actaggcgca gctgcgctta gtcattgtga 1440
aataaagttg tacaataaaa gtatatgtga cgcatgcaag gtgtgggttta tgactcatgg 1500
gcgggggctta gtcctatata agtggcaaca cctgggcact tgggcacaga ccttcaggga 1560
gttcctgatg gatgtgtgga ctatccttgc agacttttagc aagacacgcc ggcttgtaga 1620
ggatagttca gacgggtgct ccgggttctg gagacactgg tttggaactc ctctatctcg 1680
cctggtgtat acagttaaga aggattataa agaggaattht gaaaatcttt ttgctgactg 1740
ctctgggtctg ctagattctc tgaatcttgg ccaccagtcc cttttccagg aaagggtact 1800
ccacagcctt gattttttcca gcccaggggc cactacagcc ggggttgctt ttgtggtttt 1860
tctggttgac aaatggagcc aggacaccca actgagcagg ggctacatcc tggacttcgc 1920
agccatgcac ctgtggaggg cctggatcag gcagcgggga cagagaatct tgaactactg 1980

```

gcttctacag	ccagcagctc	cgggtcttct	tcgtctacac	agacaaacat	ccatgttggg	2040
ggaagaaatg	aggcaggcca	tggacgagaa	cccaggagc	ggcctggacc	ctccgtcgga	2100
agaggagctg	gattgaatca	ggtatccagc	ctgtaccag	agcttagcaa	ggtgctgaca	2160
tccatggcca	ggggagtga	gagggagagg	agcgatggg	gtaataccgg	gatgatgacc	2220
gagctgactg	ccagtctgat	gaatcggaag	cgcccagagc	gccttacctg	gtacgagcta	2280
cagcaggagt	gcagggatga	gataggcctg	atgcaggata	aatatggcct	ggagcagata	2340
aaaaccatt	ggttgaaccc	agatgaggat	tgggaggagg	ctattaagaa	gtatgccaa	2400
atagccctgc	gccagattg	caagtacata	gtgaccaaga	ccgtgaatat	cagacatgcc	2460
tgctacatct	gggggaacgg	ggcagagggtg	gtcatcgata	ccctggacaa	ggccgccttc	2520
aggtgttgca	tgatgggaat	gagagcagga	gtgatgaata	tgaattccat	gatcttcatg	2580
aacattaagt	tcaatggaga	gaagttta	ggggtgctgt	tcatggccaa	cagccacatg	2640
accctgcatg	gctgcagctt	cttcggtttc	aacaacatgt	gcgccgaggt	ctggggagct	2700
gctaagatca	ggggatgtaa	gttttatggc	tgctggatgg	gcgtgggtcg	aagacccaag	2760
agcgagatgt	ctgtgaagca	gtgtgtgttt	gagaaatgct	acctgggagt	ctctaccgag	2820
ggcaatgcta	gagtgaagca	ctgctcttcc	atggagacgg	gctgcttctg	cctgggtgaag	2880
ggcacggcct	ctctgaagca	taatatgggtg	aagggtcgca	cggatgagcg	catgtacaac	2940
atgctgacct	gcgattcggg	ggtctgccat	atcctgaaga	acatccatgt	gacctccac	3000
cccagaaaga	agtggccagt	gtttgagaat	aacctgctga	tcaagtgcc	tatgcacctg	3060
ggtgccagaa	ggggcacctt	ccagccgtac	cagtgcact	ttagccagac	caagctgctg	3120
ttggagaacg	atgccttctc	caggggtgaac	ctgaacggca	tctttgacat	ggatgtctcg	3180
gtgtacaaga	tcctgagata	cgatgagacc	aagtccaggg	tgccgcgttg	cgagtgcggg	3240
ggcagacaca	ccaggatgca	accagtggcc	ctggatgtga	cagaggagct	gagaccagac	3300
cacctggtga	tggcctgtac	cgggaccgag	ttcagctcca	gtggggagga	cacagattag	3360
aggtagggtt	tgagtagtgg	gcgtggctaa	ggtgagtata	aaggcggtgt	cttacgaggg	3420
tctttttgct	tttctgcaga	catcatgaac	gggaccggcg	gggccttcga	aggggggctt	3480
tttagccctt	atttgacaac	ccgcctgccg	ggatggggcg	gagttcgtca	gaatgtgatg	3540
ggatctacgg	tggatggggc	cccagtgcct	ccagcaaatt	cctcgaccat	gacctacgcg	3600
accgtgggga	gctcgtcgct	cgacagcacc	gccgcagccg	cggcagccgc	agccgcatg	3660
acagcgacga	gactggcctc	gagctacatg	cccagcagca	gcagtagccc	ctctgtgccc	3720
agttccatca	tcgccgagga	gaaactgctg	gccctgctgg	cagagctgga	agccctgagc	3780
cgccagctgg	ccgccctgac	ccagcagggtg	tcagagctcc	gcgagcaaca	gcagcagcaa	3840
aataaatgat	tcaataaaca	cagattctga	ttcaaacagc	aaagcatctt	tattatttat	3900
tttttcgcgc	cggttagggc	ctgggtccacc	tctcccgatc	attgagagtg	cggtggattt	3960
tttccaggac	ccggtagagg	tgggattgga	tggttaggta	catgggcatg	agcccgctcc	4020
gggggtggag	gtagcaccac	tgcatggcct	cgtgctctgg	ggtcgtgttg	tagatgatcc	4080
agtcatagca	ggggcgctgg	gcgtgggtgt	ggatgatgtc	cttaaggagg	agactgatgg	4140
ccacggggag	ccccttggtg	taggtgttgg	cgaagcggtt	gagctgggag	ggatgcatgc	4200
ggggggagat	gatgtgcagt	ttggcctgga	tcttgagggt	ggcaatgttg	ccgccagat	4260
ccgcctggg	gttcatgttg	tgcaggacca	ccaggacggg	gtagcccgtg	cacttgggga	4320
acttatcatg	caacttgga	gggaatgcgt	ggaagaattt	ggagacgccc	ttgtgcccgc	4380
ccagggtttt	catgcactca	tccatgatga	tggcgatggg	cccgtgggct	gcggctttgg	4440
caaagacgtt	tctgggggtca	gagacatcgt	aattatgtct	ctgggtgaga	tcatcataag	4500
acattttaat	gaattttggg	cggagggtgc	cagattgggg	gacgatagtt	ccctcggggc	4560
ccggggcaaa	gttccccctca	cagatctgca	tctcccaggc	tttcatctcg	gaggggggga	4620
tcatgtccac	ctgccccggc	atgaaaaaaa	cggtttccgg	ggcgggggtg	atgagctgcg	4680
aggagagcag	gtttctcaac	agctgggact	tgccgcaccc	ggtcggggcg	tagatgaccc	4740
cgatgacggg	ttgcagggtg	tagttcaagg	acatgcagct	gccgtcgtcc	cggaggagg	4800
gggccacctc	gttgagcatg	tctctgactt	ggagggtttt	ccggacgagc	tcgcccagga	4860
ggcgttcccc	gcccagcgag	aggagctctt	gcagggaagc	aaagtttttt	aggggcttga	4920
gtccgtcggc	catgggcatc	ttggcgagg	tctgcgagag	gagctccagg	cgggtcccaga	4980
gctcggtgac	gtgctctacg	gcatctcgat	ccagcagact	tcctcgtttt	gggggttggg	5040
acgactgcga	ctgtagggca	cgagacgatg	ggcgtccagc	gcggccagcg	tcatgtcctt	5100
ccagggtctc	aggggtccgc	tgagtgtgg	ctccgtcacg	gtgaaggggt	gggccccggg	5160
ctgggcgctt	gcaaggggtg	gcttgagact	cacctgtctg	gtgctgaaac	gggcacgggc	5220
ttcgccctgc	gcgtcggcga	gatagcagtt	gacctagagc	tcgtagttga	gggcctcggc	5280

```

ggcgtggccc ttggcgcgga gcttgccctt ggaagagcgc ccgcaggcgg gacagagaag 5340
ggattgcagg gcgtagagct tgggtgcgag aaaaacggac tcgggggcga aggcgtccgc 5400
tccgcagtgg gcgcagacgg tctcgactc gacgagccag gtgagctcgg gctgctcggc 5460
gtcaaaaacc agttttcccc cgttcttttt gatgcgcttc ttacctcggc tctccatgag 5520
tctgtgtccg cgctcgggtga caaacaggct gtcgggtgcc ccgtagacgg acttgatggg 5580
cctgtcctgc agggacgtcc cgcggtcctc ctcgtagaga aactcggacc actctgagac 5640
aaaggcgcgc gtccacgcca agacaaagga ggccacgtgc gaggggtagc ggtcgttgct 5700
caccaggggg tccacctttt ccaccgtgtg cagacacatg tccccctcct ccgcatccaa 5760
gaaggtgatt ggcttgtagg tgtaggccac gtgaccgggg gtccccgacg ggggggtata 5820
aaagggggcg ggtctgtgct cgtcctcact ctcttcgcg cgtctgtcca cgagcgccag 5880
ctggtggggg aggtattccc tctctagagc gggcatgacc tcggcactca ggttgtcagt 5940
ttctagaaac gaggaggatt tgatgttggc ctgccctgcc gcgatgcttt ttaggagact 6000
ttcatccatc tggtcagaaa agacaatttt tttattgtca agcttgggtg caaaggagcc 6060
atagagggcg ttggatagaa gcttggcgat ggatctcatg gtctgatttt tgtcacggtc 6120
ggcgcgctcc ttggccgcga tgttttagctg gacatactcg cgcgcgacgc acttccattc 6180
ggggaagacg gtgggtgcgt cgtcgggcac gatcctgacg cgccagccgc ggttatgcag 6240
ggtgaccagg tccacgctgg tggccacctc gcccgcgagg ggctcgttgg tccagcagag 6300
tctgccgccc ttgcgcgagc agaacggggg cagcacatca agcagatgct cgtcaggggg 6360
gtccgcacgc atggtgaaga tgcccgga caagttcctt tcaaaataat cgatttttga 6420
ggatgcatca tccaaggcca tctgccactc gggggcgggc agcgctcgct cgtaggggtt 6480
gagggcgcgga cccaggggca tgggatgcgt gagggcgagg gcgtacatgc cgcagatgtc 6540
atacacatag atgggctccg agaggatgcc gatgtagggt ggataacagc gcccccgcg 6600
gatgctggcg cgcacgtagt catacaactc gtgcgagggg gccaagaagg cggggccgag 6660
attggtgcgc tggggctgct cggcgcgga gacgatctgg cgaaagatgg cgtgcgagtt 6720
tgaggagatg gtgggcccgtt ggaagatggt aaagtgggcg tgaggcaggc ggaccgagtc 6780
gcggatgaag tgcgcgtagg agtcttgacg cttggcgacg agctcggcgg tgacgaggac 6840
gtccatggca cagtagtcca gcgtttcgcg gatgatgtca taaccgcct ctcctttcct 6900
ctcccacagc tcgcggttga gggcgctact ctcgtcatcc ttccagtact cccggagcgg 6960
gaatcctcga tcgtccgcac ggtaagagcc cagcatgtag aaatggttca cggccttgta 7020
gggacagcag cccttctcca cggggagggc gtaagcttga gcggccttgc ggagcgagg 7080
gtgcgtcagg gcgaagggtg ctctgacct gactttcaag aactgggtact tgaaatccga 7140
gtcgtcgcag ccgcccgtgt cccagagctc gaaatcgggt cgcttcttcg agaggggggt 7200
aggcagagcg aaagtgcagt cattgaagag aatcttgctt gcccgcgcca tgaaattgcg 7260
ggtgatgcgg aaagggcccg ggacggaggc tcggttgttg atgacctggg cggcgagcac 7320
gatctcgtcg aagccgttga tgttgtgccc gacgatgtag agttccatga atcgcgggcg 7380
gcctttgatg tgcggcagct ttttgagctc ctctaggtg aggtcctcgg ggcattgcag 7440
tccgtgctgc tcgagcgccc actcctggag atgtgggttg gcttgcagta aggaagccca 7500
gagctcgcgg gccatgaggg tctggagctc gtcgcgaaag agacggaact gctggcccac 7560
ggccatcttt tcgggtgtga cgcagtagaa ggtgaggggg tcccgcctcc agcgatccca 7620
gcgtaagcgc acggcgagat cgcgagcgag ggcgaccagc tcggggtccc cggagaattt 7680
catgaccagc atgaagggga cgagctgctt gccgaaggac cccatccagg ttaggtttc 7740
tacatcgtag gtgacaaaga gccgctccgt gcgaggatga gagcgattg ggaagaactg 7800
gatttctctg caccagttgg acgagtggct gttgatgtga tgaaagtaga aatcccgcg 7860
gcgaaccgag cactcgtgct gatgcttgta aaagcgtccg cagtactcgc agcgctgcac 7920
gggctgtacc tcatccacga gatacacagc gcgtcccttg aggaggaact tcaggagtgg 7980
cggccctggc tgggtggttt catgttcgcc tgcgtgggac tcaccctggg gctcctcgag 8040
gacggagagg ctgacgagcc cgcgcgggag ccaggtccag atctcggcgc ggcggggggc 8100
gagagcgaag acgagggcgc gcagttggga gctgtccatg gtgtcgcgga gatccaggtc 8160
cgggggcagg gttctgaggt tgacctcgta gaagcgggtg agggcgtgct tgagatgcag 8220
atgggtactt atttctacgg gtgagttggt ggccgtgtcc acgcatgtca tgagcccgt 8280
gctgcgcggg gccacgaccg tgccgcgggt cgctttttaga agcgtgtcgc cggacgcgct 8340
ccggcgggca gcgcggttc cggcccccg ggcagggggc gcagaggcac gtcggcgtgg 8400
cgctcgggca ggtcccgtg ctgcgcctg agagcgctgg cgtgcgcgac gacgcggcgg 8460
ttgacatcct ggaatctgcg cctctgcgtg aagaccagc gcccgtgac tttgaacctg 8520
aaagacagtt caacagaatc aatctcggcg tcattgacgg cggcctgacg caggatctct 8580

```

tgcacgtcgc ccgagttgtc ctggtagggc atctcggaca tgaactgctc gatctcctcc 8640
 tcctggagat cgcgcgggcc cgcgcgctcc acggtggcgg cgaggtcatt cgagatgcga 8700
 cccattagct gcgagaaggc gccagggccg ctctcgttcc agacgcggct gtagaccacg 8760
 tccccgtcgg cgtcgcgcgc gcgcatgacc acctgcgcga ggttgagctc cacgtgccgc 8820
 gtgaagacgg cgtagttgcg caggcgcctgg aagaggtagt tgagggtggg ggcgatgtgc 8880
 tcggtgacga agaagtacat gatccagcgg cgcaggggca tctcgcctgat gtcgccgatg 8940
 gcctccagcc tttccatggc ctctagaaaa tccacggcga agttgaaaaa ctggggcgtt 9000
 cgggcccaga ccgtgagctc gtcttccagg agccggatga gctcggcgat ggtggcgcg 9060
 acctcgcgct cgaaatcccc ggggacctcc tctcttccct cttcttccat gacgacctc 9120
 tcttctatctt cttctctctg gggcggtggg ggtggcgggg cccgacgacg acggcgacgc 9180
 accgggagac ggtcgacgaa gcgctcgatc atctccccgc ggcggcgacg catggtttcg 9240
 gtgacggcgc gaccccgttc gcgaggacgc agcgtgaaga cgccgccggg catctcccgg 9300
 taatggggcg ggtccccgtt gggcagcgat agggcgctga cgatgcattc tatcaattgc 9360
 ggtgtagggg acgtgagcgc gtcgagatcg accggtcgg agaactcttc gaggaaagcg 9420
 tctagccaat cgcagtcgca aggtaaagctc aaacacgtag cagccctgtg gacgtgtta 9480
 gaattgcggg tgctgatgat gtaattgaag taggcgtttt taaggcggcg gatgggtggc 9540
 aggaggacca ggtccttggg tcccgccttg tggatcgga gccgctcggc catgccccag 9600
 gcctggccct gacaccggct caggttcttg tagtagtcac gcatgagcct ttcaatgtca 9660
 tcactggcgg aggcggagtc tccatgcgg gtgacccga cggccctgag cggctgcacg 9720
 agcgcagggt cggcgacgac gcgctcggcg aggatggcct gttgcacgcg ggtgaggggtg 9780
 tcctggaagt catccatgtc gacgaagcgg tggtagggcc cgggtgtgat ggtgtaggtg 9840
 cagttggcca tgagcgacca gttgacggtc tgcaggccgg gttgcacgac ctcgaggtac 9900
 ctgagccgcg agaaggcgcg cgagtcgaag acgtagtcgt tgcaggtgcg cagcaggtac 9960
 tggtagccga ctaggaagtg cggcggcggc tggcggtaga ggggccagcg ctgggtggcc 10020
 ggcgcgcccg gggccaggtc ctcgagcatg aggcgggtgg agccgtagag gtagcgggac 10080
 atccaggtga tgccggcggc ggtgggtggg gcgcgcggga actcgcggac gcggttccag 10140
 atgttcgcga gcggcaggaa ataataccatg gtccggcacgg tctggccggg gagacgcgcg 10200
 cagtcattga cgctctagag gcaaaaacga aagcggttga gcgggctctt cctccgtagc 10260
 ctggcggaac gcaaacgggt taggcgcgct gtgtacctc cgacctgagc cctcgaatca 10320
 ggctggagcc gcgactaacg tggtattggc actcccgtct cgacctgagc ccatagaccg 10380
 ccaggatacg gcggagagcc ctttttgccg accgagtggg gtcgctagac ttgaaagcgg 10440
 ccgaaaaccg cgccgggtag tggctcgcgc ccgtagtctg gagaagcatc gccagggtt 10500
 agtcgcggca gaaccgggtt cgcggacggc cgcggcgagc gggacttggg caccctggcg 10560
 atttaaagac ccacagccag ccgacttctc cagttacggg agcgagcccc cttttttctt 10620
 tttgccagat gcatcccgtc ctgcgcaaaa tgcgtccac ccccccggcg accaccgcga 10680
 ccgcgcccgat aacaggcgcc ggcgctagcc agccacagac agagatggac ttggaagagg 10740
 gcgaagggct ggcgagactg gggcgccgct ccccgagcgc acacccccgc gtgcagctgc 10800
 agaaggacgt gcgcccggcg tacgtgcctg cgcagaacct gttcagggac cgcagcgggg 10860
 aggagcccga ggagatgcgc gactgccggt ttcgggcggg cagggagctg cgcgagggcc 10920
 tggaccgcca gcgctgtctg cgcgacgagg atttcgagcc gaacgagcag acggggatca 10980
 gccccgcgcg cgcgcacgtg gcggcgccca acctggtgac ggcctacgag cagacgggtg 11040
 agcaggagcg caacttccaa aagagtttca acaaccacgt gcgcacgctg atagcgcgcg 11100
 aggaggtggc cctgggcctg atgcacctgt gggacctggc ggaggccatc gtgcagaacc 11160
 cggacagcaa ggcctgacg gcgcagctgt tcctgggtgg gcagcacagc agggacaacg 11220
 aggcgttcag ggaggcgtg ctgaacatcg ccgagcccga gggctcgtgg ctgctggagc 11280
 tgatcaacat cttgcagagc atcgtagtgc aggagcgcag cctgagcctg gccgagaagg 11340
 tggcgcgat caactactcg gtgctgagcc tgggcaagtt ttacgcgcgc aagatttaca 11400
 agacgccgta cgtgcccata gacaaggagg tgaagataga cagcttttac atgcgcattg 11460
 cgctcaagggt gctgacgctg agcgacgacc tgggcgtgta tcgcaacgac cgcattccca 11520
 aggcgctgag cacgagccgg cggcgcgagc tgagcgaccg cgagctgatg ctgagcctgc 11580
 gccgggcgct ggtagggggc gccgcggcg gcgaggagtc ctacttcgac atggggggcg 11640
 acctgcattg gcagccgagc cggcgcgcc tggaggccgc ctacgggtcca gaggacttgg 11700
 atgaggatga ggaagaggag gaggatgcac ccgttgcggg gtactgacgc ctccgtgatg 11760
 tgtttttaga tgtcccagca agccccggac cccgccataa gggcggcgct gcaaagccag 11820
 ccgtccgggtc tagcatcgga cgactgggag gccgcgatgc aacgcattcat ggcctgacg 11880

accgcgaacc	ccgagtcctt	tagacaacag	ccgcaggcca	acagactctc	ggccattctg	11940
gaggcgggtg	tccctctctg	gaccaacccc	acgcacgaga	aggtgctggc	gacgtgaa	12000
gcgctggcgg	agaacaaggc	catccgtccc	gacgaggccg	ggctgggtga	caacgccctg	12060
ctggagcgcg	tgggcccgt	caacagcacg	aacgtgcagt	ccaacctgga	ccggctgggtg	12120
acggacgtgc	gcgagggcgt	ggcgacgcgc	gagcgggttca	agaacgaggg	cctgggctcg	12180
ctgggtggcgc	tgaacgcctt	cctggcgacg	cagccggcga	acgtgccgcg	agggcaggac	12240
gattacacca	actttatcag	cgcgctgcgg	ctgatgggtga	ccgaggttcc	ccagagcgag	12300
gtgtaccagt	cggggccgga	ctactttttc	cagacaagcc	ggcagggcct	gcagacgggtg	12360
aacctgagtc	aggctttcaa	gaacctgcgc	gggctgtggg	gcgtgcaggc	gcccgtgggc	12420
gaccggtcga	cggtagcag	cttgctgacg	cccaactcgc	ggctgctgct	gctgctgac	12480
gcgcccttca	ccgacagcgg	cagcgtgaac	cgcaactcgt	acctggggcca	cctgctgacg	12540
ctgtaccgcg	aggccatagg	ccaggcgcag	gtggacgagc	agaccttcca	ggagatcacg	12600
agcgtgagcc	gcgcgctggg	gcagaacgac	accgacagtc	tgagggccac	cctgaacttt	12660
ttgctgacaa	atagacagca	gaagatcccg	gcgcagtacg	cactgtcggc	cgaggaggaa	12720
aggatcctga	gatatgtgca	gcagagcgta	gggctgttcc	tgatgcagga	gggtgccacc	12780
cccagcgccg	cgctggacat	gaccgcgcgc	aacatgggaa	ctagcatgta	cgccgccaac	12840
cggccgttca	tcaataagct	gatggactac	ctgcaccgcg	cggcggccat	gaacacggac	12900
tactttacaa	acgccatatt	gaacccgcac	tggcttccgc	cgccgggggt	ctacacgggc	12960
gagtacgaca	tgcccgaccc	caacgacggg	ttcctgtggg	acgacgtgga	cagcgcgggtg	13020
ttctcgccga	cctttcaaaa	gcgccaggag	gcgccgccga	gcgagggcgc	ggtggggagg	13080
agcccctttc	ctagcttagg	gagtttgcag	agcttgccgg	gctcgggtga	cagcggcagg	13140
gtgagccggc	cgcgcttgct	ggcgagggac	gagtacctga	acgactcgct	gctgcagccg	13200
ccgcgggtca	agaacgccat	ggccaataac	gggatagaga	gtctgggtgga	caaactgaac	13260
cgctggaaga	cctacgctca	ggaccatagg	gagcctgcgc	ccgcgccgcg	gcgacagcgc	13320
cacgaccggc	agcggggcct	ggtgtgggac	gacgaggact	cggccgacga	tagcagcgtg	13380
ttggacttgg	gcgggagcgg	tggggccaac	ccgttcgcgc	atctgcagcc	cagactgggg	13440
cggcggaagt	tttgaatgca	aaataaaact	caccaaggcc	atagcgtgcy	ttctcttctt	13500
tgtagagat	gagggctgcy	gtgggtgtct	cctctcttcc	tccctcgtac	gagagcgtga	13560
tggcgagggc	gacctggaag	gttccgtttg	tgcctccgcg	gtatatggct	cctacggagg	13620
gcagaaacag	cattcggtac	tcggagctgg	ctccgcagta	cgacaccact	cgcgtgtact	13680
tgggtggacaa	caagtgcggc	gacatcgctt	ccctgaacta	ccaaaaacgac	cacagcaact	13740
tcctgaccac	ggtggtgcag	aacaacgatt	tcacccccgc	cgaggccagc	acgcagacga	13800
taaatTTTTga	cgagcggctg	cgggtggggc	gtgatctgaa	gaccattctg	cacaccaaca	13860
tgcccaatgt	gaacgagtac	atgttcacca	gcaagtTTaa	ggcgcggggtg	atgggtggcta	13920
gaaagcatcc	caaagatgta	gatgccagtg	atttaagcaa	ggatatctta	gagtataagt	13980
ggtttgagtt	taccttgccc	gagggcaact	tttccgagac	catgaccata	gacctgatga	14040
acaacgccat	cttggaaaac	tacttgcaag	tggggcgcca	gaatggcgtg	ctggagagcg	14100
atatcggagt	caagtttgac	agcaggaatt	tcagactggg	ctgggacccg	gtgaccaagc	14160
tggtagtgcc	aggggtctac	acctacgagg	ccttccaccc	ggacgtggtg	ctactgccgg	14220
gctgcggggg	ggacttcacc	gagagccggc	tgagcaacct	cctgggcatt	cgcaagaagc	14280
aaccttttca	agagggcttc	agaatcatgt	atgaggatct	agaagggggg	aacatccccg	14340
ctctcctgga	tacccaaaaa	tatctggata	gcaagaaggga	acttgaggat	gctgccaagg	14400
aagctgcaaa	gcaacagggga	gatgggtgctg	tcactagagg	cgataccccc	ctcactgtag	14460
ctcaagaaaa	agcagctgaa	aaggagctag	tgatcgtacc	aattgaaaag	gatgagagca	14520
acagaagtta	caacctgatc	aaggacaccc	atgacaccct	gtaccgaagc	tggtagctgt	14580
cctataccta	cggggacccc	gagaaggggg	tgcagtcgtg	gacgctgctc	accaccccg	14640
acgtcacctg	cggcgcgga	caagtctact	ggtcgctgcc	ggacctcatg	caagaccccc	14700
tcaccttccg	ctctaccag	caagtcagca	actaccccg	ggtcggcgcc	gagctcatgc	14760
ccttccgcgc	caagagcttt	tacaacgacc	tcgccgtcta	ctcccagctc	atccgcagct	14820
acacctccct	caccacgctc	ttcaaccgct	tccccgacaa	ccagatcctc	tgccgcccgc	14880
ccgcgcccac	catcaccacc	gtcagtga	acgtgcctgc	tctcacagat	cacgggacgc	14940
ttccgctgcy	cagcagtatc	cgcggagtcc	agcgagtgc	cgctactgac	gcccgtcgcc	15000
gcacctgtcc	ctacgtctac	aaggccctgg	gcatagtcgc	gccgcgcgtg	ctctccagtc	15060
gcaccttcta	aaaaatgtct	attctcatct	cgcccagcaa	taacaccggc	tgggggtctta	15120
ctaggcccag	caccatgtac	ggaggagcca	agaagcgcctc	ccagcagcac	cccgtccgcg	15180

tccgcgggtca	cttcccgcgct	ccctgggggag	cttacaagcg	ggggcgccact	gccaccgccc	15240
ccgcggtgcg	caccaccgtc	gacgacgtca	tcgactcggg	ggtcgcccac	gcgcgcaact	15300
acacccccgc	ccccccacc	gtggacgcgg	tcacgcacag	cgtggtggcc	gacgcgcgcg	15360
actatgccag	acgcaagagc	cggcgggcgac	ggatcgccag	gcgccaccgg	agcacgccc	15420
ccatgcgcgc	cgccccgggt	ctgctgcgcc	gcgccagacg	cacggggccgc	cggggccatga	15480
tgcgagccgc	gcgccgcgcc	gccactgcac	cccccgagg	caggactcgc	agacgagcgg	15540
ccgcgcccgc	tgcgcgggcc	atctctagca	tgaccagacc	caggcgccga	aacgtgtact	15600
gggtgcgcga	ctccgtcacg	ggcgtgcgcg	tgcgcgtgcg	cacccgtcct	cctcgtccct	15660
gatctaattgc	ttgtgtcctc	ccccgcaagc	gacgatgtca	aagcgcaaaa	tcaaggagga	15720
gatgtctccag	gtcgtcgccc	cggagattta	cggacccccg	gaccagaaac	cccgcaaaat	15780
caagcgggtt	aaaaaaaaag	atgaggtgga	cgagggggca	gtagagtttg	tgcgcgagtt	15840
cgtcccgcg	cggcgcgtaa	attggaaggg	gcgcagggtg	cagcgcgtgt	tgcggcccgc	15900
cacggcggtg	gtgttcacgc	ccggcgagcg	gtcctcggtc	aggagcaagc	gtagctatga	15960
cgaggtgtac	ggcgacgacg	acatcctgga	ccaggcgccg	gagcggggcg	gcgagttcgc	16020
ctacgggaag	cggtcgcgcg	aagaggagct	gatctcgtcg	ccgctggacg	aaagcaacc	16080
cacgccgagc	ctgaagcccg	tgaccctgca	gcaggtgctg	cccaggcg	tgctgctgcc	16140
gagccgcggg	gtcaagcgcg	agggcgagag	catgtaccgg	accatgcaga	tcatggtgcc	16200
caagcgcgg	cgcgtggagg	acgtgctgga	caccgtgaaa	atggatgtgg	agccccaggt	16260
caaggtgcgc	cccatcaagc	aggtggcgcc	gggcctgggc	gtgcagaccg	tggacattca	16320
gatccccacc	gacatggatg	tcgacaaaaa	accctcgacc	agcatcgagg	tgcagaccga	16380
cccttggtc	ccagcctcca	ccgtaccgt	ctccactttt	accgcccga	cggctaccga	16440
gcctcccagg	aggcgaagat	ggggcgccgc	cagccggctg	atgcccact	acgtgttgca	16500
tccttccatc	atcccagcgc	cgggctaccg	cggcaccgg	tactacgcca	gccgcaggcg	16560
cccagccgcc	aaacgcccgc	gccgcactgc	cacccgccc	cgtctggccc	ccgcccgcgt	16620
gcgcgcgta	accacgcgcc	ggggccgctc	gctcgttctg	cccaccgtgc	gctaccacc	16680
cagcatcctt	taatccgtgt	gctgtgatac	tgttgagag	agatggctct	cacttgccgc	16740
ctgcgcattc	ccgtcccga	ttaccgagga	agatcccgc	gcaggagagg	catggcaggc	16800
agcggcctga	accgcgcgcg	gcggcgggcc	atgcgcaggc	gcctgagtgg	cgggtttctg	16860
cccgcgctca	tcccataat	cgcgcggcc	atcggcacga	tcccgggcat	agcttccgtt	16920
gcgctgcagg	cgtcgcagcg	ccgttgatgt	gcgaataaag	cctctttaga	ctctgacaca	16980
cctggtcctg	tatatTTTTA	gaatggaaga	catcaatttt	gcgtccctgg	ctccgcggca	17040
cggcacgcgg	ccgttcattg	gcacctggaa	cgagatcgcc	accagccagc	tgaacggggg	17100
cgctttcaat	tggagcagtg	tctggagcgg	gcttaaaaa	ttcggctcga	cgtccgggac	17160
ctatgggaac	aaggcctgga	atagtagcac	ggggcagttg	ttaaggga	agctcaaaga	17220
ccagaacttc	cagcagaagg	tggtggacgg	gctggcctcg	ggcattaacg	gggtggtgga	17280
catcgcgaac	caggccgtgc	agcgcgagat	aaacagccgc	ctggacccgc	ggccgcccac	17340
ggtggtggag	atggaagatg	caactcttcc	gccgccc aaa	ggcgagaagc	ggccgcggcc	17400
cgacgcggag	gagacgatcc	tgacgggtgga	cgagccgccc	tcgtacgagg	aggccgtcaa	17460
ggccggcatg	cccaccacgc	gcacatcgc	gccgctggcc	acgggtgtaa	tgaacccgc	17520
cacccttgac	ctgcctccac	caccacgc	cgtccaccg	aaggcagctc	cggttgtgca	17580
ggccctccg	gtggcgaccg	ccgtgcgcc	cgtccccgc	cgccgcagg	cccagaactg	17640
gcagagcacg	ctgcacagta	tcgtgggccc	ggagtgaaa	agtctgaagc	gccgccgatg	17700
ctattgagag	aggaaagagg	acactaaagg	gagagcttaa	cttgatgtg	ccttaccgcc	17760
agagaacgcg	cgaagatggc	cacccctcg	atgatgccgc	agtggcgcta	catgcacatc	17820
gccgggcagg	acgctcggga	gtacctgagc	ccgggtctgg	tgcagtttgc	ccgcgccacc	17880
gacacgtact	tcagcctggg	caacaagttt	aggaaccca	cgggtggccc	gacccatgat	17940
gtgaccacgg	accggtccca	gcgtctgacg	ctgcgcttcg	tgcccggtga	tcgcgaggac	18000
accacgtact	cgtacaaggc	gcgcttca	ctggccgtgg	gcgacaaccg	ggtgctagac	18060
atggccagca	cgtactttga	catccgcggc	gtcctggacc	gcggtcccag	cttcaaacc	18120
tactcgggca	cggcttacia	cagtttgccc	cccaagggcg	cccccaactc	cagtcagtgg	18180
actgacaaag	aacggcaaaa	tgggtggaaa	ccaccacta	caaaagatgt	tacaaaaaca	18240
ttcggagtag	cagccagggg	agggcttcat	attactgata	aaggactaca	aataggagaa	18300
gatgaaaata	acgaggatgg	tgaagaagag	atatatgcag	acaaaacttt	ccagccagaa	18360
cctcaagtag	gagaggaaaa	ctggcaagat	actgatgttt	tctatggcgg	cagagcgctt	18420
aaaaaggaaa	caaaaatgaa	accatgctat	ggctcttttg	ccagacctac	caatgaaaaa	18480

```

ggagggtcaag ctaaattttt aaatggcgaa aacgggtcaac cttctaaaga tcaagatatt 18540
acattagctt tctttgatct taaacaaaat gacactggaa ctactcaaaa ccagccagat 18600
gttgtcatgt acactgaaaa tgtgtatctg gaaaccccag acacccatgt ggtgtacaaa 18660
cctggcaagg aagatacaag ctccgctgct aaccttacac aacagtccat gcccaacagg 18720
cccaactaca ttggtttcag ggacaacttt gtggggctca tgtattacaa cagcactggc 18780
aacatgggtg tgctggctgg tcaggcctct cagttgaatg ctgtggttga cttgcaagac 18840
agaaacaccg agctgtctta tcagctcttg ctagattctc tgggtgacag aaccagatac 18900
tttagcatgt ggaattctgc ggtggacagc tatgatccag atgtcaggat cattgagaat 18960
cacgggtgtt aagatgagct tccaaattat tgcttccac tggatggatc tggcagcaat 19020
accgcataac aaggtgttaa atatgaaaac ggagctggca atggaagctg gaaagtagat 19080
ggcgaagttg cttctcagaa tcagatcgcc aagggtaatc tgtatgcat ggagataaac 19140
cttcaggcca acctgtggaa gagttttctg tactcgaacg tggcgtgta tctaccagac 19200
tcctacaagt acacgccggc caacatcacg ctgcccacca acaccaacac ctacgagtac 19260
atgaacggcc gcgtggtggc accctcgctg gtggatgcct atgtcaacat cggtgcccg 19320
tggtcgctgg accccatgga caacgtcaac ccttcaacc accaccgaa cgcgggtctg 19380
cgctaccgct ccatgcttct gggcaacggc cgctacgtgc cttccacat ccaagtgc 19440
caaaagttct ttgccatcaa gaacctgctc ctgcttccg gttcctacac ctacgagtgg 19500
aacttccgca atgtgtcaa catgatcctg caaagttccc tcggcaacga cctgcgcgtc 19560
gacggcgcc cgtccgctt cgacagcgct aacctctatg ccaccttctt cccatggcg 19620
cacaacaccg cctccaccct tgaagccatg ctgcgcaacg acaccaacga ccagtcctt 19680
aacgactacc tctcgccgc caacatgctc tacccaatc cggccaaggc caccaacgtg 19740
cccatctcca tccccctcg caactgggccc gccttccgcg gctggagttt caccgcgtc 19800
aagaccaagg aaactccctc cctcggtcg ggtttcgacc cctactttgt ctactcgggc 19860
tccattccct acctcgacgg aaccttctac ctcaaccaca cttcaagaa ggtctccatc 19920
atgttcgact cctcggtcag ctggcccggc aacgaccggc tgctcacgcc gaacgagttc 19980
gagatcaagc gcagcgtcga cggggagggc tacaacgtgg ccaatgcaa catgactaag 20040
gactggttcc tcgtccagat gctctctcat tacaacattg gctaccaggg cttctacgtg 20100
cctgagggtt acaaggaccg catgtactcc ttcttccgca acttccagcc catgagcagg 20160
caggtggtcg atgagatcaa ctacaaggac tacaaggccg tcacctgcc cttccagcac 20220
aacaactcgg gcttcaccgg ctacctcgca cccacctatg gtcaggggca gccatacccc 20280
gccaaacttc cctaccgct catcgccag acagccgtgc cctccgtcac ccagaaaaag 20340
ttctctcgcg acagggtcat gtggcgcatc ccttctcca gcaacttcat gtccatgggc 20400
gccctcaccg acctgggtca gaacatgctc tacgccaaact cggcccacgc gctcgacatg 20460
accttcgagg tggaccccat ggatgagccc accctctct atcttctctt cgaagttttc 20520
gacgtggtca gagtgcacca gccgcaccgc ggcgtcatcg aggcgtcta cctgcgcacg 20580
cccttctccg ccggaaacgc caccacataa gcatgagcgg ctccagcgaa cgagagctcg 20640
cggccatcgt gcgcgacctg ggctgcgggc cctacttttt gggaaccac gacaagcgct 20700
tccttggtt cctcgccggc gacaagctgg cctgcgccat cgtcaacacg gccggccgcg 20760
agaccggagg cgtgcaactg ctcgccctcg gctggaaccc gcgctcgcg accctgtaca 20820
tgttcgaccc ctttgggttc tcggaccgcc ggctcaagca gatttacagc ttcgagtacg 20880
aggccatgct gcgccgcagc gccctggcct cctcgcccga ccgctgtctc agcctcgagc 20940
agtccacca cttcgtgcac tggcccgacc gaccttgga cgggaacccc accatgaact 21060
tgctgacggg ggtgcccac ggcattctac aatcgccaca ggtgctgcc accctccggc 21120
gcaaccagga ggagctctac cgcttctctg cgcgccactc cccttacttt cgatccacc 21180
gcgcgcccat cgaacacgcc accgcttttg aaaaaatgaa acaactgcgt gtatctcaat 21240
aaacagcact tttattttac atgcaactgga gtatatgcaa gttattttaa agtcgaagg 21300
gttctcgcg tcgtcggtt gcgccgcgt ggggaggggc acgttgcgg actggtactt 21360
gggaagccac ttgaactcgg ggcacaccag tttgggaacc ggaatctcgg ggaaggtctc 21420
gctccacatg gcgcggctca tctgcagggc gccagcatg tccggggcg agatcttgaa 21480
atgcagttg ggaccggtgc tctgcgcgcg cgagttgcgg tacacgggg tgcagcactg 21540
gaacaccatc agactggggt acttcacact ggccagcacg ctcttgctcg tgatctgatc 21600
cttgtcaagg tcctcggcgt tgctcaggcc aaacggggtc atcttgca ca gctggcgcc 21660
cagggaaggg acgctctgag gcttgtggtt aactcgcag tgaatgggca ttagcatcat 21720
ccccgcgccg cgctgcatat tcgggtagag ggccttgaca aaggccgaga tctgtttgaa 21780

```

agcttgctgg	gccttggctc	cctcgctgaa	aaacagcccc	cagctcttcc	cgctgaactg	21840
gttattcccc	caaccggcat	cctgcacgca	gcagcgcgcg	tcatggctgg	tcagttgcac	21900
cacgctccgt	ccccagcgg	tctgggtcac	cttggccttg	ctgggttgct	ccttcagcgc	21960
gcgctgtccg	ttctcgctgg	tcacatccat	ctccaccacg	tggctcttgt	ggatcatcac	22020
cgttccatgc	agacacttga	gctggccttc	cacctcggtg	cagccgtgat	cccacagggc	22080
gcatccggtg	cactcccagt	tcttatgcgc	gatcccgtg	tggctgaaga	tgtaaccttg	22140
caacaggcga	cccatgacgg	tgctaaaggc	tttctgggtg	gtgaagggtca	gttgacagacc	22200
gcgggcctcc	tcgttcatcc	aggtctggca	catcttctgg	aagatctcgg	tctgctctgg	22260
catgagcttg	taagcatcgc	gcaggccgct	gtcgacgcgg	tagcgttcca	tcagcacgtt	22320
catggtatcc	atgcccttct	cccaggacga	gaccagaggc	agactcaggg	ggttgcgcac	22380
gttcaggaca	ccgggggtcg	cgggctcgac	gatgcgtttt	ccgtccttgc	cttccttcaa	22440
cagaaccgga	ggctggctga	atcccactcc	cacgatcacg	gcgtcttcc	ggggcatctc	22500
ttcgtctggg	tctacctttg	tcacatgctt	ggtctttctg	gcttgcttct	tttttggagg	22560
gctgtccacg	gggaccacgt	cctcctccga	agaccgggag	cccaccgct	gatactttcg	22620
gcgcttgggtg	ggcagaggag	gtggcggcga	ggggctcctc	tcctgctccg	gcggatagcg	22680
cgccgacccg	tggccccggg	gcggagtggc	ctctcgctcc	atgaaccggc	gcacgtcctg	22740
actgcccggc	gccattgttt	cctaggggaa	gatggaggag	cagccgcgta	agcaggagca	22800
ggaggaggac	ttaaccaccc	acgagcaacc	caaaatcgag	caggacctgg	gcttcgaaga	22860
gccggctcgt	ctagaacccc	cacaggatga	acaggagcac	gagcaagacg	caggccagga	22920
ggagaccgac	gctgggctcg	agcatggcta	cctgggagga	gaggaggatg	tgctgctgaa	22980
acacctgcag	cgccagtccc	tcatectccg	ggacgccctg	gccgaccgga	gcgaaacccc	23040
cctcagcgtc	gaggagctgt	gtcgggccta	cgagctcaac	ctcttctcgc	cgcgctgccc	23100
ccccaaacgc	cagcccaacg	gcacctgcga	gccccacccg	cgtctcaact	tctatcccgt	23160
ctttgcgggtc	cccaggggcc	ttgccacctt	tcacatcttt	ttcaagaacc	aaaagatccc	23220
cgtctcctgc	cgcgccaacc	gcacccgcgc	cgacgcgctc	ctcgtctctg	ggcccggcgc	23280
gcgcatacct	gatatcgctt	ccctggaaga	ggtgcccaag	atcttcgaag	ggctcggctg	23340
ggacgagacg	cgcgcgggcg	acgctctgaa	agaaacagca	gaggaagagg	gtcacactag	23400
cgccctggta	gagttggaag	gcgacaacgc	caggctgggtc	gtgctcaagc	gcagcgtcga	23460
gtcacccac	ttcgccctacc	ccgcccgttaa	cctcccgcgc	aaggctcatgc	gtcgcatcat	23520
ggatcagctt	atcatgcccc	acatcgaggc	catcatgag	acccaagagc	agcgccccga	23580
ggagccccgg	ccggtggcta	gcgacgagat	gctcgcgcgc	tggctcggga	cccgcgaccc	23640
ccaggctttg	gaacagcggc	gcaagctgat	gctggccgta	gtcctgggtca	ccctcgagct	23700
cgaatgcatg	cgccgcttct	tctgcgaccc	cgagaccctg	cgcaaggctg	aggagaccct	23760
gcactacact	ttcagacacg	gtttcgctcag	gcaagcctgc	aagatctcca	acgtggagct	23820
gaccaacctg	gtctcctgcc	tggggatcct	gcatgagaac	cgcttggggc	agacagtgc	23880
ccactctacc	ctgaagggcg	aggcacggcg	ggactatgtc	cgcgactgcg	tctttctctt	23940
tctatgccac	acatggcaag	cagccatggg	cgtgtggcag	cagtgtctcg	aggacgagaa	24000
cctgaaggag	ctggacaagc	ttcttgctag	aaaccttaaa	aagttgtgga	cgggcttcga	24060
cgagcgcacc	gtcgcctcgg	acctggccga	gatcgttttc	cccagcgcgc	tgaggcatac	24120
gctgaaaggc	gggctgcccg	acttcatgag	ccagagcatg	ttgcaaaact	accgcacttt	24180
cattctcgag	cgctcgggta	tcctgcccgc	cacctgcaac	gccttcccct	ccgactttgt	24240
cccgtgagc	taccgcgagt	gtccccgcgc	gctgtggagc	cactgctacc	tcttgagct	24300
ggctaactac	atctcctacc	actcggacgt	gatcgaggac	gtgagcggcg	aggggctgct	24360
cgagtgccac	tgccgctgca	acctgtgctc	cccgcaccgc	tccttgggtct	gcaacccccca	24420
gctcctgagc	gagacccagg	tcacgcgtac	cttcgagctg	caaggctccgg	agaagtccac	24480
cgctccgctg	aaactcacgc	cggggttggtg	gacttccgcg	tacctgcgca	aatttgtacc	24540
cgaagactac	cacgcccacg	agataaagtt	cttcgaggac	caatcgcgtc	cgcagcacgc	24600
ggatctcacg	gcctgcgtca	tcacccaggg	cgcgactcctc	gccccatttc	atgccatcca	24660
aaaatccgc	caagagtttc	ttctgaaaaa	gggtagaggg	gtctacctgg	acccccagac	24720
ggcgagggtg	ctcaacccgg	gtctccccca	gcatgccgag	gaagaagcag	gagccgctag	24780
tggaggagat	ggaagaagaa	tgggacagcc	aggcagagga	ggacgaatgg	gaggaggaga	24840
cagaggagga	agaattggaa	gagggtggaag	aggagcaggc	aacagagcag	cccgtcgccg	24900
caccatccgc	gccggcagcc	ccggcggtca	cggatacaac	ctccgctccg	gtcaagcctc	24960
ctcgtagatg	ggatcgagtg	aagggtgacg	gtaagcacga	gcggcagggc	taccgatcat	25020
ggaggggcca	caaagccgcg	atcatcgcct	gcttgcaaga	ctgcgggggg	aacatcgctt	25080

tcgcccgcgc	ctacctgctc	ttccaccgcg	gggtgaacat	ccccgcgaac	gtgttgcat	25140
actaccgtca	ccttcacagc	taagaaaaag	caagtaagag	gagtcgcccg	aggaggcctg	25200
aggatcgccg	cgaacgagcc	ctcgaccacc	agggagctga	ggaaccggat	cttccccact	25260
ctttatgcca	tttttcagca	gagtcgaggt	cagcagcaag	agctcaaagt	aaaaaatcgg	25320
tctctgcgct	cgctcaccgc	cagttgcttg	taccacaaaa	acgaagatca	gctgcagcgc	25380
actctcgaag	acgccgaggg	tctgttccac	aagtactgcg	cgctcactct	taaagactaa	25440
ggcgcgccca	cccggaaaaa	aggcgggaa	tacctcatcg	ccaccatgag	caaggagatt	25500
cccacccctt	acatgtggag	ctatcagccc	cagatggggc	tggccgcggg	cgctcccag	25560
gactactcca	cccgcatgaa	ctggctcagt	gccggcccct	cgatgatctc	acgggtcaac	25620
ggggctccgta	accatcgaaa	ccagatattg	ttggagcagg	cggcggtcac	atccacgccc	25680
agggcaaagc	tcaaccgcgc	taattggccc	tccaccctgg	tgtatcagga	aatccccggg	25740
ccgactaccg	tactacttcc	gcgtgacgca	ctggccgaag	tccgcatgac	taactcaggt	25800
gtccagctgg	ccggcggcgc	ttcccgggtc	ccgctccgcc	cacaatcggg	tataaaaacc	25860
ctgatgatcc	gaggcagagg	cacacagctc	aacgacgagt	tggtagctc	ttcgatcggg	25920
ctgcgaccgg	acggagtgtt	ccaactagcc	ggagccggga	gatcatcctt	cactcccaac	25980
caggcctacc	tgaccttgca	gagcagctct	tcggagcctc	gctccggagg	catcggaacc	26040
ctccagtctc	tggaggaggt	tgtgcctctg	gtctacttca	accccttctc	gggatcgcca	26100
ggcctctacc	tgtaccgagt	cataccgaac	ttcgacgcag	tgagagaagc	ggtggacggc	26160
tacgactgaa	tgtcccatgg	tgactcggct	gagctcgctc	ggttgaggca	tctggaccac	26220
tgccgcgcgc	tgcgctgctt	cgcccgggag	agctgcggac	tcatctactt	tgagctgccc	26280
gaggagcacc	ccaacggccc	tgcacacgga	gtacggatca	ccgtagaggg	caccgccgag	26340
tctcacctgg	tcaggttctt	caccagcaaa	cccttccctg	tcgagcggga	ccggggcgcc	26400
accacctaca	ccgtctactg	catctgtcct	accccaaagt	tgcagagaaa	tttttgctgt	26460
actctttgtg	gtgagtttaa	taaaagctga	actaagaacc	tactttggaa	tcccttgctg	26520
tcatcaaata	cacaagacca	tcaacttcac	ctttgaggaa	caggtgaact	ttacctgcaa	26580
gccacacaag	aagtacgtca	cctgggtttta	ccagaacact	actctagcag	tagccaacac	26640
ctgctcgaac	gacggtgttc	ttcttccaaa	caatctcacc	agtggactaa	ctttctcagt	26700
gaaaagggca	aagctaattc	ttcatcgccc	tattgtagaa	ggaacttacc	agtgtcagag	26760
cggaccttgc	ttccacagtt	tcactttggg	gaacgttacc	ggcagcagca	cagtcgctcc	26820
agaaactaac	cttctttctg	atactaacac	tcctaatacc	ggaggtgagc	tctgggttcc	26880
ctctctgaca	gaggggggta	gtcatattga	agcggtcggg	tatttgattt	taggggtggg	26940
cctgggtggg	tgcatagcgg	tgctatatta	ccttccttgc	tgggtcgaaa	tcagggtatt	27000
tatctgctgg	gtcagacatt	gtggggaggga	accatgaagg	ggctcttgct	gattatcctt	27060
tccctgggtg	gggggttact	ggcctgccac	gaacagccac	gatgtaacat	caccacaggc	27120
aatgagagga	acgactgctc	tgtagtgate	aaatgcgagc	accagtgtcc	tctcaacatt	27180
acattcaaga	ataagaccat	gggaaatgta	tgggtgggat	tctggcaacc	aggagatgag	27240
cagaactaca	cggctactat	ccatggtagc	gatggaaatc	acactttcgg	tttcaaattc	27300
atttttgaag	tcatgtgtga	tatcacactg	catgtggcta	gacttcatgg	cttgtggccc	27360
cctaccaagg	agaacatggg	tgggttttct	ttggcttttg	tgatcatggc	ctgtgcaatg	27420
tcagggtctgc	tggtaggggc	tctagtgtgg	ttcctgaagc	gcaagcccag	gtacggaaat	27480
gaggagaagg	aaaaattgct	ataaatcttt	ttctcttcgc	agaaccatga	atactttgac	27540
cagtgtcgtg	ctgctctctc	ttttagtatt	taatgtggaa	tgtgccgac	ctattctagt	27600
tagtgtagat	tggggaaaaa	atcttacatt	agagggtcct	aaagaaacac	cagttgaatg	27660
gtgggggtgga	agaaacatac	aacaactgtg	catagggaat	caaaccaaac	ataaagagct	27720
aagtcacaga	tgtaatgtcc	agaacataac	tttactgttt	gtaaataacta	gttttaattg	27780
agactacttt	gggttttaaaa	atgataacag	cggtatgaaa	cattataaag	tcacagttat	27840
accccctaaa	ccctccactc	ggaaacctct	ttctcctcca	cactatgtaa	acgcaactat	27900
ggggcaaaaac	ctaaccattg	tggggcctgc	aaacattcca	gttacttggc	ttagtgaata	27960
tggcacgttg	tgtgagggca	aaaaaatttt	gcacaaagaa	ttaaatcaca	cctgtaacga	28020
acagaacctc	acgttactgt	ttgttaatat	gacacacaac	ggaccatatt	ttggctttga	28080
caaatacaac	attgatagag	agcagtatga	ggtttctatt	attagtttgt	ttaaagttgg	28140
cgctggaacg	aagaaaattg	ggaaaggaca	gaaaaaggag	gaaaagacaa	aaccaaactc	28200
tagtgatttg	ggacaaaagc	aatccagacc	aaagaaaaaa	gatattgttg	aagaggtcca	28260
aatcaaaaca	ggagaaaatc	gaacccttgt	tgggtccacct	ggaaaagtgt	attggattaa	28320
actttccagt	ggaaacaata	atgttcttaa	gttgtgtaat	ggcgacaagt	atattaaaca	28380

cacatgtgat	ggtcaaaatt	taacattaat	taatgtgact	agaatztatg	acggaactta	28440
ttatggttct	agcaatgatg	gctcaagtca	ttacaaagt	accatctatg	aattacacaa	28500
agttaataaa	actaaatcta	tgcttaagcc	atacactaca	aaaagaacta	cagtgaatgc	28560
aacagatgac	agtgtcaca	aaattgcttt	gcagcaggaa	aataatgggc	aaacagaaaa	28620
tgatcaagaa	tcaaaaattc	catctgtac	tgtggcaatc	gtggtgggag	tgattgcggg	28680
cttcataact	ataatcattg	tcattctgtg	ctacatctgc	tgccgcaagc	gtcccagggc	28740
atacaataat	atggtagacc	cactactcag	cttctcttac	tgagactcag	tcactttcat	28800
ttcagaacca	tgaaggcttt	cacagcttgc	gttctgatta	gcataattac	acttagttta	28860
gcagcaccta	aaccagaagt	atatacacia	gttaatgtca	ctaggggtgg	gaatgctaca	28920
ctagatggac	catttaacaa	taacacatgg	acaagatatc	atgatgatgg	gagaaaaaac	28980
ggatggatga	atatttgtaa	atggtcagac	ccatcatata	catgtcatag	taatggaagc	29040
cttagtatatt	ttgctttcaa	cattagttca	ggtaaataata	aagttcaaag	ttactactaac	29100
agttataatg	gattagatgg	ttatgaaaaa	cttgaagtta	aaatgtttta	tctaacagta	29160
attgagcctc	caaccactag	agcaccacc	acagtttagga	caactaagga	aacaacacag	29220
cctaccactg	taccactac	acatccaacc	accacagtca	gtacaactat	tgagaccact	29280
actcatacta	cacagctaga	cacaacagt	cagaatacta	ctttactgat	tgaattttta	29340
ctaagagga	atgaaagtac	tactgatcag	acagaggcta	cctcaagtgc	cttcagcagt	29400
actgcaaat	taactctgct	tgcttgagct	aatgaaaccg	gagtatcatt	gatgcatggc	29460
cagccttact	caggtttggg	tattcaaat	acttttctgg	ttgtctgtgg	gatctttatt	29520
cttgtgggtc	ttctgtactt	tgtctgtctg	aaagccagag	agaaatctag	tagggccatc	29580
tacaggccag	taatcgggga	acctcagcct	ctccaagtgg	aaggggggtc	aaggaatctt	29640
ctcttctctt	tttcagtatg	gtgatcagcc	atgattccta	ggttcttcct	atttaacatc	29700
ctcttctgtc	tcttcaacat	ctgcgtctgc	ttcgcgccg	tctcgcacgc	ctcgcccgac	29760
tgtctcgggc	ccttccccac	ctacctctc	tttgccctgc	tcacctgcac	ctgcgtctgc	29820
agcattgtct	gcctgggtcg	caccttcctg	cagctcatcg	actggtgctg	cgcgctttac	29880
aattatctcc	accacagtcc	cgaatacagg	gacaagaacg	tagccagaat	cttaaggctc	29940
atctgaccat	gcagactctg	ctgatactgc	tatccctcct	ctccctgccc	cttgctgact	30000
gtaaatattg	ggacatatgg	aatttcttag	actgttatca	agagaaaatg	gatatgcctt	30060
cctattactt	ggtgattgtg	ggtgtagtca	tggctgtctc	ctgcactttc	tttgctatca	30120
tgatctaccc	ctgttttgat	ctcggttggg	actctgttga	ggcattcaca	tacacactag	30180
aaagcagttc	actagcctcc	acgcgcgcac	ccacaccgcc	tccccgcaga	aatcagttcc	30240
ccctgattca	gtacttagaa	gagccccctc	cccggcccc	ttccactggt	agctactttc	30300
acataaccgg	cggcgatgac	tgaccaccac	ctggacctcg	agatggacgg	ccaggcctcc	30360
gagcagcgca	tcctgcaact	gcgcgtccgt	cagcagcagg	agcggggccg	caaggagctc	30420
ctcgatgcca	tcaacatcca	ccagtgcagg	aagggcatct	tctgcctggg	caaacaggca	30480
aagatcacct	acgagctcgt	gtccaacggc	aaacagcatc	gcctcaccta	tgagatgccc	30540
cagcagaagc	agaagttcac	ctgcatgggt	ggcgtcaacc	ccatagtcac	caccagcag	30600
tcgggcgaga	ccagcggctg	catccactgc	tcctgcgaaa	gccccgagtg	catctactcc	30660
ctcctcaaga	ccctttgcgg	acttcgcgac	ctcctcccca	tgaactgatt	gattaaagcc	30720
cagaaaccaa	tcaaaccccc	ttccccatca	ccccaaataa	caatcattgg	aaataatcat	30780
tcaataaaga	tactttactt	gaaatctgaa	agtatgtctc	tgggtgtagt	gttcagcagc	30840
acctcggtac	cctcctccca	gctctgggtac	tccagtcccc	ggcggggcgg	gaacttcttc	30900
cacaccttga	aagggtatgtc	aaattcctgg	tccacaattt	tcattgtctt	ccctctcaga	30960
tgtcaaagag	gctccgggtg	gaagatgact	tcaaccccg	ctacccctat	ggctacgcgc	31020
ggaatcagag	tatccccctt	ctcactcccc	cctttgtctc	ctccgatgga	ttccaaaact	31080
tcccccttgg	ggtcctgtca	ctcaaactgg	ctgatccaat	cgccatcgcc	aatgggaatg	31140
tctcactcaa	ggtgggaggg	ggactcactg	tagaacaaca	gtctggaaaa	ctgagtgtgg	31200
ataactaagg	acccttgcaa	gttgcaaatg	acaacaaatt	ggagctatct	tatgatgatc	31260
catttaagg	agagaataac	aaacttgga	ttaaagctgg	ccatgggtta	gcagttgtta	31320
ctaaagaaaa	cacaagtctt	cctagtctag	ttggaacact	tgtagtttta	actggaaaag	31380
gaataggtag	tggatcaagt	gcacatggag	gaactattga	tgtgaagact	ggtgaaggag	31440
gtgggttatc	atttgatgaa	aaaggagact	tagtagcttg	ggacaaaaaa	aatgatacac	31500
gcaccctttg	gacaacacct	gatccttctc	caaattgcaa	agttgaaaca	gcaagagact	31560
caaagctaac	cttagcactt	acaaaatgtg	gtagtcaa	tttggccact	gtatctttac	31620
ttgttggtac	gggcaaatat	gctattataa	gtgacacagt	caacccaaag	cagttctcta	31680

```

ttaagttact gtttaatgac aaggggtgttt tgттаagtga ctcaaactctt gatgggacat 31740
attggaacta tagaagcaac aataacaaca taggcactcc ttataaagag gctgttggtt 31800
ttatgccaa gacaaacagct taccctaagc caaccaacaa caccagcaca gatccggata 31860
aaaaagtgag tcaaggtaaa aataaaattg taagcaatat ttatcttgga ggagagggtat 31920
atcaaccagg atttattgtt gttaaattta atcaggaaac tgatgccaat tgtgcatact 31980
ctattacatt tgattttgga tggggtaagg tgtataagga tcctatacca tatgatacct 32040
cttcttttac tttctcatat atcgctcaag aatgaaagac caataaacgt gtttttcatt 32100
gaaaattttc atgtatcttt attgattttt acaccagcac gggtagtcag tctcccacca 32160
ccagcccatt tcacagtgtg aacaattctc tcagcacggg tggccttaaa taggggaatg 32220
ttctgattag tgcgggaact ggatttagtg tctataatcc acacagtttc ctggcgagcc 32280
aaacggggat cgggtgattga gatgaagccg tcctctgaaa agtcttccaa gcgggcctca 32340
cagtcacaagg tcacagtctg gtggaatgag aagaacgcac agattcatac tcggaaaaca 32400
ggatgggtct gtgcctctcc atcagcgcgc tcaacagtct ctgccgccgg ggctcgggtgc 32460
ggctgctgca gatgggatcg ggatcgcaag tctctctgac tatgatcccc acagccttca 32520
gcatcagtct cctgggtgct cgggcacagc acctcatcct gatctcgtct atgttctcac 32580
agtaagtgc gacataatc accatgttat tcagcagccc ataattcagg gtgctccagg 32640
caaagctcat gttggggatg atggaaccca cgtgaccatc ataccaaatg cggcagtata 32700
tcaggtgcct gcccctctg aacacactgc ccatatacat gatctctttg ggcatgtttc 32760
tgttcacaat ctggcgggtac caggggaagc gctgggtgaa catgcacccg taaatgactc 32820
tcctgaacca cacggccagc aggggtgctc ccgcccggca ctgcagggag cccggggagc 32880
aacagtggca atgcaggatc cagcgtcgt acccgtcac catctgagct ctcaccaagt 32940
ccagggtagc ggggcacagg cactctgaca tacatctttt taaaattttt atttcctctg 33000
gggtcaggat catatcccag gggactggaa actcttgagg cagggtaaag ccagcagcac 33060
atggtaatcc acggacagaa cttacattat gatattcagc atgatcacia tcgggcagca 33120
gggggtgttg ttcagtcagt gaggccctgg tctcctctc agatcgtggg aaacggggccc 33180
tgcggtatgg atgatggcgg agcgaggtcg attgttctc ggtgctcatt gtagtgaccc 33240
ctcttgcgta ccttgctgta ctctgcccag cagaaagtgg cccgggaaca gcagataccc 33300
ctcctccgct cgtcctttcg ctgctgcccg tcagtcaccc aactgaagta catccattcc 33360
cgaaggttct ggagaagttc ctctgcatct gatgaaacaa aaagcccgtc catgcgaatt 33420
cccctcatca catcagccag gactctgtag gccatcccc tccagttaat gctgccttgt 33480
ctatcattca gagggggcgg tggcaggatt ggaagaacca ttattttttt actccaaacg 33540
gtcgcgcagc aatataaaat tcaggtcacg gaggtggcac ctctctcctc cactgttttg 33600
gtggaacacg acagccaaat caaaaattat gcgattctca aggtgctcta ctgtggcttc 33660
cagcagaggc tccacacgta catccagaaa caacagcaca ttaaaagcgg gcccgccatc 33720
ctgctcctca atcatcatat tacagtccct aaccatcccc aggtaatttt cgtttttcca 33780
gccttgaatt atcgctacaa actggttggt taagtccact ccacacatgt tgaaaagctc 33840
ccacagcgcc cctccacct tcataagcag gcagaccttc ataatagcaa cagatctggc 33900
tgctccacaa cctgcagcgt gttcaaaaca acaagattca atgagtttct gccctctgcc 33960
ctgagctcgc gtctcagtgt tagctgtaaa aagtcactca agtcctcggc cactacagat 34020
gccaagtcag agccagggtc aagcgttgga ctggcaagcg tgagggagta tcttagagct 34080
ccataattag ccccaaaaa ctgcatgctg gaataagctc tctttgtgtc accggtgatt 34140
ccttccaaga ggtgagtgat aaagcgaggt aggtgctctt taatcatagc agtaatggaa 34200
aagtcctcta aataagtcac tagggcccca gggaccacaa tgtggtagct gacaccgcgc 34260
cgctcaagca tggttagtag agatgagagt ctgaaaaaca gaaagcatgc actaaaccag 34320
ggtggctatt ctactgaag gaaaaatcac tctctccagc agcagggttc ccactgggtg 34380
gccctcgcgg acatacaaaa atcggtccgt gtggttaaag agcagcacag agagatcatg 34440
tcttctcccg gcaaatatca catcggaact ggtagtatg cccctggcat gtagtcatg 34500
caaggccata aatctgccct gatatccagt aggaaccagg acgctcactc tgaggtgaag 34560
caatatcacc ccatgcggag gaatgtggaa agattcgggg caaaagaaga tatatctatt 34620
gttagccctt tcttgacgg gagcgatccc tccagggtcta tctatgaaag catacagaga 34680
ttcagccata gctcagcccg cttaccagta gacagagagc acagcagtac aagcgccaac 34740
agcagcgact gactaccac tgaccagct ccctatttaa aggcacctta cactgacgta 34800
atgaccaaa gttctaaaaa cccgccaaaa aaaaacacac acgcccctggg tgtttttcgc 34860
gaaaacactt ccgctgtgtc acttccctcg atcgatttcg tgactcaact tccgggttcc 34920
cacgttacgt cacttctgcc cttacatgtg gctccgcgt agggcgccat cttgccacg 34980

```

```

tccaaaatgg cttccatgtc agaccacgcc tccgcgccga ccgttgacgc cgtgcgtcgt 35040
gacgtcagta acggctcttc caacagccaa tcagcgtcgg cccgcacca aattcaaaag 35100
ctcatttgca tattaacttt tgtttacttt gtgggggtata ttattgatga tg 35152

```

<210> 2

<211> 33359

<212> DNA

<213> Artificial Sequence

<220>

<223> pAd36DE1DE4Ad5Orf6

<400> 2

```

aaaccatcaa taatataccc cacaaagtaa acaaaagtta atatgcaaat gagctttttga 60
atttggggcg tggccgacgc tgattggacg agaagtgggtg atgcaaatga cgtcacgacg 120
cacggccgac ggtcgccgcg gaggcgtggc ctagtcggga agcaagtgcg ggggctgatg 180
acgtataaaa aagcggactt tagaccggga aacggccgat tttcccgcgg tcacgcccgg 240
atatgaggta attctgggcg gatgcaagtg aaattaggcc attttggcgc gaaaactgaa 300
tgaggaagtg aaaagtgaaa aataccgggc ccgcccaggg cggaatatatt accgagggcc 360
gagagacttt gaccgattac gtggggggttt cgattgcggt gtttttttcg cgaatttccg 420
cgtccgtgtc aaagtccggt gtttatgtca cagcggccgc atttaaattg cgcgcctagg 480
ttttgagtag tgggcgtggc taaggtgagt ataaaggcgg tgtcttacga gggctctttt 540
gcttttctgc agacatcatg aacgggacgc gcggggcctt cgaagggggg ctttttagcc 600
cttatttgac aaccgcctg ccgggatggg ccggagttcg tcagaatgtg atgggatcta 660
cgggtggatgg gcgcccagtg cttccagcaa attcctcgac catgacctac gcgaccgtgg 720
ggagctcgtc gctcgacagc accgcccagc ccgcccagc cgcagccgcc atgacagcga 780
cgagactggc ctcgagctac atgcccagca gcagcagtag cccctctgtg cccagttcca 840
tcatcgccga ggagaaactg ctggccctgc tggcagagct ggaagccctg agccgccagc 900
tggccgcctt gaccagcag gtgtccgagc tccgcgagca acagcagcag caaaataaat 960
gattcaataa acacagattc tgattcaaac agcaaagcat ctttattatt tatttttttcg 1020
cgcgcggtag gccctggtcc acctctcccg atcattgaga gtgcggtgga ttttttccag 1080
gacccggtag aggtgggatt ggatgttgag gtacatgggc atgagcccg tccgggggtg 1140
gaggtagcac cactgcatgg cctcgtgctc tggggtcgtg ttgtagatga tccagtcata 1200
gcagggggcg tgggcgtggt gctggatgat gtccttaagg aggagactga tggccacggg 1260
gagccccttg gtgtaggtgt tggcgaagcg gttgagctgg gagggatgca tgcgggggga 1320
gatgatgtgc agtttggcct ggatcttgag gttggcaatg ttgccgccca gatccgcct 1380
ggggttcatg ttgtgcagga ccaccaggac ggtgtagccc gtgcacttgg ggaacttatc 1440
atgcaacttg gaagggaaat cgtggaagaa tttggagacg ccttgtgcc cgcccaggtt 1500
ttccatgcac tcatccatga tgatggcgat gggcccggtg gctgcggctt tggcaaagac 1560
gtttctgggg tcagagacat cgtaattatg ctccctgggt agatcatcat aagacatttt 1620
aatgaatttg gggcgagggg tgccagattg ggggacgata gttccctcgg gccccggggc 1680
aaagtcccc tcacagatct gcatctccca ggctttcatc tcggaggggg ggatcatgtc 1740
cacctgcggg gcgatgaaaa aaacggtttc cggggcgggg gtgatgagct gcgaggagag 1800
caggtttctc aacagctggg acttgccgca ccgggtcggg ccgtagatga ccccgatgac 1860
gggttgacag tggtagttca aggacatgca gctgccgtcg tcccggagga ggggggccac 1920
ctcgttgagc atgtctctga cttggaggtt ttcccggacg agctcgccga ggaggcggtc 1980
cccgccagc gagaggagct cttgcaggga agcaaagttt ttcaggggct tgagtccgtc 2040
ggccatgggc atcttggcga gggctctgca gaggagctcc aggcgggtcc agagctcgg 2100
gacgtgctct acggcatctc gatccagcag acttcctcgt ttcgggggtt gggacgactg 2160
cgactgtagg gcacgagacg atgggcgtcc agcgcggcca gcgtcatgtc cttccaggtt 2220
ctcaggggtc gcgtgagtgt ggtctccgtc acggtgaagg ggtgggcccc gggctgggcg 2280
cttgcaaggg tgcgcttgag actcatcctg ctggtgctga aacgggcacg gtcttcgccc 2340
tgcgcgtcgg cgagatagca gttgaccatg agctcgtagt tgagggcctc ggcggcgtgg 2400
cccttggcgc ggagcttgcc cttggaagag cgggacagag aagggaattg 2460
agggcgtaga gcttgggtgc gagaaaaacg gactcggggg cgaaggcgtc cgctccgag 2520

```

tgggcgcaga	cggtctcgcga	ctcgacgagc	caggtgagct	cggtctgctc	ggggcctcaaaa	2580
accagtttttc	ccccgttctt	tttgatgcgc	ttcttacctc	gcgtctccat	gagtctgtgt	2640
ccgcgctcgg	tgacaaacag	gctgtcgggtg	tccccgtaga	cggacttgat	gggcctgtcc	2700
tgcagggacg	tcccgcgggtc	ctcctcgtag	agaaactcgg	accactctga	gacaaaggcg	2760
cgcgctccacg	ccaagacaaa	ggaggccacg	tgcgaggggt	agcgggtcgtt	gtccaccagg	2820
gggtccacct	tttccaccgt	gtgcagacac	atgtccccct	cctccgcctc	caagaagggtg	2880
attggcttgt	aggtgtaggc	cacgtgaccg	gggggtccccg	acgggggggt	ataaaaagggtg	2940
gcgggtctgt	gctcgtcctc	actctcttcc	gcgtcgtgt	ccacgagcgc	cagctgttgg	3000
ggtaggtatt	ccctctctag	agcggggcatg	acctcggcac	tcagggttgtc	agtttctaga	3060
aacgaggagg	atttgatgtt	ggcctgccct	gccgcgatgc	tttttaggag	actttcatcc	3120
atctggtcag	aaaagacaat	ttttttattg	tcaagcttgg	tggcaaagga	gccatagagg	3180
gcgttggata	gaagcttggc	gatggatctc	atggtctgat	ttttgtcacg	gtcggcgcg	3240
tccttggccg	cgatgtttag	ctggacatac	tcgcgcgcga	cgcacttcca	ttcggggaag	3300
acgggtggtg	gctcgtcggg	cacgatcctg	acgcgccagc	cgcggttatg	cagggtgacc	3360
aggtccacgc	tgggtggccac	ctcgccgcgc	aggggctcgt	tgggtccagca	gagtctgccg	3420
cccttgccgc	agcagaacgg	gggcagcaca	tcaagcagat	gctcgtcagg	gggggtccgca	3480
tcgatggtga	agatgcccg	acagagtctc	ttgtcaaaat	aatcgatttt	tgaggatgca	3540
tcattccaagg	ccatctgcca	ctcgccggcg	gccagcgctc	gctcgtaggg	gttgaggggc	3600
ggaccccagg	gcattgggatg	cgtgagggcg	gaggcggtaca	tgccgcagat	gtcatacaca	3660
tagatgggct	ccgagaggat	gccgatgtag	gtgggataac	agcgcctccc	gcggatgctg	3720
gcgcgcacgt	agtcatacaa	ctcgtgcgag	ggggccaaga	aggcggggcc	gagattggtg	3780
cgctggggct	gctcggcgcg	gaagacgatc	tggcgaaaga	tggcgtgcga	gtttgaggag	3840
atggtgggct	gttggaagat	gttaaagtgg	gcgtgaggca	ggcggaccga	gtcgcggatg	3900
aagtgcgcgt	aggagtcttg	cagcttggcg	acgagctcgg	cggtgacgag	gacgtccatg	3960
gcacagtagt	ccagcgtttc	gcggatgatg	tcataaccgg	cctctccttt	cttctcccac	4020
agctcgcggg	tgagggcgta	ctcctcgtca	tccttccagt	actcccggag	cgggaatcct	4080
cgatcgtccg	cacggtaaga	gccagcatg	tagaaatggt	tcacggcctt	gtaggacacg	4140
cagcccttct	ccacggggag	ggcgtaagct	tgagcggcct	tgccggagcga	ggtgtgcgtc	4200
agggcgaaag	tgtctctgac	catgactttc	aagaactggt	acttgaaatc	cgagtcgtcg	4260
cagccgccgt	gctcccagag	ctcgaaatcg	gtgcgcttct	tcgagagggg	gttaggcaga	4320
gcgaaagtga	cgtcattgaa	gagaatcttg	cctgcccggg	gcattgaaat	gcgggtgatg	4380
cggaaaggcg	ccgggacgga	ggctcgggtg	ttgatgacct	gggcggcgag	cacgatctcg	4440
tcgaagccgt	tgatgttgtg	cccagacgatg	tagagtcca	tgaatcgcgg	gcggcctttg	4500
atgtgcggca	gcttttttgag	ctcctcgtag	gtgaggtoct	cggggcattg	cagtccgtgc	4560
tgctcgcgag	cccactcctg	gagatgtggg	ttggcttgca	tgaaggagc	ccagagctcg	4620
cgggccatga	gggtctggag	ctcgtcgcga	aagagacgga	actgctggcc	cacggccatc	4680
ttttcgggtg	tgacgcagta	gaagggtgag	gggtcccgt	cccagcgatc	ccagcgtaag	4740
cgcacggcga	gatcgcgagc	gagggcgacc	agctcggggg	ccccggagaa	tttcatgacc	4800
agcatgaagg	ggacgagctg	cttgccgaag	gacccccatc	aggtgtagg	ttctacatcg	4860
taggtgacaa	agagccgctc	cgtgcgagga	tgagagccga	ttgggaagaa	ctggatttcc	4920
tgccaccagt	tggacgagtg	gctgttgatg	tgatgaaagt	agaaatcccg	ccggcgaaacc	4980
gagcactcgt	gctgatgctt	gtaaaagcgt	ccgcagtact	cgcagcgctg	cacgggctgt	5040
acctcatcca	cgagatacac	agcgcgtccc	ttgaggagga	acttcaggag	tggcggtcct	5100
ggctggtggt	tttcatgttc	gcctgcgtgg	gactcaccct	ggggctcctc	gaggacggag	5160
aggctgacga	gcccgcgcgg	gagccaggtc	cagatctcgg	cgcggcgggg	gcggagagcg	5220
aagacgaggg	cgcgagtttg	ggagctgtcc	atggtgtcgc	ggagatccag	gtccgggggc	5280
agggttctga	ggttgacctc	gtagaagcgg	gtgagggcgt	gcttgagatg	cagatggtac	5340
ttgatttcta	cgggtgagtt	ggtggccgtg	tccacgcatt	gcattgagccc	gtagctgcgc	5400
ggggccacga	ccgtgccgcg	gtgcgctttt	agaagcgggtg	tcgcggagcg	gtccccggcg	5460
gcagcgggcg	ttccggcccc	gcgggcaggg	gcggcagagg	cacgtcggcg	tggcgctcgg	5520
gcagggtccc	gtgctgcgcc	ctgagagcgc	tggcgtgcgc	gacgacgcgg	cgggttgacat	5580
cctggatctg	ccgcctctcg	gtgaagacca	cgggccccgt	gactttgaac	ctgaaagaca	5640
gttcaacaga	atcaatctcg	gcgtcattga	cggcgccctg	acgcaggatc	tcttgacagt	5700
cgcccagagtt	gtcctggtag	gcgatctcgg	acatggaactg	ctcgatctcc	tcctcctgga	5760
gatcgcgcgc	gcccgcgcgc	tccacgggtg	cggcgagggtc	attcgagatg	cgaccattta	5820

gctgcgagaa ggcgcccagg ccgctctcgt tccagacgcg gctgtagacc acgtccccgt 5880
 cggcgctcgcg cgcgcgcatg accacctgcg cgagggttag ctccacgtgc cgcgtgaaga 5940
 cggcgtagtt gcgcaggcgc tggaagaggt agttgagggg ggtggcgatg tgctcggtag 6000
 cgaagaagta catgatccag cggcgccagg gcatctcgt gatgtcgccg atggcctcca 6060
 gcctttccat ggctcgttag aaatccacgg cgaagttgaa aaactgggcg ttgctgggccc 6120
 agaccgtgag ctgctcttcc aggagccgga tgagctcggc gatgggtggcg cgcacctcgc 6180
 gctcgaaatc cccggggacc tcctcctctt cctctcttct catgacgacc tcttcttcta 6240
 tttcttctc tgggggcggt ggtggtggcg gggcccgacg acgacggcga cgcaccggga 6300
 gacggctcgac gaagcgctcg atcatctccc cgcggcgggc acgcatgggt tcggtgacgg 6360
 cgcgaccccc ttccgcgagga cgcagcgtga agacgcgcgc ggtcatctcc cggtaatggg 6420
 gcggttcccc gttgggcagc gatagggcgc tgacgatgca tcttatcaat tgcggtgtag 6480
 gggacgtgag cgcgtcgcga tcgaccggat cggagaatct ttcgaggaaa gcgtctagcc 6540
 aatcgcagtc gcaaggtaag ctcaaacacg tagcagccct gtggacgctg ttagaattgc 6600
 ggttgctgat gatgtaattg aagtaggcgt ttttaaggcg gcggatgggt gcgaggagga 6660
 ccaggctcctt ggggtcccgt tgctggatgc ggagcgcgtc ggccatgcc caggcctggc 6720
 cctgacaccg gctcagggtt ttgtagtagt catgcatgag cctttcaatg tcatcactgg 6780
 cggaggcggga gtcttccatg cgggtgaccc cgacgcccc ctgagcgctgc acgagcgcca 6840
 ggtcggcgac gacgcgctcg gcgaggatgg cctgttgcac gcgggtgagg gtgtcctgga 6900
 agtcatccat gtcgacgaag cgggtgtagg ccccggtgtt gatgggttag gtgcagttgg 6960
 ccatgagcga ccagttgacg gtctgcaggc cgggttgcac gacctcggag tacttgtagc 7020
 gcgagaaggc gcgcgagtcg aagacgtagt cgttgtaggt gcgcacgagg tactggtagc 7080
 cgactaggaa gtgcggcggc ggctggcggt agagcggccca gcgctgggtg gccggcgccg 7140
 ccggggccag gtctcgcgag atgaggcggt ggtagccgta gaggtagcgg gacatccagg 7200
 tgatgccggc ggcggtgggt gaggcgcgcg ggaactcgcg gacgcggttc cagatgttgc 7260
 gcagcggcag gaaataatcc atggtcggca cggctcggcc ggtgagacgc gcgcagtcac 7320
 tgacgctcta gaggcaaaaa cgaagcggt tgagcgggct ctctctcgt agcctggcgg 7380
 aacgcaaacg ggttaggcg cgtgtgtacc cgggttcgag tcccctcgaa tcaggctgga 7440
 gccgcgacta acgtggtatt ggcaactccc tctcgacccg agcccgatag ccgccaggat 7500
 acggcgagga gccctttttg ccgaccgagt ggggtcgccta gacttgaaag cggccgaaaa 7560
 ccccgccggg tagtggtctg cgcctcgtagt ctggagaagc atcgccaggg ttgagtcgcg 7620
 gcagaaccgc gttcgcggac ggccgcggcg agcgggactt ggtcaccctc ccgatttaaa 7680
 gaccacagc cagccgactt ctccagttac gggagcgcgc cccctttttt ctttttgcca 7740
 gatgcatccc gtcttgcgcc aaatgcgtcc ccccccccg gcgaccaccg cgaccgcggc 7800
 cgtaacaggc gccggcgcta gccagccaca gacagagatg gacttggaag agggcgagg 7860
 gctggcgaga ctggggggcg cgtccccgga gcgacacccc cgcgtgcagc tgcagaagga 7920
 cgtgcgcccg gcgtacgtgc ctgcgcagaa cctgttcagg gaccgcagcg gggaggagcc 7980
 cgaggagatg cgcgactgcc ggtttcgggc gggcagggag ctgcgcgagg gcctggaccg 8040
 ccagcgcgtg ctgcgcgacg aggttttcga gccgaacgag cagacgggga tcagccccgc 8100
 gcgcgcgcac gtggcgggcg ccaacctggt gacggcctac gagcagacgg tgaagcagga 8160
 gcgcaacttc caaaagagtt tcaacaacca cgtgcgcacg ctgatatgcg gcgaggaggt 8220
 ggccctgggc ctgatgcacc tgtgggacct ggcggaggcc atcgtgcaga acccgacag 8280
 caagcctctg acggcgcgagc tgttctcgtt ggtgcagcac agcagggaca acgaggcgtt 8340
 cagggaggcg ctgctgaaca tcgcccagcc cgagggtcgc tggctgctgg agctgatcaa 8400
 catcttgtag agcatcgtag tgcaggagcg cagcctgagc ctggccgaga aggtggcggc 8460
 gatcaactac tcggtgctga gcctgggcaa gttttacgcg cgcaagattt acaagacgcc 8520
 gtacgtgccc atagacaagg aggtgaagat agacagcttt tacatgcgca tggcgctcaa 8580
 ggtgctgacg ctgagcgacg acctgggcgt gtatcgcaac gaccgcatcc acaaggccgt 8640
 gagcacgagc cggcggcgcg agctgagcga ccgcgagctg atgctgagcc tgcgccgggc 8700
 gctggtaggg ggcgccgcgc gcggcgagga gtctacttcc gacatggggg cggacctgca 8760
 ttggcagccg agccggcgcg ccttggaggc cgcctacggt ccagaggact tggatgagga 8820
 tgagggaagag gaggaggatg cacccttgc ggggtactga cgcctcctg atgtgtttt 8880
 agatgtccca gcaagccccg gaccccgcca taagggcggc gctgcaaagc cagccgtccg 8940
 gtctagcacc ggacgactgg gaggccgcga tgcaacgcac catggccctg acgaccgcga 9000
 accccgagtc ctttagacaa cagccgcagg ccaacagact ctcggccatt ctggaggcgg 9060
 tggctccctc tcggaccaac cccacgcacg agaaggtgct ggcgatcgtg aacgcgctgg 9120

cggagaacaa ggccatccgt cccgacgagg ccgggctggt gtacaacgcc ctgctggagc 9180
 gcgtggggccg ctacaacagc acgaacgtgc agtccaacct ggaccggctg gtgacggacg 9240
 tgcgcgaggc cgtggcgagc cgcgagcggg tcaagaacga gggcctgggc tcgctggtgg 9300
 cgctgaacgc cttcctggcg acgcagccgg cgaacgtgcc gcgagggcag gacgattaca 9360
 ccaactttat cagcgcgctg cggctgatgg tgaccgaggt tccccagagc gaggtgtacc 9420
 agtcggggccc ggactacttt ttccagacaa gccggcaggg cctgcagacg gtgaacctga 9480
 gtcaggcttt caagaacctg cgcgggctgt ggggcgtgca ggcgccctg ggcgaccggt 9540
 cgacggtagag cagcttgctg acgcccact cgcggctgct gctgctgctg atcgcgccct 9600
 tcaccgacag cggcagcgtg aaccgcaact cgtacctggg ccacctgctg acgctgtacc 9660
 gcgagggccat agggcaggcg caggtggacg agcagacctt ccaggagatc acgagcgtga 9720
 gccgcgcgtt ggggcagaac gacaccgaca gtctgagggc caccctgaac tttttgctga 9780
 caaatagaca gcagaagatc ccggcgagc acgcactgtc ggccgaggag gaaaggatcc 9840
 tgagatatgt gcagcagagc gtagggctgt tcctgatgca ggaggggtgcc acccccagcg 9900
 ccgcgctgga catgaccgcg cgcaacatgg aacctagcat gtacgccgcc aaccggccgt 9960
 tcatcaataa gctgatggac tacctgcacc gcgcggcgcc catgaacacg gactacttta 10020
 caaacgccat attgaacctg cactggcttc cgccgcggg gttctacacg ggcgagtagc 10080
 acatgccga ccccaacgac ggggtcctgt gggagcaggt ggacagcgcg gtgttctcgc 10140
 cgacctttc aaagcgccag gaggcgccg cgagcgagg cgcggtgggg agggagccct 10200
 ttcctagctt agggagtgtt catagcttgc cgggctcggg gaacagcggc aggggtgagc 10260
 ggccgcgctt gctgggagc gacgagtacc tgaacgactc gctgctgcag ccgccgcggg 10320
 tcaagaacgc catggccaat aacgggatag agagtctggt ggacaaactg aaccgctgga 10380
 agacctacgc tcaggaccat agggagcctg cgcccgcgcc gcggcgacag cgccacgacc 10440
 ggcagcgggg cctgggtgtg gacgacgagg actcgccga cgatagcagc gtgttggact 10500
 tgggcgggag cgggtggggc aaccggttcg cgcactctga gccagactg gggcgccgga 10560
 tgttttgaat gcaaaataaa actcaccaag gccatagcgt gcgttctctt ccttgttaga 10620
 gatgaggcgt gcggtggtgt cttcctctcc tcctccctcg tacgagagcg tgatggcgca 10680
 ggcgacctg gaggttccgt ttgtgcctcc gcggtatatg gctcctacgg agggcagaaa 10740
 cagcattcgt tactcggagc ttggtccgca gtacgacacc actcgctgt acttgggtgga 10800
 caacaagtcg cgggacatcg cttccctgaa ctacaaaaac gaccacagca acttccctgac 10860
 cacggtggtg cagaacaacg atttcacccc cgccgaggcc agcacgcaga cgataaattt 10920
 tgacgagcgg tcgcggtggg gcggtgatct gaagaccatt ctgcacacca acatgccaa 10980
 tgtgaacgag tacatgttca ccagcaagtt taaggcgcg gtgatggtg ctagaaagca 11040
 tcccaaagat gtagatgcca gtgatttaag caaggatata ttagagtata agtggtttga 11100
 gtttacctg cccgagggca acttttccga gaccatgacc atagacctga tgaacaacgc 11160
 catcttgaa aactacttgc aagtggggcg gcagaatggc gtgctggaga gcgatatcg 11220
 agtcaagttt gacagcagga atttcagact gggctgggac ccggtgacca agctggtgat 11280
 gccaggggtc tacacctacg aggccttcca cccggacgtg gtgctactgc cgggctgcgg 11340
 ggtggacttc accgagagcc gcctgagcaa cctctgggc attcgcaaga agcaaccttt 11400
 tcaagagggc ttcagaatca tgtatgagga tctagaaggg ggtaacatcc ccgctctcct 11460
 ggataccaaa aaatatctgg atagcaagaa ggaacttgag gatgctgcca aggaagctgc 11520
 aaagcaacag ggagatggtg ctgtcactag aggcgatacc cacctcactg tagctcaaga 11580
 aaaagcagct gaaaaggagc tagtgatcgt accaattgaa aaggatgaga gcaacagaag 11640
 ttacaacctg atcaaggaca cccatgacac cctgtaccga agctggtacc tgtcctatac 11700
 ctacggggac ccgagaagg ggggtgcagtc gtggacgctg ctcaccaccc cggacgtcac 11760
 ctgcgggcgcg gagcaagtct actggtcgct gccggacctc atgcaagacc cgtcacctt 11820
 ccgctctacc cagcaagtca gcaactaccc cgtggtcggc gccgagctca tgcccttcg 11880
 cgccaagagc ttttacaacg acctcgccgt ctactccag ctctgccgcc cgcccgcgcc 11940
 cctcaccacac gcttcaacc gcttcccga caaccagatc ctctgccgcc cgcccgcgcc 12000
 caccatcacc accgtcagtg aaaacgtgcc tgctctcaca gatcacggga cgcttccgct 12060
 gcgacgagc atccgcggag tccagcgagt gaccgtcact gacgcccgtc gccgcacctg 12120
 tccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtgctctcca gtcgcacctt 12180
 ctaaaaatg tctatttca tctcgcccag caataacacc ggctggggtc ttactaggcc 12240
 cagcaccatg tacggaggag ccaagaagcg ctcccagcag caccocgtcc gcgtcccgcg 12300
 tcacttccgc gctccctggg gagcttaca gggggggcg actgccacc cgcccgccgt 12360
 gcgcaccacc gtcgacgacg tcatcgactc ggtggtcgcc gacgcgcgca actacacccc 12420

cgccccctcc accgtggacg cgggtcatcga cagcgtgggtg gccgacgcgc gcgactatgc 12480
 cagacgcaag agccggcgccg gacggatcgc caggcgccac cggagcacgc ccgccatgcg 12540
 cgccgccccg gctctgctgc gccgcgccag acgcacgggc cgccggggcca tgatgcgagc 12600
 cgcgcgccgc gccgccactg cccccccgc aggcaggact cgcagacgag cggccgcccgc 12660
 cgctgcgcgc gccatctcta gcatgaccag acccaggcgc ggaaacgtgt actgggtgcg 12720
 cgactccgtc acgggcgtgc gcgtgcccgt gcgcacccgt cctcctcgtc cctgatctaa 12780
 tgcttggtgc ctcccccgca agcgcacgat tcaaagcgca aaatcaagga ggagatgctc 12840
 cagggtcgtc ccccgagat ttacggaccc cggaccaga aaccgcgca aatcaagcg 12900
 gttaaaaaaa aggatgaggt ggacgagggg gcagtagagt ttgtgcgcga gttcgctccg 12960
 cggcgccgcg taaattggaa ggggcgcagg gtgcagcgcg tgttgccggc cggcacggcg 13020
 tggtgtgtca cgcccgccga gcggtcctcg gtcaggagca agcgtagcta tgacgaggtg 13080
 tacggcgacg acgacatcct ggaccaggcg cgggagcggg cgggcgagtt cgcctacggg 13140
 aagcggtcgc gcgaagagga gctgatctcg ctgccgtgg acgaaagcaa cccacgcgcg 13200
 agcctgaagc ccgtgaccct gcagcaggtg ctgcccagg cgggtgctgt gccgagccgc 13260
 ggggtcaagc gcgagggcga gagcatgtac ccgaccatgc agatcatggt gcccaagcg 13320
 cggcgcggtg aggcagtgct ggacaccgtg aaaatggatg tggagcccga ggtcaaggtg 13380
 cgccccatca agcaggtggc gccgggcctg ggcgtgcaga ccgtggacat tcagatcccc 13440
 accgacatgg atgtcgacaa aaaaccctcg accagcatcg aggtgcagac cgaccctgg 13500
 ctcccagcct ccaccgctac cgtctccact ttaccgccc ccacggctac cgagcctccc 13560
 aggaggcgaa gatggggcgc cgccagcccg ctgatgccca actacgtgtt gcatccttcc 13620
 atcatcccga cgccgggcta ccgcggcacc cgggtactac ccagccgcag gcgccagcc 13680
 gccaaacgcc gccgcgcac tgccaccgc cgccgtctgg ccccgcccg cgtgcgcgcg 13740
 gtaaccacgc gccggggccg ctgctcgtt ctgcccaccg tgcgctacca cccagcatc 13800
 ctttaaatccg tgtgctgtga tactgttgca gagagatggc tctcacttgc cgcctgcgca 13860
 tccccgtccc gaattaccga ggaagatccc gccgcaggag aggcattggc ggcagcgcc 13920
 tgaaccgcgc cggcgggcg gccatgcgca ggccgctgag tggcggttt ctgcccgcg 13980
 tcatccccat aatcgccgcg gccatcgcca cgatcccggt catagcttcc gttgcgctgc 14040
 aggcgctcga gcgcgcttga tgtgcgaata aagcctcttt agactctgac acacctggtc 14100
 ctgtatattt ttagaatgga agacatcaat ttgctcctc tggctccgcg gcacggcacg 14160
 cggccggttca tgggcacctg gaacgagatc ggcaccagcc agctgaacgg gggcgccctc 14220
 aattggagca gtgtctggag cgggcttaaa aatttcggct cgacgctccg gacctatggg 14280
 aacaaggcct ggaatagtag cacggggcag ttgttaaggg aaaagctcaa agaccagaac 14340
 ttccagcaga aggtggtgga cgggctggcc tcgggcatta acggggtggt ggacatcgcg 14400
 aaccaggccg tgcagcgca gataaacagc cgcctggacc cgcgcccgcc caccggtggtg 14460
 gagatggaag atgcaactct tccgcgcgcc aaaggcgaga agcgcccgcg gcccgacgcg 14520
 gaggagacga tctgcaggt ggacgagccg ccctcgtacg aggagccgt caaggccggc 14580
 atgcccacca cgcgcatcat cgcgcgcgtg gccacgggtg taatgaaacc cgccaccctt 14640
 gacctgcctc caccaccac gccgcctcca ccgcccgcgc aggcagaa ctggcagagc 14700
 ccggtggcga ccgcgtgcg cgcgtcccc gcccgccgc aggcagaa ctggcagagc 14760
 acgctgcaca gtatcgtggg cctgggagtg aaaagtctga agcgccgcg atgctattga 14820
 gagaggaaaag aggacactaa agggagagct taacttgat gtgccttacc gccagagaac 14880
 gcgcgaagat ggccaccccc tcgatgatgc cgcagtgggc gtacatgcac atcgccgggc 14940
 aggacgcctc ggagtacctg agcccgggtc tgggtcagtt tgcccgcgcc accgacacgt 15000
 acttcagcct gggcaacaag tttaggaacc ccacggtggc cccgaccat gatgtgacca 15060
 cggaccggtc ccagcgtctg acgctgcgtc tcgtgcccgt ggatcgcgag gacaccagct 15120
 actcgtacaa ggcgcgcttc actctggcgc tgggcgacaa ccgggtgcta gacatggcca 15180
 gcacgtactt tgacatccgc ggcgctcctg accgcgtcc cagcttcaa ccctactcg 15240
 gcacggctta caacagttt gcccccagg gcgccccaa ctccagtcag tggactgaca 15300
 aagaacggca aatgggtgga caaccacca ctacaaaaga tgttacaaaa acattcggag 15360
 tagcagccag gggagggtt catattactg ataaaggact acaaatagga gaagatgaaa 15420
 ataacgagga tgggtgaaga gagatatatg cagacaaaac tttccagcca gaacctcaag 15480
 taggagagga aaactggcaa gatactgatg ttttctatgg cggcagagcg cttaaaaagg 15540
 aaacaaaaat gaaaccatgc tatggctctt ttgccagacc taccaatgaa aaaggaggtc 15600
 aagctaaatt tttaaatggc gaaaacggtc aaccttctaa agatcaagat attacattag 15660
 ctttctttga tcttaacaa aatgacactg gaactactca aaaccagcca gatgttgtca 15720

tgtacactga	aaatgtgtat	ctggaaaccc	cagacaccca	tgtggtgtac	aaacctggca	15780
aggaagatac	aagctccgct	gctaaccctta	cacaacagtc	catgccaac	aggccaact	15840
acattgggtt	cagggacaac	tttgtggggc	tcatgtatta	caacagcact	ggcaacatgg	15900
gtgtgctggc	tggtcaggcc	tctcagttga	atgctgtggt	tgacttgcaa	gacagaaaca	15960
ccgagctgtc	ttatcagctc	ttgctagatt	ctctgggtga	cagaaccaga	tacttttagca	16020
tgtggaattc	tgcggtggac	agctatgata	cagatgtcag	gatcattgag	aatcacgggtg	16080
ttgaagatga	gcttccaaat	tattgcttcc	cactggatgg	atctggcagc	aataccgcat	16140
atcaagggtg	taaatatgaa	aacggagctg	gcaatggaag	ctggaaagta	gatggcgaag	16200
ttgctttctc	gaatcagatc	gccaagggtg	atctgtatgc	catggagata	aaccttcagg	16260
ccaacctgtg	gaagagtttt	ctgtactcga	acgtggcgct	gtatctacca	gactcctaca	16320
agtacagccc	ggccaacatc	acgctgccca	ccaacaccaa	cacctacgag	tacatgaacg	16380
gccgcgtggg	ggcacctctc	ctgggtggatg	cctatgtcaa	catcggtgcc	cgctggctcg	16440
tggaccccat	ggacaacgtc	aaccccttca	accaccaccg	caacgcgggt	ctgcgctacc	16500
gctccatgct	tctgggcaac	ggccgctacg	tgcccttcca	catccaagt	ccccaaaagt	16560
tctttgccat	caagaacctg	ctcctgcttc	ccggttccca	cacctacgag	tggaaacttcc	16620
gcaaggatgt	caacatgata	ctgcaaagtt	ccctcgga	cgacctgcgc	gtcgacggcg	16680
ctcccgctcg	cttcgacagc	gtcaacctct	atgccacctt	cttccccatg	gcgcacaaca	16740
ccgcctccac	ccttgaagcc	atgctgcgca	acgacaccaa	cgaccagtcc	ttcaacgact	16800
acctctcggc	cgccaacatg	ctctacccaa	tcccggccaa	ggccaacca	gtgcccatct	16860
ccatccccctc	ctcgcaactgg	gccgccttcc	gcggtcggag	tttcaaccgg	ctcaagacca	16920
aggaaactcc	ctccctcggc	tcgggtttcg	acccctactt	tgtctactcg	ggctccattc	16980
cctacctcga	cggaaccttc	tacctcaacc	acaccttcaa	gaaggctctc	atcatgttcg	17040
actcctcggg	cagctggccc	ggcaacgacc	ggctgctcac	gccgaacgag	ttcgagatca	17100
agcgagcgt	cgacggggag	ggctacaacg	tggcccaatg	caacatgact	aaggactggg	17160
tcctcgtcca	gatgctctct	cattacaaca	ttggctacca	gggttcttac	gtgcctgagg	17220
gttacaagga	ccgcatgtac	tccttcttcc	gcaacttcca	gcccattgagc	aggcaggtag	17280
tcgatgagat	caactacaag	gactacaagg	ccgtcacctt	gcccttcacg	cacaacaact	17340
cgggcttcac	cggtacacct	gcaccacca	tgctcagg	gcagccatac	cccgccaaact	17400
tcccctaccc	gctcatcgcc	cagacagccg	tgccctccgt	caccagaaa	aagttcctct	17460
gcgacagggt	catgtggcgc	atcccccttct	ccagcaactt	catgtccatg	ggcgccctca	17520
ccgacctggg	tcagaacatg	ctctacgcca	actcgcccca	cgcgctcgac	atgaccttcg	17580
agggtggacc	catggatgag	cccaccttcc	tctatcttct	cttcgaagtt	ttcgacgtgg	17640
tcagagtga	ccagccgcac	cgcggcgtca	tcgaggccgt	ctacctgcgc	acgcccttct	17700
ccgcccga	cgccaccaca	taagcatgag	cggtccagc	gaacgagagc	tcgcgcccat	17760
cgtgcgcgac	ctgggctgcg	ggccctactt	tttgggaacc	cacgacaagc	gcttccctgg	17820
cttccctcgcc	ggcgacaagc	tgccctgcgc	catcgtcaac	acggccggcc	gcgagaccgg	17880
aggcgtgcac	tggctcgcc	tcggctggaa	cccgcgctcg	cgcacctgct	acatgttcga	17940
cccccttggg	ttctcggacc	gccggctcaa	gcagatttac	agcttcgagt	acgaggccat	18000
gctgcgcgcg	agcgccctgg	cctcctcgcc	cgaccgctgt	ctcagcctcg	agcagtcac	18060
ccagaccgtg	caggggcccg	actccgcgcg	ctgcggactt	ttctgttgca	tgttcttgca	18120
tgccctcgtg	cactggcccg	accgacctat	ggacgggaac	cccaccatga	acttgctgac	18180
gggggtgccc	aacggcatgc	tacaatcgcc	acaggtgctg	cccaccttcc	ggcgcaacca	18240
ggaggagctc	taccgcttcc	tcgcgcgcca	ctcccccttac	tttcgatccc	accgcgcgcg	18300
catcgaacac	gccaccgctt	ttgacaaaat	gaaacaactg	cgtgtatctc	aataaacagc	18360
actttttattt	tacatgcact	ggagtatatg	caagttatatt	aaaagtcgaa	gggggtctctg	18420
cgctcgtcgt	tgtgcgccgc	gctggggagg	gccacgttgc	ggtactggta	cttgggaagc	18480
cacttgaact	cggggatcac	cagttttggga	accggaatct	cggggaagg	ctcgctccac	18540
atgcgcgggc	tcactctgag	ggcgccagc	atgtccgggg	cgagatctt	gaaatcgag	18600
ttgggaccgg	tgctctgcgc	gcgcgagttg	cggtacacgg	gggtgcagca	ctggaacacc	18660
atcagactgg	ggtacttcac	actggccagc	acgctcttgt	cggtgatctg	atccttgtca	18720
aggctcctcg	cggtgtctag	gccaacggg	gtcatcttgc	acagctggcg	gccaggaag	18780
ggcacgctct	gaggcttgtg	gttacctcg	cagtgaatgg	gcattagcat	catccccg	18840
ccgcgctgca	tattcgggta	gagggccttg	acaaaggccg	agatctgttt	gaaagcttgc	18900
tgggccttgg	ctccctcgct	gaaaaacagc	ccgcagctct	tcccgtgaa	ctgggtattc	18960
ccgcaaccgg	catcctgcac	gcagcagcgc	gcgtcatggc	tggtcagttg	caccacgctc	19020

```

cgccccagc ggtttctgggt caccttggcc ttgctgggtt gctccttcag cgcgcgctgt 19080
ccgtttctgc tggtcacatc catctccacc acgtgggtctt tgtggatcat caccgttcca 19140
tgcagacact tgagctggcc ttccacctcg gtgcagccgt gatccacag ggcgcacccg 19200
gtgcactccc agttcttatg cgcgatcccg ctgtggctga agatgtaacc ttgcaacagg 19260
cgacccatga cgggtgctaaa ggctttctgg gtgggtgaagg tcagttgcag accgcggggc 19320
tcctcgttca tccaggtctg gcacatcttc tggaagatct cgggtctgctc tggcatgagc 19380
ttgtaagcat cgcgcaggcc gctgtcgacg cggtagcggt ccatcagcac gttcatggta 19440
tccatgccct tctcccagga cgagaccaga ggcagactca ggggggttgcg cacgttcagg 19500
acaccggggg tcgcggggtc gacgatgcbt tttccgtcct tgccttcctt caacagaacc 19560
ggaggctggc tgaatccac tcccacgatc acggcgctctt cctggggcat ctcttcgtct 19620
gggtctacct ttgtcacatg cttgggtcttt ctggcttgct tcttttttgg agggctgtcc 19680
acggggacca cgtcctcctc cgaagacccg gagccaccc gctgatactt tcggcgcttg 19740
gtgggcagag gagggtggcg cgagggggtc ctctcctgct ccggcggata gcgcgccgac 19800
ccgtggcccc ggggcggagt ggctctcgc tccatgaacc ggcgcacgtc ctgactgccg 19860
ccggccattg tttcctaggg gaagatggag gagcagccgc gtaagcagga gcaggaggag 19920
gacttaacca cccacgagca acccaaaatc gagcaggacc tgggcttcga agagccggct 19980
cgtctagaac cccacagga tgaacaggag cagcagcaag acgcaggcca ggaggagacc 20040
gacgtggggc tcgagcatgg ctacctggga ggagaggagg atgtgctgct gaaacacctg 20100
cagcgccagt ccctcatcct ccgggacgcc ctggccgacc ggagcgaaac cccctcagc 20160
gtcagaggagc tgtgtcgggc ctacgagctc aacctcttct cgcgcgcgct gcccccaaa 20220
cgccagccca acggcacctg cgagcccaac ccgcgtctca acttctatcc cgtctttgctg 20280
gtccccgagg cccttgccac ctatcacatc tttttcaaga accaaaagat ccccgctctc 20340
tgccgcgcca accgcacccg cgcgcagcgc ctctcgtctc tggggcccg cgcgcgcata 20400
cctgatatcg ctcccttgga agaggtgccc aagatcttcg aagggtcgcg tcgggacgag 20460
acgcgcgcgg cgaacgctct gaaagaaaca gcagaggaag agggtcacac tagcgccttg 20520
gtagagttgg aaggcgacaa cgccaggctg gtcgtgctca agcgcagcgt cgagctcacc 20580
cacttcgctt acccgcgct taacctcccg cccaaggctc tgcgtcgcat catggatcag 20640
cttatcatgc cccacatcga ggccatcgat gagacccaag agcagcgcgc cgaggacgcc 20700
cggcccgctg tcagcgacga gatgctcgcg cgctggctcg ggaccgcga ccccaggct 20760
ttggaacagc ggcgcaagct gatgctggcc gtagtcttg taccctcga gctcgaatgc 20820
atgcgcgctt tcttctgca ccccgagacc ctgcgcaagg tcgaggagac cctgcactac 20880
actttcagac acggtttctg caggcaagcc tgcaagatct ccaacgtgga gctgaccaac 20940
ctggtctcct gcctggggat cctgcatgag aaccgcctgg ggcagacagt gctccactct 21000
accctgaagg gcgaggcacg gcgggactat gtccgcgact gcgtctttct ctttctatgc 21060
cacacatggc aagcagccat gggcggtgtg cagcagtgct tcgaggacga gaacctgaag 21120
gagctggaca agcttcttgc tagaaacctt aaaaagttgt ggacgggctt cgacgagcgc 21180
accgtcgctt cggacctggc cgagatcggt ttccccgagc gcctgaggca tacgctgaaa 21240
ggcgggctgc ccgacttcat gagccagagc atgttgcaaa actaccgcac ttctattctc 21300
gagcgctcgg gtatcctgcc cgccacctgc aacgccttcc cctccgactt tgtcccgctg 21360
agetaccgag agtgtcccc gccgctgtgg agccactgct acctcttgca gctggctaac 21420
tacatctcct accactcgga cgtgatcgag gacgtgagcg gcgaggggct gctcgagtgc 21480
cactgcccgt gcaacctgtg ctccccgcac cgctccctgg tctgcaacct ccagctcctg 21540
agcgagaccc aggtcatcgg taccttcgag ctgcaaggct cggagaagtc caccgctccg 21600
ctgaaactca cgccgggggt gtggacttcc gcgtacctgc gcaaatttgt acccgaagac 21660
taccacgccc atgagataaa gttcttcgag gaccaatcgc gtccgcagca cgcggatctc 21720
acggcctgcg tcatcaccca gggcgcgatc ctgcgccaat tgcagtccat ccaaaaatcc 21780
cgccaagagt ttcttctgaa aaagggtaga ggggtctacc tggaccccca gacgggagag 21840
gtgctcaacc cgggtctccc ccagcatgcc gaggaagaag caggagccgc tagtggagga 21900
gatggaagaa gaatgggaca gccaggcaga ggaggacgaa tgggaggagg agacagagga 21960
ggaagaattg gaagaggtgg aagaggagca ggcaacagag cagcccgctc ccgcaccatc 22020
cgcgcgggca gcccggcggt tcacggatac aacctccgct ccggtcaagc ctctcgtag 22080
atgggatacga gtgaaggggt acggtaagca cgagcggcag ggctaccgat catggagggc 22140
ccaaaagcc cctgcttgca cctgctcggg agactccgag ggaacatcg ctttcgccc 22200
ccgtacctg ctcttcacc gcgggggtgaa catccccgc aacgtgttgc attactacc 22260
tcaccttcac agctaagaaa aagcaagtaa gaggagtcgc cggaggaggc ctgaggatcg 22320

```

cggcgaacga gccctcgacc accaggggagc tgaggaaccg gatcttcccc actctttatg 22380
 ccatTTTTca gcagagtcga ggtcagcagc aagagctcaa agtaaaaaat cggctctctgc 22440
 gctcgctcac ccgcagttgc ttgtaccaca aaaacgaaga tcagctgcag cgcactctcg 22500
 aagacgccga ggctctgttc cacaagtact gcgcgctcac tcttaaagac taaggcgcg 22560
 ccacccggaa aaaaggcggg aattacctca tgcgcccat gagcaaggag attccaccc 22620
 cttacatgtg gagctatcag cccagatgg gctggccgc gggcgccctc caggactact 22680
 ccacccgcat gaactggctc agtgccggcc cctcgatgat ctacagggtc aacgggggtc 22740
 gtaaccatcg aaaccagata ttgttgagc aggcggcggt cacatccacg cccagggcaa 22800
 agctcaaccc gcgtaattgg ccctccaccc tgggtgatca ggaaatcccc gggccgacta 22860
 ccgtactact tccgcgtgac gcactggccg aagtccgcat gactaactca ggtgtccagc 22920
 tggccggcgg cgcttccccg tgcgcgctcc gccacaatc ggggtataaaa accctgatga 22980
 tccgaggcag aggcacacag ctcaacgacg agttggtgag ctcttcgatc ggtctgcgac 23040
 cggacggagt gttccaaact gccggagccg ggagatcatc cttcactccc aaccaggcct 23100
 acctgacctt gcagagcagc tcttcggagc ctgcctccgg aggcacgga accctccagt 23160
 tcgtggagga gtttgtgccc tcggtctact tcaacccctt ctcgggatcg ccaggcctct 23220
 acccggaacga gttcataccg aacttcgacg cagtgcagaga agcgggtggac ggctacgact 23280
 gaatgtccca tgggtgactcg gctgagctcg ctcggttgag gcatctggac cactgccgc 23340
 gcctgcgctg cttcgcccgg gagagctgcg gactcatcta ctttgagctg cccgaggagc 23400
 accccaacgg ccctgcacac ggagtagcga tcaccgtaga gggcaccgcc gagtctcacc 23460
 tggtcagggt cttcaccacg caacccttcc tggtcgagcg ggaccggggc gccaccacct 23520
 acaccgtcta ctgcatctgt cctaccccaa agttgcatga gaatttttgc tgtactcttt 23580
 gtggtgagtt taataaaagc tgaactaaga acctactttg gaatcccttg tcgtcatcaa 23640
 atccacaaga ccatcaactt cacctttgag gaacagggtga actttacctg caagccacac 23700
 aagaagtacg tcacctgggt ttaccagaac actactctag cagtagccaa cacctgctcg 23760
 aacgacgggt ttcttcttcc aaacaatctc accagtggac taactttctc agtgaaaagg 23820
 gcaaagctaa ttcttcatcg ccctattgta gaaggaaact accagtgtca gagcggacct 23880
 tgcttccaca gtttcaactt ggtgaacggt accggcagca gcacagtcgc tccagaaact 23940
 aaccttcttt ctgatactaa cactcctaaa accggagggt agctctgggt tccctctctg 24000
 acagaggggg gtagtcatat tgaagcggtc gggttattga ttttaggggt ggtcctgggt 24060
 ggggtgcatag cgggtgctata ttaccttctt tgctgggtcg aaatcagggt atttatctgc 24120
 tgggtcagac attgtgggga ggaaccatga aggggctctt gctgattatc ctttccctgg 24180
 ttgggggttt actggcctgc cacgaacagc cagcatgtaa catcaccaca ggcaatgaga 24240
 ggaacgactg ctctgtagt atcaaatgcg agcaccagtg tctctcaac attacattca 24300
 agaataagac catgggaaat gtatgggtgg gattctggca accaggagat gagcagaact 24360
 acacggtcac tatccatggt agcgatggaa atcacacttt cggtttcaaa ttcatttttg 24420
 aagtcatgtg tgatatcaca ctgcatgtgg ctagacttca tggcttgtgg cccctacca 24480
 aggagaacat gggtgggttt tctttggctt ttgtgatcat ggcctgtgca atgtcaggte 24540
 tgctggtagg ggctctagt tggttcctga agcgcaagcc caggtagcga aatgaggaga 24600
 aggaaaaatt gctataaaat tttttctctt cgcagaacca tgaatacttt gaccagtgtc 24660
 gtgctgctct ctcttttagt tattaatgtg gaattgtccg atcctattct agttagtgt 24720
 gattggggaa aaaatcttac attagagggt cctaagaaa caccagttga atgggtgggt 24780
 ggaagaaaca tacaacaact gtgcataggg aatcaaacca aacataaaga gctaagtcac 24840
 agatgtaatg tccagaacat aactttactg tttgtaataa ctagttttaa tggagactac 24900
 tttgggttta aaaatgataa cagcggtagt aaacattata aagtcacagt tataccccct 24960
 aaacctcca ctcggaaccc tctttctctt ccacactatg taaacgcaac tatggggcaa 25020
 aacctaacat tagtggggcc tgcaaacatt ccagttactt ggcttagtga atatggcacg 25080
 ttgtgtgagg gcaaaaaaat tttgcacaaa gaattaaatc acacctgtaa cgaacagAAC 25140
 ctacagttac tgtttgttaa tatgacacac aacggaccat attttggctt tgacaaatac 25200
 aacattgata gagagcagta tgaggtttct attattagtt tgtttaaagt tggcgctgga 25260
 cagaagaaaa ttgggaaagg acagaaaaag gaggaaaaaga caaaaccaa ctctagtgat 25320
 ttgggacaaa gacaatccag accaaagaaa aaagatattg ttgaaggagt ccaaatacaa 25380
 acaggagaaa atcgaacctt tgttgggtcca cctggaaaag ttgattggat taaactttcc 25440
 agtggaaaca ataattgtct taagtgtgt aatggcgaca agtatattaa acacacatgt 25500
 gatggtaaaa atttaacatt aattaatgtg actagaattt atgacggaac ttattatggt 25560
 tctagcaatg atggctcaag tcattacaaa gttaccatct atgaattaca caaagttaat 25620

aaaactaaat	ctatgcttaa	gccatacact	acaaaaagaa	ctacagtgaá	tgcaacagat	25680
gacagtgc	acaaaattgc	tttgcagcag	gaaaataatg	ggcaaacaga	aaatgatcaa	25740
gaatcaaaaa	ttccatctgc	tactgtggca	atcgtggtgg	gagtgattgc	gggcttcata	25800
actataatca	ttgtcattct	gtgctacatc	tgctgccgca	agcgtcccag	ggcatacaat	25860
aatatggtag	accactact	cagcttctct	tactgagact	cagtcacttt	catttcagaa	25920
ccatgaaggc	tttcacagct	tgctgtctga	ttagcataat	tacacttagt	ttagcagcac	25980
ctaaaccaga	agtatatata	caagttaatg	tcactagggg	tgggaatgct	acactagatg	26040
gaccatttaa	caataacaca	tggacaagat	atcatgatga	tgggagaaaa	aacggatgga	26100
tgaatatattg	taaatggtca	gacccatcat	acacatgtca	tagtaatgga	agccttagta	26160
tttttgc	caacattagt	tcaggtaaat	ataaagttca	aagttacact	aacagttata	26220
atggattaga	tggttatgaa	aaacttgaag	ttaaaatggt	taatctaaca	gtaattgagc	26280
ctccaaccac	tagagcacc	accacagtta	ggacaactaa	ggaaacaaca	cagcctacca	26340
ctgtaccac	tacacataca	accaccacag	tcagtacaac	tattgagacc	actactcata	26400
ctacacagct	agacacaaca	gtgcagaata	ctactttact	gattgaattt	ttactaagag	26460
ggaatgaaag	tactactgat	cagacagagg	ctacctcaag	tgcttcagc	agtactgcaa	26520
atttaacttc	gcttgcttgg	actaatgaaa	cgggagtatc	attgatgcat	ggccagcctt	26580
actcagggtt	ggatattcaa	attacttttc	tggttgtctg	tgggatcttt	attcttgttg	26640
ttcttctgta	ctttgtctgc	tgcaaagcca	gagagaaatc	tagtagggcc	atctacaggc	26700
cagtaatcgg	ggaacctcag	cctctccaag	tgggaagggg	tctaaggaat	cttctcttct	26760
ctttttcagt	atgggtgatca	gccatgattc	ctagggttct	cctattttaac	atcctcttct	26820
gtctcttcaa	catctgcgct	gccttcgctg	cgtctcgca	cgcctcgccc	gactgtctcg	26880
ggcccttccc	cacctacctc	ctctttgccc	tgctcacctg	cacctgcgtc	tgacgattg	26940
tctgcctgg	cgtcaccttc	ctgcagctca	tcgactgggt	ctgcgcgct	tacaattatc	27000
tccaccacag	tcccgaatac	agggacaaga	acgtagccag	aatcttaagg	ctcatctgac	27060
catgcagact	ctgctgatac	tgctatccct	cctctccctt	gcccttgctg	actgtaaatt	27120
tgcggacata	tggaaattct	tagactgtta	tcaagagaaa	atggatatgc	cttcttatta	27180
cttgggtgatt	gtgggtgtag	tcattggtctg	ctcctgcact	ttctttgcta	tcattgctta	27240
cccctgtttt	gatctcggct	ggaactctgt	tgaggcatte	acatacacac	tagaaagcag	27300
ttcactagcc	tccacgcgc	caccacacac	gcctccccgc	agaaatcagt	tccccctgat	27360
tcagtactta	gaagagcccc	ctccccggcc	cccttccact	gttagctact	ttcacataac	27420
cggcggcgat	gactgaccac	cacctggacc	tcgagatgga	cggccaggcc	tccgagcagc	27480
gcaccttgca	actgcgcgct	cgtcagcagc	aggagcgggc	cggcaaggag	ctcctcgatg	27540
ccatcaacat	ccaccagtgc	aagaagggca	tcttctgcct	ggtcaaacag	gcaaagatca	27600
cctacgagct	cgtgtccaac	ggcaaacagc	atcgccctac	ctatgagatg	ccccagcaga	27660
agcagaagtt	cacctgcatg	gtgggcgtca	accccatagt	catcacccag	cagtcggggc	27720
agaccagcgg	ctgcatccac	tgctcctgct	aaagccccga	gtgcatctac	tccctcctca	27780
agaccctttg	cggacttcgc	gacctcctcc	ccatgaactg	attgattaaa	gcccagaaac	27840
caatcaaacc	cccttcccca	tcaccccaaa	taacaatcat	tggaaataat	cattcaataa	27900
agactactta	cttgaaatct	gaaagtatgt	ctctgggtga	gttggttcagc	agcacctcgg	27960
taccctcctc	ccagctctgg	tactccagtc	cccggcgggc	ggcgaacttc	ctccacacct	28020
tgaaggggat	gtcaaattcc	tgggtccacaa	ttttcatgtt	cttccctctc	agatgtcaaa	28080
gaggctccgg	gtggaagatg	acttcaaccc	cgtctacccc	tatggctacg	cgcggaatca	28140
gaatatcccc	ttcctcactc	ccccctttgt	ctcctccgat	ggattccaaa	acttcccccc	28200
tggggctcctg	tactcaaac	tggctgatcc	aatcgccatc	gccaatggga	atgtctcact	28260
caaggtggga	gggggactca	ctgtagaaca	acagtctgga	aaactgagtg	tggatactaa	28320
ggcacccttg	caagttgcaa	atgacaacaa	attggagcta	tcttatgatg	atccatttaa	28380
ggtagagaat	aacaaacttg	gaattaaagc	tggccatggt	ttagcagttg	taactaaaga	28440
aaacacaagt	cttcctagtc	tagttggaac	acttgtagtt	ttaactggaa	aaggaatagg	28500
tactggatca	agtgcacatg	gaggaactat	tgatgtaaga	cttgggtgaag	gagggtgggt	28560
atcattttgat	gaaaaaggag	acttagtagc	ttgggacaaa	aaaaatgata	cacgcacctt	28620
ttggacaaca	cctgatcctt	ctccaaattg	caaagttgaa	acagcaagag	actcaaagct	28680
aaccttagca	cttacaaaaa	gtggtagtca	aattttggcc	actgtatctt	tacttgttgt	28740
tacgggcaaa	tatgtattta	taagtgcac	agtcaaccca	aagcagttct	ctattaagtt	28800
actgtttaat	gacaagggtg	ttttgttaag	tgactcaaat	cttgatggga	catattggaa	28860
ctatagaagc	aacaataaca	acataggcac	tccttataaa	gaggctgttg	gttttatgcc	28920

aagcacaaca	gcttatccta	agccaaccaa	caacaccagc	acagatccgg	ataaaaaagt	28980
gagtcagggt	aaaaataaaa	ttgtaagcaa	tatttatctt	ggaggagagg	tatatcaacc	29040
aggattttatt	gttggttaaat	ttaatcagga	aactgatgcc	aattgtgcat	actctattac	29100
atlttgatttt	ggatggggta	agggtgtataa	ggatcctata	ccatatgata	cctcttcttt	29160
tacttttctca	tatatcgctc	aagaatgaaa	gaccaataaa	cgtgtttttc	attgaaaatt	29220
ttcatgtatc	tttattgatt	tttacaccag	cacgggtagt	cagtctccca	ccaccagccc	29280
atlttcacagt	gtaaacaatt	ctctcagcac	gggtggcctt	aaatagggga	atgttctgat	29340
tagtgcgggg	actggagtcg	acctacatgg	gggtagagtc	ataatcgtgc	atcaggatag	29400
ggcgggtggtg	ctgcagcagc	gcgcgaataa	actgctgccg	ccgccgctcc	gtcctgcagg	29460
aatacaacat	ggcagtggtc	tcctcagcga	tgattcgcac	cgcccgagc	ataaggcgcc	29520
ttgtcctccg	ggcacagcag	cgcaccctga	tctcacttaa	atcagcacag	taactgcagc	29580
acagcaccac	aataattgttc	aaaatcccac	agtgcaggcc	gctgtatcca	aagctcatgg	29640
cggggaccac	agaaccacg	tggccatcat	accacaagcg	caggtagatt	aagtggcgac	29700
ccctcataaa	cacgctggac	ataaacatta	cctcttttgg	catgttgtaa	ttcaccacct	29760
cccgggtacca	tataaacctc	tgattaaaca	tggcgccatc	caccaccatc	ctaaaccagc	29820
tggccaaaac	ctgcccgcgg	gctatacact	gcagggaacc	gggactggaa	caatgacagt	29880
ggagagccca	ggactcgtaa	ccatggatca	tcatgctcgt	catgatatca	atgttggcac	29940
aacacaggca	cacgtgcata	cacttcctca	ggattacaag	ctcctccgc	gttagaacca	30000
tatcccaggg	aacaacccat	tcctgaatca	gcgtaaatcc	cacactgcag	ggaagacctc	30060
gcacgtaact	cacgttggtc	attgtcaaag	tgttacattc	gggcagcagc	ggatgatcct	30120
ccagtatggg	agcgcggggt	tctgtctcaa	aaggaggtag	acgatcccta	ctgtacggag	30180
tgcgccgaga	caaccgagat	cgtgttggtc	gtagtgtcat	gccaaatgga	acgccggacg	30240
tagtcatagc	tagccccgct	taccagtaga	cagagagcac	agcagtacaa	gcgccaacag	30300
cagcgactga	ctacccactg	accagctcc	ctattttaaag	gcaccttaca	ctgacgtaat	30360
gaccaaagggt	ctaaaaaccc	cgccaaaaaa	aaacacacac	gccttggggtg	tttttcgcga	30420
aaacacttcc	gcgttggtcac	ttcctcgat	cgatttcgtg	actcaacttc	cgggttccca	30480
cgttacgtca	cttctgcccc	tacatgtggc	tccgcgtag	ggcgccatct	tgcccacgtc	30540
caaaatggct	tccatgtcag	accacgcctc	cgcgccgacc	gttgacgcgg	tgcgctctga	30600
cgtcagtaac	ggctcttcca	acagccaatc	agcgtcggcc	ccgccccaaa	ttcaaaagct	30660
catttgcata	ttaaactttg	tttactttgt	ggggtatatt	attgatgttt	aaacaagctt	30720
ggcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	30780
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaattgag	tgagctaact	30840
cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagct	30900
gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	cgtattgggg	gctcttccgc	30960
ttcctcgctc	actgactcgc	tgcgctcggg	cgttcggctg	cgcgagcgcg	tatcagctca	31020
ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	31080
agcaaaaggc	cagcaaaagg	caggaaccg	taaaaggcc	gcgttgctgg	cgtttttcca	31140
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtacaga	gggtggcgaaa	31200
cccgacagga	ctataaagat	accaggcggt	ttcccctgga	agctccctcg	tgcgctctcc	31260
tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	31320
gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	taggtcggtc	gctccaagct	31380
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	31440
tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	31500
gattagcaga	gcgagggtatg	taggcgggtg	tacagagttc	ttgaagtggg	ggcctaacta	31560
cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	31620
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtggtttttt	31680
tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	31740
ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	taaggagattt	tggtcatgag	31800
attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	ttaaatcaat	31860
ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	tgcttaataca	gtgaggcacc	31920
tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	tgactccccg	tcgtgtagat	31980
aactacgata	cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	cgcgagaccc	32040
acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	ccgagcgag	32100
aagtggctct	gcaactttat	ccgcctccat	ccagtcctatt	aattgttgcc	gggaagctag	32160
agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	32220

```

gggtgtcacgc tcgtcgtttg gtatggcctc attcagctcc gggtcccaac gatcaaggcg 32280
agttacatga tcccccatgt tgtgcaaaaa agcgggttagc tccttcggtc ctccgatcgt 32340
tgtcagaagt aagttggccg cagtgttatc actcatgggt atggcagcac tgcataattc 32400
tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc 32460
attctgagaa tagtgtatgc ggcgaccgag ttgctccttg ccggcgtcaa tacgggataa 32520
taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg 32580
aaaactctca aggatccttac cgctgttgag atccagttcg atgtaacca ctctgtcacc 32640
caactgatct tcagcatctt ttactttcac cagcgtttct ggggtgagcaa aaacaggaag 32700
gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt 32760
cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt 32820
tgaatgtatt tagaaaaata aacaaatagg gggtccgcgc acatttcccc gaaaagtgcc 32880
acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata ggcgtatcac 32940
gaggcccttt cgtctcgcgc gtttcgggtga tgacgggtgaa aacctctgac acatgcagct 33000
cccggagacg gtcacagctt gtctgtaagc ggatgccggg agcagacaag cccgtcaggg 33060
cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac tatgcggcat cagagcagat 33120
tgtactgaga gtgcaccata tgcggtgtga aataccgcac agatgcgtaa ggagaaaaata 33180
ccgcatcagg cgccattcgc cattcaggct gcgcaactgt tgggaagggc gatcgggtgcg 33240
ggcctcttcg ctattacgcc agctggcgaa agggggatgt gctgcaaggc gattaagttg 33300
ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagtg aattcgttt 33359

```

<210> 3

<211> 1521

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence for HIV-1gag

<400> 3

```

atgggtgcta gggcttctgt gctgtctggt ggtgagctgg acaagtggga gaagatcagg 60
ctgaggcctg gtggcaagaa gaagtacaag ctaaagcaca ttgtgtgggc ctccaggagg 120
ctggagaggt ttgctgtgaa ccctggcctg ctggagacct ctgaggggtg caggcagatc 180
ctgggccagc tccagccctc cctgcaaaaa ggctctgagg agctgaggtc cctgtacaac 240
acagtggcta ccctgtactg tgtgcaccag aagattgatg tgaaggacac caaggaggcc 300
ctggagaaga ttgaggagga gcagaacaag tccaagaaga agggccagca ggctgctgct 360
ggcacaggca actccagcca ggtgtcccag aactacccca ttgtgcagaa cctccagggc 420
cagatgggtg accaggccat ctccccccgg acctgaatg cctgggtgaa ggtggtggag 480
gagaaggcct tctcccctga ggtgatcccc atgttctctg cctgtctgta ggtgcccacc 540
ccccaggacc tgaacacctg gctgaacaca gtggggggcc atcaggctgc catgcagatg 600
ctgaaggaga ccatcaatga ggaggctgct cagtgggaca ggctgcatcc tgtgcacgct 660
ggccccattg cccccggcca gatgaggagg cccaggggct ctgacattgc tggcaccacc 720
tccaccctcc aggagcagat tggctggatg accaacaacc ccccatccc tgtgggggaa 780
atctacaaga ggtggatcat cctgggcctg aacaagattg tgaggatgta ctccccacc 840
tccatcctgg acatcaggca gggccccaag gagcccttca gggactatgt ggacagggtc 900
tacaagaccc tgagggctga gcaggcctcc caggaggtga agaactggat gacagagacc 960
ctgctgggtg agaatgcaa ccctgactgc aagaccatcc tgaaggccct gggccctgct 1020
gccaccctgg aggagatgat gacagcctgc caggggggtg ggggccctgg tcacaaggcc 1080
agggtgctgg ctgaggccat gtcccagggt accaactccg ccaccatcat gatgcagagg 1140
ggcaacttca ggaaccagag gaagacagtg aagtgttca actgtggcaa ggtgggccac 1200
attgccaaaga actgtagggc ccccaggaag aagggctgct ggaagtgtgg caaggagggc 1260
caccagatga aggactgcaa tgagaggcag gccaaattcc tgggcaaaat ctggccctcc 1320
cacaagggca ggcctggcaa cttcctccag tccaggcctg agcccacagc ccctcccag 1380
gagtccttca ggtttgggga ggagaagacc accccagacc agaagcagga gccattgac 1440
aaggagctgt accccctggc ctccctgagg tccctgtttg gcaacgacct ctcctcccag 1500
taaaataaag cccgggcaga t

```

<210> 4
 <211> 2550
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> gag expression cassette

<400> 4
 ccattgcata cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaca 60
 ttaccgccat gttgacattg attattgact agttattaat agtaatcaat tacgggggtca 120
 ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcc 180
 ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta 240
 acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac 300
 ttggcagtac atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacgg 360
 aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag 420
 tacatctacg tatttagtcat cgctattacc atgggtgatgc ggttttggca gtacatcaat 480
 gggcgtggat agcgggtttga ctacacgggga tttccaagtc tccaccccat tgacgtcaat 540
 gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc 600
 ccattgacgc aaatgggccc taggcgtgta cgggtgggagg tctatataag cagagctcgt 660
 ttagtgaacc gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga 720
 caccgggacc gatccagcct ccgcgccggc gaacgggtgca ttggaacgcg gattccccgt 780
 gccaagagtg agatctacca tgggtgctag ggcttctgtg ctgtctggtg gtgagctgga 840
 caagtgggag aagatcaggc tgaggcctgg tggcaagaag aagtacaagc taaagcacat 900
 tgtgtgggccc tccagggagc tggagagggt tgctgtgaac cctggcctgc tggagacctc 960
 tgagggggtgc aggagatcc tgggccagct ccagccctcc ctgcaaacag gctctgagga 1020
 gctgagggtcc ctgtacaaca cagtggctac cctgtactgt gtgcaccaga agattgatgt 1080
 gaaggacacc aaggaggccc tggagaagat tgaggaggag cagaacaagt ccaagaagaa 1140
 ggcccagcag gctgctgctg gcacaggcaa ctccagccag gtgtcccaga actaccccat 1200
 tgtgcagaac ctccagggcc agatgggtgca ccaggccatc tccccccgga ccctgaatgc 1260
 ctgggtgaag gtgggtggagg agaaggcctt ctcccttgag gtgatcccca tgttctctgc 1320
 cctgtctgag ggtgccaccc ccaggacct gaacaccatg ctgaacacag tggggggcca 1380
 tcaggctgcc atgcagatgc tgaaggagac catcaatgag gaggctgctg agtgggacag 1440
 gctgcacccct gtgcacgctg gcccattgc ccccgccag atgaggggag ccaggggctc 1500
 tgacattgct ggcaccacct ccaccctcca ggagcagatt ggctggatga ccaacaacct 1560
 ccccatccct gtgggggaaa tctacaagag gtggatcatc ctgggcctga acaagattgt 1620
 gaggatgtac tccccacct ccactctgga catcaggcag ggcccgaagg agcccttcag 1680
 ggactatgtg gacaggttct acaagacct gagggctgag caggcctccc aggaggtgaa 1740
 gaactggatg acagagaccc tgctgggtgca gaatgccaac cctgactgca agaccatcct 1800
 gaaggccctg ggcctgctg ccaccctgga ggagatgatg acagcctgcc aggggggtggg 1860
 gggccctggt cacaaggcca ggggtgctggc tgaggccatg tcccaggatga ccaactccgc 1920
 caccatcatg atgcagaggg gcaacttcag gaaccagagg aagacagtga agtgcttcaa 1980
 ctgtggcaag gtggggccaca ttgccaagaa ctgtagggcc ccaggaaga agggctgctg 2040
 gaagtgtggc aaggaggggcc accagatgaa ggactgcaat gagaggcagg ccaacttcct 2100
 gggcaaaaatc tggccctccc acaagggcag gcctggcaac ttcctccagt ccaggcctga 2160
 gccacagacc cctcccagag agtccttcag gtttggggag gagaagacca ccccagcca 2220
 gaagcaggag ccattgaca aggagctgta cccctggcc tccctgaggt ccctgtttgg 2280
 caacgacccc tcctcccagt aaaataaagc ccgggcagat ctgatctgct gtgccttcta 2340
 gttgccagcc atctgttgtt tgccctccc ccgtgccttc cttgaccctg gaagggtgcca 2400
 ctcccactgt cctttcctaa taaaatgagg aaattgcac gcattgtctg agtaggtgtc 2460
 attctattct ggggggtggg gtggggcagc acagcaaggg ggaggattgg gaagacaata 2520
 gcaggcatgc tggggatgcg gtgggctcta 2550

<210> 5

<211> 49
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> short synthetic polyA signal

<400> 5
 aataaaagat ctttattttc attagatctg tgtgttggtt ttttgtgtg 49

<210> 6
 <211> 2577
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized pol sequence with deletions in
 protease

<400> 6
 agatctacca tggcccccat ctccccatt gagactgtgc ctgtgaagct gaagcctggc 60
 atggatggcc ccaaggtgaa gcagtggccc ctgactgagg agaagatcaa ggccctgggtg 120
 gaaatctgca ctgagatgga gaaggagggc aaaatctcca agattggccc cgagaacccc 180
 tacaacacccc ctgtgtttgc catcaagaag aaggactcca ccaagtggag gaagctgggtg 240
 gacttcaggg agctgaacaa gaggacccag gacttctggg aggtgcagct gggcatcccc 300
 caccgctg gcctgaagaa gaagaagtct gtgactgtgc tggatgtggg ggatgcctac 360
 ttctctgtgc ccctggatga ggacttcagg aagtacactg ccttcacccat cccctccatc 420
 aacaatgaga cccctggcat caggtaccag tacaatgtgc tgccccaggg ctggaagggc 480
 tccccggcca tcttccagtc ctccatgacc aagatcctgg agcccttcag gaagcagaac 540
 cctgacattg tgatctacca gtacatggat gacctgtatg tgggctctga cctggagatt 600
 gggcagcaca ggaccaagat tgaggagctg aggcagcacc tgctgaggtg gggcctgacc 660
 acccctgaca agaagcacca gaaggagccc cccttctctgt ggatgggcta tgagctgcac 720
 cccgacaagt ggactgtgca gccattgtg ctgcctgaga aggactcctg gactgtgaat 780
 gacatccaga agctggtggg caagctgaac tgggcctccc aaatctaccc tggcatcaag 840
 gtgaggcagc tgtgcaagct gctgaggggc accaaggccc tgactgaggt gatccccctg 900
 actgaggagg ctgagctgga gctggctgag aacagggaga tcctgaagga gcctgtgcat 960
 ggggtgtact atgaccctc caaggacctg attgctgaga tccagaagca gggccagggc 1020
 cagtggacct accaaatcta ccaggagccc ttcaagaacc tgaagactgg caagtatgcc 1080
 aggatgaggg gggcccacac caatgatgtg aagcagctga ctgaggctgt gcagaagatc 1140
 accactgagt ccattgtgat ctggggcaag accccaagt tcaagctgcc catccagaag 1200
 gagacctggg agacctggtg gactgagtac tggcaggcca cctggatccc tgagtgggag 1260
 tttgtgaaca cccccccct ggtgaagctg tggtagcagc tggagaagga gcccattgtg 1320
 ggggctgaga ccttctatgt ggatggggct gccaacaggg agaccaagct gggcaaggct 1380
 ggctatgtga ccaacagggg caggcagaag gtggtgaccc tgactgacac caccaaccag 1440
 aagactgagc tccaggccat ctacctggcc ctccaggact ctggcctgga ggtgaacatt 1500
 gtgactgact cccagtatgc cctgggcatc atccaggccc agcctgatca gtctgagtct 1560
 gagctggtga accagatcat tgagcagctg atcaagaagg agaaggtgta cctggcctgg 1620
 gtgcctgccc acaagggcat tgggggcaat gagcaggtgg acaagctggt gtctgctggc 1680
 atcaggaagg tgctgttctt ggatggcatt gacaaggccc aggatgagca tgagaagtac 1740
 cactccaact ggagggctat ggcctctgac ttcaacctgc cccctgtggt ggctaaggag 1800
 attgtggcct cctgtgacaa gtgccagctg aagggggagg ccatgcatgg gcaggtggac 1860
 tgctcccctg gcatctggca gctggactgc acccacctgg agggcaaggt gatcctgggtg 1920
 gctgtgcatg tggcctccgg ctacattgag gctgaggtga tcctgtctga gacaggccag 1980
 gagactgcct acttctgct gaagctggct ggcaggtggc ctgtgaagac catccacact 2040
 gacaatggct ccaacttcac tggggccaca gtgagggctg cctgctggtg ggctggcatc 2100

```

aagcaggagt ttggcatccc ctacaacccc cagtcccagg ggggtggtgga gtccatgaac 2160
aaggagctga agaagatcat tgggcagggtg agggaccagg ctgagcacct gaagacagct 2220
gtgcagatgg ctgtgttcat ccacaacttc aagaggaagg ggggcatcgg gggctactcc 2280
gctggggaga ggattgtgga catcattgcc acagacatcc agaccaagga gctccagaag 2340
cagatcacca agatccagaa cttcagggtg tactacaggg actccaggaa cccctgtgg 2400
aagggccctg ccaagctgct gtggaagggg gagggggctg tggatgatcca ggacaactct 2460
gacatcaagg tgggtgccag gaggaaggcc aagatcatca gggactatgg caagcagatg 2520
gctgggggatg actgtgtggc ctccaggcag gatgaggact aaagcccggg cagatct 2577

```

<210> 7

<211> 850

<212> PRT

<213> Artificial Sequence

<220>

<223> pol with altered protease

<400> 7

```

Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
1      5      10      15
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
20      25      30
Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
35      40      45
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
50      55      60
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
65      70      75      80
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
85      90      95
Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp
100     105     110
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
115     120     125
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
130     135     140
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
145     150     155     160
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
165     170     175
Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly
180     185     190
Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
195     200     205
Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
210     215     220
Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
225     230     235     240
Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
245     250     255
Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
260     265     270
Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr
275     280     285

```

Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
 290 295 300
 Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
 305 310 315 320
 Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
 325 330 335
 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
 340 345 350
 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
 355 360 365
 Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
 370 375 380
 Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
 385 390 395 400
 Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
 405 410 415
 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
 420 425 430
 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala
 435 440 445
 Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
 450 455 460
 Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu
 465 470 475 480
 Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
 485 490 495
 Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
 500 505 510
 Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
 515 520 525
 Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
 530 535 540
 Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys
 545 550 555 560
 Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
 565 570 575
 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
 580 585 590
 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
 595 600 605
 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
 610 615 620
 Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
 625 630 635 640
 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
 645 650 655
 Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
 660 665 670
 Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
 675 680 685
 Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
 690 695 700
 Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu
 705 710 715 720

Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
 725 730 735
 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
 740 745 750
 Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
 755 760 765
 Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
 770 775 780
 Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
 785 790 795 800
 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
 805 810 815
 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
 820 825 830
 Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
 835 840 845
 Glu Asp
 850

<210> 8

<211> 2577

<212> DNA

<213> Artificial Sequence

<220>

<223> sequence for IA-Pol

<400> 8

```

atgacttaccac tggcccccat ctccccatt gagactgtgc ctgtgaagct gaagcctggc 60
atggatggcc ccaagggtgaa gcagtggccc ctgactgagg agaagatcaa ggccctgggtg 120
gaaatctgca ctgagatgga gaaggagggc aaaatctcca agattggccc cgagaacccc 180
tacaacaccc ctgtgtttgc catcaagaag aaggactcca ccaagtggag gaagctgggtg 240
gacttcaggg agctgaacaa gaggaccag gacttctggg aggtgcagct gggcatcccc 300
caccctgctg gcctgaagaa gaagaagtct gtgactgtgc tggctgtggg ggatgcctac 360
ttctctgtgc ccctggatga ggacttcagg aagtacactg ccttcaccat cccctccatc 420
aacaatgaga ccctggcat caggtaccag tacaatgtgc tgccccaggg ctggaagggc 480
tccctgcca tcttcagtc ctccatgacc aagatcctgg agccctcag gaagcagaac 540
cctgacattg tgatctacca gtacatggct gccctgtatg tgggctctga cctggagatt 600
gggcagcaca ggaccaagat tgaggagctg aggcagcacc tgctgaggtg gggcctgacc 660
accctgaca agaagcacca gaaggagccc cccttctgtg ggatgggcta tgagctgcac 720
cccgacaagt ggactgtgca gccattgtg ctgcctgaga aggactcctg gactgtgaat 780
gacatccaga agctgggtgg caagctgaac tgggcctccc aaatctaccc tggcatcaag 840
gtgaggcagc tgtgcaagct gctgaggggc accaaggccc tgactgaggt gatccccctg 900
actgaggagg ctgagctgga gctggctgag aacagggaga tcctgaagga gcctgtgcat 960
ggggtgtact atgacccctc caaggacctg attgctgaga tccagaagca gggccagggc 1020
cagtggacct accaaatcta ccaggagccc ttcaagaacc tgaagactgg caagtatgcc 1080
aggatgaggg gggccacac caatgatgtg aagcagctga ctgaggctgt gcagaagatc 1140
accactgagt ccattgtgat ctggggcaag accccaagt tcaagctgcc catccagaag 1200
gagacctggg agacctgggt gactgagtac tggcaggcca cctggatccc tgagtgggag 1260
tttgtgaaca cccccccct ggtgaagctg tggtagcagc tggagaagga gccattgtg 1320
gggctgaga ctttctatgt ggctggggct gccaacaggg agaccaagct gggcaaggct 1380
ggctatgtga ccaacagggg caggcagaag gtggtgacct tgactgacac caccaaccag 1440
aagactgccc tccaggccat ctacctggcc ctccaggact ctggcctgga ggtgaacatt 1500
gtgactgcct ccagtatgc cctgggcata atccaggccc agcctgatca gtctgagtct 1560
gagctggtga accagatcat tgagcagctg atcaagaagg agaaggtgta cctggcctgg 1620

```

```

gtgcctgccc acaagggcat tgggggcaat gagcaggtgg acaagctggt gtctgctggc 1680
atcaggaagg tgctgttcct ggatggcatt gacaaggccc aggatgagca tgagaagtac 1740
cactccaact ggagggctat ggccctctgac ttcaacctgc cccctgtggt ggctaaggag 1800
attgtggcct cctgtgacaa gtgccagctg aagggggagg ccatgcatgg gcagggtggac 1860
tgctccccctg gcatctggca gctggcctgc acccacctgg agggcaaggt gatcctggtg 1920
gctgtgcatg tggcctccgg ctacattgag gctgaggtga tccctgctga gacaggccag 1980
gagactgcct acttcctgct gaagctggct ggcaggtggc ctgtgaagac catccacact 2040
gccaatggct ccaacttcac tggggccaca gtgagggctg cctgctggtg ggctggcatc 2100
aagcaggagt ttggcatccc ctacaacccc cagtcccagg ggggtggtggc ctccatgaac 2160
aaggagctga agaagatcat tgggcaggtg agggaccagg ctgagcacct gaagacagct 2220
gtgcagatgg ctgtgttcat ccacaacttc aagaggaagg ggggcatcgg gggctactcc 2280
gctggggaga ggattgtgga catcattgcc acagacatcc agaccaagga gctccagaag 2340
cagatcacca agatccagaa cttcaggggtg tactacaggg actccaggaa cccctgtggt 2400
aagggccctg ccaagctgct gtggaagggg gagggggctg tggatgatcca ggacaactct 2460
gacatcaagg tgggtgccag gaggaaggcc aagatcatca gggactatgg caagcagatg 2520
gctgggggatg actgtgtggc ctccaggcag gatgaggact aaagcccggg cagatct 2577

```

<210> 9

<211> 850

<212> PRT

<213> Artificial Sequence

<220>

<223> IA-Pol

<400> 9

```

Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
1           5           10           15
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
20           25           30
Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
35           40           45
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
50           55           60
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
65           70           75           80
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
85           90           95
Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Ala
100          105          110
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
115          120          125
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
130          135          140
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
145          150          155          160
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
165          170          175
Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly
180          185          190
Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
195          200          205
Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
210          215          220

```

Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
 225 230 235 240
 Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
 245 250 255
 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
 260 265 270
 Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr
 275 280 285
 Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
 290 295 300
 Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
 305 310 315 320
 Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
 325 330 335
 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
 340 345 350
 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
 355 360 365
 Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
 370 375 380
 Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
 385 390 395 400
 Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
 405 410 415
 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
 420 425 430
 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala
 435 440 445
 Asn Arg Gly Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
 450 455 460
 Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala
 465 470 475 480
 Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
 485 490 495
 Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
 500 505 510
 Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
 515 520 525
 Lys Lys Gly Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
 530 535 540
 Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys
 545 550 555 560
 Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
 565 570 575
 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
 580 585 590
 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
 595 600 605
 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
 610 615 620
 Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
 625 630 635 640
 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
 645 650 655

Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
 660 665 670
 Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
 675 680 685
 Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
 690 695 700
 Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu
 705 710 715 720
 Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
 725 730 735
 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
 740 745 750
 Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
 755 760 765
 Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
 770 775 780
 Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
 785 790 795 800
 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
 805 810 815
 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
 820 825 830
 Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
 835 840 845
 Glu Asp
 850

<210> 10

<211> 671

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence for HIV-1 jrf1 nef

<400> 10

gatctgccac catgggcgcc aagtgggtcca agagggtccgt gcccggctgg tccaccgtga 60
 gggagaggat gaggagggcc gagcccgccg ccgacagggg gaggaggacc gagcccgccg 120
 ccgtgggctg gggcgccgtg tccagggacc tggagaagca cggcgccatc acctcctcca 180
 acaccgccc caccaacgcc gactgcccct ggctggaggc ccaggaggac gaggaggtgg 240
 gcttccccgt gaggccccag gtgcccctga ggcccatgac ctacaagggc gccgtggacc 300
 tgtcccactt cctgaaggag aagggcgccc tggagggcct gatccactcc cagaagaggc 360
 aggacatcct ggacctgtgg gtgtaccaca ccaggggcta cttccccgac tggcagaact 420
 acacccccgg ccccgccatc aggttcccc tgaccttcgg ctggtgcttc aagctgggtgc 480
 ccgtggagcc cgagaagggt gaggaggcca acgagggcga gaacaactgc ctgctgcacc 540
 ccatgtccca gcacggcatc gaggaccccc agaaggaggt gctggagtgg aggttcgact 600
 ccaagctggc cttccaccac gtggccaggg agctgcaccc cgagtactac aaggactgct 660
 aaagcccggg c 671

<210> 11

<211> 216

<212> PRT

<213> Artificial Sequence

<220>

<223> HIV-1 jrfl nef

<400> 11

```

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
 1           5           10           15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
          20           25           30
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
          35           40           45
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
          50           55           60
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
65          70          75          80
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
          85          90          95
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
          100         105         110
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
          115         120         125
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
          130         135         140
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
145         150         155         160
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His
          165         170         175
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
          180         185         190
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
          195         200         205
His Pro Glu Tyr Tyr Lys Asp Cys
          210         215

```

<210> 12

<211> 651

<212> DNA

<213> Artificial Sequence

<220>

<223> wild type nef sequence

<400> 12

```

atgggtggca agtgggtcaaa acgtagtgtg cctgggatggt ctactgtaag ggaaagaatg 60
agacgagctg agccagcagc agataggggtg agacgaactg agccagcagc agtaggggtg 120
ggagcagtat ctcgagacct ggaaaaacat ggagcaatca caagtagcaa tacagcagct 180
accaatgctg attgtgcctg gctagaagca caagaggatg aggaagtggg ttttccagtc 240
agacctcagg tacctttaag accaatgact tacaaggagg ctgtagatct tagccacttt 300
ttaaagaaa aggggggact ggaagggtc attcactcac agaaaagaca agatatcctt 360
gatctgtggg tctaccacac acaagggtac ttccctgatt ggcagaacta cacaccaggg 420
ccaggaatca gatttccatt gacctttgga tggtgcttca agctagtacc agttgagcca 480
gaaaaggtag aagaggccaa tgaaggagag aacaactgct tgttacaccc tatgagccag 540
catgggatag aggaccggga gaaggaagtg ttagagtgga ggtttgacag caagctagca 600
tttcatcacg tggcccagaga gctgcacccg gactactaca aggactgctg a 651

```

<210> 13

<211> 671

<212> DNA

<213> Artificial Sequence

<220>

<223> sequence for opt nef (G2A, LLAA)

<400> 13

```

gatctgccac catggccggc aagtgggtcca agaggtccgt gcccggctgg tccaccgtga 60
gggagaggat gaggagggcc gagcccggcg ccgacagggt gaggaggacc gagcccggcg 120
ccgtgggcgt gggcgccgtg tccagggacc tggagaagca cggcgccatc acctcctcca 180
acaccggcgc caccaacgcc gactgcgcct ggctggaggc ccaggaggac gaggaggtgg 240
gcttccccgt gaggccccag gtgcccctga ggcccattgac ctacaagggc gccgtggacc 300
tgtcccactt cctgaaggag aagggcggcc tggagggcct gatccactcc cagaagaggc 360
aggacatcct ggacctgtgg gtgtaccaca cccagggcta cttccccgac tggcagaact 420
acacccccgg ccccgccatc aggttcccc tgacctcgg ctggtgcttc aagctggtgc 480
ccgtggagcc cgagaagggt gaggaggcca acgaggcgga gaacaactgc gccgcccacc 540
ccatgtccca gcacggcatc gaggaccccg agaaggaggt gctggagtgg aggttcgact 600
ccaagctggc cttccaccac gtggccaggg agctgcaccc cgagtactac aaggactgct 660
aaagcccggg c                                     671

```

<210> 14

<211> 216

<212> PRT

<213> Artificial Sequence

<220>

<223> opt nef (G2A, LLAA)

<400> 14

```

Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
 1              5              10              15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
      20              25              30
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
      35              40              45
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
      50              55              60
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
      65              70              75              80
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
      85              90              95
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
      100             105             110
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
      115             120             125
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
      130             135             140
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
      145             150             155             160
Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Ala Ala His
      165             170             175
Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
      180             185             190
Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
      195             200             205

```

His Pro Glu Tyr Tyr Lys Asp Cys
210 215

<210> 15

<211> 671

<212> DNA

<213> Artificial Sequence

<220>

<223> sequence for opt nef (G2A)

<400> 15

```

gatctgccac catggccggc aagtgggtcca agaggtccgt gcccggctgg tccaccgtga 60
gggagaggat gaggagggcc gagcccgccg cgcacagggg gaggaggacc gagcccgccg 120
cagtgggctg gggcgccgtg tccagggacc tggagaagca cggcgccatc acctcctcca 180
acaccgcccgc caccaacgcc gactgcgcct ggctggaggc ccaggaggac gaggaggtgg 240
gttccccctg gaggccccag gtgccccctg gggccatgac ctacaagggc gccgtggacc 300
tgtcccactt cctgaaggag aagggcgggc tggagggcct gatccactcc cagaagaggc 360
aggacatcct ggacctgtgg gtgtaccaca cccagggccta cttccccgac tggcagaact 420
acaccccccg ccccgccatc aggttcccc tgaccttcgg ctggtgcttc aagctggtgc 480
ccgtggagcc cgagaagggt gaggaggcca acgagggcga gaacaactgc ctgctgcacc 540
ccatgtccca gcacggcatc gaggaccccg agaaggaggt gctggagtgg aggttcgact 600
ccaagctggc cttccaccac gtggccaggg agctgcaccc cgagtactac aaggactgct 660
aaagcccggg c

```

671

<210> 16

<211> 216

<212> PRT

<213> Artificial Sequence

<220>

<223> opt nef (G2A)

<400> 16

```

Met Ala Gly Lys Trp Ser Lys Arg Ser Val Pro Gly Trp Ser Thr Val
1      5      10      15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Arg Val Arg Arg
20     25     30
Thr Glu Pro Ala Ala Val Gly Val Gly Ala Val Ser Arg Asp Leu Glu
35     40     45
Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr Asn Ala Asp
50     55     60
Cys Ala Trp Leu Glu Ala Gln Glu Asp Glu Glu Val Gly Phe Pro Val
65     70     75     80
Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Gly Ala Val Asp
85     90     95
Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly Leu Ile His
100    105    110
Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu Trp Val Tyr His Thr Gln
115    120    125
Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro Gly Ile Arg
130    135    140
Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val Pro Val Glu Pro
145    150    155    160

```

Glu Lys Val Glu Glu Ala Asn Glu Gly Glu Asn Asn Cys Leu Leu His
 165 170 175
 Pro Met Ser Gln His Gly Ile Glu Asp Pro Glu Lys Glu Val Leu Glu
 180 185 190
 Trp Arg Phe Asp Ser Lys Leu Ala Phe His His Val Ala Arg Glu Leu
 195 200 205
 His Pro Glu Tyr Tyr Lys Asp Cys
 210 215

<210> 17

<211> 2662

<212> DNA

<213> Artificial Sequence

<220>

<223> SEAP expression cassette

<400> 17

```

ccattgcata cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaca 60
ttaccgccat gttgacattg attattgact agttattaat agtaatacat tacgggggtca 120
ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgct 180
ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta 240
acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac 300
ttggcagtac atcaagtgtg tcatatgcc agtacgcccc ctattgacgt caatgacggg 360
aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag 420
tacatctacg tattagtcac cgctattacc atggtgatgc ggttttggca gtacatcaat 480
gggctgtgat agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat 540
gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc 600
ccattgacgc aaatgggcgc taggcgtgtg cggtgggagg tctatataag cagagctcgt 660
ttagtgaacc gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga 720
caccgggacc gatccagcct ccgcgcccg gaacgggtgca ttggaacgcg gattccccgt 780
gccaagagtg agatctaagt aagcttcctg catgctgctg ctgctgctgc tgctgggcct 840
gaggctacag ctctccctgg gcatcatccc agttgaggag gagaaccccg acttctggaa 900
ccgcgaggca gccgagggcc tgggtgccgc caagaagctg cagcctgcac agacagccgc 960
caagaacctc atcatcttcc tgggcgatgg gatgggggtg tctacgggtg cagctgccag 1020
gatcctaaaa gggcagaaga aggacaaact ggggcctgag atacccttg ccatggaccg 1080
cttcccatat gtggctctgt ccaagacata caatgtagac aaacatgtgc cagacagtcg 1140
agccacagcc acggcctacc tgtgcggggc caagggcaac ttccagacca ttggcttgag 1200
tgcagccgcc cgctttaacc agtgcaacac gacacgcggc aacgaggtca tctccgtgat 1260
gaatcgggcc aagaaagcag ggaagtcagt gggagtggta accaccacac gagtgcagca 1320
cgcctcgcca gccggcacct acgcccacac ggtgaaccgc aactggtact cggacgccga 1380
cgtgcctgcc tccgcccgcc aggaggggtg ccaggacatc gctacgcagc tcatctccaa 1440
catggacatt gacgtgatcc taggtggagg ccgaaagtac atgtttcgca tgggaacccc 1500
agaccctgag taccagatg actacagcca aggtgggacc aggctggacg ggaagaatct 1560
ggtgcaggaa tggctggcga agcgccaggg tgcccgggtat gtgtggaacc gactgagct 1620
catgcaggct tccctggacc cgtctgtgac ccattctcat ggtctctttg agcctggaga 1680
catgaaatac gagatccacc gagactccac actggacccc tccctgatgg agatgacaga 1740
ggctgccctg cgctgtctga gcaggaaccc ccgcggttc ttcctcttcg tggaggtgg 1800
tcgcatcgac catggtcatc atgaaagcag ggcttaccgg gactgactg agacgatcat 1860
gttcgacgac gccattgaga gggcgggcca gctcaccagc gaggaggaca cgctgagcct 1920
cgtcactgcc gaccactccc acgtcttctc cttcggaggc taccctctgc gaggagctc 1980
catcttcggg ctggcccctg gcaaggcccg ggacaggaag gcctacacgg tcctcctata 2040
cggaacgggt ccaggctatg tgctcaagga cggcgcccg ccgatgtta ccgagagcga 2100
gagcgggagc cccgagtatc ggcagcagtc agcagtgccc ctggacgaag agaccacgc 2160
aggcgaggac gtggcggtgt tcgcgcgcgg cccgcaggcg cacctgggtc acggcgtgca 2220

```

```
ggagcagacc ttcatagcgc acgtcatggc cttcgccgcc tgcctggagc cctacaccgc 2280
ctgcgacctg gcgccccccg ccggcaccac cgacgccgcg cacccggtt aacccggtgt 2340
ccccgcgttg cttcctctgc tggccgggac atcagggtggc ccccgctgaa ttggaatcga 2400
tcagaattca gtcgacgata tctgatcacg atctgatctg ctgtgccttc tagttgccag 2460
ccatctgttg tttgcccctc ccccgtcct tccttgacct tggaagggtg cactcccact 2520
gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt 2580
ctgggggggtg ggggtggggca gcacagcaag ggggaggatt gggaagacaa tagcaggcat 2640
gctgggggatg cggtgggctc ta 2662
```