**CHIMPANZEE ADENOVIRUS VECTORS**

A recombinant vector comprises chimpanzee adenovirus sequences and a heterologous gene under the control of regulatory sequences. A cell line which expresses chimpanzee adenovirus gene(s) is also disclosed. Methods of using the vectors and cell lines are provided.
FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>Albania</td>
<td>ES</td>
<td>Spain</td>
<td>LS</td>
<td>Lesotho</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>Armenia</td>
<td>FI</td>
<td>Finland</td>
<td>LT</td>
<td>Lithuania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>Austria</td>
<td>FR</td>
<td>France</td>
<td>LU</td>
<td>Luxembourg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
<td>GA</td>
<td>Gabon</td>
<td>LV</td>
<td>Latvia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AZ</td>
<td>Azerbaijan</td>
<td>GB</td>
<td>United Kingdom</td>
<td>MC</td>
<td>Monaco</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>Bosnia and Herzegovina</td>
<td>GE</td>
<td>Georgia</td>
<td>MD</td>
<td>Republic of Moldova</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
<td>GH</td>
<td>Ghana</td>
<td>MG</td>
<td>Madagascar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
<td>GN</td>
<td>Guinea</td>
<td>MK</td>
<td>The former Yugoslav Republic of Macedonia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
<td>GR</td>
<td>Greece</td>
<td>ML</td>
<td>Mali</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>HU</td>
<td>Hungary</td>
<td>MN</td>
<td>Mongolia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
<td>IE</td>
<td>Ireland</td>
<td>MR</td>
<td>Mauritania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
<td>IL</td>
<td>Israel</td>
<td>MW</td>
<td>Malawi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
<td>IS</td>
<td>Iceland</td>
<td>MX</td>
<td>Mexico</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
<td>IT</td>
<td>Italy</td>
<td>NE</td>
<td>Niger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
<td>JP</td>
<td>Japan</td>
<td>NL</td>
<td>Netherlands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
<td>KE</td>
<td>Kenya</td>
<td>NO</td>
<td>Norway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
<td>KG</td>
<td>Kyrgyzstan</td>
<td>NZ</td>
<td>New Zealand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
<td>KR</td>
<td>Republic of Korea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
<td>KZ</td>
<td>Kazakhstan</td>
<td>PL</td>
<td>Poland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU</td>
<td>Cuba</td>
<td>LC</td>
<td>Saint Lucia</td>
<td>PT</td>
<td>Portugal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
<td>LI</td>
<td>Liechtenstein</td>
<td>RO</td>
<td>Romania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
<td>LK</td>
<td>Sri Lanka</td>
<td>RU</td>
<td>Russian Federation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
<td>LR</td>
<td>Liberia</td>
<td>SD</td>
<td>Sudan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>Estonia</td>
<td></td>
<td></td>
<td>SE</td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SG</td>
<td>Singapore</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>Slovenia</td>
<td>SK</td>
<td>Slovakia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>Senegal</td>
<td>SZ</td>
<td>Swaziland</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>Chad</td>
<td>TG</td>
<td>Togo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TJ</td>
<td>Tajikistan</td>
<td>TM</td>
<td>Turkmenistan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>Turkey</td>
<td>TT</td>
<td>Trinidad and Tobago</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UA</td>
<td>Ukraine</td>
<td>UG</td>
<td>Uganda</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>United States of America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UZ</td>
<td>Uzbekistan</td>
<td>VN</td>
<td>Viet Nam</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YU</td>
<td>Yugoslavia</td>
<td>ZW</td>
<td>Zimbabwe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHIMPANZEE ADENOVIRUS VECTORS

This invention was supported by the National Institute of Health Grant No. DK47757. The United States government has rights in this invention.

Field of the Invention

The present invention relates to the field of vectors useful in somatic gene therapy and the production and use thereof, and also to the field of vaccines.

Background of the Invention

I. Gene Therapy

Gene therapy is an approach to treating disease, generally human disease, that is based on the modification of gene expression in cells of the patient. It has become apparent over the last decade that the single most outstanding barrier to the success of gene therapy as a strategy for treating inherited diseases, cancer, and other genetic dysfunctions is the development of useful gene transfer vehicles.

Eukaryotic viruses have been employed as vehicles for somatic gene therapy. Among the viral vectors that have been cited frequently in gene therapy research are adenoviruses. Adenoviruses are eukaryotic DNA viruses that can be modified to efficiently deliver a therapeutic or reporter transgene to a variety of cell types. Human adenoviruses are composed of a linear, approximately 36 kb double-stranded DNA genome, which is divided into 100 map units (m.u.), each of which is 360 bp in length. The DNA contains short inverted terminal repeats (ITR) at each end of the genome that are required for viral DNA replication. The gene products are organized into early (E1 through E4) and late (L1 through L5) regions, based on expression before or after the initiation of viral DNA synthesis [see, e.g., Horwitz,

Recombinant adenoviruses types 2 and 5 (Ad2 and Ad5, respectively), which cause respiratory disease in humans, are currently being developed for gene therapy. Both Ad2 and Ad5 belong to a subclass of adenovirus and are not associated with human malignancies.

Recombinant adenoviruses are capable of providing extremely high levels of transgene delivery to virtually all cell types, regardless of the mitotic state. High titers (10^{13} plaques forming units/ml) of recombinant virus can be easily generated in an adenovirus-transformed, human embryonic kidney cell line 293 [ATCC CRL1573]. The 293 cell line contains a functional adenovirus E1a gene which provides a transacting E1a protein. It can be cryo-stored for extended periods without appreciable losses.

the transduction of genes into hepatocytes in vivo has previously been demonstrated in rodents and rabbits [see, e.g., Kozarsky II, cited above, and S. Ishibashi et al, J. Clin. Invest., 92:883-893 (1993)]. Further support of the safety of recombinant adenoviruses for gene therapy is the extensive experience of live adenovirus vaccines in human populations.

However, many humans have pre-existing immunity to human adenoviruses as a result of previous natural exposure, and this immunity is a major obstacle to the use of recombinant human adenoviruses for gene therapy protocols.

II. Vaccines

Replication competent, recombinant adenovirus (Ad) containing a variety of inserted genes have been used as vaccine compositions with some success [see, e.g. Davis, U.S. Patent No. 4,920,309]. Others have described the insertion of a foreign gene into a live [L. Prevac, J. Infect. Dis., 161:27-30 (1990)] and a replication-defective adenovirus for putative use as a vaccine [See, e.g. T. Ragot et al, J. Gen. Virol., 74:501-507 (1993); M. Eliot et al, J. Gen. Virol., 71:2425-2431 (1990); and S. C. Jacobs et al, J. Virol., 66:2086-2095 (1992)]. Jacobs et al, cited above, describes a recombinant E1-deleted, E3 intact, Ad containing encephalitis virus protein NS1 under the control of a heterologous cytomegalovirus (CMV) promoter. When mice were immunized with the recombinant Ad vaccines and challenged with virus, Jacobs et al obtained partial protection (at most a 75% protection) for an average survival of 15 days. Eliot et al, cited above, describe a recombinant E1-deleted, partially E3-deleted Ad with pseudorabies glycoprotein 50 inserted into the E1 deletion site under the control of a homologous Ad promoter. In rabbits and mice, after immunization and
challenge, only partial protection was obtained (i.e., about one-third). Ragot et al, cited above, describe a recombinant E1-deleted, partially E3-deleted Ad with Epstein Barr virus glycoprotein gp340/220 inserted into the E1 deletion site under the control of a homologous Ad promoter. In marmosets (tamarins) after three high dose (5x10^9 pfu, 1x10^{10} pfu and 2x10^{10} pfu), intramuscular immunizations and viral challenge, full protection was obtained.

For certain highly infectious diseases, there is a demand for an effective vaccine. Desirably, a vaccine should be effective at a low dosage to control the occurrence of side effects or to enable sufficient amounts of vaccine to be introduced into the animal or human.

There exists a need in the gene therapy art for the development of additional adenovirus vector constructs that do not stimulate immediate immune responses which quickly eliminate the recombinant virus and the therapeutic transgene from the patient. There also exists a need in the vaccine art for new vaccine carriers, which are safe and effective in humans and other mammals.

Summary of the Invention

The present invention meets the need in the art by providing adenovirus nucleotide sequences of chimpanzee origin, a variety of novel vectors, and cell lines expressing chimpanzee adenovirus genes.

In one aspect the invention provides the nucleotide sequence of a chimpanzee C1 adenovirus. See SEQ ID NO: 1.

In another aspect the invention provides the nucleotide sequence of a chimpanzee C68 adenovirus. See SEQ ID NO: 2.
In a further aspect, the invention provides a recombinant adenovirus comprising the DNA sequence of a chimpanzee adenovirus and a selected heterologous gene operatively linked to regulatory sequences directing its expression. The recombinant virus is capable of infecting a mammalian, preferably a human, cell and capable of expressing the heterologous transgene product in the cell. In this vector, the native chimpanzee E1 gene, and/or E3 gene, and/or E4 gene may be deleted. A heterologous gene may be inserted into any of these sites of gene deletion. The heterologous transgene may encode a normal or therapeutic gene which, upon expression, replaces or modifies an inherited or acquired genetic defect. The heterologous gene may be an antigen against which a primed immune response is desired (i.e., a vaccine).

In another aspect, the invention provides a mammalian cell infected with the viral vector described above.

In still a further aspect of this invention, a novel mammalian cell line is provided which expresses a chimpanzee adenovirus gene or functional fragment thereof.

In still a further aspect, the invention provides a method for delivering a transgene into a mammalian cell comprising the step of introducing into the cell an effective amount of a recombinant virus described above.

Another aspect of this invention is a method for delivering to a mammalian patient having a disorder related to an inherited or acquired genetic defect a desired transgene. The method comprises the step of administering to the patient by an appropriate route an effective amount of an above-described recombinant
chimpanzee adenovirus containing a normal or therapeutic transgene, wherein the transgene product is expressed in vivo.

Still another aspect of this invention provides a method for eliciting an immune response in a mammalian host to protect against an infective agent. The method comprises the step of administering to the host an effective amount of a recombinant chimpanzee adenovirus comprising a heterologous gene that encodes an antigen from the infecting organism against which the immune response is targeted.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

Brief Description of the Drawings

Fig. 1A is a diagrammatic bar graph illustrating the structure of the chimpanzee adenovirus C1 (also referred to as C-1) and the location of the adenovirus genes thereon by nucleotide position and by map unit numbers appearing under the bar graph. The locations of the late genes (L-1 through L-5) are represented by arrows below the graph with molecular weight indications above the arrows and nucleotide positions below the arrows. The location of the E2a region early TATA box and transcriptional start site was not determined. The E2a region is estimated to begin approximately at nucleotide 27,100. The position of the translation initiation codon for the E2a encoded DNA binding protein is indicated by an asterisk.

Fig. 1B is a line graph showing the correlation between map units and nucleotide (base) pairs of the sequence of C1 [SEQ ID NO: 1].
Fig. 1C is a bar graph illustrating the various Bam HI clones obtained for the C1 Ad, indicating nucleotide numbers, fragment size in nucleotides, clone numbers, and fragment boundaries in nucleotides.

Fig. 2 is a tabular comparison of C1 and C68 predicted amino acid sequences examined for homology to previously described adenoviral protein sequences, Ad4, Ad5, Ad7, Ad12, and Ad40. Symbol "a" indicates that comparison of fragments of different size resulted in an underestimate of homology. Symbol "b" indicates a 95% identity from Ad-4 aa 1-95. A possible mistake in sequence apparently resulted in a frameshift and premature termination in this comparison. Symbol "c" indicates that Ad-5 has 2 small ORF's in this region encoding proteins of 64 and 67 residues with approximately 50% amino acid identity with, respectively, the amino and carboxy halves of the chimp Ad homologs. Symbol "d" indicates that Ad-3 and Ad-7 fragments were not sequenced for this protein. Symbol "e" indicates that Ad-35 and Ad-4 were not sequenced for this protein. Symbol "f" indicates that the reported sequence for Ad-7 pVIII is 197aa, and the homology begins at aa30 of the chimp Ad sequences. The homology between the chimp Ad's and Ad-7 for the 197 aa region is 98% for C-1 and 90% for C-68.

Fig. 3A is a diagrammatic bar graph illustrating the structure of the chimpanzee adenovirus C68 and the location of the adenovirus genes thereon by nucleotide position and by map unit numbers appearing under the bar graph. The locations of the late genes are represented as described for Fig. 1A. The location of the E2a region early TATA box and transcriptional start site was not determined. The E2a region is estimated to begin approximately at nucleotide 26,800. The position of the translation initiation codon for the E2a encoded
DNA binding protein is indicated by an asterisk. Although the entire genome of C68 has been cloned, certain of the fragments in Fig. 3 have been individually cloned (white bars) or not cloned (shaded bars).

Fig. 3B is a line graph showing the correlation between map units and nucleotide (base) pairs of the sequence of C68 [SEQ ID NO: 2]. White and shaded boxes are defined as in Fig. 3A.

Fig. 3C is a bar graph illustrating the various Pst fragments obtained for the C68 Ad, indicating nucleotide numbers, fragment sizes in nucleotides, clone numbers and fragment boundaries in nucleotides. White and shaded boxes are defined as in Fig. 3A.

Fig. 3D is a bar diagram illustrating Bam HI fragments of the C68 genome indicating nucleotide numbers, fragment size in nucleotides, clone numbers, and fragment boundaries in nucleotides. White and shaded boxes are defined as in Fig. 3A.

Fig. 3E is a bar diagram illustrating the HindIII-B fragment and its nucleotide boundaries and size. White and shaded boxes are defined as in Fig. 3A.

Fig. 4A is a more detailed schematic drawing of pC68-CMV-LacZ.

Fig. 4B is a schematic representation of pBS-Notx2.

Fig. 5A is a schematic drawing of plasmid pGPGK. The arrow indicates the direction of the murine PGK promoter. Restriction sites and marker genes are conventionally labeled.

Fig. 5B is a schematic drawing of plasmid pNEB-C68BamE. This plasmid contains fragments of the LacZ gene (small arrow) flanking either side of the bar indicating the C68 Ad BamE fragment. The large arrow illustrates the AmpR gene. Restriction sites and marker genes are conventionally labeled.
Fig. 5C is a schematic drawing of plasmid pGPGK-C68BamE in which the BamE fragment from pNEB-C68BamE has been cloned downstream from the PGK promoter of pGPGK.

Fig. 5D is a representation of the PCR amplification of the C68 sequence from pNEB-C68BamE, illustrating the use of primers to introduce a KpnI site just upstream of the C68 E1 region translation initiation codon at nucleotide 576 of the C68 genomic DNA and reduce the sequence distance between the promoter and C68 coding sequence. Location of the primers is indicated.

Fig. 5E is a schematic drawing of plasmid pGPGK-C68E1-ATG, in which the ATG translational start codon was moved closer to the PGK promoter.

Fig. 5F is a schematic drawing of plasmid pBS-C68BamF, in which the BamF fragment was cloned into the BamHI site of pGPGK-C68E1-ATG to generate pGPGK-C68E1 (Fig. 5G).

Fig. 5G is a schematic drawing of plasmid pGPGK-C68E1, containing the complete chimpanzee C68 Ad E1 region under the control of the murine PGK promoter.

Fig. 6A is a schematic drawing of plasmid pGPGK, a duplication of Fig. 5A for purposes of explaining construction of the C1 Ad E1 expression plasmid.

Fig. 6B illustrates the isolation of the 5' end of the C1 E1 region as a 1.9kb SnaBI - XbaI fragment.

Fig. 6C illustrates the use of primers to introduce by PCR amplification a KpnI site just upstream of the C1 E1 region translation initiation codon E1-ATG at nucleotide 577 of the C1 genomic DNA.

Fig. 6D is a schematic drawing of plasmid pGPGK-C1 mul.3-6.6 (7.4kb).

Fig. 6E is a schematic drawing of plasmid pGPGK-C1-E1ATG.
Fig. 6F is a schematic drawing of plasmid pBS-C1BamI.

Fig. 6G is a schematic drawing of plasmid pGPGK-C1E1, containing the complete chimpanzee C1 Ad E1 region under the control of the murine PGK promoter.

Fig. 7A is a schematic drawing of plasmid pSP72-Pac with indicated restriction endonuclease enzyme cleavage sites.

Fig. 7B is a schematic drawing of plasmid pNEB-C1-BamG.

Fig. 7C is a schematic drawing of plasmid pSP-C1-mu0-1.3.

Fig. 7D is a schematic drawing of plasmid pCMV-β.

Fig. 7E is a schematic drawing of plasmid pSP-C1-mu0-1.3-CMV-β.

Fig. 7F is a schematic drawing of plasmid pGEM-3Z.

Fig. 7G is a schematic drawing of plasmid pBS-C1-BamI.

Fig. 7H is a schematic drawing of plasmid pGEM-C1-mu9-10.

Fig. 7I is a schematic drawing of plasmid pBS-C1-BamE.

Fig. 7J is a schematic drawing of plasmid pGEM-C1-mu9-17.

Fig. 7K is a schematic drawing of plasmid pC1-CMV-LacZ, illustrating C1 Ad mu 0 to 1.3, followed by the CMV promoter, a splice donor/splice acceptor sequence (SD/SA), the LacZ gene, a SV40 poly A sequence and C1 Ad mu 9-17, and additional plasmid sequence. The plasmid also contains an ori and AmpR sequence.

Fig. 8A is a schematic drawing of pSP72-Pac with indicated restriction endonuclease enzyme cleavage sites.
Fig. 8B is a schematic drawing of pNEB-C68-BamE.

Fig. 8C is a schematic drawing of pSP-C68-mu 0-1.3.

Fig. 8D is a schematic drawing of pCMV-B.

Fig. 8E is a schematic drawing of pSP-C68-mu 0-1.3-CMV-B.

Fig. 8F is a schematic drawing of pGEM-3Z.

Fig. 8G is a schematic drawing of pBS-C68-BamF.

Fig. 8H is a schematic drawing of pGEM-C68-mu9-10.

Fig. 8I is a schematic drawing of pBS-C68-BamB.

Fig. 8J is a schematic drawing of pGEM-C68-mu9-16.7.

Fig. 8K is a schematic drawing of pC68-CMV-LacZ, illustrating C68 Ad mu 0 to 1.3, followed by the CMV promoter, an SD/SA, the LacZ gene, a SV40 poly A sequence and C68 Ad mu 9-16.7, and additional plasmid sequence. The plasmid also contains an ori and an AmpR sequence.

Fig. 9A is a schematic drawing of pEGFP-1 (Clontech, Palo Alto, CA).

Fig. 9B is a schematic drawing of a Not-I synthetic linker (New England Biolabs).

Fig. 9C is a schematic drawing of pEGFP-Notx2.

Fig. 9D is a schematic drawing of pC1-CMV-LacZ (from Fig. 7K).

Fig. 9E is a schematic drawing of pC68-CMV-LacZ (from Fig. 8K).

Fig. 9F is a schematic drawing of pC1-CMV-GFP, in which the GFP coding region replaces the LacZ gene of pC1-CMV-LacZ.

Fig. 9G is a schematic drawing of pC68-CMV-GFP, in which the GFP coding region replaces the LacZ gene of pC68-CMV-LacZ.
Fig. 10A is a schematic drawing of pC68-CMV-GFP as discussed in Fig. 9G.

Fig. 10B is a schematic drawing of the C68 genome.

Fig. 10C is a schematic drawing of the C68-SspI-A fragment, which is 35,199 nucleotides.

Fig. 10D is a schematic drawing of the C68-CMV-GFP genome, which is formed by homologous recombination between the C68 mu 9-16.7 sequence in pC68-CMV-GFP and the homologous sequence in the C68-SspI-A fragment.

Fig. 11A is a schematic drawing of pNEB-C1-BamG.

Fig. 11B is a schematic drawing of the C1 genome.

Fig. 11C is a schematic drawing of pNEB-C1-AscI-B.

Fig. 11D is a schematic drawing of a Not-I synthetic linker (New England Biolabs).

Fig. 11E is a schematic drawing of pNEB-C1-AscI-B-NotI.

Fig. 11F is a schematic drawing of the C1 genome.

Fig. 11G is a schematic drawing of the AscI-A fragment of the C1 genome.

Fig. 11H is a schematic drawing of the C1 genome engineered to have a unique NotI site replacing the Spe-I site in the E1B 21K protein coding region.

Detailed Description of the Invention

The present invention provides novel adenovirus vectors and packaging cell lines to produce those vectors for use in the in vitro production of recombinant proteins or fragments or other reagents, and for use in the treatment of inherited or acquired genetic disorders and abnormalities in humans and other mammals. The
present invention also provides novel vaccine compositions which comprise those vectors, the vectors comprising an inserted heterologous gene encoding an antigen from an infectious agent.

The methods of the invention involve delivering one or more selected heterologous gene(s) to a mammalian patient by administering a vector of the invention. Because the various vector constructs are derived from chimpanzee rather than from human adenoviruses, the immune system of the patient is not primed to respond immediately to the vector as a foreign antigen. A similar response would be expected where the patient was any mammal other than chimpanzee.

Use of the compositions of this invention thus permits a more stable expression of the selected transgene when administered to a non-chimpanzee, preferably human patient. Use of the compositions of this invention for vaccination permits presentation of a selected antigen for the elicitation of protective immune responses. The recombinant chimpanzee adenoviruses of this invention may also be used for producing heterologous gene products in vitro.

I. Cloning of Chimpanzee Adenovirus Sequences

Chimpanzee adenovirus, strain Bertha or C1 [ATCC Accession No. VR-20] and chimpanzee adenovirus, strain Pan-9 or CV68 [ATCC Accession No. VR-594] were obtained from the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD. For convenience, the virus CV68 is referred to throughout this specification as "C68". The viruses were originally isolated from feces [C1, Rowe et al, Proc. Soc. Exp. Med., 91:260 (1956)] or mesenteric lymph node [C68, Basnight et al, Am. J. Epidemiol., 94:166 (1971)] of infected chimpanzees.
Little is known about these viruses. However, limited restriction and immunological analyses have been published. For example, C1 was shown to be most similar to Subgroup B human adenoviruses, but it was not neutralized by heterologous sera, and no hemagglutination inhibition was observed [Wigand et al, *Intervirology*, 30:1 (1989)]. Restriction analysis demonstrated that C68 was most similar to human Ad4 serotype (Subgroup E), but only 1 in 16 enzymes tested did not distinguish C68 and Ad4 [Kitchingman, *Gene*, 20:205 (1982)].

Both chimpanzee adenoviruses grow well in human cells and were propagated in human embryonic kidney 293 cells. As described in detail in Examples 1 and 2 below, genomic DNA was isolated from purified virus stocks and digested with a panel of restriction enzymes and the restriction fragments cloned and sequenced. The genomic nucleotide sequence of C1 adenovirus is set out in SEQ ID NO: 1. The genomic nucleotide sequence of C68 adenovirus is set out in SEQ ID NO: 2.

Preliminary analysis of the sequence homology between C1, C68 and human adenoviruses was in agreement with the previously mentioned immunologic or restriction enzyme analysis. By reference to Figs. 1A-1C and 3A to 3D, it is shown that the putative E1 region of C1 occurs between about nucleotides 480 and about 3958; and of C68 between about nucleotides 480 and about 3956.

Other gene regions of C1 are identified by homology of the C1 sequence of SEQ ID NO: 1 to the known sequences of human adenoviruses Ad3, Ad5 and Ad7. Similarly, other gene regions of C68 are identified by homology of the C68 sequence of SEQ ID NO: 2 to the known sequence of human adenovirus Ad4 and Ad5. The genomic regions encoding early gene functions for E2a, E2b, E3,
E4, as well as the regions of C1 and C68 encoding late adenoviral gene products, are identified in Tables I and II below.

Table I

<table>
<thead>
<tr>
<th>Gene</th>
<th>Nucleotides</th>
<th>Map Units</th>
<th>Size (nucl./mu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1A</td>
<td>480-1540</td>
<td>1.4-4.3</td>
<td>1060/3.0</td>
</tr>
<tr>
<td>E1B</td>
<td>1566-3958</td>
<td>4.4-11.1</td>
<td>2392/6.7</td>
</tr>
<tr>
<td>E2A</td>
<td>23665-22065</td>
<td>66.6-62.1</td>
<td>1600/4.5</td>
</tr>
<tr>
<td>E2B</td>
<td>10379-3959</td>
<td>29.2-11.1</td>
<td>6420/18.1</td>
</tr>
<tr>
<td>E3</td>
<td>27181-31375</td>
<td>76.5-88.3</td>
<td>4194/11.8</td>
</tr>
<tr>
<td>E4</td>
<td>35228-32535</td>
<td>99.2-91.6</td>
<td>2693/7.6</td>
</tr>
<tr>
<td>L1</td>
<td>10893-13864</td>
<td>30.7-39.0</td>
<td>2971/8.4</td>
</tr>
<tr>
<td>L2</td>
<td>13925-17591</td>
<td>39.2-49.5</td>
<td>3666/10.3</td>
</tr>
<tr>
<td>L3</td>
<td>17641-22083</td>
<td>49.7-62.2</td>
<td>4442/12.5</td>
</tr>
<tr>
<td>L4</td>
<td>23697-27813</td>
<td>66.7-78.3</td>
<td>4116/11.6</td>
</tr>
<tr>
<td>L5</td>
<td>31556-32551</td>
<td>88.8-91.6</td>
<td>995/2.8</td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>Gene</th>
<th>Nucleotides</th>
<th>Map Units</th>
<th>Size (nucl./mu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1A</td>
<td>480-1521</td>
<td>1.3-4.2</td>
<td>1041/2.9</td>
</tr>
<tr>
<td>E1B</td>
<td>1560-3956</td>
<td>4.3-10.8</td>
<td>2396/6.6</td>
</tr>
<tr>
<td>E2A</td>
<td>23370-21787</td>
<td>64.0-59.7</td>
<td>1583/4.3</td>
</tr>
<tr>
<td>E2B</td>
<td>10346-3957</td>
<td>28.3-10.8</td>
<td>6389/17.5</td>
</tr>
<tr>
<td>E3</td>
<td>26806-31877</td>
<td>73.4-87.3</td>
<td>5071/13.9</td>
</tr>
<tr>
<td>E4</td>
<td>36193-33486</td>
<td>99.1-91.7</td>
<td>2707/7.4</td>
</tr>
<tr>
<td>L1</td>
<td>10823-13817</td>
<td>29.6-37.8</td>
<td>2994/8.2</td>
</tr>
<tr>
<td>L2</td>
<td>13884-17431</td>
<td>38.0-47.7</td>
<td>3547/9.7</td>
</tr>
<tr>
<td>L3</td>
<td>17480-21804</td>
<td>47.9-59.7</td>
<td>4324/11.8</td>
</tr>
<tr>
<td>L4</td>
<td>23399-27439</td>
<td>64.1-75.1</td>
<td>4040/11.1</td>
</tr>
<tr>
<td>L5</td>
<td>32134-33502</td>
<td>88.0-91.7</td>
<td>1368/3.7</td>
</tr>
</tbody>
</table>

Our preliminary experiments demonstrated that human antisera do not neutralize the chimpanzee adenoviruses in neutralizing antibody assays (see, e.g., International patent application PCT95/03035), thus indicating the desirability of vectors prepared from these sequences for gene therapy in humans. As further described in the examples, plasmids establishing chimpanzee adenovirus E1-expressing cell lines and
recombinant E1-deleted adenoviruses expressing a transgene are prepared.

The viral sequences used in the vectors and cell lines described below may be generated by using the teachings and references contained herein, coupled with standard recombinant molecular cloning techniques known and practiced by those skilled in the art.

II. E1-Expressing Complementation Cell Lines

To generate recombinant chimpanzee adenoviruses (Ad) deleted in any of the genes described above, the function of the deleted gene region, if essential to the replication and infectivity of the virus, must be supplied to the recombinant virus by a helper virus or cell line, i.e., a complementation or packaging cell line. For example, to generate a replication-defective chimpanzee adenovirus vector, a cell line is needed which expresses the E1 gene products of the chimpanzee adenovirus. The protocol for the generation of the cell lines expressing the chimpanzee E1 gene products (Examples 3 and 4) is followed to generate a cell line which expresses any selected chimpanzee adenovirus gene.

Conventional assays were not useful in identifying the chimpanzee adenovirus E1-expressing cell line and a novel AAV augmentation assay was developed to identify the chimpanzee adenovirus E1-expressing cell line. This assay is useful to identify E1 function in cell lines made by using the E1 genes of other uncharacterized adenoviruses, e.g., from other species. That assay is described in Example 4B below.

According to this invention, the selected chimpanzee adenovirus gene, e.g., E1, is under the transcriptional control of a promoter for expression in a selected parent cell line. Inducible or constitutive promoters may be employed for this purpose. Among
inducible promoters are included the sheep metallothionine promoter, inducible by zinc, or the mouse mammary tumor virus (MMTV) promoter, inducible by a glucocorticoid, particularly, dexamethasone. Other inducible promoters, such as those identified in International patent application WO95/13392, published May 18, 1995, and incorporated by reference herein may also be used in the production of packaging cell lines according to this invention. Constitutive promoters in control of the expression of the chimpanzee adenovirus gene may be employed also. The promoter used to express E1 as exemplified below is the well-known constitutive murine PGK promoter.

A parent cell is selected for the generation of a novel cell line expressing any desired C1 or C68 gene. Without limitation, such a parent cell line may be HeLa [ATCC Accession No. CCL 2], A549 [ATCC Accession No. CCL 185], KB [CCL 17], Detroit [e.g., Detroit 510, CCL 72] and WI-38 [CCL 75] cells. These cell lines are all available from the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA. Other suitable parent cell lines may be obtained from other sources.

The present invention provides an exemplary cell line which contains and expresses the chimpanzee C68 or C1 Ad E1 gene, as described in detail in Examples 3 and 4 below. Briefly described, the entire chimpanzee adenovirus E1 region was cloned and, by a series of plasmid manipulations, it was placed under the control of a murine PGK promoter in a desired shuttle vector. See Figs. 5A-5G and 6A-6G.

After the desired shuttle vector containing the adenoviral sequences (i.e., pGPGK-C68 E1 described in Example 3) was transfected into the selected parental cell line (e.g., HeLa), expression of the E1
gene was detected. Conventional G418 selection as described in Example 4A was used to generate stable clones of these E1-expressing cells. The resulting cell line is thus able to provide chimpanzee Ad E1 gene products to the replication-defective recombinant virus (see Example 5) to allow productive infection and recovery of the recombinant virus.

The E1-expressing cell lines are useful in the generation of recombinant chimpanzee adenovirus E1 deleted vectors. Cell lines constructed using essentially the same procedures that express one or more other chimpanzee adenoviral gene products are useful in the generation of recombinant chimpanzee adenovirus vectors deleted in the genes that encode those products.

Further, cell lines which express other human Ad E1 gene products are also useful in generating the chimpanzee recombinant Ads of this invention.

III. Recombinant Viral Particles as Vectors

The compositions of this invention comprise desirable viral vectors, that deliver a functional, normal or therapeutic gene to cells. Such vectors comprise chimpanzee adenovirus DNA sequence and a selected heterologous gene operatively linked to regulatory sequences which direct expression of the gene. The vector is capable of expressing the gene product in an infected mammalian cell. The vector is preferably functionally deleted in one or more viral genes. A minigene comprises the heterologous gene under the control of regulatory sequences. Optional helper viruses and/or packaging cell lines supply to the chimpanzee viral vectors any necessary products of deleted adenoviral genes.

The term "functionally deleted" means that a sufficient amount of the gene region is removed or otherwise damaged, e.g., by mutation or modification, so
that the gene region is no longer capable of producing functional products of gene expression. If desired, the entire gene region may be removed.

The viral sequences, helper viruses, if needed, and recombinant viral particles, and other vector components and sequences employed in the construction of the vectors described herein are obtained as described above. The DNA sequences of the two chimpanzee adenoviruses are employed to construct vectors and cell lines useful in the preparation of such vectors.

Modifications of the nucleic acid sequences forming the vectors of this invention, including sequence deletions, insertions, and other mutations may be generated using standard molecular biological techniques and are within the scope of this invention.

A. The "Minigene"

The methods employed for the selection of the transgene, the cloning and construction of the "minigene" and its insertion into the viral vector are within the skill in the art given the teachings provided herein. By "minigene" is meant the combination of a selected heterologous gene and the other regulatory elements necessary to transcribe the gene and express the gene product in a host cell. The gene is operatively linked to regulatory components in a manner which permits its transcription. Such components include conventional regulatory elements necessary to drive expression of the transgene in a cell transfected with the viral vector.

Thus the minigene also contains a selected promoter which is linked to the transgene and located, with other regulatory elements, within the selected viral sequences of the recombinant vector.

Selection of the promoter is a routine matter and is not a limitation of this invention.
Useful promoters may be constitutive promoters or regulated (inducible) promoters, which will enable control of the amount of the transgene to be expressed. For example, a desirable promoter is that of the cytomegalovirus immediate early promoter/enhancer [see, e.g., Boshart et al, Cell, 41:521-530 (1985)]. Another desirable promoter includes the Rous sarcoma virus LTR promoter/enhancer. Still another promoter/enhancer sequence is the chicken cytoplasmic β-actin promoter [T. A. Kost et al, Nucl. Acids Res., 11(23):8287 (1983)]. Other suitable or desirable promoters may be selected by one of skill in the art.

The minigene may also desirably contain nucleic acid sequences heterologous to the viral vector sequences including sequences providing signals required for efficient polyadenylation of the transcript (poly-A or pA) and introns with functional splice donor and acceptor sites. A common poly-A sequence which is employed in the exemplary vectors of this invention is that derived from the papovavirus SV-40. The poly-A sequence generally is inserted in the minigene following the transgene sequences and before the viral vector sequences. A common intron sequence is also derived from SV-40, and is referred to as the SV-40 T intron sequence. A minigene of the present invention may also contain such an intron, desirably located between the promoter/enhancer sequence and the transgene. Selection of these and other common vector elements are conventional [see, e.g., Sambrook et al, "Molecular Cloning. A Laboratory Manual.", 2d edit., Cold Spring Harbor Laboratory, New York (1989) and references cited therein] and many such sequences are available from commercial and industrial sources as well as from Genbank.

As above stated, the minigene is located in the site of any selected deletion in the viral
vector, such as the site of the E1 gene region deletion or E3 gene region deletion, among others which may be selected.

B. Construction of The Viral Plasmid Vector

The chimpanzee adenovirus vectors useful in this invention include recombinant, defective adenoviruses, that is, chimpanzee adenovirus sequences functionally deleted in the E1a or E1b genes, and optionally bearing other mutations, e.g., temperature-sensitive mutations or deletions in other genes. It is anticipated that these chimpanzee sequences are also useful in forming hybrid vectors from other adenovirus and/or adeno-associated virus sequences. Homologous adenovirus vectors prepared from human adenoviruses are described in the published literature [see, for example, Kozarsky I and II, cited above, and references cited therein, U. S. Patent No. 5,240,846].

In the construction of useful chimpanzee adenovirus vectors for delivery of a gene to the human (or other mammalian) cell, a range of adenovirus nucleic acid sequences can be employed in the vectors. A vector comprising minimal chimpanzee adenovirus sequences may be used in conjunction with a helper virus to produce an infectious recombinant virus particle. The helper virus provides essential gene products required for viral infectivity and propagation of the minimal chimpanzee adenoviral vector. When only one or more selected deletions of chimpanzee adenovirus genes are made in an otherwise functional viral vector, the deleted gene products can be supplied in the viral vector production process by propagating the virus in a selected packaging cell line that provides the deleted gene functions in trans.
1. **Recombinant Minimal Adenovirus**

A minimal chimpanzee Ad virus is a viral particle containing only the adenovirus cis-elements necessary for replication and virion encapsidation, which cis-elements flank the heterologous gene. That is, the vector contains only the cis-acting 5' and 3' inverted terminal repeat (ITR) sequences of the adenoviruses of this invention (which function as origins of replication) and the native 5' packaging/enhancer domains (that contain sequences necessary for packaging linear Ad genomes and enhancer elements for the E1 promoter). See, for example, the techniques described for preparation of a "minimal" human Ad vector in International Patent Application WO96/13597, published May 9, 1996, and incorporated herein by reference.

2. **Other Defective Adenoviruses**

Recombinant, replication-deficient adenoviruses of this invention may also contain more than the minimal chimpanzee adenovirus sequences defined above. These other Ad vectors can be characterized by deletions of various portions of gene regions of the virus, and infectious virus particles formed by the optional use of helper viruses and/or packaging cell lines, as described herein.

As one example, suitable vectors may be formed by deleting all or a sufficient portion of the adenoviral immediate early gene E1a and delayed early gene E1b, so as to eliminate their normal biological functions. Replication-defective E1-deleted viruses are capable of replicating and producing infectious virus when grown on a chimpanzee adenovirus-transformed, complementation cell line containing functional adenovirus E1a and E1b genes which provide the corresponding gene products in trans. Based on the homologies to known adenovirus sequences, it is
anticipated that, as is true for the human recombinant E1-deleted adenoviruses of the art, the resulting recombinant chimpanzee adenovirus is capable of infecting many cell types and can express a transgene, but cannot replicate in most cells that do not carry the chimpanzee E1 region DNA unless the cell is infected at a very high multiplicity of infection.

As another example, all or a portion of the adenovirus delayed early gene E3 may be eliminated from the chimpanzee adenovirus sequence which forms a part of the recombinant virus. The function of chimpanzee E3 is believed to be irrelevant to the function and production of the recombinant virus particle.

Chimpanzee adenovirus vectors may also be constructed having a deletion of the E4 gene. Still another vector of this invention contains a deletion in the delayed early gene E2a.

Deletions may also be made in any of the late genes L1 through L5 of the chimpanzee adenovirus genome. Similarly, deletions in the intermediate genes IX and IVa2 may be useful for some purposes. Other deletions may be made in the other structural or non-structural adenovirus genes.

The above discussed deletions may be used individually, i.e., an adenovirus sequence for use in the present invention may contain deletions of E1 only. Alternatively, deletions of entire genes or portions thereof effective to destroy their biological activity may be used in any combination. For example, in one exemplary vector, the adenovirus sequence may have deletions of the E1 genes and the E4 gene, or of the E1, E2a and E3 genes, or of the E1 and E3 genes, or of E1, E2a and E4 genes, with or without deletion of E3, and so on. As discussed above, such deletions may be used in
combination with other mutations, such as temperature-sensitive mutations, to achieve a desired result.

The minigene containing the transgene may be inserted optionally into any deleted region of the chimpanzee Ad virus. Alternatively, the minigene may be inserted into an existing gene region to disrupt the function of that region, if desired.

The construction of exemplary E1-deleted chimpanzee Ad virus vectors is described in detail in Example 5 below. Desirably, such a vector contains chimpanzee adenovirus sequences Ad m.u. 0-1.3, followed by a minigene containing the transgene of interest (e.g., a therapeutic gene for the correction of a genetic defect in a patient or a marker gene to visualize infected cells) and the sequence Ad m.u. 9 to 100 of C1 or C68. These recombinant adenoviruses are functionally deleted of E1a and E1b.

C. Production of the Recombinant Viral Particle

1. Helper Viruses

Depending upon the chimpanzee adenovirus gene content of the viral vectors employed to carry the minigene, a helper adenovirus or non-replicating virus fragment may be necessary to provide sufficient chimpanzee adenovirus gene sequences necessary to produce an infective recombinant viral particle containing the minigene.

Useful helper viruses contain selected adenovirus gene sequences not present in the adenovirus vector construct and/or not expressed by the packaging cell line in which the vector is transfected. A preferred helper virus is desirably replication-defective and contains a variety of adenovirus genes in addition to the sequences described above. The helper
virus is desirably used in combination with the E1-expressing cell lines described herein.

Most preferably for C68, the "helper" virus is a fragment formed by clipping the C-terminal end of the C68 genome with SspI, which removes about 1300 bp from the left end of the virus. This clipped virus is then co-transfected into the E1-expressing cell line with the plasmid DNA, thereby forming the recombinant virus by homologous recombination with the C68 sequences in the plasmid.

Because there is no similarly unique restriction site in the 5' end of C1, to create a recombinant virus, the SpeI site at position 1733 is replaced with a unique NotI site, generating the modified C1 NotI genome of about 35,526 bp. See, e.g., Figs 12A-12F.

Helper viruses may also be formed into poly-cation conjugates as described in Wu et al, *J. Biol. Chem.*, 264:16985-16987 (1989); K. J. Fisher and J. M. Wilson, *Biochem. J.*, 299:49 (April 1, 1994). Helper virus may optionally contain a second reporter minigene. A number of such reporter genes are known to the art. The presence of a reporter gene on the helper virus which is different from the transgene on the adenovirus vector allows both the Ad vector and the helper virus to be independently monitored. This second reporter is used to enable separation between the resulting recombinant virus and the helper virus upon purification.

2. **Assembly of Viral Particle and Infection of a Cell Line**

Assembly of the selected DNA sequences of the adenovirus, and the transgene and other vector elements into various intermediate plasmids and shuttle vectors, and the use of the plasmids and vectors
to produce a recombinant viral particle are all achieved using conventional techniques. Such techniques include conventional cloning techniques of cDNA such as those described in texts [Sambrook et al, cited above], use of overlapping oligonucleotide sequences of the adenovirus genomes, polymerase chain reaction, and any suitable method which provides the desired nucleotide sequence. Standard transfection and co-transfection techniques are employed, e.g., CaPO₄ precipitation techniques. Other conventional methods employed include homologous recombination of the viral genomes, plaquing of viruses in agar overlay, methods of measuring signal generation, and the like.

For example, following the construction and assembly of the desired minigene-containing viral vector, the vector is transfected in vitro in the presence of a helper virus into the packaging cell line. Homologous recombination occurs between the helper and the vector sequences, which permits the adenovirus-transgene sequences in the vector to be replicated and packaged into virion capsids, resulting in the recombinant viral vector particles. The current method for producing such virus particles is transfection-based. However, the invention is not limited to such methods.

The resulting recombinant chimpanzee adenoviruses are useful in transferring a selected transgene to a selected cell. In in vivo experiments with the recombinant virus grown in the packaging cell lines, the E1-deleted recombinant chimpanzee adenovirus demonstrates utility in transferring a transgene to a non-chimpanzee, preferably a human, cell.
IV. Use of the Recombinant Virus Vectors

The resulting recombinant chimpanzee adenovirus containing the minigene (produced by cooperation of the adenovirus vector and helper virus or adenoviral vector and packaging cell line, as described above) thus provides an efficient gene transfer vehicle which can deliver the transgene to a human patient in vivo or ex vivo.

The above-described recombinant vectors are administered to humans according to published methods for gene therapy. A chimpanzee viral vector bearing the selected transgene may be administered to a patient, preferably suspended in a biologically compatible solution or pharmaceutically acceptable delivery vehicle. A suitable vehicle includes sterile saline. Other aqueous and non-aqueous isotonic sterile injection solutions and aqueous and non-aqueous sterile suspensions known to be pharmaceutically acceptable carriers and well known to those of skill in the art may be employed for this purpose.

The chimpanzee adenoviral vectors are administered in sufficient amounts to transduce the human cells and to provide sufficient levels of gene transfer and expression to provide a therapeutic benefit without undue adverse or with medically acceptable physiological effects, which can be determined by those skilled in the medical arts. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, direct delivery to the liver, intranasal, intravenous, intramuscular, subcutaneous, intradermal, oral and other parental routes of administration. Routes of administration may be combined, if desired.

Dosages of the viral vector will depend primarily on factors such as the condition being treated, the age, weight and health of the patient, and may thus
vary among patients. For example, a therapeutically effective human dosage of the viral vector is generally in the range of from about 20 to about 100 ml of saline solution containing concentrations of from about $1 \times 10^9$ to $1 \times 10^{11}$ pfu/ml virus vector. A preferred human dosage is estimated to be about 50 ml saline solution at $2 \times 10^{10}$ pfu/ml. The dosage will be adjusted to balance the therapeutic benefit against any side effects and such dosages may vary depending upon the therapeutic application for which the recombinant vector is employed. The levels of expression of the transgene can be monitored to determine the frequency of dosage administration.

An optional method step involves the co-administration to the patient, either concurrently with, or before or after administration of the viral vector, of a suitable amount of a short acting immune modulator. The selected immune modulator is defined herein as an agent capable of inhibiting the formation of neutralizing antibodies directed against the recombinant vector of this invention or capable of inhibiting cytolytic T lymphocyte (CTL) elimination of the vector. The immune modulator may interfere with the interactions between the T helper subsets ($T_{H1}$ or $T_{H2}$) and B cells to inhibit neutralizing antibody formation. Alternatively, the immune modulator may inhibit the interaction between $T_{H1}$ cells and CTLs to reduce the occurrence of CTL elimination of the vector.

The recombinant chimpanzee adenoviruses may also be employed as vaccines or immune response-inducing compositions. The present invention provides a recombinant replication-defective chimpanzee Ad which can contain in any of its adenovirus sequence deletions a gene encoding a desired antigen. The chimpanzee adenovirus is likely to be better suited for use as a live recombinant virus vaccine in different animal species compared to an adenovirus of human origin. The recombinant adenoviruses can be used as prophylactic or therapeutic vaccines against any pathogen for which the antigen(s) crucial for induction of an immune response and able to limit the spread of the pathogen has been identified and for which the cDNA is available.

Because the recombinant chimpanzee adenoviruses described above are deleted in the E1 sequences, the adenoviruses are replication defective and thus highly unlikely to spread within a host or among individuals. The recombinant virus lacks oncogenic potential because the E1 gene, that can function as an oncogene in some adenovirus strains, has been deleted.

With respect to efficacy, the recombinant, replication-defective adenoviruses of this invention are expected to be highly efficacious at inducing cytolytic T cells and antibodies to the inserted heterologous antigenic protein expressed by the virus. This has been demonstrated with a recombinant, replication-defective human Ad containing a sequence encoding the rabies virus glycoprotein as the heterologous gene. See, e.g., Z. Q. Xiang et al., *Virol.*, 219:220-227 (1996).

As described above and in the examples below, in the site of the E1 deletion of either of the two chimpanzee adenoviruses of this invention, and under control of a promoter heterologous to adenovirus, a sequence encoding a protein heterologous to the
adenovirus is inserted using techniques known to those of skill in the art. The heterologous nucleic acid encodes a protein which is desirably capable of inducing an immune response to a pathogen when administered to an immunocompetent host. Such a protein may be a protein from, among others, rabies virus, human papilloma virus, human immunodeficiency virus (HIV), and respiratory syncytial virus (RSV), as well as antigens associated with diseases of other mammals.

It is also anticipated that the vaccine method of the present invention may be employed with a tumor-associated protein specific for a selected malignancy. These tumor antigens include viral oncogenes, such as E6 and E7 of human papilloma virus, or cellular oncogenes such as mutated ras or p53. Particularly, where the condition is human immunodeficiency virus (HIV) infection, the protein is preferably HIV glycoprotein 120 for which sequences are available from GenBank. Where the condition is human papilloma virus infection, the protein is selected from the group consisting of E6, E7 and/or L1 [Seedorf, K. et al, Virol., 145:181-185 (1985)]. Where the condition is respiratory syncytial virus infection, the protein is selected from the group consisting of the glyco- (G) protein and the fusion (F) protein, for which sequences are available from GenBank. In addition to these proteins, other virus-associated proteins, including proteins which are antigens for disease-causing agents of other mammals, e.g., domestic animals, horses, farm animals, etc., are readily available to those of skill in the art. Selection of the heterologous proteins is not a limiting factor in the design of vaccine compositions of this invention.
A recombinant replication-defective chimpanzee adenoviral vector bearing a gene encoding an immunogenic protein may be administered to a human or other mammalian patient, preferably suspended in a biologically compatible solution or pharmaceutically acceptable delivery vehicle. A suitable vehicle is sterile saline. Other aqueous and non-aqueous isotonic sterile injection solutions and aqueous and non-aqueous sterile suspensions known to be pharmaceutically acceptable carriers and well known to those of skill in the art may be employed for this purpose.

Optionally, a vaccinal composition of the invention may be formulated to contain other components, including, e.g. adjuvants, stabilizers, pH adjusters, preservatives and the like. Such components are well known to those of skill in the vaccine art.

The recombinant, replication defective adenoviruses are administered in a "pharmaceutically effective amount", that is, an amount of recombinant adenovirus that is effective in a route of administration to transfect the desired cells and provide sufficient levels of expression of the selected gene to provide a vaccinal benefit, i.e., some measurable level of protective immunity.

Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, intranasal, intramuscular, intratracheal, subcutaneous, intradermal, rectal, oral and other parental routes of administration. Routes of administration may be combined, if desired, or adjusted depending upon the immunogen or the disease. For example, in prophylaxis of rabies, the subcutaneous, intratracheal and intranasal routes are preferred. The route of administration primarily will depend on the nature of the disease being treated.
Doses or effective amounts of the recombinant replication-defective Ad virus will depend primarily on factors such as the condition, the selected gene, the age, weight and health of the animal, and may thus vary among animals. For example, a prophylactically effective amount or dose of the Ad vaccine is generally in the range of from about 100 μl to about 10 ml of saline solution containing concentrations of from about $1 \times 10^4$ to $1 \times 10^7$ plaque forming units (pfu) virus/ml. A preferred dose is from about 1 to about 10 ml saline solution at the above concentrations. The levels of immunity of the selected gene can be monitored to determine the need, if any, for boosters. Following an assessment of antibody titers in the serum, optional booster immunizations may be desired.

An additional use of the recombinant adenovirus vectors described herein resides in their use as expression vectors for the production of the products encoded by the heterologous genes. For example, the recombinant adenoviruses containing a gene inserted into the location of an E1 deletion may be transfected into an E1-expressing cell line as described above. The transfected cells are then cultured in the conventional manner, allowing the recombinant adenovirus to express the gene product from the promoter. The gene product may then be recovered from the culture medium by known conventional methods of protein isolation and recovery from culture.

The following examples illustrate the cloning of the chimpanzee adenoviruses and the construction and testing of the chimpanzee Ad E1 expressing cell line and the construction of exemplary recombinant adenovirus vectors of the present invention. These examples are illustrative only, and do not limit the scope of the present invention.
Example 1 - Virus Stocks and Propagation

The C1 [ATCC Accession No. VR-20] and C68 [ATCC Accession No. 594] virus stocks were obtained and propagated in 293 cells [ATCC CRL1573] cultured in Dulbecco's Modified Eagles Medium (DMEM; Sigma, St. Louis, MO.) supplemented with 10% fetal calf serum (FCS) [Sigma or Hyclone, Logan, UT] and 1% Penicillin-Streptomycin (Sigma). Infection of 293 cells was carried out in DMEM supplemented with 2% FCS for the first 24 hours, after which FCS was added to bring the final concentration to 10%. Infected cells were harvested when 100% of the cells exhibited virus-induced cytopathic effect (CPE), collected, and concentrated by centrifugation. Cell pellets were resuspended in 10 mM Tris (pH 8.0), and lysed by 3 cycles of freezing and thawing.

Virus preparations were obtained following two ultra centrifugation steps on cesium chloride density gradients and stocks of virus were diluted to $1 \times 10^{12}$ particles/ml in 10 mM Tris/100 mM NaCl/50% glycerol and stored at -70°C.

Example 2 - Cloning and Sequencing of Viral Genomic DNA

Genomic DNA was isolated from the purified virus preparations of Example 1, following standard methods [see, e.g., M. S. Horwitz et al, "Adenoviridae and Their Replication", Virolology, second edition, pp. 1712, ed. B. N. Fields et al, Raven Press Ltd., New York (1990); B. J. Carter, in "Handbook of Parvoviruses", ed. P. Tijssser, CRC Press, pp. 155-168 (1990)] and digested with a panel of 16 restriction enzymes following the manufacturers' recommendations. Enzymes that cut the DNA 10-15 times were utilized for cloning of the viral DNA into pBluescript SK+. Except as noted, all restriction
and modifying enzymes used in this and the following examples were obtained from Boehringer Mannheim, Indianapolis, IN.

Manipulation of the genomic DNA to remove the covalently attached terminal protein was performed [Berkner and Sharp, *Nucleic Acids Res.*, 11: 6003 (1983)]. Taking advantage of the absence of Pac-I restriction sites, synthetic PacI linkers (New England Biolabs, Beverly, MA) were ligated onto the ends of the genomic DNA. Genomic DNA was digested with BamHI, PstI, SalI or XbaI and the restriction fragments (all but the genomic terminal fragments) were cloned into pBluescript SK+ (Stratagene, La Jolla, CA). Fragments containing the left and right genomic termini were cloned into pNEB-193 (New England Biolabs, Beverly, MA) as Pac-I/BamHI or Pac-I/Pst-I fragments.

The clones generated for C1 and C68 are illustrated in Figs. 1C and 3C, respectively. The cloned fragments are described in Table III(C1) [nucleotide sequence numbers correspond with SEQ ID NO: 1] and Table IVA-IVB (C68) [nucleotide sequence numbers correspond with SEQ ID NO: 2].
<table>
<thead>
<tr>
<th>Construct Name</th>
<th>Insert Size</th>
<th>Clone #</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS:Cl-Bam-A</td>
<td>8477</td>
<td>250,260</td>
<td>6135-14611</td>
</tr>
<tr>
<td></td>
<td></td>
<td>281</td>
<td></td>
</tr>
<tr>
<td>PBS:Cl-Bam-B</td>
<td>8253</td>
<td>285</td>
<td>24678-32930</td>
</tr>
<tr>
<td>PBS:Cl-Bam-C</td>
<td>3990</td>
<td>252</td>
<td>17259-21248</td>
</tr>
<tr>
<td>PBS:Cl-Bam-D</td>
<td>3429</td>
<td>263,269</td>
<td>21250-24677</td>
</tr>
<tr>
<td></td>
<td></td>
<td>275</td>
<td></td>
</tr>
<tr>
<td>PBS:Cl-Bam-E</td>
<td>2537</td>
<td>251</td>
<td>3598-6134</td>
</tr>
<tr>
<td>PBS:Cl-Bam-F</td>
<td>2203</td>
<td>267,270</td>
<td>14612-16814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>pNEB:Cl-Bam-G</td>
<td>1927</td>
<td>516</td>
<td>1-1927 left end</td>
</tr>
<tr>
<td>PBS:Cl-Bam-H</td>
<td>1632</td>
<td>486,487</td>
<td>32931-34562</td>
</tr>
<tr>
<td>PBS:Cl-Bam-I</td>
<td>1538</td>
<td>288-293</td>
<td>2060-3597</td>
</tr>
<tr>
<td></td>
<td></td>
<td>483,485</td>
<td></td>
</tr>
<tr>
<td>pNEB:Cl-Bam-J</td>
<td>962</td>
<td>519</td>
<td>34563-35524</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>right end</td>
</tr>
<tr>
<td>PBS:Cl-Bam-K</td>
<td>288</td>
<td>256,295</td>
<td>16971-17258</td>
</tr>
<tr>
<td></td>
<td></td>
<td>296,298</td>
<td></td>
</tr>
<tr>
<td>PBS:Cl-Bam-L</td>
<td>156</td>
<td>260</td>
<td>16815-16970</td>
</tr>
<tr>
<td>PBS:Cl-Bam-M</td>
<td>132</td>
<td>259,261</td>
<td>1928-2059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>PBS:Cl-Bam-A/Pst</td>
<td></td>
<td>423-428</td>
<td>subclone of 250</td>
</tr>
<tr>
<td>PBS:Cl-Bam-B/HindIII</td>
<td></td>
<td>429-434</td>
<td>subclone of 285</td>
</tr>
<tr>
<td>pNEB:C-1AscB</td>
<td>7937</td>
<td>955</td>
<td>1-7937 left end</td>
</tr>
</tbody>
</table>

Table III
<table>
<thead>
<tr>
<th>Construct Name</th>
<th>Size</th>
<th>Clone #</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBS:C68-Pst-A</td>
<td>6768</td>
<td>133,141,213-217,303-305</td>
<td>24790-31554</td>
</tr>
<tr>
<td>pBS:C68-Pst-B</td>
<td>6713</td>
<td>219-221</td>
<td>4838-11550</td>
</tr>
<tr>
<td>pBS:C68-Pst-C</td>
<td>5228</td>
<td>78,140</td>
<td>14811-20038</td>
</tr>
<tr>
<td>pBS:C68-Pst-D</td>
<td>2739</td>
<td>127,129,146,151</td>
<td>20039-22685</td>
</tr>
<tr>
<td>pBS:C68-Pst-E</td>
<td>2647</td>
<td>138,149,222-224</td>
<td>32046-33996</td>
</tr>
<tr>
<td>pBS:C68-Pst-F</td>
<td>1951</td>
<td>502,505</td>
<td>1-1874 left end</td>
</tr>
<tr>
<td>pNEB:C68-Pst-G</td>
<td>1874</td>
<td>506</td>
<td>30182-36536</td>
</tr>
<tr>
<td>pBS:C68-Pst-H</td>
<td>1690</td>
<td>128,135,145,152</td>
<td>23094-24783</td>
</tr>
<tr>
<td>pBS:C68-Pst-I</td>
<td>1343</td>
<td>222-224</td>
<td>33997-35339</td>
</tr>
<tr>
<td>pNEB:C68-Pst-J</td>
<td>1180</td>
<td>508</td>
<td>35340-36519</td>
</tr>
<tr>
<td>pBS:C68-Pst-K</td>
<td>1111</td>
<td>87,131,132,136,225-230</td>
<td>2763-3873</td>
</tr>
<tr>
<td>pBS:C68-Pst-L</td>
<td>964</td>
<td>320,321,323,324</td>
<td>3874-4837</td>
</tr>
<tr>
<td>pBS:C68-Pst-M</td>
<td>888</td>
<td>319,322</td>
<td>1875-2762</td>
</tr>
<tr>
<td>pBS:C68-Pst-N</td>
<td>408</td>
<td>84,125,130</td>
<td>22686-23093</td>
</tr>
<tr>
<td>pBS:C68-Pst-O</td>
<td>380</td>
<td>31666-32045</td>
<td></td>
</tr>
<tr>
<td>pBS:C68-Pst-P</td>
<td>285</td>
<td>79,126</td>
<td>11551-11835</td>
</tr>
<tr>
<td>pBS:C68-Pst-Q</td>
<td>236</td>
<td>31552-31665</td>
<td></td>
</tr>
<tr>
<td>pBS:C68-Pst-R</td>
<td>114</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>
Table IVB

<table>
<thead>
<tr>
<th>BamHI Fragments</th>
<th>Size</th>
<th>Clone #</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS:C68-Bam-A</td>
<td>16684</td>
<td></td>
<td>19836-36519 right end</td>
</tr>
<tr>
<td>5</td>
<td>PBS:C68-Bam-B</td>
<td>8858</td>
<td>95, 99, 101-103, 119-121, 165, 166, 169, 171</td>
</tr>
<tr>
<td>10</td>
<td>PBS:C68-Bam-C</td>
<td>4410</td>
<td>104, 106, 167, 179, 171</td>
</tr>
<tr>
<td></td>
<td>PBS:C68-Bam-D</td>
<td>2986</td>
<td>195-197</td>
</tr>
<tr>
<td></td>
<td>pNEB:C68-Bam-E</td>
<td>2041</td>
<td>537, 545</td>
</tr>
<tr>
<td>15</td>
<td>PBS:C68-Bam-F</td>
<td>1540</td>
<td>198-200</td>
</tr>
</tbody>
</table>

HindIII Fragments

| pBR:C-68-Hind-B | 9150 | 489, 419, 492, 23471-32620 |

---

Cloned restriction fragments were ordered in the genome by comparison to known adenoviral sequences. The nucleotide sequence of both viruses was determined [Commonwealth Biotechnologies Incorporated, Richmond, VA]. The nucleotide sequence of the top strand of C1 DNA is reported in SEQ ID NO: 1. The nucleotide sequence of the top strand of C68 DNA is reported in SEQ ID NO: 2. Restriction maps were generated using a number of enzymes and compared to data obtained from restricted genomic DNA following electrophoreses on agarose gels.

Regulatory and coding regions in the viral DNA sequences were identified by homology to known adenoviral sequences using the Mac Vector program (Oxford Molecular Group) and a MacIntosh Quadra 610 computer (Apple Computer, Cupertino, CA). See Tables I and II. Open
reading frames were translated and the predicted amino acid sequences examined for homology to previously described adenoviral protein sequences, Ad4, Ad5, Ad7, Ad12, and Ad40. See Fig. 2 below.

The C1 E1 coding region is defined as the sequences between the E1A translation initiation site at nucleotide 576 of SEQ ID NO: 1 and the E1B translation termination signal at nucleotide 3507 of SEQ ID NO: 1. The corresponding sequences in the C68 genome are located at nucleotides 577 and 3510 of SEQ ID NO: 2. Other open reading frames and regulatory elements of the viruses are being examined for homology with other adenoviral sequences.

Our preliminary experiments have demonstrated that human antisera do not neutralize the chimpanzee adenoviruses in neutralizing antibody assays.

Example 3 - Generation of Plasmid Vectors Expressing the C1 and C68 E1 Genes

Plasmid vectors were constructed which encode the C1 and C68 E1 region genes, and these plasmids were used to generate stable cell lines expressing viral E1 proteins.

A. pGPGK-C68 E1

pGPGK (gift of Gaung Ping Gao, University of Pennsylvania, Philadelphia, PA) is illustrated in Fig. 5A. pGPGK is a 5.5 kb plasmid containing the known murine PGK promoter (indicated by the arrow on Fig. 5A), followed by a multiple cloning site, a growth hormone polyA sequence, an SV40 ori, a neomycin resistance gene, an SV40 polyA sequence and an ampicillin resistance gene. The remainder of the plasmid is additional plasmid sequence.
As shown in Fig. 5B, the 5' end of the C-68 E1 region was derived from clone 245 which contains a defective version of the C-68 BamHI-E fragment (2042 base pairs) in pNEB-193, i.e., clone 245 was shown to lack approximately the first 30 base pairs of the C-68 genomic sequence, a region not included in the final product of this construction scheme, pPGPK-C68 E1. This plasmid pNEB-C68BamE was digested with BamHI and HindIII and the 2.1kb fragment was ligated with similarly digested pPGPK DNA. The resulting plasmid is designated pPGPK-C68 BamE, illustrated in Fig. 5C.

PCR primers SF-34
(GCAGGTACCAGCTGAGCTACAT) [SEQ ID NO: 4] and SF-35
(CTGTCGAGCTCAGTAC) [SEQ ID NO: 5] were designed to introduce a KpnI restriction site 31 base pairs upstream of the E1A translation initiation site (nucleotide 577 of SEQ ID NO: 2). Using clone 245 as template, a 293bp PCR product was obtained using reagents from Perkin Elmer (Foster City, CA) under the following conditions: 94 = B0C x 5 minutes; 25 cycles of 94 = B0C x 1 minute; 54 = B0C x 1 minute; 72=B0C x 2 minutes; and a final extension cycle of 72= B0C x 7 minutes. The PCR product was purified and is indicated by the hatched bar in Fig. 5D.

The PCR product was digested with KpnI and NheI, yielding a 253bp fragment, which was purified and ligated with similarly digested pPGPK-C68 BamE (Fig. 5C) DNA to yield pPGPK-C68 E1-ATG (Fig. 5E).

The region derived from the PCR step was sequenced for several isolates and the adenovirus insert in pPGPK-C68E1-ATG was shown to match the expected sequence derived from C-68 genomic DNA. pPGPK-C68 E1-ATG (Fig. 5E) was digested with BamHI and the linearized plasmid treated with calf intestinal phosphatase. The purified/phosphatased backbone was ligated with the
1544bp C-68 BamF fragment isolated from PBS-C68 BamF (Fig. 5F) to yield the final plasmid, designated pGPKG-C68 E1 (Fig. 5G).

The C-68 derived sequence in plasmid pGPKG-C68 E1 ends at the BamHI site corresponding to nucleotide 3581 of SEQ ID NO: 2 in the C-68 genomic sequence, which is 80bp downstream of the end of the E1B coding region. This expression plasmid contains from about nucleotide 546 to nucleotide 3581 of SEQ ID NO: 2 which encodes Ela and E1b of chimpanzee Ad C68 under the control of the PGK promoter.

B. pGPKG-C1 E1

The C1 Ad E1 expression plasmid was constructed in a manner similar to that described above for the C68 E1 expression plasmid. Refer to Figs. 6A through 6G.

The 5' end of the C-1 E1 region is isolated as a 1.9kb SnaBI - XbaI fragment (Fig. 6B) and is cloned into pGPKG (Fig. 6A) digested with XbaI and EcoRV. The resulting pGPKG-C1 (map units 1.3-6.6) (Fig. 6D) is used as the template for PCR. Primers are designed to introduce a KpnI site just upstream of the C1 E1 region translation initiation codon (E1-ATG) at nucleotide 578 of the C1 genomic DNA. (See Fig. 6C).

The PCR product is double digested with KpnI and KspI and ligated with similarly digested pGPKG-C1 (m.u. 1.3-6.6) to yield pGPKG-C1 E1-ATG. Partial digestion of pGPKG-C1 E1-ATG (Fig. 6E) with BamHI and isolation of the full length linear DNA, followed by XbaI digestion and isolation of the full length band, followed by ligation with similarly digested PBS-C1 Bam-I (Fig. 6F) yields the final product, pGPKG-C1 E1 (Fig. 6G). The C-1 derived sequence in plasmid pGPKG-C1 E1 ends at the BamHI site corresponding to nucleotide 3599 in the C-1 genomic sequence, which is 90bp downstream of
the end of the E1B coding region. This expression plasmid contains from about nucleotide 548 to about nucleotide 3581 of SEQ ID NO: 1 which encodes E1a and E1b of Ad C1 under the control of the PGK promoter.

Example 4 - Generation of Cell Lines Expressing Chimpanzee Adenovirus E1 Proteins

Cell lines expressing viral E1 proteins were generated by transfecting HeLa (ATCC Acc. No. CCL2) and A549 (ATCC Acc. No. CCL185) cell lines with either pPGK-C1 E1 or pPGK-68 E1 of Example 3. These cell lines are necessary for the production of E1 deleted recombinant chimpanzee adenoviruses by co-transfection of genomic viral DNA and the expression plasmids described above. Transfection of these cell lines, as well as isolation and purification of recombinant chimpanzee adenoviruses therefrom were performed by methods conventional for other adenoviruses, i.e., human adenoviruses [see, e.g., Horwitz, cited above and other standard texts].

A. Cell lines expressing C1 and C68 E1 proteins

HeLa and A549 cells in 10cm dishes were transfected with 10 μg of pPGK-C1-E1 DNA or pPGK-C68-E1 DNA using a Cellphect™ kit (Pharmacia, Uppsala, Sweden) and following the manufacturer's protocol. 22 hours post-transfection, the cells were subjected to a three minute glycerol shock (15% glycerol in Hepes Buffered Saline, pH 7.5) washed once in DMEM (HeLa) or F12K (A549; Life Technologies, Inc., Grand Island, NY) media supplemented with 10% FCS, 1% Pen-Strep, then incubated for six hours at 37°C in the above described media. The transfected cells were then split into duplicate 15cm plates at ratios of 1:20, 1:40, 1:80, 1:160, and 1:320. Following incubation at 37°C overnight, the media was
supplemented with G418 (Life Technologies, Inc.) at a concentration of 1µg/ml. The media was replaced every 5 days and clones were isolated 20 days post-transfection.

Thirty-two A549 and 16 HeLa C1 E1 cell clones and 40 A549 and 37 HeLa C68 E1 cell clones were isolated and assayed for their ability to augment adeno-associated virus (AAV) infection and expression of recombinant LacZ protein as described below.

B. AAV Augmentation Assay for Screening E1 Expressing Cell Lines

AAV requires adenovirus-encoded proteins in order to complete its life cycle. The adenoviral E1 proteins as well as the E4 region encoded ORF-6 protein are necessary for the augmentation of AAV infection.

A novel assay for E1 expression based on AAV augmentation is disclosed herein. Briefly, the method for identifying adenoviral E1-expressing cells comprises the steps of infecting in separate cultures a putative adenovirus E1-expressing cell and a cell containing no adenovirus sequence, with both an adeno-associated virus (AAV) expressing a marker gene and an AAV expressing the ORF6 of the E4 gene of human adenovirus, for a suitable time. The marker gene activity in the resulting cells is measured and those cells with significantly greater measurable marker activity than the control cells are selected as confirmed E1-expressing cells. In the following experiment, the marker gene is a lacZ gene and the marker activity is the appearance of blue stain.

For example, the cell lines described above, as well as untransfected control cells (A549 and HeLa) are infected with 100 genomes per cell of an AAV vector bearing a marker gene, e.g., AV.LacZ [K. Fisher et al., J. Virol., 70:520 (1996)] and an AAV vector expressing the ORF6 region of human Ad5 (AV.orf6) (see SEQ ID NO: 3). The DNA sequence [SEQ ID NO: 3] of the
plasmid pAV.CMVALP.GRE-ORF6, also called AV.orf6, generates a novel recombinant adeno-associated virus (rAAV) containing the LacZ transgene and the Ad E4 ORF 6, which is an open reading frame whose expression product facilitates single-stranded (ss) to double-stranded (ds) conversion of rAAV genomic DNA. In SEQ ID NO: 3, the AAV 5' inverted terminal repeat (ITR) is at nucleotides 53-219; the cytomegalovirus (CMV) enhancer/promoter is at nucleotides 255-848; the human placenta alkaline phosphatase cDNA (ALP) is at nucleotides 914-2892; the SV40 polyadenylation (polyA) signal is at nucleotides 2893-3090; the glucocorticoid dependent (GRE) promoter is at nucleotides 3114-3393; the Ad5 E4-ORF6 cDNA is at nucleotides 3402-4286; the SV40 polyA signal is at nucleotides 4315-4512; and the 3' AAV ITR is at nucleotides 4547 - 4713. All other nucleotides are plasmid-derived. These vectors are incubated in medium containing 2% FCS and 1% Pen-Strep at 37°C for 4 hours, at which point an equal volume of medium containing 10% FCS is added. It should be understood by one of skill in the art that any marker gene (or reporter gene) may be employed in the first AAV vector of this assay, e.g., alkaline phosphatase, luciferase, and others. An antibody-enzyme assay can also be used to quantitate levels of antigen, where the marker expresses an antigen. The assay is not limited by the identity of the marker gene. Twenty to twenty-four hours post-infection, the cells are stained for LacZ activity using standard methods. After 4 hours the cells are observed microscopically and cell lines with significantly more blue cells than the A549 or HeLa cell controls are scored as positive.

Eight A549 (A-2,3,8,13,15,18,23,38) and five HeLa (H-3,4,15,16,20) cell clones are significantly positive in the AAV augmentation assay and the three best
of each cell type (A-18, A-23, A-13 and H-16, H-4, H-20), when tested, support the growth of E1 deleted recombinant C68 viruses.

Four A549 (A-3, 6, 19, 22) and nine HeLa (H-2,5-7, 11-16) cell clones are significantly positive in the AAV augmentation assay and the three best of each cell type (A-3, A-19, A-22 and H-5, H-12, H-14), when tested, support the growth of E1 deleted recombinant C1 viruses.

Example 5 - Generation of Recombinant Chimpanzee Adenoviruses

Recombinant chimpanzee adenovirus vectors are prepared using the C1 and C68 sequences described herein and HEK293 cells. The cell lines described in Example 4 may also be used similarly. Plasmids used to construct C68 and C1 recombinant adenovirus vectors are illustrated in Figs. 7A through 7K, and 8A through 8K, respectively. See also Figs. 11A-11K.

A. pC1-CMV-LacZ

pSP72 (Promega, Madison, WI) is modified by digestion with BglII, followed by filling-in of the ends with Klenow and ligation with a synthetic 12bp PacI linker (New England Biolabs, Beverly, MA) to yield pSP72-Pac (Fig. 7A), which contains a large multiple cloning site with conventional restriction enzyme cleavage sites.

pSP72-Pac is digested with PacI and EcoRV and ligated with the 465bp PacI-SnaBI fragment isolated from pBSC1-BamG (Fig. 7B) to yield pSP-C1-MU 0-1.3 (Fig. 7C). The CMV promoter-driven LacZ gene is isolated from pCMV-β (Clontech, Palo Alto, CA; Fig. 7D) as a 4.5kb EcoRI/SalI fragment and ligated with similarly digested pSP-C1-MU 0-1.3 DNA to yield pSP-C1-MU 0-1.3-CMV-β.
For the initial step in the isolation of the C1 Ad map units 9-16 region, pGEM-3Z (Promega, Madison, WI; Fig. 7F) and pBS-C1-BamI (Fig. 7G) are digested with BamHI and SphI and the 310bp fragment from pBS-C1-BamI is ligated with the pGEM-3Z backbone to form pGEM-C1-MU9-10 (Fig. 7H). C1 map units 10-17 are isolated from pBS-C1 BamE (Fig. 7I) by digestion with BamHI. The 2.5 kb fragment is ligated with BamHI-digested pGEM-C1-MU9-10 to form pGEM-C1-MU9-17 (Fig. 7J). The 2.9 kb fragment containing C1 map unit 9-17 region is isolated from pGEM-C1-MU9-17 by digestion with HindIII and ligated with pSP-C1-MU 0-1.3-β (Fig. 7E) digested with HindIII to form the final plasmid, pC1-CMV-LacZ (Fig. 7K).

pC1-CMV-LacZ (Fig. 7K) thus contains C1 Ad mu 0 to 1.3, followed by the CMV promoter, an SD/SA, the LacZ gene, a SV40 poly A sequence and C1 Ad mu. 9-17, as well as additional plasmid sequence. This plasmid is cotransfected into the E1-expressing cell line with a left terminal clipped C1 Ad fragment (or a replication-defective C1 Ad helper virus) to produce by homologous recombination a recombinant chimpanzee adenovirus carrying the LacZ gene.

C. pC68-CMV-LacZ

pSP72-Pac (Fig. 8A; also Fig. 7A) is digested with PacI and EcoRV and ligated with the 465bp PacI-SnaBI fragment isolated from pBS-C68-BamE (Fig. 8B) to yield pSP-C68-MU 0-1.3 (Fig. 8C). As above, the CMV promoter-driven LacZ gene is isolated from pCMVB (Clontech; Fig. 8D; also Fig. 7D) as a 4.5kb EcoRI-SalI fragment and ligated with similarly digested pSP-C68-MU 0-1.3 DNA to yield pSP-C68-MU 0-1.3-CMVB (Fig. 8E).

For the initial step in the isolation of the map unit 9-16 region of C68, pGEM-3Z (Fig. 8F; also Fig. 7F) and pBS-C68-BamF (Fig. 8G) are double digested
with BamHI and SphI and the 293bp fragment from pBS-C68-BamF is ligated with the pGEM-3Z backbone to form pGEM-C68-MU9-10 (Fig. 8H). C68 map units 10-16.7 are isolated from pBS-C68 BamB (Fig. 8I) by digestion with XbaI, followed by filling in of the ends and digestion with BamHI. The 2.4 kb fragment is ligated with BamHI/EcoRV-digested pGEM-C68-MU9-10 to form pGEM-C68-MU9-16.7 (Fig. 8J). The C68 map unit 9-16.7 region is isolated from pGEM-C68-MU9-16 by digestion with EcoRI, filling in of the ends with Klenow and then digestion with HindIII. The 2.7 kb fragment is ligated with pSP-C68-MU 0-1.3-CMVβ (Fig. 8E), digested with HindIII and PvuII to form the final plasmid, pC68-CMV-LacZ (Fig. 8K). pC68-CMV-LacZ (Fig. 8K) thus contains C68 Ad mu 0 to 1.3, followed by the CMV promoter, an SD/SA, the LacZ gene, a SV40 poly A sequence and C68 Ad mu 9-16.7, as well as additional plasmid sequence. This plasmid is co-transfected into the E1-expressing cell line with another C68 Ad to produce by homologous recombination a recombinant chimpanzee adenovirus carrying the LacZ gene.

D. pBS-Notx2

The LacZ gene is removed from either pC1-CMV-LacZ (Fig. 7K) or pC68-CMV-LacZ (Fig. 8K) by digestion with NotI, and replaced by the coding sequence of any desired gene. This cloning step is facilitated by having the gene of interest flanked by NotI restriction sites, preferably with the upstream site in the 5' untranslated region of the gene.

Such a cloning vector is derived from pBluescript SK+ (Stratagene, La Jolla, CA) by digestion of SK+ with SalI, followed by filling in of the ends and
ligation with a synthetic 8bp NotI linker (New England Biolabs, Beverly, MA): GCGGCGCG.

CGCCGCGG

The resulting PBS-Notx2 shuttle vector (Fig. 4B) is thus designed to facilitate cloning of cDNAs into pCI-CMV-LacZ (Fig. 7K) and pC68-CMV-LacZ (Fig. 8K; see also Fig. 4A) as a NotI fragment. pBS-Notx2 has two NotI sites flanking a number of restriction sites suitable for cloning the cDNA to be expressed in the recombinant adenoviruses and the LacZ ORF from pBluescript is maintained, allowing blue/white screening of clones in pBS-Notx2.

E. Homologous Recombination with Helper Virus

To generate the recombinant adenoviruses from the plasmids described above, the appropriate E1-expressing packaging cell line, such as 293 cell line or a cell line of Example 4, is co-transfected with a replication defective C1 or C68 helper virus, or a left-end clipped C1 or C68 fragment, as appropriate. These helper viruses may be deleted of other non-essential genes. The infected cell line is subsequently transfected with an adenovirus vector as described above bearing the transgene of interest. Homologous recombination occurs between the helper and the plasmid, which permits the adenovirus-transgene sequences in the vector to be replicated and packaged into virion capsids, resulting in the recombinant adenovirus.

Transfection is followed by an agar overlay for 2 weeks, after which the viruses are plaqued, expanded and screened for expression of the transgene. See, for example, Figs. 10A-10D. Several additional rounds of plaque purification are followed by another expansion of the cultures. Finally the cells are harvested, a virus extract prepared and the recombinant chimpanzee adenovirus containing the desired transgene is
purified by buoyant density ultracentrifugation in a CsCl gradient. All of the above procedures are known to those of skill in the art.

F. Another Cl Recombinant Adenovirus

Another set of plasmids used to construct a Cl recombinant adenovirus is described as follows. Figs. 11A-11H illustrate the scheme employed to generate a unique restriction site in the left end of the Cl genome. A unique site is necessary in the procedure employed in generating a recombinant adenovirus, but Cl has no such site. There are two Spe-I restriction sites, including one at position 1733, within the E1B 21K coding region. To replace this Spe-I site with a unique Not-I site, plasmid pNEB-Cl-BamG (Fig. 11A), containing the left end of the Cl genome, was digested with Spe-I and Asc-I, and ligated to the 6204 bp Spe-I/Asc-I fragment from the Cl genome (Fig. 11B). The resulting plasmid, pNEB-Cl-AscI-B (Fig. 11C) is then digested with Spe-I, filled in with Klenow enzyme and ligated to the synthetic 8bp Not-I linker (Fig. 11D) described above, to yield pNEB-Cl-AscI-B-NotI (Fig. 11E).

This plasmid is digested with Pac-I and Asc-I and the purified fragment is ligated overnight with the Cl-Asc-I-A fragment (Fig. 11G). The ligation reaction is extracted with phenol:chloroform:iso-amyl alcohol, then chloroform, and then 3 μg of sheared salmon sperm DNA is added and the DNA is ethanol precipitated. The resuspended DNA is used to transfect 293 cells and DNA from viral plaques is tested for a Not-I site (11H).

G. GFP as a Transgene

Plasmids used to construct exemplary C68 expression plasmids containing the bacterial green fluorescent protein (GFP) gene are illustrated in Figs. 9A through 9G, respectively. To facilitate the cloning of the GFP gene into the chimp Adeno expression vectors,
pEGFP-1 (Fig. 9A, Clonitech, Palo Alto, CA) was digested with Sma-I and ligated to the previously described 8bp Not-I linker (Fig. 9B). The resulting plasmid, pEGFP-Notx2 (Fig. 9C) has the GFP gene flanked by Not-I sites.

The purified pEGFP-Notx2 Not-I fragment is ligated to Not-I digested pC1-CMV-LacZ (Figs. 7K and 9D) or pC68-CMV-LacZ (Figs. 8K and 9E) to yield the GFP expression vectors pC1-CMV-GFP (Fig. 9F) and pC68-CMV-GFP (Fig. 9G and Fig. 10A), respectively.

Example 6: Delivery of Transgene to Host Cell

The resulting recombinant chimpanzee adenovirus described in Example 5 above is then employed to deliver the transgene to a mammalian, preferably human, cell. For example, following purification of the recombinant C68-CMV-GFP virus of Example 5G, human embryonic kidney 293 cells and A549 cells were infected at an MOI of 50 particles per cell. GFP expression was documented 24 hours post-infection.

In vivo studies have tested the infectivity of the virus in murine liver (tail vein injection), lung (intratracheal injection) and muscle (intramuscular injection). Preliminary data indicate that the C68-CMV-GFP recombinant virus transduces all three tissues, and GFP expression can be detected.

When administered in vivo, a less severe immune response is produced by the human immune system (which is naive to the chimpanzee adenovirus sequences) than to a human adenovirus construct, thereby permitting subsequent administration of the same or another vector.

All references recited above are incorporated herein by reference. Numerous modifications and variations of the present invention are included in the scope of the above-identified specification and are expected to be obvious to one of skill in the art. Such
modifications and alterations to the compositions and processes of the present invention, such as selections of different minigenes or selection or dosage of the vectors or immune modulators are believed to be within the scope of the claims appended hereto.
51

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Trustees of the University of Pennsylvania
   Wilson, James M.
   Farina, Steven F.
   Fisher, Krishna J.

(ii) TITLE OF INVENTION: Chimpanzee Adenovirus Vectors

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:
   (A) ADDRESSEE: Howson and Howson
   (B) STREET: Spring House Corporate Cntr., P.O. Box 457
   (C) CITY: Spring House
   (D) STATE: Pennsylvania
   (E) COUNTRY: United States of America
   (F) ZIP: 19477

(v) COMPUTER READABLE FORM:
   (A) MEDIUM TYPE: Floppy disk
   (B) COMPUTER: IBM PC compatible
   (C) OPERATING SYSTEM: PC-DOS/MS-DOS
   (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
   (A) APPLICATION NUMBER: WO
   (B) FILING DATE:
   (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
   (A) APPLICATION NUMBER: US 60/024,700
   (B) FILING DATE: 06-SEP-1996

(viii) ATTORNEY/AGENT INFORMATION:
   (A) NAME: Bak, Mary E.
   (B) REGISTRATION NUMBER: 31,215
   (C) REFERENCE/DOCKET NUMBER: GNPKN.021CIP1PCT

(ix) TELECOMMUNICATION INFORMATION:
   (A) TELEPHONE: 215-540-9200
   (B) TELEFAX: 215-540-5818

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 35524 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: double
   (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCATCATCAA TAAATACCT TAAACTTTTG GTGGGTGTGA ATATGCAAAT GAGGGCTTTG  
AATTTGGGGA GGTTGAAAAGG TGGATTGGCTG TGGGGACGGCG GACGGTTAGG GCGGGGGCGG  120
CTGACGTTTTT GATGACGCTGG TGCTTGGCGG GAGTTGCTTT GCACGTTCTC GTGGGAAAG  180
TGACGTCAAA CGAGGCTGGTG TTTTGAACACG GAAATACCTGA ATTTTCGCCG GCTCTCTGAC  240
AGGAAGATGAT GCTGTTTTTTGG GCGGATGCAA GTGAAAAATTC CTCATTTTGG CGCGAAAACT  300
AAAAGGAGAA GTGAATTTCTT GAGAATATTTC GTGTTTATGA CAGGGTAGG TATTTACCGA  360
GCGCCGAGTA GACTTTGACCC GATTACCTGG AGGTTCGATT TACGCTGTGT TTCACCTAAAA  420
TTTCCGCGTAA CGGCTGTCAA GTCTCTGGTT TTTACGTAGG TGTCAGCTGA TGCTTAGAGT  480
ATTAAAACCT GAGGAGTTCC GCTAAGAGGC CACTCTTGGAG TGCCAGGGAGG AAGATTTGCC  540
TCCTCCGCGGT TGCGAGTCAG ATCTCCACTT TTGAAATGAG ACACCTGCGG TCTCTGTCC  600
AGGAGATAGT CCTCAGTTTG ACGGGAAGAA AAAATCTGCA ATTTCTGTGA AATACAGCTGA  660
TGCGAGACCA TCCAGAGGCG CCTGAGCCAC CTCTGTATA TCCCAGGCTT CATGAATTAT  720
ATGAGATTTAG GCTAGAGCGA CGCAGGAGCC AGCTATGAAA CGACGTGAAAT GGGTTTTTT  780
CTGATTCTAT GTTATTAGCT GCATATAGGG GAGTGGATTT AGACCCACCT TCTGAACTTT  840
TTGATACTCC AGGGGTGATT TGGGAAGCGT ACATAGATGG GAAAAATTTA CTCGATTTGG  900
TGCTGCTGTA ATPTGACTTA TACTGCTATT AGAAGGGTTT TCCTCAGAGT GATGATGAAG  960
ATGCTGGAGA TGCGACTGCA ATTCGAGCCT CGCGGGCTGG CGGAGTGAAA GCCTCGGATG 1020
ATGCGTTTAA GCGGCGCTTC TGAGGACATG CGTGAAGTCG TTGTAATATC  1080
ACAGGAAAAA TACTGGAGTA AAAGAAATAT TAGCTTCGCT TTGTATATAG AGAGCGCATT  1140
GCCACCTTTAT TCAACTGAGA TGTTTTAAA GTTAAATTTA AAGGAACAGT ACGCTTTTTA  1200
ATAACTTTTG ATGGGATGTG TTTAGTTTTG CTGATTTTTA TAGGTCTCTG GTCTGATGCT  1260
GATGAACTCC CTCTCCTCGA TTCAACTACC TCAGCTCCTG AAAATCCGGG ACCCGTTTCT  1320
GCACACGATG CAGGCCAATT TTCTGTGAAG CTTAAGCCGT GGAAAACGCC TGCTGTGAAT  1380
AAACTTGAGG ATTTGCTGGA GGGTGGGAT GACCCCTTGG ACTTTGTGAC CGCGAAATAA  1440
CAAGGCCAAT GAGTGCTCCG CACCTGTTGT TATCTAATGT GACGTCTCGT TTGGTTGTAG  1500
AGTGCTCATG ATAAAAATTA TGTCAGCACG TGAGTGTTTT ATGTTTTATT GGCTGGGACT  1560
TGCGGTAATAT AAGTAGAGAC AGACCTGTTG GTTAGCTCA CAGCAGCTTG CTGCCATCCA  1620
TGGAGGTTGG GCCCATATGA AGAAGACCTA GCCAGCTTAG GCAACTGCTA GAARAACGCT  1680
CGGCAAGAGT CTCGCGCTCT GGAGAGATTG GTGTTCTGTTG TATCTGACTG AGACAGTACT  1740
TTGGATAAA CGAGGATTAC AGCCAGAAAT TTGAAAAAGT ATTTGACACC TGTCAGGACG  1800
TTTTTGAGGC TCTAATCTTG GGCACCAGAG CTCTATTTAA GGAAGAGGTT TTATCGTTTT  1860
TGATTTTTTC TACCCCTGTT AGAACGTGCTG CTGCTGGAGC TTCCCCTACA TTATATATTTG  1920
ATAAAAGGAT CCCACAGAAC CACTCCAGCA AGGAGATGCT TTTGGATTCC ATAGGACGAG  1980
CTTGTTGAG AACATGGGAAG GCTCGCAGGA TGAGGACAAT CTTTAGATTAC TGCCAGTAC  2040
ACTGTAATGA ATCTTTATTT GATTTTTCGC GCACGGGTATG CCCTGGACCA CGGCTCTCAG
TCATTGAGAA CTCCGGTGAG TTTTTCACGG ACCCTGGTAGA GTGGGTTATT AATGTTTAGA
TACATGGGCA TTAGGCCGTC TCAGGGGGTG AGATAGCCAC ATAGGAAGAC CTCAATGTCCT
GGGGTGATAT TATAAACATC CCAACTTACA CAAGTGTTGG GTGGCATGTG TGGCAACATA
TCTTTAAAGA GCAGGCTGTAT TGCAACTGAG AGCCCCTGGG TGATATGTGT TACAATAACTG
TTAAGCTGAG ATGGATGCACT TCTGGGTAAG ATATTATGCA TTTTTTACTG TAATCTTGAGG
TTGGCAATGT TGCGCCGCCAG ATCCGGCTCTC GGTTTCAAGT TATGCAGGAC CACCAAGAGG
GTGTATGGCC TGCACCTTGG AAATTTATCA TGCAAGTTAG ATGGAAAAGC ATGAAAAAT
TTGGAGAGCC CTGTGTGCTC GGCAGATTTG CTTATGACTG CATCATGATG GATAGCCAGT
GGGCCTGGG CCGCGGCACGG GCACCATACCA TTGCCGCTTG CTGACACATC ATAGTTATGC
TCTGAGACCA GCTCATCATT GAACACCTTG GGCGAGGCTGG GCACATTGGC
GGTATATAAG TACCCCTGGG CCCCGGACCA TAGTTTCCCC CACAGATTGG CATTCCCAA
GCTTTCAATT CAGAGGGGGG GATCATGTTCC ACCGGAAGGG CTAATAAAAA TACCGTTTCT
GGGCTGGGG TGGATAACTG TGATGATAGG AAATTTCTGA GCAGCTGGTA CTTGGCCACAC
CCAGTGGGGC GGTAAGTACG CCGGATCAGG GTGGTCAGATT GGGAGGGGAC
CTCGGGCTCT TCGGGAGGAG GGGGGGCAGG TCGTTATCTC TTTTCTTCA ATGGATATTT
TCCCGACACA AGTGCGTGTAG GAGGGCTCTC CCACCTAGGCG ATAAAAATTGG CTGAGGGGAG
GGAAAGTTTT TGAGCCGCTT TAGCCGCTCA GACATTGGCA TTTGGAAGAG ATGTGTGTTG
AAGAGCTCAA GCGGGTGTAG GAGCTGTATTCA ATGTGTCTCTA TGGCATCTCG ATCCAGCAGA
CCCTCCTGTT TCAGGGGTGG GACAGGCCTC TCGAGATTAG TATCAGACCA TGGGGCGCCA
GGCTGGCAGG GTCTGGCTCT TCTACGGGTC GCCAGCGTCC AGTCAGGTGT GCTTTGGCTCA
CAGTGAAGGG GTGGGCGGCTC GTTGCGGCGC TGGCGAGGCTG CCGCTTCAAG CTCATCCTCG
TGGTGGAGAA CCAGCTGGGCA TCGGCGGCTC CCATGCTCAGC CATGAGTACG TTACACTGA
GTTGCTAGTT GAGTGCTCCG GTGCCTGTGAC TTTTGGGCGG GAGCTTACCT TGGAGAGTTT
TCCTGGAGGC AGGGACTTAG AGACACCTGA GGGCAATTAG TCTGCGGCGG AGGAGGATGG
ATTCGGGGGA GTATCGATCC GCACGGCAGG AGGGCGGATG GCTTGGCATC TCCACGAGGC
AGGTGACATC CGGCCATCGG GGCTGAAAAA CAAGTTTACC GCGATGTTTT TGAGTCCGCT
TCTTACTTTT GCTCTACATG AGTGGCTCTC CCGCTGGGGT GACAAAGAGG CTGTCGCTGTT
CCCGTGAGGC AGATTATGGGG GGTCTCTCC TGGCGGGAGT GCCTGGCTTC TCTCGCTAGA
GGCAACCGGA CCACCTCTGAT ACAAGGGCGG GGCCTGGGCG CAGTACAAAA GAGGGCCAGT
GGGAGGGGTA GCGTGCGTTA TCAACCAAGG GGCTCAGGCT CTCACAGTAA TGGAAACACA
TGGCCCCCTC CTCACCATCC AAGAAGGTGA TTGGCTTGTGA AGTGTAAGCC ACGTGACCAG
TTCCAGCCGG TGGGTGATAAA AGGGGGGGG GTCCTCTGCTC GTCTCCTAGT TCTTCCGGAT 5940
CGCTGTCCAG GAGGGCCAGC TGTTGGGTAA GGTATTTCCCT TCTGAAAGGG GGGATAACCT 6000
CTGCACTACG GTTGTCAGTG TCTAAGAAGC AGAGAGAATT GATAATTGAC GTGCGAAGTTG 6060
AGATGCCCTTT CATGAGACTC TGCTCCTATT GGTGCGAAAGA GACAATTTTC TTGGTGTCAA 6120
GCTTGTTGCG AAAGAGTCCG TATAGGGAAT TGGATAAAAG CTGGCCGATG GAGGCCATGG 6180
TTTTGATTCTT ATCTTCTGTCC GCACGCTCTT TGCCAGCAAT GTTGAGTTAG AGCTACTGGC 6240
GCCCCCGCCA CTCCTTCTAC GGAAGAAGTG TGCTGACATT ATCTCGCACG ATCTGACCTC 6300
GCCGACCGCA ATATGACAGG GTGATTAGAG CCACACTGGT GCCCACCTCG CCTGCGAGGG 6360
GCTGTTGTTT CCAGCAGAGT CGACCCCTTT TCTCTGACAA GAAAGGGGGG AGGGGTCTTA 6420
GCATGAGTTTC ATCGAGCGGG TCTGCTACAT TGTTGAATAT TCTGGGAGGC AGATCTTTGT 6480
CAAAATAGCT AAATGGAGCG GTGTCATCCA AAGCCCATCG CCATTTCTGA GCTGCCAGCG 6540
CCGGTTCTAA GGGATTGGAT GGGGTTCCCC ATGGCAGATGG GTGGGTGAGT GCAAGAGCAT 6600
ACATGCCACA AGATGTCAATG ACATACAGTG TTTCTCCAGG GATGCCGATG TAGGGTGAGT 6660
AACAGCCGCC CCTCTCAGTT CTGGCTCCAG CATAGCTACA GATTTCAATG GAGGGGCGCA 6720
GAAGACCGCG GCGCCAGATTG GTACGGTGGG GTTTTTACAG TCTGTAACCG ATCTGGGCAA 6780
AGATGGGCAAG GGAATGGGAA GAGATGGTAG TCTCTGAAAA GATGTTAAAA TGGGCATGAG 6840
GCAGGCACCAC AGACTGTTTG ACGAAGTGCC CATAGACCTC TTGCAAGCTG GCCACACAGT 6900
CGCGGGTGA GAGCACATCC AGGGGCAGTT AGCTCAAGGGT CTCTGGAAAT ATGTCAAAAC 6960
CTGGTTGGTT TTTCTTCTCC CAGACGCTCG GGTGGAAGAG GTATTCTCCG CGATCTTTCC 7020
AGTACCTTAC GAGGGGAACC CGCTTTTGTG CTGCAAGGTA AGAGCACCAG ATGTGACACT 7080
AGTTAGCTGC CTTTGAGGGG CAGCATCCCT TCTC ACCGGG GAGAGGATG CTTGGGGGCG 7140
CCTGCCGAG AGAGGTATAG GTGAGGCGCA AGGTGTCCCT GACCATGACT TTAAGGAACT 7200
GATACCTTGAA GTGAGATGCA TTCAGGCCCC CCTGTTCCTCA GAGTTGGGAG TCTACGCGCT 7260
TCTTGTAGGC GGGATTGGGC AAACCGGAAG TAAACCTCGT GAAGAGTATC TTGCTGCCCC 7320
TGGGCTAGGA ATCGGCGGTG ATGGGGAAAG CCTGGGGCAG TTTTGCTCTG TTATTTGATC 7380
CCTGAGGGCC TAGAGCTGTC TCAAACAGCC TCTGATGATT GTGCCCCACT ATGATCAAGT 7440
CTATGAAATCG AGGGGTGCCG TTGACATGAG GCAGCTTTCT CAGTTCTCCG AAAATTTAGG 7500
CTGTTGGGCTC AGAGAGGACA TAGTGTGCGA GGGCCATTGC GTGACGGTTA GGGTGCGCAT 7560
TGAGGAAGGA GGGCAAAGAG TCCACTGCCA GTGCTGGTTTG TAACTGGTCC GGGTACTGGC 7620
GAAAATGCTG GCCGACTGCC ATCTTTCTGT GGGTGACACA TAGAAGGTTG TTGGGGTCTGT 7680
GCTGCCAGGC ATCCCCATTT AGTTTTCAAG CAGGCTCATA GGGCAGTGTTG AGGACCGCGT 7740
CGTCCCCAGA GAATTTGTAG ACCAGCATGA AGGTTATAGG TTGGTCGAGA AAGGACCCCA 7800
TCCAGGTTGTA GGTTCACACA TGAGATGATGAG GGAAGACCGCT TCCCGTGGAG GGAACGACGC 7860
CGATCGGGGAA GAACTGGGCTG CCTGCCACCC ATGGGAGAGA ATGAGCTTTG ATGATGGGA 7920
AGTAGAAAAATC TCTGGCGGCG GCGAGGACTT CATGGCTTGT CTTGTACAGA CGCGGCGAGT 7980
ACGCGGCGCC TGCGACGGGA TGCACTCCTG GAAATGAGTG TACCTGGCTT CTTTGGACGA 8040
GAAATTTCAG TGGGAAGTTG AAGCGCTTGGC CTTAGAATCT CTCCTCTACT ATGTTTATCTG 8100
CATCCGGCTT GGCATCTCTT GTCTGAGATT GCTGATGCTG ACAAAGCCCG CGCGGGAGGC 8160
AAAGTCCGAC CGCGGCTAAG AAGCGGACGC GTCGAGGGCA GAGAGCGGC AGGCGGACGC 8220
TGCGGAGGTT CCTGAGACGC TGCAGGTCAAG GATTTGAGAG TACGATGGGG ATGATTGACTT 8280
GCATGATCCT TTGAGGCGCA CGCCGAGGTA TGCTAGTGGAA TTTGATCTCC ACGGTCGGGT 8340
TGTTGGAAGT GCAGATGGGCT TCCAGGGCTT GCGGCCCTCT GGGCGCACC ACCGTGGCCC 8400
TGTGGTTCTC TTGCGGGCGC GCGCGTGTTG TAGCTTTGCTG CAGGGTACGA AGCGTGGCCG 8460
AGCGGGCGGC CTTGGATGGA AGGCGGCTCT CGCGGCGCGC GCATGATGCG AGATGGCCAC 8520
GTCCGGGGCG CGCGGCGGGA GGTTCCTGGTA CTGGCGCCCTG AGAGAGCCGCC GTGCGGCAC 8580
AAGCCCGGCG TGAGCTTCTT GCAGCTCTCG GCTGGGTATT AAAGCTAGCG GCCCGGTTAG 8640
CTTGAACCTCT AAAGAGATTT CAACAGAAATC AAAATCGCTA TGTTGAGGGG CGGACGGCT 8700
TAGATGCTCT GTAGAAGTGC CTGAGTGGTCG TGTGGGGATC ATCTCCGGCA TGAACGTCTC 8760
GATTTCTTCC TCGTGAAGAT CTCCGGCGCC TGGTCTCTCG ACGGGTCGCG CGAGGTCGTT 8820
GGAGATGCGA CCCATGAGTT GAGAAGATGC ATTCATGCTC GCCTGCTTCC AGACCGCGCT 8880
GTGATCCAGC GGGCGTTCGG GATCTCTCAG GCGGAATGACC ACCGAGGCGG GTGAGAGCTC 8940
CGCTTGGGGG GTGAAGACCG CATAGTTGGCA TAGCGGTCGG AAGAGGAGCT TGAGTGTGTT 9000
GGCGATGTGC TGGGTACCCA AGAATACAT GATCAGATCT CTCAGGGCGA TCTCCGCTAC 9060
ATGCGCGAGG GCTCTCAGAG GTCGAGATTG ACGGAGGCA TAGCGGCAMAG AGTTGAAAAA 9120
CTGGGAGATCG GGGCGAGGCA CGCGCAACTC CTCTCCGACG AGAGGATGAG GTGGCGCGATG 9180
GGTGCGCGGC ACCTGGCGCT CGAAATCTCCT CGAGATTTCT TCCTTCTCTT TCCTACTCTC 9240
CTCTTTCCAT AACAATCTTT CCTTCCTCTTC AGCCGCGGAG GGAGAGGAGA GGGCGACCGG 9300
GGAGACGCGG CGCGCCACCG CAAGAAGCCT GATGAAATCT TCAATGACCT CTCCGCGCGG 9360
GGCGCGCATG GTCTGGGGTA GCGCAGCGGCT GTGTCTCTCTG GCCTAGCGAG TGAAGAAGGG 9420
TCCGGCGATC TCCCTAAAGT GTGGAGAGGG GGGCTCCTTG TTGGGCGAGG ACAAGGCGCT 9480
GATTATGCAT TTTATCAATT CGCGGCGAGG CTTCAAGGCG AAGGCCGCTG TGCTCTCAAG 9540
ATCCAGGCGA TGGGAAGGAA TTTGAGGCGA AGCGTCAACT CAGTCCGAGG CCGAAGGCTG 9600
GCTAGCGACT GATTCTGTGA CGCGGCGGTG GCTACAGGCT CGTGCGGGGG TCTCTATTTT 9660
TTCTCTCTCC TCCCTCCTGG AGGGTGAGAC GATGCTGCTG GTGATGAAT TAAAATAGGC 9720
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGTTCTGAGA CGCCGAGATGG TTGGCAGGG CACCCAGTCT TTGGGACCGG CTTGCTGGAT</td>
<td>9780</td>
</tr>
<tr>
<td>GGGCAGCCGA TTGGCCATTC CCCAAGCATT ATCTCTGGAC CTTGCCAGAT CTTTGAGTA</td>
<td>9840</td>
</tr>
<tr>
<td>GTCTTGCAATA AGTGGCTGCC CGGCCACTTC TTCTTCGGCC GCTCTGGCAT GACATGGCGT</td>
<td>9900</td>
</tr>
<tr>
<td>GACCCCCAAC CGGCCGATAGG GCTGGAACAG TGGCCAGGTCC GCTACGACCC TTTCTGGGAG</td>
<td>9960</td>
</tr>
<tr>
<td>GATGGGCTTGTC TGCACTCTGG CGAGGTGGCC TTGGAGTGCC TCAAACTCCA CAAAGCAGTG</td>
<td>10020</td>
</tr>
<tr>
<td>GTAGGCCCCCG GTTTGATGAG TGTAACGAGCA GTTGGCCATAT ACTGACCAT TGACGTGTCTG</td>
<td>10080</td>
</tr>
<tr>
<td>GTCCCCGGGG CGCCAACTCT CGGTTGACCTA AGGGGCAGGT TAGGGCGAGG TTCTAAAGAT</td>
<td>10140</td>
</tr>
<tr>
<td>GTAATCGCTTG CAGGGCCGCA CCAGGTACGT GTAGGGCGATT AGAAAATGTG GTGGCGGTCTG</td>
<td>10200</td>
</tr>
<tr>
<td>GCGGTATAGGG GGCCTATCGCT CTGTACCAGG CGCCGCCAAG GAGCGAGTCTT CCAGCATGAG</td>
<td>10260</td>
</tr>
<tr>
<td>GGGTGTATCA ACGTATGACG ACTCTGGACAT CCAAGGGCATC CAGGAGGGCC TGGTTAGGCT</td>
<td>10320</td>
</tr>
<tr>
<td>CCGGGAGAAC TGCGTATCCGG GTTTCCAGAT GTTGGCCAGC GCACATGAGT ATGTCATGAT</td>
<td>10380</td>
</tr>
<tr>
<td>AGGCACCGTT TGCCCCGTGA GGGGCGGACA GTCTGGTATG CTCTAGACAT CAGGGCAAAA</td>
<td>10440</td>
</tr>
<tr>
<td>AGCAAGACGG GCACCGCTGTG GACCTGCGTG CCTGAGGACT AAGCAGAAGT GTCGGGGTCTG</td>
<td>10500</td>
</tr>
<tr>
<td>GCCTGTACCC CGGTTGGAAT CTGGAGTATT GCTGGAGCCG CAGCTAAAGT GTGACTGGCA</td>
<td>10560</td>
</tr>
<tr>
<td>CTCCCCGCTC AAGGCCAAGG TGCCAAAAC CGCCAGGAGTC CGCTTTTCTTG</td>
<td>10620</td>
</tr>
<tr>
<td>TTTTTTGTCT TTTCTCGGAT GGAGGCCGATG GCTCAGTCAAC GTCTTAGAAG CTCAGTTCTG</td>
<td>10680</td>
</tr>
<tr>
<td>CCGGGCTGGG AGTGGCTGCC GCCGCTAGTC TGGAGAAATCA ATGCCAACAGG TTGGCGTGGG</td>
<td>10740</td>
</tr>
<tr>
<td>GCATTCCCCG GTTGGACGTG TAGGGCGGGG GATCGGCAGCG TTTGCCGCGG AAACGAGGGT</td>
<td>10800</td>
</tr>
<tr>
<td>TTGGGAGCCC GCTCATTTCT AAGACGCGCC TAGCCGACTT CTCCAGTTTA CGGAGGGGAG</td>
<td>10860</td>
</tr>
<tr>
<td>CCCCCCCATTT TTTTTTGTGT TTTGGTGGCC AGATGCATCC GCTGGCTGGA CAGATGGGCC</td>
<td>10920</td>
</tr>
<tr>
<td>CCAAGCACA CGCCCTTCTT CAGCCAGCAG CACAGCAACA GCCCAAAAG GCTCTTTCTG</td>
<td>10980</td>
</tr>
<tr>
<td>CTCTGTACAC TACTGGCAGC CCACGCGTCA CGCGCGCCGG AGACCCCCGG TATGACTGGG</td>
<td>11040</td>
</tr>
<tr>
<td>ACCTGGAGAG GGGCGAGAGG CTTGGCCGCTC TTGGTGCCAC ATGCGCGGAG CGCCACCCGG</td>
<td>11100</td>
</tr>
<tr>
<td>GGGTGCAACT GAAAGAGGAT TCTGGCGAGG CGTACGTCGC GCACAGAAGC CGTTGCAAGG</td>
<td>11160</td>
</tr>
<tr>
<td>AGGAGGCCGG TGAGAGCGCC GAGGAATATC GACCTGCCGC TTTTACCGGC GGTGCGGCAG</td>
<td>11220</td>
</tr>
<tr>
<td>TGCGTCTAGG CTGGGCACAT TGGTGCCTAGA TGATGTCAGA GTCGATGAAG</td>
<td>11280</td>
</tr>
<tr>
<td>TGACAGGGAT AAGCTCTGCT AGGCCAATAG TGGTCTGGGC CAACCTAGTA TCAGCGCTACG</td>
<td>11340</td>
</tr>
<tr>
<td>AGCAGAGCGT GAAAGAGGAG CGCAACTTCT AAAATCTTTT CAAAACCATG GTGGCGACCC</td>
<td>11400</td>
</tr>
<tr>
<td>TGATTGCCCG CGAGAGGGTG ACATGGGTGT TAATGCACTT GGGAGACTG ATGGAAAGCTA</td>
<td>11460</td>
</tr>
<tr>
<td>TTACCCAGAA CCCACCAGGC AAACCTCTGA CGCGTCAGCT GTTTCAGTAG GTGCGAGACA</td>
<td>11520</td>
</tr>
<tr>
<td>GCAGAGCACA TGAGGCAATTT AGGGGAGGGC TGTTGACAC TACGAGCCCC GAGGCGGAGT</td>
<td>11580</td>
</tr>
<tr>
<td>GGTTGTATGA TCTTATCTAT ATCTCAGAAA GATCTCAGAT GCAAGAACGT AGCCTGGGTC</td>
<td>11640</td>
</tr>
</tbody>
</table>
TGGCTGAGAA GGTTGCTGCT ATTTAACAAGT CCTGCTTAAG CCGTGGCAAG CACTACGGTC 11700
GCAAGATCTA TAAACCCCA TACCTACCTA TAGACAGATAA GTGTAGGTAG TCTTTTCTCT 11760
ATATGCTCCAT GACTTCAGCT GTGCTGACTCT TACAGACGAC TCTGAGATGT TACGGCAAGG 11820
ACAGGAGGCA CCGGTGCATTG AGGCGGCCGA CAAGGCTGCA GCTGAGCGAC AGAGAACTTA 11880
TGCCACAGCTT GCAAAGAGCT CTGAGCAGGG CTTGAAACGGA GGGGAGAAAC TACTTTGACA 11940
TGGGACGGGA TTTGCAATGG CAGCCGCTTC GAGGCGGCGG CAGCCGACGA GGGTAGGAC 12000
TTCTTCTACT CTAGAGGGGGA TATGACGAGG GCAGCTGTAAC CTGGAGAAGC 12060
GATGGCGCAGA CATCGGTTAT TTTTTTGGTA CGCAGCAATAC GCGACCTCCT CAGCGCCAAA 12120
TGGGGCCGCC GTGCGAGCGC CAGCGCTCAG GCAGATTACCT CTGAGACGAT TGGACACGGG 12180
CCATGCAAAC CATCATGGCG CTGAGCGGCC GCAACCGGCA AGACCTTTAGA CAGCGACCCC 12240
AGGCCAGCAG CTTCTTGGCG ATCTTGTGGC GTGGTGGCGTT GAAACCTTTA CTGACGACC 12300
AGGAAGAGCT CTTGGCTATC TGTTAAGCCG TGGTGAGGAA CAAACGCAAAA CCTGGGGCTG 12360
AGGCTGGACT GTATACAAAT GCCCTATTTG AGGCGGTAGCC CCCTAACAAC AGCAGCAAGG 12420
TGCAGACCAA CTTTGACCGG ATGTGAGCGC ATGTGCGGCC GGTCTGCTCT CAGCGGAGGC 12480
GGTGCCAGCG AGACTCCGTA CTAGGGGCTC GTGGTGGCGTT GAAACCTTTA CTGACGACC 12540
AGCCTTGCAA CGTGCCCTGG GCGCCAGCAA ACTACAACAA CTTCTTGAATG GCAATTAGAC 12600
TCATGTTGCC CGAAGTCCCT CAAATGGAGG TGTACAGGTC CGGCGAGAC CAGACGGGGC 12660
AGACCCAGAG CAGGGCTTTG CAGACGCTGA ACCGTAGCCCA GCCCTTAAAA ACGCTGAAATG 12720
GCTCTGGGGG AGTGGCTGGCC CCAAGTGGCA GCTGCTGACCT TTGCTACGGC 12780
CCACCTTGGG CTTCCTACTG CCTCTCTGATG CCCATTTCAA TGCAGCGCGT ACATCCGACC 12840
GCAATTCTTA CTTGGGCTAT TTTGTAACC TTGACGCGCA GCCCATAGGG CAAACCTCGG 12900
TAGATGAGCA AACCTTACA GAAATTCGGA AATGACGCGC CGCTCTGGGT CAGGAGAGCA 12960
CTGCGACCTT GGAAGCACCAC TAAACTTCTC TGTGACCCAA CCGTGGCGAG AGATCCCTCC 13020
CTCGATAGCG GCTTACCACG GAGGAAGAAC GAATCTGAGA ATACGCGAGC CAGAGCTGGG 13080
GAGTCTCCCT AATGCAGGAG GGGGCAGCTC TACTGCTGCG GCTAGATATCG ACAGCCCCAA 13140
ACATGGAGCC CAGACGTGAT CAGCTGAAAC GGGCCTTTAT CAATAAATAC CTAGACCTAC 13200
TACACAGGGC CGCTGCTATT AACTCTCACA TTCTCTACCA ATGCATTACT AACCAGCTTT 13260
GCGTGGCCAC ACCTGCTGTC TATACCCCGG ATATGACGAC GCGCCAGCACC AATGACGGGT 13320
TTTTATGGGA CGATGGGTAAC AGTAGTGGTTT TCTCCTCGGTC TCTGAGTTAT AACCCTGGA 13380
AGAAGGAAGG GGGCAGTGAAG AGGCATGTCT CCTATCTGCT GTGGCGGCGCA AGGGTGCTGG 13440
TCGCGAGGCT GCCGCGGCTGG CAAAGCTCTT TCTCTAGTTC GCCATTGGGC CTAAAACAGTG 13500
TACCGAAGCAG TGAGCTGGCC AGGATACAGG GTCCCGGCTTT GATGGCGGAG GAGGAGTACT 13560
TGATGACTC GCTGTTGGG CCAGAGCGGG AGAAGAAGCTT CCCAATAAAC GGGGATAGAAG
GCTGTTGGGA TAAGTAGACG CGCTGGAAGA GTAGCGCGCA CGAAGCAGAG GACGAGCCCC
GAACACGACG CACGGGCGGG GCCGGTAGAC GCCAGGGGCA GATAGGCCAG GGGAACGACTTG
TGAGGGAGGG TAAGAGGTAC CGAGCGGGCC GCAGGGGCTGG GAGACTGCTTG GGGAGTGTGTG
GTGGTACCCC GTTGCTCAC CTGGGCCCTCC GGATTGCGCC CTCGATGTAA AAACCAGAAAA
TAAATGCTAC TCACCAAGGC CATGGGAGAC AGCTGCGTCTT GCTTCTCTCT CTGTGTGATA
TAGTATGAGT AGGGAGAACG TGCTAAGGAG ACGGGTTGTTA ATACGCCGAGG GTCCTCTCCC
TTCCTTTGTC CCCCTCTGGT ACCTGACTAC CCCAGAGGGG AGAAACAGCA TTGATTACTC
GAGCGCTGCA CACTGATATT ATACACCCGG CTGTTATTTG TGACACACCG AAACGAGGGA
CATGGCCCTCA CTGAAACTAC AGAAGCCACCA CAGCAACTTCCTCACCAGGG TGTTGCAAAA
CAATGACTTT ACCCCCGACC AGGCGAGCAC CCAGAACATC AACTTTGAGC AGCGGTCCCG
ATGGGTTGTG CACGTGAGAA CTGCTACTGA CACCAAGATG CCCAACACTA AGCAATGATAC
GTGATAGCAAC AGTTCCTGAA ATCGGCTGATG GCTTCAGAGA AAGGTCCTAC
AGTATGAGAC AATATGTATC ACAACAGAGA TATTTTGAGA TATGAGTTG TTAGTTTATC
CTCACCGCAAG GCGAAGATTA CAGCAACTAC CTAATGACAC ATGCAACTCC
TGATATACCA CTTGAGTGGT GCACAGTGTTA TGGAGTTGAT GGAATGAGCA TGGGTGTTAA
ATCTGACTCC AGGAACCTTTA AACTGTTGTT GATACGGAGA ACTAAGTGGT TTAGCCTGG
GGTTTACACC TATAGGGGAT TCCATTCCGA CATGGATTTA TGAGTTGCTT GGGGGTTGGA
CTTTACTGAA ATGGCGCCTA GTAATCTACCT TGTTTACAGG AAAAGCACC CATTCCAGGA
GGTTTTAGA ATCTTGATAG AGGATCTTGA AGGGGTTAAT ATCCCGACCC TTTTGAGGTA
AGAGCCTATG GAGAAGAGTA AGAAAGCCTCA AGAAGCCCGA CTAAAGCTGC
TGCTATGCTG AAGGCAATAA TAGTGTCAG CGACCTCTGC AGGGTGCGTA ATGCCGAAAG
AGTCAGAGG AACAACCTGA CAGCTCACTC TGTCGCAACT GAAGAATGGC TATGAGTAC
TGCTGCGACT GGGACCACAA ATACAGGAGC AGGACTCCTC ATCAAAACCTG TAAAACAGGA
TAGCAGAGTA AAGAGGTCAC ATGCTTGAGA ATAAAGGATT AATACGAGCT ACCGACGTG
GTATCTGCTC TACAACATTG GCCACCCCTAA AAAAGGAGTT CAAGCCGCTGAA CACTGCTCAC
CACTGCGAT GTCACCCTGT GACAGAGCGA GGTTGACTGG TCACCTCACG ACATGATGCA
GGACCGGCTC ACAATCGCTT CCAGAGACCA AGTCAGCAAC TATCGAGTTG TGGTGAGAGA
GCTCGCTTACT GCTTCCTCTAA AAGATTCTTA CAAGCACGAA GGCGTGATCT CCCAGAGCT
TGCCAGCTCC ACCTGCTTTC CGACAGCTTT CAACCGCTTC CCTGAGACCC ACATGCTCA
CCGCGCGCCA GCGCGCGCAC TATACCCCGGT CAGTGAAAAC GTGCTTGTGCT TCACAGATCA
CGGGACCCTG CCGTTGGCGA GCAGTATCCGG GGGAGTGACC CGCGTGACC GTTACTGACGC 15540
CAGACGCGCG ACCCTGGGGCT ACGTCTACAA GGCCTACTGGG ATAGTGCGGC CGCGTGCTCT 15600
TTCAGCCGGC ACTTTCTAAA AAAAAAAAAA TGTCCTGACC TATCTAGGCT AGATAATAACA 15660
CGGGTTGGGG CCTGGCGGCAG GCAACAGAAG TGGTACGGAG TGGTGCCGAA CGCTCTACAC 15720
AGCACCCGCTG CGGAGTCCGG GGACACTTCC GGGAGTCCGT GGCGGCGCCTG AAGGGCGTGA 15780
TCCGACTAG AACCCCTGCC GTAGTGTGTA TGCCAGCGGT GGTTGGGGAT GTCTGCTAATT 15840
ATACTCTTAC TGGACCTTCA TCTTACTGGG ATGGAATTTA AGCAGCGTAA GCTAGCTGAG 15900
CCTGCCCTAT TGGTGCCGGG AAGACAGGCC GAGAGCGCAT CGCGGAGCGC CACCGGGCTA 15960
CTGGCGCTATG AGGACCGAGA AGAGCTGGTC TACGGAGAGC CAAACCGGGT GGGGCAAGAG 16020
CTATGGCTAG AAGCAGCCAG CGCCGGGTCT CAGGGTCGAC TCTGGCCGAG TCCCGCAGGC 16080
GCGCGCCACT TGGACGCGGT CCAACGGGTC CCAACGCGAT AGAGGCAATT 16140
TGTTGCCTAGG GGGCGCGGCG ACCACCGCCG AGGGTGCTGG CTTGCGGACC GTGCCCGTC 16200
GCTCGTGGAA GATATGCGCC AGTCCTCAGG GTTCAGTCGG ACCGGTACAG TCCAGGCGCA 16260
AATACAAGAG AGGATGTCTC AGGGTGATCG CCGCTGAAT ATCGCCTGCG CGGCGTGAAG 16320
ATGAAAAAAA GCCCCCGGAA ATCAAGCGGG TCAAAAAGGA CAAAAAGAA GAGATGGCGG 16380
ATGAGTGTTG GGTGTGAGGT TTGGCCTGGT TGCCGCAAG GCGCGGTAAG CATGGCTGGTG 16440
GACGCAAAGG GCCGGCTGTG TCTGAGACCG GAAACAGCGT GCTGCTTACG CCCGGGCAGC 16500
GCTCGACGAC TGCTTTTAAAG CGTCTGCTATG ATGAGAGTGA TGGGGAGATG GATATTCTGG 16560
AGCAGGCGGC TGACCGCGCTG GGGAGGTGGG CTATTTGGGAA CGCGCTGGGG TCCAGCTGCG 16620
AGGGAGGAGC GCAGTGGCATT CCACTGCGACA ATGGGATACC CACCCTGAGC CTCAAGCCAG 16680
TCACCGCTGA CAAAGCTCGT CCGGTGCGCTC CAGCCAGGAG CATCAAGGCA GAGGCTGAGG 16740
ATCGTGATCC CACTGTCGAA TTGAGACGGC CCAAGCGGCA GCAGCCTGGG GAGCTGGCTGG 16800
AGAAATAGAA AGTGAGCTCC GATATACACA CTGAGGTCGA ATGAGAGGCC ATCAAGGCAGG 16860
TGCGGGCCAG TTTGGAGGTA CAAACCGTAG ACATCAAGAT TCCAGCAGGC TCCATGGAAG 16920
TCCAAACCAG ACCTGCAAAG CCCACACCA CCAAGTGGGA GCTACAAAGC GATATGCTGA 16980
TGTCAGCCA CTTTAAACT CGAGTTGCCG TCAGAACAAC CGAGAGTACC TCGAGGCCAGT 17040
ACGGTCCAGC AGTGTTGCTG ATGCCAATGG TCTGCTGACA CCCCCTATTG ATTCGACCCT 17100
CGGTTACGG AGGACAGGG TACTCGGCTA CGCCGGACAG TACCTCGGGC GTGGCGGCGA 17160
AAACACCTAC AGTCTGACTG CACCGTGTGC GCCGTCGGCC CACCGAAGT CTGAGCTCAGG 17220
CTGCTCTGGT GCCGAGAGTG TATCCGAGTG CGCGCGGCGA TCCATGAGCG TTGCCACCGG 17280
TAGCTACCA CCCAGGACTG ACAACTTAAAC GACCTGGTGCC GCCTCCTGCT TGCAGATATTG 17340
GCCCTCAGTT CGCGCCTTCTG TGTCCTGGAT ACTGGAATCC GAGGAAGAAA CTGCACGGGC 17400
AGAAAGAGGA TGTTGGGGCG CGGGAGTCCGA CGCCACAGGC GGCNGGCGGC TATCAGCAAG 17460
AGGCCTGGGG GTGGCTTCTCT GTCCGTCTGT CTGCCTCGATCA TACCCGCCCCG GCCGCCATGC GATGCTGGGCG 17520
ATACACACAG TACCGTCCGC GGCTGGTCGAC GCCTCCAGGC GCCACTGACA TTGAGAAAC 17580
TAAAAATAGAA CACTCAGTAC ACTGTCATAC CAGTGTGATCC CTGCATATGT TTTTGTAGAC 17640
ATGGAAGACAC CAAATCTTGAC ATCGCTTCGCT CCGCGACAGC GCAGCACGGC GTACATGGGC 17700
ACTGGAGCGG ACGTGGACAG CCAGCACAGG AACGGGCGCG CTTTGAATTT GAGCAGTATG 17760
TTGGAGGCGG AGTTAAAATTT TGGGCTGCTAC ATAAAAACTC ATGGGAAACA AGCTTGGAAC 17820
AGCAGCGACG GGCGGCCAGT TGAAGAAGAT CTTAGAAGAC AAAAATTCCA ACAAGAGTG 17880
GTGGATGGGA TGCCCTCTTG TATTAATGGT GTGGGTGACAT TGCCCAACCA GCGGCTGGAG 17940
AAACAGATATA AAGCCGGCGT GGAGCCGGCG CGCTCGCCACGG GGGTGGAAAT GGAAGTGGAG 18000
GAAGGATCTCC CTCCCCCTTG AAAAGGGGAC GACACCGTGC CGGCGCCCCA TGCTGGGAGG 18060
AAGACTAGCA CAGGTGCTAGA CGAACCCGCG TCCCTACTAG AGCCAGTGAA GCTTGGAAATG 18120
CCACCACCA GCCGTGTAGC CCCATGGCT CGCCGGGTGA TGAAACTCCT TCAGTCACCAC 18180
CGACCCGGCTT CTTTGCATCT GCCCTGCTCC CGTCTGTGCTG AGCCCGCTCG TGCAAGGCTT 18240
GTGCCCTACC CGAAGGCACC CACCGTCTAG CCGGCGGCGG TAGCCCGACC GGTCTCGGGG 18300
GGGGGCGGC ACCGCAATTG AAATGCGCAG ATGCTGCTGA ACAGCATCGT TGGTCTGGGC 18360
GTGCRRAAGTG TAAAACGCGG TGCGCTCCTTT TAAAATTAAT ATGGAGTACC GCTTAAACTTG 18420
CTGCTCTCGT CGTATGTCCT ACATCAGGC CGCTGGCAGG CCAACACGAG AAGGAAAGAG 18480
AAGAGGTGGG CGGGGGACGG TGAGTCGTCT CTCAAGATGCG CACCCCATCG ATGCTGCCCC 18540
AGTGGCCATA CATGCACATC GCGGGACAGG ATGCTTGCGA GTACCTGAGT CGGGCTCTGGA 18600
TGCAAGTCGC CGCGGGCAAC GACACCTACT TCAATCTCGG GAACAAGGTTT AGGAACCCCA 18660
CGTGGGCAG CGCCCATGAT GTGACCACCG ACGGCGTACTG CTGGCGTCTTG 18720
TACCGCTGTA CGGGGAGGAC AATACCTACT CATACAAGGT TGATAGACCT TGCGGTGGTGG 18780
GGCACAACAG ATGTGCTGATG ATGCCCAGCA CTTCTTGTGA CATTGGGGGT GTGGTGATGG 18840
GAGGCCCTAG CTCAAGGCAA TATTCTCTGG CTGCTTACAA CTCAATGCGG CCTAAAGGCGG 18900
CTCCCAATAC ATCTCAGTGG CTGGATAAGG GAGTCAACAAC ACTGTAATAT ATACGTAAA 18960
ACGGAGATGA AGAAGATGAA GTGCCGGAGG AAGGGGAAAG AGAAGAAAAA GCTACATACA 19020
CTTTTGCAAA TGGCCCGTGA AAAGCCGAAG CTGAATTAC AAAAGAAGGAA CGCAGAATTAG 19080
GTTTTGAAAG TCCATCTGAA GGGTCACCTA AACCCTTTAA TGCTGATAAA CTGTATCACG 19140
CAGAACCTCA GGGGGGAGAG GAATCTTTGA CTGATACGGA TGGCACAGAT GAAAAATATG 19200
GAGGCAAGGC ACTTAAACCT GAATACAAAA TGAACCCCTG CTACGGGTCTT TTTGTAAAC 19260
CTACTAATGT TAAAGGGCGC CAAAGAAAGT TGAAGAAAGT AGAAGAAGGC AAGGTTGAAT 19320
ATGACATTGA CATGAACTTT TTCGACCTAA GATCACAAAA GACTGGTCTC AAGGCTAAAA
19380
TTGTAATGTA TGCAAAATTT GTGGATCTAG AAATCCCGAA CACTCACTTGG GTGTAACAAA
19440
CTGGACCTTC AGATGCTGAT TCTCATCCAA ACCTTGTTCA AGACCTCATG CCCAATGAC
19500
CTAACGTATA TTGGCTCCGG GACAACTTCA TCCGACTCAT GTACTATAAC AGTACTGGCG
19560
ACATGGGAGT GCTGCTTGGCA CAAGGCTCTC AGCTAAATGC AGTGGTTGAC TTGCAAGAGA
19620
GAAACACAGA ATTTGTCATA TCAACTCTTG CTTGATCTCT GGGAGACAGA ACCAGATATT
19680
TCAGGATGTA GAATCAAGCA TTGAGATGCT AGGACCGAGA TGCGTCGTTT ATGGAAACCC
19740
ATGTTGTTGGA AGATGAACCTT CCAACTATT TTTTTCTATT GGAGGCTTTA GGTGCCGGAA
19800
CAGACAGTTA CAAGGGAATTT GAGACAAATG GTAGCAGAAA CACTACTTGG AAAGATTAG
19860
ATCCAAATGG CATAGTGAA CTGACTAAGG AAGAACCTTT GGCAGACAGA ATCAACATCC
19920
AAGCTAATCT CTGAGAAGAT TTTCTTATT CCAAGTGCAG TTTCTTATCT CCGACACTGT
19980
ACAATAACAC TCCAACCAAT GTTACTCTCC CAGAAAAACA AAACACCTAT GACTACATGA
20040
ATGGGCCGCT GTTCCCCCCC TCTCTGGTGAG TACAGCTAGT AAACATTGGGC GCCAGATGCT
20100
CTTTGAGATGC CATGGACACG TCTAACCCTCT TCAACATCTA CCGAAGCCGT GCGGTCGAGT
20160
ACGGTTCTCA TCTCTTGCGC AATGTTCTCT ACGTGCTCTT CCAACTCACA GTGACTCAGA
20220
AATTCTTGTG TGTAAGAAAAC CTGGTCTCTC TACCTGTTTT TTAACACCTA CAGTGGAACG
20280
TCAGAAAGGA TGTAACATG GTCTGCAGA GTTCCCCCTT GAAATGATCC CGAGTTGATG
20340
GGGCGCAGAT CAGTTTACC AGCATCAATC TCTATGCCAC TTTCTTTCCC TGGGAGCGAGG
20400
ACACTGCCTC CACCTTGGAA GCCAGTGGCG GCAAGCAGAC CAAATGATCAG ATCATCACATG
20460
ACTACCTTTC TCAGCCTAAC ATGGTCTCCCA CCATCCCCGC CAATGTACATT AAGGTTCCCA
20520
TCTCCATTC CTCCTGGACAG TGGGGCCTGC TCAAGGGCTG TCTCCTTACC AGACTGAAAA
20580
CCAGGAGAC TCCCTCTTTC GATTAGGGT TGATCCCCCT TCTTCTTTAC TGGTTTTACT
20640
TACCCTAAGT GGATGTACCC TCTTACTCTA ACCACACTTT CAGAAAGTG TCTATACAGT
20700
TTGACTCTTC AGTCAGCTGG CCTGGTAATG ACAGATTGCT AACTCCAAAC GAGTTCGAAA
20760
TCAAGGCAC AGTTGATGGG GAAGGCTACA ATGTTGGCCA ATGTAACATG ACCAAAGACT
20820
GGTTTCTGCT GCAAAGTACT GCAGATTGAA CCAGGTGTTAT TATGTTCTCT
20880
AGGGTTAAGG GATGCGATCT TATTCTCTCT TCGAAACATT CCGACCAATG AGTAGACAGG
20940
TTGTTGATGA GATTAACACAA AAAGACTATA AAGCTGTCGG CGTACCCCTC CAGCATATAA
21000
ACTCTGCTTT TGGGGTATTG ATGGCTCCTAG CCATCGCTCGA GGGTCAAGCC TACCTGCTCT
21060
ACTACCCATA TCCCCAAATT GGAACCGTCT CAGTTAACCAG TGCTACCCAG AAAAAATTCC
21120
TGGTGACAG GACCATGTTGG CCGCATCCAC TCTCTAGCAA TCTCATGCTC ATGGGTGCCCA
21180
TTAGCAGACT GGGACAGAAC TTGCTGTAGC CCAACTCAGC CCATGCGCTG GACATGACTT
21240
TTGAGGTGGA TCCCCATGGAT GAGCCCCACC TGCTTTTACTCT TCTTTTGGAA GTAATTGACG
21300
TGTCAGAAGT GCCAACACCA CATCCGGGGG CTATCTGGAGGC GTCTCTACGT GGCAACACCGT
21360
TCTCCGGCTGG TAAGCGCACC ACATAAGAAA CTCGCTTCTT GTAAGGGGCA GCAATTGACCT
21420
GGTGACGGGG AAAAGGGCTCC AGCGGAAGCC AGCTCAAGGC CATCGTCCGA GACCTTGCGT
21480
GTGGACCCTCTA TTTTCTGGCA ACCCTTCTCC AAGCTCTCCC GGGTTTTATG GCTCCGACAA
21540
AGCTGCGCTTG CCACATGGTC AACACAGCGG GTGCGGAGAC GGGGGAGGAC CACTGCTTGCG
21600
CTTCTGAGTG GAACCGGGGC TCCAACACTG GCTACCTTTT TGATCCATTT GGGATTCTCGG
21660
ATGACGCTCT TAAACGACATC TACCAGTTTG AATACGAGGG GTTGCTGCCC GGGTGCCGCC
21720
TTGCTACTAA GGAAGCTCTGC ATAAAGCTCGAC CCAAGATCCGAG GGGCTGCCGC
21780
GTCACCCGGC TTGGTGGACTT TTTTGCTGCA TGTCTTCTCA TGCCGCTGTGA CACTGGCCAG
21840
ACGCTCCCAAT GAACGGTAAC CCCCCACAGG AGTTGCTTAC GGGAGCTCCC ACAACATGC
21900
TCCAGTCACC CCAAGTCGAC CCGAATTCGC GCAGGAACCA GGGGGCTGTC TACACGTTC
21960
TCAACACACAG TCTCATTATC TTCCGTTTTC ACCGGCCACG TATCGAAAGG GCTACTGTGGGT
22020
TGATGTATG GGGATATATT AAGTCTATGA AAACCGTTGT CAATAAACAG AACTTTTTAT
22080
TTTCACATCA CTGGTGTTTT GCTATTTATG TGCCCCAGAA TGCCGAGGGG TTTGCGGGG
22140
AATCAGAGTG ACCCGCGGCG AGGGATACGT TTGGGAACGT GAACTGAGCT TGCCACTATA
22200
ATTCCGGGAT CACCAATCTTG GGAATCTGGA GGTGCGGCTC GACCTTCTAC
22260
GGGTAGTTAG CAGGGCTCACC AACAGGTACG GGCTGAGAAAT CTAAAGACAG CAAATTGCGG
22320
CCCGTCCCGG AGCCGGGAGG TTGGAGTACA CAGGGTTCAG ACACTGGAAC ACCATAGCGG
22380
ACGGTTATTT CACACTCGGC AGCAAGTGAG GGGTCTGGAT AATTCGCCA CAACGCTTTT
22440
GGCAATTGCC CAGTCGAAGA GGGTGCTACT TGGCTCTACG TCTGCCCCA TGGCTGATAC
22500
AGGCCGCGTTT TGTTGTGCAA TGCCAGGGCA GAGGGTAGG AATCATCTTG GGGCTGTCGGG
22560
ATCTCATACC TGGATTACACA GCCTCTGATGA AAGCTTACAT TGTTTGGAAA GCGCTTGCGG
22620
CCTGTGCTAC CTCAATGTAG AACATCCCAAT ATGAATATGC TTAGAATCGG TTGGCAAGCA
22680
ACCCGGCATC ATTCAACAA CAGGGAGGCT GTTGTCTGAC TATTTGTTAC AACTCCCCGC
22740
CCACCGGTTT CTGGGTGTAC TGTTTCTCC CAGGGTTCAT TGCCAGGGCC GTGGGACCCG
22800
TTTGGCTGCC CACATCCATT TCTATGATAT GCTCCCTTCTG GATCATGAGT TGCCATTGCA
22860
AACAATTTAG CTTGGCCTCA AAATCATTAC ATCCCATGTGA CCACAACGGC CATCCCCAGAC
22920
ACTCCGCAGTT ATTTGAGGCC ATCTCAGAAG AGGAGTGAGC CAACCCCGCTG AGGAATCTTC
22980
CCATCTAGGT TGAGAGGGGT TGTTAACCTG TAAGATCAGG GGGAGGCTGC CAGTGCTTCT
23040
CATTCCACAT CGGTGGGCAA ATTCGCTTGT ACGTTTCATG CTGCTCTGCG ATTAGCTTGA
23100
AAGAGGTCTTC TAGGTCATTG TCCACCGCTCT ACTCTTCCCA CAGCAACAGC ATTAATTCCA
23160
TGCCCTTTTC CCAGGCAGAA ACCAGGGGTA GCACATGGC AATTCCAACA GAAATAGCAG 23220
CTACTTTCGC CAGAGGCTCA TCTCTGTCGA TCTCTCAAAC TTCTCTTTTT CCATCCTCTT 23280
CAGTGTAGCG CACCAGGTTGG TAGCTGAAGGC CACCAGCCCAC CAGCTGGGCTC TTCTCCTTTT 23340
CTTCTTCGCT GTCTCTGACTGT ATGTCCTTGTAA AAGGGATGATCG TCTGCTTTTG CTCGACTCTCT 23400
TTTGGGGGGG TATTGGGCGGA GGGCTGCTGC TCCGCTCGCG AGAAGATGAG GACCGCGAAG 23460
TTTGCGTCAC CAGTACGCCA TGGGCTTGGG TAAAGAACCAG CAGACCCCAA GGCGCGTGAAG 23520
TGGTTCCTCTT CGGGGCGAGA GCGCGAGGCTG GCTGCGATGG TGCCGCCTCT GCCTGGGAG 23580
GGCGATGACT GCCAGAGCCCG CTCTGGGTGT CGGGGGTGTTG GCTGCTGGGG CCGTGGCTTG 23640
ACTGATTTCC TCCGCGGCTG GCCATTGTGT TCTCCTAGGA AGAGAAAACA AGACGATGAG 23700
AGACTCAGCC ATCTGTCGCA ACACCCGTGC ACGACATCTC AGCCATGCAC TCCAGGCGAG 23760
AGGGAGAGGA ACAAAAGCTTA ACAGCCCCCA ACACAGCTCC GCACCCACAC ACCTCCTCACC 23820
TCGAGGATTGAGGAGAGTGCAC GCACCCCCAGG AGATAAGGCA GCAGAGTAGATG AAGGATGAGA 23880
AAAGCCGAGA CATTGAGCCTA GATATCGAGCG AGGACCAGGT CTATGTAACA CGCGCCGCAGC 23940
ACAGAGGAGAA GTGAGACCGCC TTCTATAGAGA AAGATGATGA CAACCGCTCA GAACAGCAAG 24000
CAGATGGGGA TCAACAGAAG CGCTGGGCTCG GTGGTCTAGT GCAGCAGATAC CTCACCGGCC 24060
TTGGTGGGGG GGAATGTGCTCTGATCAACC TACAGAGGCA GTCAGATCATA ATCAAAGAGC 24120
CAGTCTGCTGCA GTCAGGCGA TAATGCGGCTA GTGTTGCAAAG GCTACGCAGCC GCCCTAGACC 24180
TCAATGTCGTT CTGCCTGCTG TCAGCTCCAC CGCTCCCACA AAGCGCTAGC TGGAGCGCCA 24240
ACCGTCGGCT CAACTTCTCAT CGCGGATTTCA CGCTGGGGCTG GCTTGCGTGGC ACCTACGACA 24300
TATTATTATTAA AAACCCAAAA ATCCCCATTT CCTCGCGGCG CACGACAAGT CGCGGGAGTG 24360
CCCTGTTCACA CTTTGGGACC GTGGCCTTGT TACCTGATAT GCACCTCTTG GAAGAGGGCC 24420
CAAAGATATT CAGAAGCTTG CGACGTGTAG AGACTGCGCG CGCAAAATGTG CTGCAACAGG 24480
GAGAGAGTAGGATTGAGAA CATCAGAGCGC TTCTGGTGGA GTTGGAGGGC GATAATGCAC 24540
GACCTTACGT ACTCAAGGGC AGTATCGAAG TGACCCATTG GTCATACCC CGCTGCAACC 24600
TGCGCTCCCAA AGTCTAGGCG CTCGCTTCGG ATCAGATACT CATTAAACCG GCAAGTCCCC 24660
TATCAAGAAA CATGCGACGCT CCGAGTGGATG GCTATGGAGG CAAACAGCTG GTCAGTATG 24720
AACAGGCTATC TGCTGGCTCG GCAGCAGACTC CCCACTAGTA CCTGAAAGAC CGCGCCAGCC 24780
TCAGATGGCC CGTGTGGCTCA TTACTGTTGGA AATGGAGTTG TCTGCGCGCC TTCTTCACTG 24840
ACCCGGAGACTTGGCAAG ACTGAGGAGA ACCTGCACTA CACTTTTACA CTTGGATTTG 24900
TGCGACAGCC ATGCAAGATC TCACAGGCTG AGCCTACCGA CCTGTTTTCC TACATGGGCA 24960
TTTTGCAATAA AAAACAGACTC GCCAGACAGCG TTGGTGCACAC CACCTGAGAG GCTGAACGCC 25020
GTGCCGACTTA CATCCGCAGAC ACTGCTTACC TCTACCTCTC CCATACCTGG CAGACTGTTA 25080
TGGGTGTGTG GCAGCAGTGT TTGGAGAGAC AGAACCAGAA AGAGCTTGAC AAGCTCCTAC 25140
AAAGATCCCC CAATCTCTG TGGACGAGGG TTGACGAGCG CACAGTCGCC TCTGATCTGG 25200
CAGATCTCAT CTCCCCCGG AGTCTCAGGA CCCACCTGCG CAACGGGCTG CTTGACTTCA 25260
TGACCCAGAG CATGCTTAC AACTTTGCTT CTTTCATCCT GGAAGCGCTC GTATATCTGC 25320
CCGCCAACCTG CTGTGGGCTA CCATCCGACT TTGTGCTCTC GACCTACGCG GATGGCCAC 25380
CACCACCTAT GACGCAACTC TACCTGTCCG GCTTGCGCAA CTACATCTA TACCATGCCG 25440
ATGATGCTCG GGAATGCAGG GGAGATGGCC TGCTTGAAGT CCACTCGCCG TGTAATCTCT 25500
GCTCAACCAC TCACCTCCCCT GTCTGTAACC CCAAGCTGTG TACTGAAACC CAATTATCAG 25560
GCACCTTCGA ATTCGAGGCT CCCAGCCGCC AAGGCGATGG GTCTCTCTCT GGGCAAGATT 25620
TGAAACTGAC CCCGGGACTG TGGACCTGCC CCTACCTGCG CAAGTTTCCG CCGGAGGCAC 25680
ACCACCCCCTA TGAGATCGAG TTCTATGAGG ACCAATCACG GCCCGCCCAA GCCGAGCTAT 25740
CAGCATGGCT CATCACCACG GGGCCATTT TGGCCCCATT GCAAGCCATC AAAAATCCCC 25800
GCCAGAATT TGGTCGAAA AAGGTTAACG GAGCTCAGCT CGACCCCCAG ACTGTTGAGG 25860
AGCTCAACAC AAGGTTTCTT CAGGATGCTG CAGGGCCGAG GAAGCAAGAA GTGAAAAGTG 25920
CAGCTGGCGGC CCCAGAGGGA TAGAGAGGAA GACCTGGGAC ATCGAGCAAG AGAGTGGGAA 25980
GATTGGGACA GCCAGCCAGA GGGAGGAGG GACAGCCTGG AGGAAGACG TTTTGGAGGA 26040
GAAGACGGAG AGCCAGAGGA GGTGGAAAGA GCAAGGCCGC CCAAACAGTT GTCCCTGCGA 26100
GCCAGAGCAGA CAAAGGCCCAC AGACAGTACC ACAGCTCACCA TCTCCGCTCC GGGCCGGG 26160
GCCAGAAACCT GTGCAAGGGAG TAGATGAGAT GAGACGGGCG GACTCCCGAA TGGCAACCAC 26220
GCTTCTAAGA CTGTTAAGGA GCGCCAGGGA TACAAGTCTT GGGCCGCGCA TAAGAACGCT 26280
ATCATATCTC GGTGGCATGA ATGCGGGCCC AACATATCTC TCAACCCCAG CTACCCCTCT 26340
TTCCACACCG GGGTGAACTC CCCGCCAAAT GCTCTTGACT ACTACCGTCA CTCACACAGC 26400
CCCTATTACA GCCGCAGAGC CTGGCAAGAA AAAGACAACA GCAGCAGAGA CCTCCAGCAG 26460
AAAAACAGCA GCAGTCTGAA AACCACGCAG AGGTGCAGGA GGACTGAAAC TCACACCGGA 26520
CGAGCCAGCG CAGACCCAGG ATCGTAGAAA CCAGATTTTT CCAACCCCTT ATGCCATCTT 26580
CCAAACAGGTT CCCGGGAGAG AGCAAGGATT TAAGATTAAA AACGACATCTT TGGTCTCGCT 26640
CACCAGAGTT TGTTTATAGC ACAAGAGCCA AGACCAACTT CAGCCGACTC TGAGGAGCGC 26700
CGAGGCTCTC TTCAACAGGT ACTGCGCGCT CACTCTAAA GAGTAGCCCG CGCCCGCGCT 26760
ATCTCAGAAA AGGGCGGATAT TACGTCCCCT TGGCGCGGCC TCTTTTGGCC TGCTACGAG 26820
TAAAGAAATT CCCAGGGCTT ACTGTTGAG GTATCAAGCC CAAATGGGAC TGGCAGCAGG 26880
CGCTCCAGAG GACTACCTCA CCAGTATGA TGGTCGACCC GCGGCTCTCT CGATGATCTC 26940
ACGGTGAAT ATGATCAGAG CTTATCGAAA CCAATATTCTC CTAGAACAGT CAGCCTTTAC 27000
CACCACCCCA AGACAAACAC 1CAACATCCCC GCCGCGCCTGG TGTTACCAGGA 27060
AACCACCGCT CCCACACCAG TACTACTTCC TGGGAGCAGCC CAGGGCGAGG TTTGAGATGAC 27120
TAACAGGGTT GTACGAGTGG GCGGCGGTTC CCCGCTTCTTC GTGTTACGGCC CTTGACAGAG 27180
TATATAAAGCC CTTGCGTGCA GAGCGCGGAA TTATCCAGTC AAGACAGGAT CGTTGAGCTC 27240
TTCGGCTTGG GTTGGACCGAG TGGGAGCTTT CGGCAATGCG CAGCTGCGGG GAATCTCTCTT 27300
CATACTCTGT CGCCGCTTAC TCAGGCGGTT CAGGTGGTGG TGCCAGCGCC GCTCGGGTTG 27360
CAGTGCGACT CTCCGACTTGG TGGGAGGATT TACTCCTCTGT GTCTACTCTCA AGGGCGTCTC 27420
CGGATCTTACT GGGCAGTTG CATAACCAAT TTGGACGCAA TCAAGCGGAT GC 27480
AGGTGATGTT TATGATGATT GTCTAATGAGG GGGCGGGCTT AGCTAGCTCG AGGCGACAC 27540
CTGAGCAGCT GCCGCGGCTT GGCGCTGTGG TACGAGAAAC TCACCGGAGGT CCACCTACTT 27600
GAAATAACCC AGGAGCCACC TCAGGGACGG GCCACGAGAG TGTGATATAC CATCGAGGAG 27660
GGATAGACT CTGGCCCTGCA TGGGATTCTC TGGGAGGCGG GCGTCACTGC AGGCGGGAAC 27720
CAGGAAAGCA CCACAGTTCT CATCTACTGC ATCTGTAAAC ACCGGGAGTT GCCGAAAGGC 27780
CTTTGCTGTGC TTATGTTGTC TGGATTAAAT AAAACTGAG TTAGACTCTC CTTACGGACT 27840
ACCAATTCTT CAAACCCGAC TTATTAACAA TCAGACCCCT TCACCAATGC AGAAGACCCC 27900
AACCCTTCCT CTGATCGCAGG ACTCTAAATCT CACCTCCCCA GCACCACACT TTAGTACCT 27960
TCCCGAAACT AACAACCTCG GAGCTAAAAT GACCCCGTTT TCCGAAGACC TCCTCTCTGC 28020
CAATACTACC ACTCCCCAGA CCGAGGTTGA GTCGCGTATG CTCTCAAATA ACAACCCCTG 28080
GGTGTTAACG GGGTTTGGAA CATTAGGTTG AGTGGGCGGGT GGGCTTGGTC TTATGCTTGT 28140
CTACCTATACT ACACCGTGTG CGTTATTATT AAGTAACTTTG TGGTGTGGTT TTAAGAAATG 28200
GGGCCCTAC TAGGCGGCTG TGCTTTATTCT CACTTTTATT AGCCGCTGCTG TACTATGCCTA 28260
GGTCAGGGTG TACTTATTTGA TCCATTGGCT CAAATTTGATCG CAGACAAATG CACACTGACT 28320
TTTGCTCCAG AGGGTGACCG TGTTGAGGTT CTTGATTGG ATGGGACGAGG TACGGCTCCC 28380
ATTGAAATAC ACCCAAAATAA CAACCTTTGG AAAATACAT TATTCACCCAC ATGGGACGCA 28440
GGAGACCTGT AGTGTTATAC TGCTCTGCTG CTTGTTGCTGG AGGTTTTTAC CGGCGACGCT 28500
AAATAACACTT TTTATTTTTG TGAGTGTGTCC AGCTATGGCA TTGCTGAGTG CAAAGAGAT 28560
AACCTATGGC CTCCAGGCAA GGAGAACATT GTGGCAATCTT CCGCTTTTA TTTCTTGT 28620
AAGTGTCTCA 1TAATGTCAT TCTGTGTATC TGCTACACT TGGTATTGCG CATCGGCACG 28680
AGAAACAGCA ATAGGGAAAA AGAGAAATTG CTTTGAGCTT TTATCTGACT ATGTGTTTTTT 28740
TTTTGGTAC AGAGACGGCT TCAGTTGATA CTCTATTATG TGCCACGATT CCTACGTGCG 28800
CACACGGGCA AAAAACTTCT TATATTATCT TAGTGTAATA CCACACCTCT ATAGGACCCC 28860
AAATTAGTTA ACAGGTTATA TGACGCAAAC TTGGAGAGTTG TGTATTATTT GACATAATCT 28920
<table>
<thead>
<tr>
<th>DNA Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAAACAGAA</td>
<td>28980</td>
</tr>
<tr>
<td>CAAACAAATA</td>
<td>29040</td>
</tr>
<tr>
<td>TTTTGGTACCT</td>
<td>29100</td>
</tr>
<tr>
<td>GTAAACAAACA</td>
<td>29160</td>
</tr>
<tr>
<td>AAATCTTTGAG</td>
<td>29220</td>
</tr>
<tr>
<td>CTACAAACTC</td>
<td>29280</td>
</tr>
<tr>
<td>AACATGTTAGG</td>
<td>29340</td>
</tr>
<tr>
<td>TGCTGACTGCA</td>
<td>29400</td>
</tr>
<tr>
<td>GGTGCGCTCAC</td>
<td>29460</td>
</tr>
<tr>
<td>ACCATCTGTA</td>
<td>29520</td>
</tr>
<tr>
<td>ACCCATGTT</td>
<td>29580</td>
</tr>
<tr>
<td>CTACAAATGAA</td>
<td>29640</td>
</tr>
<tr>
<td>GGGCGTCTCC</td>
<td>29700</td>
</tr>
<tr>
<td>CTGTTGGCCA</td>
<td>29760</td>
</tr>
<tr>
<td>GTGTCATGAA</td>
<td>29820</td>
</tr>
<tr>
<td>ACATGAACAA</td>
<td>29880</td>
</tr>
<tr>
<td>AAAACGCGATC</td>
<td>29940</td>
</tr>
<tr>
<td>CTTTGGTAA</td>
<td>30000</td>
</tr>
<tr>
<td>ATGCCAGCA</td>
<td>30060</td>
</tr>
<tr>
<td>ATGCTGATA</td>
<td>30120</td>
</tr>
<tr>
<td>CATTGACGAG</td>
<td>30180</td>
</tr>
<tr>
<td>ACTTACTTAT</td>
<td>30240</td>
</tr>
<tr>
<td>ATGGAAGCTT</td>
<td>30300</td>
</tr>
<tr>
<td>TTTTTGGCT</td>
<td>30360</td>
</tr>
<tr>
<td>GACGTGCAT</td>
<td>30420</td>
</tr>
<tr>
<td>TTTTGGTCGA</td>
<td>30480</td>
</tr>
<tr>
<td>TTTTGGTACCT</td>
<td>30540</td>
</tr>
<tr>
<td>CACTTGGAGG</td>
<td>30600</td>
</tr>
<tr>
<td>TCTACCGGATC</td>
<td>30660</td>
</tr>
<tr>
<td>ATGGTACG</td>
<td>30720</td>
</tr>
<tr>
<td>TGCGGACGCA</td>
<td>30780</td>
</tr>
<tr>
<td>ATGCCACTC</td>
<td>30840</td>
</tr>
</tbody>
</table>
CAATAGCTTT AATAGATTAT GACAATGAGC CACAGCCCTC GCTGCTTCTT GCTATTAGTT 30900
ACTTCAACCT ACGCGGTGGA GATGACTGAC CCACCTTCGG CCTCCACTGC TGGCAGGAAA 30960
CTGCTTTGATA TGGACGCGCG CACCTTCAAGA CAGGACTCGG CCCAATCGGC CATACGCCAG 31020
CAGCAGGAAC GTGCCGGCAA GGAGCTCAGG GATGCTATAG AAATTCACCA GTGCCAAAAA 31080
GGCATATCTT GTCTGGTGAA ACAAGCCCAAG ATTTTCTACG AGATCACCAC TACTGACCATT 31140
CCGCTCTCAT AGAGACTCGG TCCGAGCGGC CAAAATAATCA GTGTATGGTG GGAATACAC 31200
CCCCATAGTCA TTTACCCAAG GCCGGAGAGAT ACTAAAGGTT GCATCCACTG TTCCTCCGTT 31260
TCCACCGAGT GCATCTACAC CTTACTTAAG ACCCTCTCGG GCCTTGAGA CATCCTCCAC 31320
ATGAACTGAT CAACTTTCTC TCCCCCATT CAAAACAAACA TAAATAAAAT CRCTTACTTG 31380
AAATCAGCAA TCATGCTTCC GTCAAAATTG TCTCCTAGCA GCACCTCACG TCCCTCTCTC 31440
CAACTCTGCT GTCTCAAACC CGCGTCCCGCA GCATACACTT TCCACAACTT AATAGGAAAT 31500
TCAAAATTTA GTTCCCTTTT TCTACCACAA ATCTTACTCT TTTATTCTCT CCCAGATGCG 31560
CAAAACGAACGG TTGGTAGCAGA GCTCTCTCAA CCCGGTCTAC CCCATAGGAAT GAAAAGACG 31620
CTCACCCACC TTTAAACC GTGTTTTGAT TCCCCATAAT GGGTTTACAC AAAGCCCGAA 31680
GGAGTCTCTG ACACTAAATT GTGTTGTCCT CCTTACAACC GCTATAGGCC CCGTAGATT 31740
CAAGTAGGGA GGAGGCTTTA AAGTAGAATC AACTGATGGA TTTCTTGAAG AAAACATAAA 31800
CATCACATCA CCACCTACAA AGTCTAACCA TTCTTAGGTT TTAGAATTGA GCCATGGGTT 31860
ACACACAAAC GAAGCCAGGC TCCTGTCTCA AGCTTTGAAAA GGTCTTGATG TTTAGCTTTC 31920
CGAGTGCTATTT CCAATGGAAA ATACACCTTT TTGGACAGGT GCAGAACCAAA GTGCCACCTG 31980
TGAAATTTAA GAGGGAGAGA ATCCCCAGAG CTGTAAGCTC ACTTTAGGTT TAGTGAAGAA 32040
TTGAGGACTG GTAAATGGAT ACATAACATT AATGGGAGAC TCAGAAATATA CTAAACACCTT 32100
GTATTTAAAAAC AAACACAGTTA CAATAGATGT AAACCTCGCA TTGGATAATA CGGCCAAAA 32160
TATCAGTTAC CTATCATCTG TAAAAAGTAA CCTGACTTTT AACAGCAACCC AAAAATAGGC 32220
TACTGGAACC ATAAACGAGT CCAAAGGTTT CATGCCAGGC ACCACCGCCT ATCCATTATAT 32280
AACATACGCC ACTCAGTCCCA TAATTGAGAA TTACATTTAT GGAGATGGTT ACTCAAATCC 32340
TACCAATGGC ATCTCTCTCA GACCAAAGT GCTATGCTCAA CTAACAGAGC GTATGTGACG 32400
TTCTGGAGATG GCCATTGCTA TGAACCTTTT ATGGCTCTTA ATACGAGAGG AGGGCCCCGA 32460
AACTACGCAA GTCACTCTCA TTACCTCCCC CTTTTTTTTTT TCTTTATATCA GAGAAGACGA 32520
CTGACACACA AAAATAAAGA TAAACTTTTT TATTGAAACT ATTTTTACAAG ATTCGAGTAG 32580
TTATTTTGGG CCCCTCTCCC CATTTTATAG ATACACAAAT CCTCTCCCCA CCGCAGACTT 32640
TGAACATATTG AATACCGTTA GAGATAGAGA TAGTTTCTGA TTCCACATTC CACAGGTTT 32700
CAGACCGGCG CAATCTTGGGA CTAGTGATAG ATATAAAAGC ATCGGAAAGC TCTTTCGAGG 32760
TGGTTCACA GTCGACTGC TGGGCTGGG GCTCGGAGGT TGAGTATTAGA TCTCATCTGGA 32820
AGAAGAACGA TGGGAGTCAT ATTCGGAGAA CCGGATCCGA CGGTTGTGGC TCAAACTCTGG 32880
AAGCAGTGGC TGTCTGGGCC GTCTGCTGGG ACTGCTGGTG ATGGGATCAG GATCCACAGT 32940
CTCTCTAAGC ATGATTTTAA TTAGCTCTAA CATTAACATC CTGTGCGGAT GTGCAAAACA 33000
ACGCAATTCA ATCTCGCTTA GCTCAGTGCA GTGGTACAAA CACATACCA CAATGGTGGT 33060
TAACTAGCCA TAATATTAGG TGCTCCAGCC AAAACTCATC TGCGGATAA TCGTGGCCGC 33120
GTGACCCATCA TACAGAGATC TTAATGTAAT CAAATGGGCC CCCCTCCCAA ACAACTTGGC 33180
CACATACATA ATCTCTCTGG GCTATGCGAT GTTCACAATC TCTCTGTAAC ATGAGACGGC 33240
CTGTTAAATC ATACAGCCCC TAAATAACCTT CCGGAACCAA ATAGCCAGCA CTGCTCCCCC 33300
ACGCAATACAT TGAAGAGAAC CCGGCTGTGTT ACAGTGCAAA TGAAGAACC ATCTCTCTGG 33360
CCCATGGGATC ATCTGGAAT GAAATATATC TATAGTGCCA CAACACAAAC ATAAATGGCAT 33420
GCATCTTTTC ATACCCCTTA ACTCTCTGGG GTTGAACAC ATATCCCGGA AATGGGGAAG 33480
CTCTTGCAAA ACACTAAACG TGCGAAACAAG AGAAAGGGGC AGAACAAATC TTACACTGTTG 33540
CAGGTTCAGG GTATTCAAT CTTGGAACGG TGGATGCTCT TACGTCATAG AGCTCTCTGG 33600
TTCTTTTCTC TACAGGCGGG GTAAGGGGCC CCTCAATAGA GGCTGATAGA TGATGCTGGG 33660
AGTCTCTGGG CATGAGCTGG ATCTGGCACG CGACCTCGTG GTAATGGGAC TGCTTCTCTGA 33720
CATTCTGTA TTTTGCAATG CAAAACCTAG CCTTAGCACA ACAAACATCT CTCTGCTTTC 33780
TATCCCTGGC CTAACGGGTC TCTGTTGAT ATTTGAAGTA CAGTACATTCC GCTAGATGTTG 33840
TCGAAGTGGC TCCGGCTCTCA GTTTTATAGA AAACCTCCATC ATGTCTGACT GCTCTGATTA 33900
AACTATTTAC TGTGAAATGG GCAATACCCA ACCATGCAAT ACAATTAGCT TGAATTAAAA 33960
TCAAGGAGGG GGGAGGAGAA CATGGAAGAA CCATAATTAA TTTTTATTTT CAGACAGATT 34020
CGCAGTATTG CTAATGGAAG ATCAGCAAGA GTGCACCTTG CGCCCCCCTT GCTGTTAGATG 34080
AAAAAATACCG CTAAGTGCAA CACGATGCAG TTCTCAAGAT GCTCAGATGT GGTCTCAAGC 34140
AAAGCCTCCTA CGCCCATCAT CAAAACAAA AAGAAGCACA AAGAGGGGAC ATGTTCTAAT 34200
TCTTGAATAC TCAATATTCA TTGTCTTACA ATTTCCGAGAT ATTTTTATCC TCAATGCGCTT 34260
TGAAATTAAC GTGTCTCCTC TTCTGTTAAA TCCAATCTCAG AATGAAAAAA CAGCTCCTGG 34320
AGGGGACCTT CACACCACCA CTTAAGACAC ACCCTCATAA TGACAAAAAT TCTGTTCTCT 34380
GTGTCACCTG CAGCAATTTG AGAAAGGGCA CATCAAGGAG CATGGCATTG TCTCTAAGCT 34440
CTCTCTAAAG TTCAGAGTTT AAAAATCCTCT TCAAATCTACG GCAAACACTG TGGCCCATAG 34500
GTGCCGCCAGG AATAAGGAGG GGGGAGCTTA CTTGTCAGAA CAAACGGGCA CCGCCTCAAT 34560
GGATACGCAC AAAAGTGGGG TTCAATAAGG CATACTGAGA ACCTCCAGTG ATATCATCCA 34620
GAGTGTGCTGGG AACATAATCA GCGCAGATTTT CTGTATTTAA ATTAATAAAA GAAAATTCTG 34680
70

CCAGATGAAC ATTTAAAA TTCTGGAATAC AGATGCAATA AGTTACC CGGGCTGCCA 34740
ACATGTGTTAG TAGATTAGT CGTAAAAAAA AGACGACAAAG AGTTATTACA TCATGCTAGC 34800
CTGGCGGAACG GATGGATAAA TCACTCTCTCT CAACACCGG CAGGCTCAAG GGTCTCCACC 34860
ACGCACCTCG TAAAAACTCTG CAGATGATT AAAAGCATTG ACCGAAAAG AGCTTTGATG 34920
AGCAGCAAAT ATATTGTGCG ATGAGACATA CAATCCGAA GGTGTATGAT CATTAAAGA 34980
AAAAAAGCCT CCAATATAGC ATCCTGGGAC AATTATGCTC AATCTCAAAT GCAGCAGACC 35040
GACACCTCTG GGATGCAAGA TAAAATCCAC AGGACGATAA AAAATGTAAT TATTTCCCTC 35100
TTGCCACAGCG AGCCCTAGTC CCGCCCCCTAC CAAATACCA TACAAAACCT CAGCCATAGC 35160
TTACCGACA AACTAGGAGG AGCAGACAGG AGAAGATAA ACTGAAGTCC GCCTGTGGCG 35220
AAATAAGATG CAACCTGATAC ACTGACTTAA TCCGATAAG TCTAAAAATA CCGGCAAAAA 35280
CCAGACACGG CCCAGAAAAT GCTGCATCCCG CGAGAAAATT TCACCTTCCGC ATTTATTCCC 35340
GGAAAGACATG CACTTCTCCT TTCCACCGAA TGCTACTTCC CGTGAATTCT TGAACGTCAC 35400
CCTCCCAGCC CGCCCTTCAC GTCCCCGTGC CCCACAGCCA ATCACCTTTC ACCCTCCCCA 35460
AATTCACGGC CCTCATTGTC ATATAACAC GCAACAAAAA TTGAAGTTAT ATTATTGATG 35520
ATGG 35524

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36519 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCACTTCAAA TAATATACCT CAACAATTTTT GTGCGCGTFA ATATGCAAAAT GAGCGGTGTTG 60
AATTTGGGGA GGAAGGCGGG TGATTTGGTG AGGGATGAGC GACCGTATGG GGCAGGGCGGA 120
GTGACCGTTTT GATGACGTGG TTGGCGGAGA GAGGCCAGTTT GCAAGTTCTC GTGGGAAAG 180
TGAGCTCAAAG CGAGTTGTTG TTGAACAAGC GAAATACGTA ATTTTTGCCG GCTCTCTGAC 240
AGGAATAATGAG GTTTTCTGG GCGGATGTCA GAAAAACGG GCATTATGG GCAGAAAACCT 300
GAATGAGGAAT GTGAAAATCT GATGAAATTCC GCCTTTATGG CAGGAGGAG TATTTCGCGA 360
GGCGCGAGTA GACTTGGACC GATTACGTGG GGGTTTCTAG TACCCTGGTT TACCTACCAA 420
TTCCCGCTTA GCCTTGCAAA GTCCGCGTTG TTTACGTTAG TGCAGCTGCA TGCCGAGGTT 480
ATTTAAACCT GCCTCTCCAA GTGAAGGGCC GACTTTTGAG TGGCGAGGAG AAGAGTTTTC 540
TCCTCAGCGGC CGCGGATCAG ATCTACACTT TGAAAGATGA GCCACCTGAG AGAACCTGCCG 600
GATGAGAAAA TCATCATCAG TTCCGGGAA GAGATTTCTGG AACTGTTGCTT AAATGCCATG 660
ATGGGGCGAG ACCCTCCGGA GCCCCCCACC CCAATTGAGA CACCTTGCCT TCAAGGATTTTG 720
TATGATCTTG AGGTTGGATG TGCAGGGAGG GATCCCAATG AGGGGCGGTT AACATGATTTT 780
TTAGGGCGATG CCGGCTGCTG AGCTGGCGAG GAGGGTTCCGA GCTCTAGTCT AAGCAGGCGAC 840
TTTTCATAGC ATATCCCTATG ACCCGGCAAG GGTGAAGAAA AGATCCCGGA GCTTAAAGGG 900
GAAAGAGTGG ACTTGGCCTG CTATGAGGAA TGGTTGCCCG CGAGGAGTGA TGAGGACGAG 960
CAGGGCGGTA AGAAAGCAGC GAGCCCGGAG GTGCAAGCGG CCGCGGAGAG CTGTCGCGTG 1020
GACTGCCCGT CTGTCCCGGA AAGCGGCTGT AAGCTCTTTG AATTTCATCG CATGAATCAT 1080
GGAGATAAAG CTGGGTTGAG TGCCATCTTG CATATGAGAG CTGCCACCAA TTGTGGTTAC 1140
AGTAAGGTGTG ATTAAGTGAAG ACTTGTAGAG GAGGCGAGCA GCCGGCGTAC TGCGCGCTGA 1200
CTGGTTTATT TAGTTATATA TGTTTTTATT ATAGGTCCCC TCTCTGACGC AGATGATGAG 1260
ACCCCCACCTA CAAAGTCACCG TTGTCGACCC CCAAGAAATG GCACATCTCC ACCTGAAGAT 1320
ATTGGTAGAC CAGTTTGCTT TAGAGCCTAT GGGAGGAGAG CAGGTCGAGG ATGTTGGGAT 1380
GACTTTGCTAC AGGGTTGGGT TGAACCTTTG GACTTTGCTT GCGCGAAACG CCCCGGCCAC 1440
TATGTTGCGAC ACATGTGGTT TTACTCTGAG TGTATCGACT ATTTATAGGG TGTTAGTTGC 1500
AAATAAAATA ATGTGTGACT GGAGGTGTGG GATTATGACT CAGGGTTGGA GACTGTGAGT 1560
ATATAAGCCAG GTGCTAGACGT TTGTTGGTTAG ATCAGAGCAG GATGAGATT TGGCGGTCT 1620
TGAGAGACTT TCACAGACCT AGACAGCTCC TAGAGAAGCC CTGGAAGGGA GTCCTTTACC 1680
TGTAAGAGTT ACAGTTGGTT GCCGACCTATG CTAGGCTATG CTACAGGCGCC AACAGGATT 1740
ATATTTGAACA ATTTAGGATT AAGTTGAGAT AGTGGTTGCT TCTTTTTGAC GCTCTTTTAC 1800
TGCCCCAACG GTCTCCACTT AACCAGAGGA TTTGAGAGCC CTTGATTTTT AACTCTCTTG 1860
GCAGAGACAC TCAGACAGTT GGCTTTTTTG TTTTTTTCT TGACAAATGG AGTCAAGAAA 1920
CCCATTTCAG CAGGGATAC CAGGCTGATT TCTTAGCAGT AGGCTGTTGG AGAACCATGGA 1980
AGTGCCACCG CCTGAATCGA ATCTCCGGGT ACTGGCGGGT ACAGGGCGTA GAACTCTGA 2040
GGATCCCGGA TTCCACAGGA AAGCCCGGGA CAGGGACAGG TTGCGCAGCG AGGAGCAGG 2100
AGGGAGATCG AGAAGAAGAC CGGAGAGCGC GCCTGGGCCCT GCGGCGAGAG GGGAGGAGT 2160
AGCTGACCTG TTCTCTGACG TGCGCGGCTG GCCTGACTAGG TCTTTGTAGG GTGGGAGAG 2220
GGGGTTAAAG CGGGAGAGCC ATGAGACAGT TAATCAGACG ACTGAACTGAG CTGTTGGCTC 2280
GATGCTCGC AAGCCCGAG AAACAGGTGTT GTGGCAGTAGG GTGCACGCTG GCGCCCAGA 2340
TGGATGTTGG GTGATGACTG AGAGGGTTTT TCTAGAAGAC GTCAGACTCT GTTGGTTGAG 2400
GCCTGAGGAT GATGGGGAGG TAGCCATCAG GAATTATGCC GAAGCTGGCTG TGAGGCGAGA 2460
CAAGAAGATG AAGTAACTCA AGCTGATAAA CATCAGAAAT GCCTGCTACA TCTCAGGAAA 2520
TTGGCGACGC CTTTGTGCC GCCCAGTTTT TCCATGCACT CATCCATGAT GATGCGGATG 4500
GCCCCGTGGG CGCAGGCGCTG GCCAACAGAC TGTTGGGGGT CGGACACATC ATAGTGTGGG 4560
TCTCTGGTGA GTGTACTATA GCCCATTTTA ATGAATTGG TGGCAGAGGT GCGGGACTGG 4620
GGGACAAAGG TACCTCTGAA CCCGGGGGCC TAGTTCCTCC CACAGATCTG CATCTCCAG 4680
GCTTTGAGCT CGGAGGGGGG GATCATGTCG ACCTGCGGGG CGATAAAAGA CACGGTTTCC 4740
GGGGGCAGGG AGATGAGCTG GGCCCAGAGGC AGAGTCGCGG GCAGCTGGGA CTTGCCGCGA 4800
CGGTGGGGC CGTATAGTCG CCCAGTGACC GGGTCGGATT GGATGTTGAG GGAGAGACAG 4860
CCTCCGTCTT CCCGGAGGAG GGGGGGACCC GCTGTTCACG TCTCCGCGCAG GTGATGTTCC 4920
TCCGCCACCA GTCGCCGGTT CAGGGCCTGC CCCCCCCGAG ATAGGAACCT CTGGAGCCAG 4980
GGCAAGTTTT TACGCGGCTT GAGTCCGGCG GGGATGGGCA TTGGGAGAGG GGTGTGTGCC 5040
AGAGATTCGA GCCGCTCCCA GAGCTCGGTT ATGTTGTCTA CCGCAGTCTG ATCCAGCAGA 5100
CTCCTCTGTT TCCGCGGTG CGAGGGCTGC GGGAGTAAGG CACCCAGAGA TGGGCGCAGA 5160
GCCGAGCCAG GTCCGGTCTC TTTCAGGTTG GAGCCGTGCG CTGGAGGTTG TGGTCGTCCG 5220
CGGTGAAGGG GTGCGGCAGG GTGCTGGGGG TGGCAAGGTT GCCGTCCAGG CTCCAGCGGC 5280
TGTCGAAAAG CGCTCGCCCA TCAGCCGGCT GGGTCGCGCA CAGTGAAGCA TTGACCATGA 5340
GTTTCGATTT GAGCCGCTCG GCCGGGTGCC CTGTGGGGCG GAGTTACCT TTGAAGTCTT 5400
GCCGCCAGGC GGGACAGAGG AGGGACTTGA GGGCGTAGAG CTTGGGGCG AGGAAGACGG 5460
ACTCGGGGGC GTAGCCGCTC GGGCGCGAGT GGGCGCAGAC GTGCCTGCGC TCCAGCGACC 5520
AGTGAAGGTC GGCTTGTGGC GGGTCAAAAA CCAATCTCCC GCCGTTTTCT TTGATGCGTT 5580
TCTTACTTTT GTCTCCCATG AGCTCGGTGC CCCCCTGGGT GACAAAGAGG CTGTCGCTGT 5640
CCCCGTAGAC GCACTTTTATG GGCGCGTCCT CGAGCGGTGT GCCGGCGGTT CCCTCCTGTA 5700
GGAACCGCCG CCACTCCGGG AGCAGACGCG GGCTCGCGCG CAGCAGACAA GAGGCCAGGT 5760
GGGACGGGTA GGGTCGTTTG GCCACAGGCC GGTCCACCTT TTTCCAGGTA TGCAACACCA 5820
TGTCCTCCCTC GCAGGACATG AGGAAGGTC TTGGGTTCTA AGTGTAGGCC AGTGACCCGG 5880
GGTCCCGGC CGGGGGGTTA TAAAGGGGTT CGGCTGGCTG CTGCTCCCTA CTGCCTCCCG 5940
GATCGCTGTC CAGGAGCGTG AGCTTGTTG GTAGTTATTG CTCCTCGAGG GGGGCTGATGA 6000
CCTCCGCACT CAGGGTGCTA GTTTCTAGAA AGCAGGAGGA TTGATATAG ACQGTGCGGG 6060
CGGAGATGCC TTTCaAGAAC CCGCCTCCTG ACTGAGTCAA AAAGCAGATC TTTTGGTGGT 6120
CGAGCCTTGG GTCCAGAGAG CGCTAGAAGG GTGGGGAGAG GAGCTTTGAG ATGGAGCAGA 6180
TGCTCTGGTT TTTTTCTCTG TCCGCGCGCT CTTGGCGCGC GATGTTGAGC TGCACTCAGT 6240
CGCCGCACCC GCACTCCCCG TCCGGAGAGA CGGTGCCGCG CTGTGCGGGC ACQGATCTGA 6300
CCTGCCACCC CGGATTTGC AGGGTGATGA GTGTCGCACT GTGGGCGCACC TCCCGCGCGA 6360
GGGGCTCATTT AGTCACAGG AGGCCCTCAGC CCGTGCCGGA GCAGAAGGGG GGCAGGGGCTT 6420
CCACATGAC CTGTCGGGGA GGGTCGCGCAT GAGAGTGAA GATGCCGGGC AGGGGCTCGG 6480
GGTCAAAGATTA GTGATGGGAA GTGGAGGACAT GCTCCAGGC CGACTTCCCAT TCCGGCAGCGG 6540
CCACGCGCCG CTCTGAGGGA CTAGGGGGCG TGCAGCCAGG CATGGAAGTG GTAAGCGCGG 6600
AGGCGTACAT GCCGCAAGTG TCGTAGAGCT AGAGGGGCTC TCGAAGGATG CCGATGTAGA 6660
TGGGCTAGCA GCCCCCCCG CGAGTGCTGG CGAGACTGTA ATCATACAGC TCGTGCCAGG 6720
GGCGGAGAGG CGCCGCCCCG AGGGTGGTGCC GACTGCGGCT TTGGGCGCGCG TAGGAGATCT 6780
GCCGGGAAAT GCCATCGGAG TGGAGAGGAG TGGTGCGCCT TTGGAGAATG TTGAAGTTGGG 6840
CGTGGGGCGC TCGACGCGAG TCCTGCGATGA ATGTGGCCGT AGATCTTGCC AGCTTTGGCGA 6900
CGAGCTCGGC GTGAGCTAGG AGCTCAGCGG GGAGTGATGC TGGCTCGGCTT TGATTAGGT 6960
CATACCTGAG CTGTCCTCCTC TGGTCCACAG GTCTGCGGTT GGAGAAGGAC TCTTCCGCGG 7020
CCTCCCCGATA CTCTCAGGAG GGGAAACCCTG CCTGATCTGC ACGTAAAGAG CTTACCATGT 7080
AGAACTCTGT GAAGGCCTTG TAGGCGGCCAG AGCCCTTCTC CACGGGGAGG GCTGAGGCCT 7140
GGCCGGGGCT CGGCGGGGAG GTGGTCGTCGA GGCGAAAGGT GTCCCTGACC ATGACCTTGA 7200
GGAAGTGTGTG GTTAGAAGTG ATATCGTGCC AGGCCCTCCT GTTCCACAGC TGGAAAGTCG 7260
TGGCGCTTC TTAGGGCGGTC TGGGCAAAG CGAAGTAAAC ATCTGGTGAAG AGGATCTCGG 7320
CCGGCGGGGG CATTTAAGTG CGAGTGATGC GGAAGATAGG GGCCATCTCG GCCGGCTTTG 7380
TGAGTGACCT CGGCCGCGGAC ACAGATCGGT CGAGACGGTTG TAGTTGTCGG CCCACGATGT 7440
AGAGTTTCCAG GAAATCGGGA CGCCCTCTGA CGTTGGGCGG TTTCTTGAGC TCTCTCTAGG 7500
TGAGCTCGTC GGGTGCGCTG AGCGCGGTCT TCTCGAGCGC ACATGGCGGC ATGGGGGGGT 7560
TGCCGGCGAG GAAGGAATGC CAGACTCAA CGCCAGGGCC GCTTTTGCGA CGCTCCCGGT 7620
ACTGACGGAA CTGCTGCGGG ACGGCCATTT TTTCTGGGGG TAGCGACTAG AAGTGCGGCG 7680
GGTCCCCGGTG CGAGCATGCC CATTGAGGCT GAGGAGGCGG ATCGAGGCGA AGCTGCAGGA 7740
GCCGCTCGTC CCCGGAGAGT TTCTAGCACG GCTAGAAGGG GAGAGGCTCG TTGGCGAAGG 7800
ACCCCATCCA GGTCTTTTTG TCCACACTGT AGTGGAGGAA GACCGTTTGG GTGCGAGAGAT 7860
CGGACGGGAT GGGAAAGAAC TGGATCTCGGT GCAACCATTT GAGGAAATGG CTGTTGTAGT 7920
GATGGAAGTA GAATCGGGAA GCGCGGCGGG AACACTCGTG TTGTTGTTTA TACAGGCAGC 7980
CAGAGTCGTC GCAAGCTCGC AGCGGATGCA CGTGGCTCAC GAGCTGTACC TGGATTTTCTT 8040
TGAGCAGGAA TTTGATGGGA AGTGGAGTGC CGTCGCGGCT CACGCTGTGCC TGACTACCGT 8100
CTGTGGGCTG CGGCTGCGCC TCTTCTGCTC GATGAGGTTG CTGACTCGAC CGCCCGCGCG 8160
GGAGGCGGAT CGAGACACTGG GGCGCGCGGG GTCTGGAGACT GAGGACGAGG GGCGGCGCGG 8220
CGGACGCTGTC CAGGCTTCCAG AGACGTGCGG GAGTGACCAG GGGCGGCGCG 8280
GTTTGACTTG CAGGAGTTTT TCCAGGGCCC GCGGGAGGTC CAGATGCTTC TTGATCTCCA 8340
CCGGCCATT GGGCCGGCGG TCGATGCTTG CCAGGGCTCCG GCCGGGCTGG GGGTGGACCA 8400
CGTGCCCCGG TTTCTTCTTG GCGGGCTGGG GCGAGGGGCG GGGTGGTTCT TCAGTGTTTA 8460
GAAGGCCGGG CGAGGCAGCG CGCCGGCGGC CAGGGGCGGG TCGGGGCGCC GAGGGAGGGG 8520
CCGCGGGGCG ACGTCGCCGC CGCGCGGCAG TAGTGTCTGG TACTGGCGGC GAGAAGAGCT 8580
GCCGTGAGCG ACGCAGCGAC GGGTGGAGTC CTGGATCTGA CGCTCTGTGG TGAAGGGCCAC 8640
GGGACCCGTGC AGTTTGAAAC TGAAAGAGAG TTGCAGCAAGA TCAATCTCGG TATGTTGAC 8700
GCCGGCTTGC CGCGAGATCT CTGTCAGCTGC GCCCGAGGTC TCCTGCTGATG CGATCTCGGT 8760
CATGAACTGC TGATCTCTCT CCTCTTGAGA GTCTGGCCGC GCGGGGCGCT CACCGGTGGC 8820
CCGCGACTGC TTGGAGATGC GGGCCCATGA CTGGAGAAGA GGGTGTACAC CGGGCCTGTT 8880
CCAGAGCGGG CGTGTAGACCA CGACCGCTCT GGGATCGCGG GGGCGCATGA CACCTGGGC 8940
GAGGTTGAGC TCCACGTTGC GCGTGAAAAC CAGGGTAGTG CAGAGGCCCT GGTAGAGGTA 9000
GTTGAGCTTG GTGCCGATGT GCTCGGATGC GAAGAAATAC ATGATCCAGG GGGGGAGGGG 9060
CATCTCGCTG AGCTCGCCCA GGGCCTCCTA AGCTGTCAAT GCCTGTTAAA AGTCCACGGC 9120
GAGATTCGAA AAATGGGAGT TGCGGGCGGA GACGTCACAC TCTCTCCTCA GAAGACCGAT 9180
GAGCTGCAGC ATGGGGGCCC GCACCTCGGC CTGGAAGCCC GCCGGGAGGT CCTCCACTTC 9240
CTCTCTCTCC TCCCATCCTCA ACATCTCTTC TACTTCCTCC TGAGGCGCAG GTGGGGGCGG 9300
GGGAGGGGGC CGTGGTCGCC GCAGGGCAGC GGGCAGACCG TCGATGACGC GCTCGATGGT 9360
CTGCGGCCGG CGCGTGCGCA TGCTCTGGTG GACGCGGCAG CGGTCTCCGC GGGGCCGCAG 9420
CGTGAAAGCG CGGCCCGCGA TCTCGATGGT GGGGGGGGG CCCCCGGTGG CAGGGAGGAG 9480
GGGCGCTGACG ATGCACTTGA TCAATTGCCG GTAGGGACGT CGCGGGCAAG ACCTGAGCGT 9540
CTCGAGATCC ACGGTGACTG AAAAAAGCGT AAGGAAGAGT TGCGAGCGAT CGCAGTCCCA 9600
AGGTAGGTGT AGCAGGTTTT CTTCTGGCGG GTCAATGTTG TGGGAGCGG GGGGCGGCGAT 9660
GCTGCTGGTG ATGAAGTGGG AATAGGCGGT TCTGAGACGG CGATGGTGG CAGGAAGCAC 9720
CAGTGTTTGT GGGCCCGGTT CTTGGATGCG CAGACCGGCG CCATGGCTCC ACCGGCGGTG 9780
CGTCACCTTG GCCAGGTCTCT TGATAGTTGC CTCACTGACGG GCCTCAGCG GACCTCCTCC 9840
CTGCGGCCGG CGGGCGTGCA TGCGCGTATG CCCGAAGCCG CGCTGGCGCT GACAGGCGAG 9900
CAGGTCGGGC ACGACGGCCT CGGGAGAGAT GCTGTGCTGG ATCGTGCTGGA GGGTGGTCTG 9960
GAAGTCCATCA AAGTGGAGCA AGCGGTGTTA GGGTCCGTTG TGATGATTGT AGGAGGCATT 10020
GCCATGACGC GACAGTGGG CAGGTGCTGT GCCGGACGCC AGAGCTGCTG GTTACTGTGA 10080
GGCGAGATGG GCCGGCGGCG GAGAAGATGA TGCGTGACAG GTGCGACCAA GGTACTGTTA 10140
GGCGATGAGG AAGTGCGGCG GCCGCTGGGC TAGAGGCAGC CATCGTCTGC TGCCGGGGGC 10200
GCCGGGCGCG AGGTCTCTGA GCATGGTGCG GTGGTAGCCG TAGATGTACC TGGACATCCA 10260
GGTGATGCCG GCGCGGTTGG TGGAGGGCGC GGCGAACCTGG CCGACGCGGT TCCGATAGTT 10320
GCGCAGGCGG AGGAACTGAG TCTAGGCTGG CAGGTCTGGG CCGTGATGGC GCCGACGATC 10380
GGTGATGCTC TATACGCGCA AAAACGGAAG CGGTAGGCGG CTCGACTCCG TGGGCCTGAG 10440
GCTAACGGAA CGGGTTGGGC TGGCGCTTGA CCCGCGTTGC AATCTGAAT CAGGCTGAG 10500
CCGCAGCTAA CTGTGTATTG GACTCCTGGG CTCGACCCAA GCTTGCACCA ACCCTCCAGG 10560
ATACGGAGGC GGTCGTTTCT GCAACTTTTT TTTGGAGGCG GGTGATGGCT AGTACGGGCCG 10620
GAAACGGGAC GACCGGAGTC ATGGCCTGCG GTCTCGCTGA GAAGAATCGC CAGGGTGCCG 10680
TTGGCTGATG CCAGCTTGCC AGGGCCGGCG GATTCCGGCT CTAACGAGGG CTTGGCTGCC 10740
CCGCTGTATT CAAGACCCCA TAGCCAGCGG ACTTCTCCAG TTACGCGGAC AGGCCCTCTT 10800
TTGTTTTGTT GTGGTTGCCC AGATGCAATCC GTACTGCGG CAGATGGCAG CCCACCCACC 10860
TCCACGGGCA CACAGGCCCT CTCCAGACGC GCAGCTTTCTG CCCCGCGGCC AGACGCAACT 10920
TCCAGGCCAG ACCGCGCGGC CCAGCGTAGG CGGGCTGAG CCAGCTTTAG ATCAGACAGT 10980
GGCTCTGGAA GAGGGCGAGG GCTGGCGCCG CTTGGGGGGC TCTGGCGCCG AGCGGCACCC 11040
GGCGGTGCCG ATGAAAGGG AGCTCGCGGA GCGCTAGTGG CCCAACCCGAG ACCTGTTGCA 11100
AGACAGGAGC GGCAGGAGAC CCGAGAGGAT CGGCGGGGCC CGGTTCACCG CGGGCGGGGA 11160
GCTCGGCGGC GGCCTGGACCC GAAGAGGAGT GCCTAGGGAC GAGGATTTCG AGGCGAGCAG 11220
GCTGAGCGGG ATCGAGGCCC CGCGCGCGCC CGTGGGCGCG GCCAACTTCG TCACGCCCCA 11280
CGAGCACAGC GTGAAAGAGG AGAGCAACTT CCAAAATCCT TCTCAACACCC AGCTGGCGCAC 11340
CCTGATCGCG GCGAGAGGAG TCACCGTTGG CCGTGATGCAC CTGTGGGACC TTCGAGGCCA 11400
CATCGTGCGA AACCCACCCA GCAAGCCGCC GACGGCGAGC CGTGTCTTGG TGGTGCAAGC 11460
TAGCTGGAGC AACGAGGCTT TCAGGGAGGC GCTGCTGACT ATCAGCGAGC CGAGGGCCCG 11520
CTGGCTCTGT GACCTGCTGA ACATTTCTCA GAGCATCTTG GTGGAGCCAG GCAGGGCTGC 11580
GCTGGCAAGG AAGCTGGCGG CCAATCAACTT CTGGGTGCTG AGTGTGGCGA AGTACTACCG 11640
TAGGAAGATC TACAGACCC CGCTACGTGG CATAAGCAGA GAGTGGAAGA TCGACGTTTT 11700
TTACATGGCC ATGACCCCTG AAGTGTGGAC TGCTAGGGAC GATCTGGGGG TCTACCGCAA 11760
CGCAAGAGTG CACGGTGCGG TGAAGGGCGC CAGGCGGCGC GAGTTGACG ACCAGGACCT 11820
GATGGCTAGT CTGGCAGGGC CCCTGACCGG GGCGGGGACC GGAGGGGAGA GCTACTTTGA 11880
CATGGGCCGG GACCTGACTG GCCAGCCCGG CCGGCGGGCC TTGGAGGCGG CGCAGGGACC 11940
CTACGGCTAA GAGGGTGGCA ATGAGGCTGA CGAGGGAGGC GACTACCTGG AAGACTCATG 12000
GGCGAGGCTT ATTTTCTGCTA GATGCAACAA CAACAGCCAC CTCTGTGACTC CGCGATTGGG 12060
GCGGCGTGGC AGAGCCAGCC GTGCCGCAATT AACTCTCTCG ACGATGGGAC CCAGGCCCCA 12120
<table>
<thead>
<tr>
<th>DNA Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAACGCATCA TGGGCGTGAC GACCCGCAAC CCAGGAAGCT TTAGACAGCA GCCCCAGGCC</td>
<td>12180</td>
</tr>
<tr>
<td>AACCAGCTCT CGGCCATCCT GGAGGGGCTG GTGCCCTGCC GCTCCACCCC CAGCAGAGCG</td>
<td>12240</td>
</tr>
<tr>
<td>AAGGTCCCTGG CCATCGTGAAC CGGCGCTGAG GAGAAACAGG CCAATCCCGG CGACAGGCGC</td>
<td>12300</td>
</tr>
<tr>
<td>GCCCTGGGTG ACAACCGCGT GCTGGAGGCG GTGGCCCGCTT ACAACAGCAC CAACGTTCAG</td>
<td>12360</td>
</tr>
<tr>
<td>ACCAACCCTGG ACCCGCATGGT GACGAGACGTT CGCAGGGCCG TGGCCCAAGG CGACGGTTTC</td>
<td>12420</td>
</tr>
<tr>
<td>CAGCAGGGT CCAACCTGGG ATCCATCGGT GCGCTGAACG CTTCTCCTCG CACCCAGCCC</td>
<td>12480</td>
</tr>
<tr>
<td>GCCCAAGCTC CCAGGGGCA CGAGAGGCTA ACCAATCTCA TACGCGCCCTT GCGCTGATG</td>
<td>12540</td>
</tr>
<tr>
<td>GTGACCCAGG TGCCCCAGAG CGAGGTTACG AGTCCGCCAC CGGACACTCT TTCCAGGACC</td>
<td>12600</td>
</tr>
<tr>
<td>AGTCGCCAGG GCTTGGCAGAC GCTGAACTTG AGCCAGGGCT TGAAAGACCTTG GCAAGGCTTG</td>
<td>12660</td>
</tr>
<tr>
<td>TGGGGCGTGC AGGCCCCGGT GGGGAGCGGC GCAGCCTGGG TGAGCTGCTC GACGGCCAGAC</td>
<td>12720</td>
</tr>
<tr>
<td>TGGGCGCCCTG TGCTGCTGCT GTGGGCCCCC TTTACGGAAC GCGGACAGAT CAACCCGCCC</td>
<td>12780</td>
</tr>
<tr>
<td>TGGTACCTGG GTAACCTGTA TAACCTGTAC CGAGAAGGCC TGGGCAAAGG GACGTTGGAC</td>
<td>12840</td>
</tr>
<tr>
<td>GAGCGGACCT ACCAGGAGAT CACCCACGCG AGGCCGGCCGC TGGGCAAGAG CGACGGGCGG</td>
<td>12900</td>
</tr>
<tr>
<td>AACCTGGGAG CCAACCCGGA CTTTTTGTCT ACCAAGCCTG CGCAGAAGAT CCGGCCGACTA</td>
<td>12960</td>
</tr>
<tr>
<td>TACGCCGCTCA GCACCGGAGG GGGGAGCATC CTGGTCCACC TGCCGAAAAG CTGGGGCCTG</td>
<td>13020</td>
</tr>
<tr>
<td>TTCCTGATGC AGGAGGGGGC CACCCCCGAC GGCGGCCCTGC ACATGAGCAG GCCGAACATG</td>
<td>13080</td>
</tr>
<tr>
<td>GACCCACGCA TGCTAGCCCAT CAACCGCCGC TTCATCAATA AACTGATGGA CTACTTGAT</td>
<td>13140</td>
</tr>
<tr>
<td>CGGGCGGGCG CCATGAACTC TGACTTATTC ACCAAGCGCA TCTCTGAATCC CCACTGGGCTC</td>
<td>13200</td>
</tr>
<tr>
<td>CGGGCGGGCG GTTTCTACAC GGGGAGTAAC GACATCGCCC ACCCAATGCA CGGTTCTGTC</td>
<td>13260</td>
</tr>
<tr>
<td>TGGGACGTAG TGGAAGACAG CTTGGTACCC CCGCGGCGCG GTGCTAACGA GCGCCCCTTG</td>
<td>13320</td>
</tr>
<tr>
<td>TGAGAAGAGG AAGGGCAAGC CGAGGGCCGCG TCTCCGCCGC TGCGGGCTGG CGGAGGCTGC</td>
<td>13380</td>
</tr>
<tr>
<td>GCGCGGGGGG TGCCCGAGGC GGCCAGGCTC TTCCGGAGCT TGGGCTTCTC TCTGGACAGT</td>
<td>13440</td>
</tr>
<tr>
<td>ATCCAGCAGA GGAGCTGGG CAGATGACG AGCCCGCGCTG TGCTGGGCGA AGAGAGAGATC</td>
<td>13500</td>
</tr>
<tr>
<td>TTGAATTGCT CCCTGTTAGA ACCCGAGGCG GAGAAGAAGT TCCCCATATT CGGGGATAGAA</td>
<td>13560</td>
</tr>
<tr>
<td>AGCGCTTGGT ACAAGATGAG CCGCTCGGGA AGCTGAGCCC AGCAGATCCC GAGCGGACCAG</td>
<td>13620</td>
</tr>
<tr>
<td>CGGGCGTCCG AGGGGGGCGG GACGGGGGCG ACGCCCGCGC GAAAGCCCGC GTAAACCGGG TGCGCAGCAG</td>
<td>13680</td>
</tr>
<tr>
<td>AGGCACCGGG GACAGATGTT GGAGATGACG GACTCCCGCC ACCAAGCAGC CGTGGTTGAC</td>
<td>13740</td>
</tr>
<tr>
<td>TGGGAGTTAGA TGTTGACATG CAGATGATGAC GGGGGCTGG GCGGGCGCTT CTGTGCTCCTT</td>
<td>13800</td>
</tr>
<tr>
<td>GAGAAACCGA AAATAAATAG TACTACACCA GCGCGCGGAC ACCAAGCTGC GTGCTGTTTCT</td>
<td>13860</td>
</tr>
<tr>
<td>TCTCTGTTGG TGTTGATCTC AGTATGATGAA GGCGTGACTA CCCGGAGGCT CTTCTCTCTT</td>
<td>13920</td>
</tr>
<tr>
<td>CGTCCAGAGG GTGGATGGCAG CGCGGATGGA CGCGCGCGCG GCATCGAGCG CCCGGCTGAGG</td>
<td>13980</td>
</tr>
<tr>
<td>CTCTTACGTT GGCCCCCGGG AACGCTGGCGT CTACGGAGGG AGCGGACAGC ATTCTGTACT</td>
<td>14040</td>
</tr>
</tbody>
</table>
CGGAGCTGGC ACCCTTGTAC GATACCACCC GGTGTGACCT GGTGGACAAAC AGTGGGCGCG
ACATCGCCTC GCTGAACTAC CAGAAAGACC ACAGCAACTT CCTGACCACC GTGGTGCAGA
ACAATGACTT CACCCCCAGG GAGGCCAGCA CCCAGACCAT CAACCTTGAC GACGGCTGCG
GGTGGGGCGG CGACGTTAAA ACCATCATGC ACACCAACAT GCCCAACGTT AACGAGTCTCA
TGACAGCAAA CAAGTCGAAAG GGGCAGGTGA TGGTCTCCCG CAAGACCCCA AATGGGCTGA
CAGTGACAGA GGGAATTGAT GTAGTCCAGG ATGAGCTGAA GTATGAAATG GTGGAATTGTG
AGCTGCCCGC AAGGAAGCTTC TGCGTCGAGC TCGACATGGA CCTGATGAAAC AAGGCCATCA
TGCAAAATTT CGTGGGCGTG GGGCGCGGAG ACGGGGTGCT GGAAGAGGCAC ACGCGCGTG-
AGTTCGACAC TAGGAACTCTC AGGCTGGGCT GGGACCCGGT GACGAGCTG GTCATGCGCG
GGGTGTCAC CAAACGGGCT TTTCCATCCGG ATATTTGCTT GTCGCGCGGC TGGCGCGGT-
ACTTCACCCG GAGCGGCGCTC AGCAACCTGC TGGCCATCTG CAAAGGGGAG CACCTTCCAG
AAGGCTTCCA GATCATGCTAC GAGGATCTGG AGGGGCGCA CATCCCCGGC CTCTGTGGAT
TGACGCCTTA TGAGAAAAGC AAGGGTAGTG CAGCGAGCTGA AGCAACTGCA GCGTAGCTA
CCGCTTCTAC CAGAGCTACGG GGGATAATT TTGCAAGCGG CCGAGACATG GAGCGGCGC
AGGGCGCTGA AAGCAGAAATG AGAAGATCGA TTCCAGCGGT GGAAGGAAAGT AGCAAGAAC
GGAGCTCAAA GCTATACCCAG GACAGAGATA ACACGCCTCA CCGAGCTGTT TACCTAGCTT
ACAACATGGG CGACCCCCAGG AAGGCCTGGC CCTCCTGGAC GTGCTGTCACC ACCTCGGACG
TCACCTGGCG CGTGACCGGAA GTCTACTGGT GCGTTGGCAG CATGATGCAA GACCGGCCTA
CCTTCCGGTC CACGGTGCAA GTAGACCACT ACCGGGTGGT GGGGCGCGAG CTCCTGCGCG
TCTACTCCAA GAGCTTCTTC AAGGAGCAAG CGGCTACTGC GCAGCAGCTG GGGCCCTCTA
CCTCGGTTAC GCACGCTTTCC AACCAGCTCC CCGAGAACCA GATCCTGTCG CGCAGCGCCG
CGCCACCAT TACACCAGTC AGTAAAAAC TTCTCTGCTC CAAAGATACG GGGACCTCTC
CGCTGCGCAG CAGTACCCGG GAGATCCAGC GGGTAGAGGT TACTGAGCCC AGACGGCGGA
CCTGCCCTCA TGCTTACAAG GCCCTGGGCA TAGTCCGGCC GGGCGTCCTC TCGAGCCGCA
CCTTCTAAAT GTCCATTCTC ATCTGGCCCA GTAATACAC GGGTTGGGCG CTGGCGCGCG
CCACGAAAGT GTACGCGCAAC GTCTCCACCGA CACCTACGGC GGGTGGCGG
GGCCTTTGGGGCGGTCCTTG GGGGGGCGGT GGCGTCGCGC ACCACCGTGC
ACGAGCTGAT CGACGCTGTC GTGGCCGACG CGCGCAACTA CACCCCCGCC GGGCGCGCCG
TCTCCAGGTC GCACGCCGTGC TAGACAGCCG TGGTGGGCGG GGGCGCGCGG TACGCCGCGG
CAGAGGCGCG CGCAGGGGCGG ATCGCAGGTCGG CACCGGCGAG CACCCCGCCT ATGGCGCGCG
CGCGAGCTTT GTGCGCGAGG GCGAGGCGCA CGGGAGCGAG GGGCAATGCTC AAGGGGCGCCA
GAGCGCCGGG TCCAGGGCGC AGCGCGCGCA GGACCCGGGAG ACAGCGCGCGG ACAGCGCGCGG

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGCGGCCAT CGCCAGCATG TCCC GCCG GCCGAGGGAA CTTGC TACTGG GTGCCG CGGAC</td>
<td>16020</td>
</tr>
<tr>
<td>CGCCACCCGG TGTGCGCTGG CCGT GCGGCA CGCCGCCCGC TCGA CTJGTA GAC</td>
<td>16080</td>
</tr>
<tr>
<td>TTTGCGATGT TGTGATGCC GAGC CGCGGACGAG GTAGAATCTA CTGAGAACA ATGAGGGAAGA AGGA</td>
<td>16140</td>
</tr>
<tr>
<td>GATGCTCCAG GTCA TCAGGGCG CTGAGATCTA CGGCGGCGG GTGTGAAGG AGGAAAGAAA</td>
<td>16200</td>
</tr>
<tr>
<td>GCCCGCAAAA ATCAAGCGGG TCAAAAAGGA CAAAAGGGAA GAAAGAAGAT AAGTGGAAGCAG</td>
<td>16260</td>
</tr>
<tr>
<td>ATGGGCTTGGAG TTTGGTCGCG AGTTCCGCAC CGCGGC CGGGGAGTC GAGGGCACG</td>
<td>16320</td>
</tr>
<tr>
<td>GGTGGCAAGCG GTTGGACAGG CGGCGCAGCA CTTGCTGCTTC AGCGCGGCGG AGCGCGGGCGG</td>
<td>16380</td>
</tr>
<tr>
<td>CACGGCCTTC AAGGGCTGCT ATGACGGAAT GTACGCGG GTAGATATTC TGAGCGAGGC</td>
<td>16440</td>
</tr>
<tr>
<td>GGCCGAGGAG CTGGCGGACT TTGGTTACGG CAAGCGGAGC GTTCCCGCAG CAGAGGAGAAG</td>
<td>16500</td>
</tr>
<tr>
<td>GCCCGCTTCC ATCCGGCTGG ACCAGCGGAA CCCACAGCGG ACCGCTCAAC GC</td>
<td>16560</td>
</tr>
<tr>
<td>GCCGAGGGTG CTCGCGGGCG CGGCGCGCGC CGGGGGGTTCC AAGGGCGAGG GCAGGAAGGTCT</td>
<td>16620</td>
</tr>
<tr>
<td>GTACCCCGCC ATGACGCTGA TTGTGCACCA GCGACAGAG CTGAGAAGCG GTCTGAGAGC</td>
<td>16680</td>
</tr>
<tr>
<td>CATGAGGTTG GACCGCGAGC TCGACGCGCC GGTAAGGTG CGCCGCACTAC AGAGGTTGGC</td>
<td>16740</td>
</tr>
<tr>
<td>CCGCGGCTTG GCAGTTCGAG CATTGACGACC ACCGTTCTAG CTGAGCGGG</td>
<td>16800</td>
</tr>
<tr>
<td>GACCGAGGCTT AGTACCAAC GCACGGCACTG CACGGAGGAC ATCCGCTGCTA</td>
<td>16860</td>
</tr>
<tr>
<td>GCCATCGGCT CCTAGTGGAA GACCGCGCC CAAGTGACCG GGCGCGAGCC TGCTGAGACG</td>
<td>16920</td>
</tr>
<tr>
<td>CAACTACGGG CTGACATCTCT CCATCATCCC CAACGGGGAG TCACGCGGGCA CGCGCTTCTA</td>
<td>16980</td>
</tr>
<tr>
<td>CGCGGCGCAT ACCAGCGGCG GCCCGCGCA ACGAGCCACT GCCGCGCAGCT GC</td>
<td>17040</td>
</tr>
<tr>
<td>CGCGGCCTGA ACCCGCCGCG CGCCCGTGGT GCAGAAGGAT TACCAGCGCG CGCGCGCCAC</td>
<td>17100</td>
</tr>
<tr>
<td>TCTGACACCTG CCGCGGCGGC GCTACCGCCC GACATCGCC ATTAAACTTT TGCCAGCTTT</td>
<td>17160</td>
</tr>
<tr>
<td>TGAGATCAA TGGCCCTCAG ATGCGGCTCTT CGCTCCCTAA TTACGGGGCTA CGAGGAAG</td>
<td>17220</td>
</tr>
<tr>
<td>AAACCGCGCC GTAGAAGGCTT GCGGGAGAAG CAGATCGTGC GCGCACCCA CGGCGCGCCG</td>
<td>17280</td>
</tr>
<tr>
<td>CGCGCGCCCTA GCAAGGCGTTT GGGGAGGGCC TCTCTGGCCG CGGTATCCC CATCATCGCC</td>
<td>17340</td>
</tr>
<tr>
<td>GGCCGAGTCG GGCGGATCCC GCACATTGT TCCCTGCGCG TGGCGGCTTC TCAGCGCCAC</td>
<td>17400</td>
</tr>
<tr>
<td>TGAGACACA TGGAAACCAC CTTGGAAATTA ACCACAGGAC TCTGACGCTC CTGTTCTGCT</td>
<td>17460</td>
</tr>
<tr>
<td>GATGTGTTT CTTGACCGAG TGGAGACAT GCATTTCCTTT CTTCTGGCTC GCAGACACGG</td>
<td>17520</td>
</tr>
<tr>
<td>CAGCGCGCGCG CCTGATCGCC CCTGAGCGCA CATCAGGCGC AGCGACTGAA ACGGGCGCGG</td>
<td>17580</td>
</tr>
<tr>
<td>CTTCAATGAG AGGAGTTGCT CGAGGCGGCT TAAAGGTTGC GGGTCGCGCT TAGAATTCTA</td>
<td>17640</td>
</tr>
<tr>
<td>TGCCCGACAAG GCCTGGACAG GCACACCGAG CGCGGCGCTG AGGATTGACA TGAAAGAGCA</td>
<td>17700</td>
</tr>
<tr>
<td>GGAGGCTTCCA GCAGAGTGGT TGGTGGACTG CTCTGGCGGG GATCAACGGG TGGTGGAACCT</td>
<td>17760</td>
</tr>
<tr>
<td>GCACCGCCAG GCCCGAGGAG GCCGAGATCAA CAGCGCGCCTG GACCGGTGCG CGCCGC CGCGCG</td>
<td>17820</td>
</tr>
<tr>
<td>TCCTGGAGAG ATGCCGCGAGG TGGAGAGGAG GCTGCCTTCCCC CGGACACGG GCCGCGGGAA</td>
<td>17880</td>
</tr>
</tbody>
</table>
GGGAGCCCCC CGGATGGCGG AGGAGACGGT GCTGACGCCC ACAGGACGAGC CGCCCCCCTA 17940
CAGGGAGCGG GTGAAGCTGG GTCTGCCACC CAGGGCCGCC ATCGGGCCCC TGGGACCAGG 18000
GTCGGTCGGAA CGGAAAAGCG CGGACCGGTC GCAGTGTCCT CTCCCAGCAG CTCCCGCCC 18060
CCTCAACAGG TGTAAGCCCC TGCCCGCGCT CGGCCGTGCG CGCGGCGGAC CGGGGGCAC 18120
CGCCGCCGCT CATGGGAAGT GCACAGACAT CTGTAACAGC ATCTGGGCTG TGGAGTGCA 18180
GAAGTGAAGG CCGCGGCGCT GCTTAAACA CTACCGTACC GCTTAACACT CTGTCTCTTG 18240
TGTTGTATGA TTAAGCGGCC GCAGCGGTGTC TACCAGCAAG GAGGGAGAGG AGAGGGCGGT 18300
CGGACAGGGA CGGCGGCGAT TACCTGAGTC CGGTTGGGTG GCAGTTGGCC CGCGGACGAG 18420
ACACCTACCT CGACTGGGGG AACAGTATTTA GGAGGCCCCG CTGGGCCGCTG AGCAGAGTAC 18480
TGACGACAGA CGCACCGGAC AGCTCCGCTAT CGGTCTCTGT GCCCGGTCAG CGGGGGGAC 18540
ACACCTACCT GCCACACCTT GGGCGGTGGG CGAAAGCAGG GTGCTGGGCA 18600
TGCCGAGCAG CTACCATGAC ATCCGCGGCG TGCITGAGTGC GGGCCCTAGC TTCAACACCT 18660
ACTCCGGCAG CGCTCAGTCA ACTGGCAGCC CCAAGGCAGC ACCAAACACT TGTCATCTGA 18720
CATATAAGGC CGATTGGGAG ACTGCCCATG AAAAACAATC TATCATATGA AATTGACCCC 18780
TGCAAGGCAAT TAAACAGGAC AAGATGTTTA TTCAACATTG AACTGACACC GATGACTAAC 18840
CAATCTACGC AGATAAAACC TATCAGCCTG AACCTCAAGT GGTTGATGCT GAATGCGATG 18900
ACATCACTGG TACTGGGTAA AAGTATGGAG GAGAGGGATG GCCGCTGACG TTACAGTGA 18960
AGCCCTGTTA TGTGTCTTTT GCAAGGCTTA CTAATAAGGA AGAAGGGCA GCAAAATGGA 19020
AAACAGGGAAC AGCCACTACT AAAGAATAG ACATAGACAT GGCTTTTTT GCAAAACAGA 19080
GTGGGGCTGC TGCTGGCTTA GCTCGAGAAA TTGGTTCTCTA TACTGAAAT GTGGAATTGG 19140
AAACTCCAGA TACCCATATT GTATACAAGG CGGACGACAG TGACAGCAGC TCTTCTATT 19200
ATTGGGGCAA CGAACGCTAG CCAACAGCAC TCAACTACAT TGTTTCGAGA GCAAACTTTA 19260
TGCGGCTCAT GTACCTACAC AGCAGCGGCA ATATGGGGGT GCTGGCCGGT CAGGCTTCTC 19320
AGCTGAATTG GTGGGTTGAC TGCGAAGAAC GAAACAACGA GCTGTCTTAC CAGCTCTCTC 19380
TGACCTCTCT GTGGGCTGAC ACCCGTTATT GAAGAGTTGA GTGACAGCGC 19440
ATGATCTGCA TGGTGGCTATT ATGAAAAATC ATGGGTGAGG GGATGAACTT CCAACACTTT 19500
GGTCCCTCTG GGAATCTGGT GGCGAAGAAC ATACCTATAC GGAATTAAG GCAATAGGAA 19560
CTGATCAACG CAGTTAGGGC AAAGATGACA GTGCTAAATG AATGGAAGG AGTACGGGGA 19620
GTTATCCATT CGCCATGGGA ACTCAAATGG ACAGCAGCGC TGAGGGAAC TCCCTCTTACG 19680
CAGCCGAGGC CGGTCGCTCT CGGACGCTTT AAACGATGCA CGCGGCAAAAT GTGACCTTGTC 19740
CCACAAACAG CAAACACCTAC GATTACATGA ACAGGCGGTT GCTGGGCCCC TGCTGGTGG 19800
WO 98/10087

PCT/US97/15694

81

ACTCCTACCAT CAACATCGGG GCCGGCTTGTT CGCTGGATCCC CATGGACAACC GTGAACCCCT 19860
TCAACCGACCA CGCGATCGGC GGGTCTGGCTT ACGCTCCTAT GCTCTGGGCC AAGGGGGCGGTT 19920
ACGGCTGTCTT CAACATGAGCC GTGCCCCAGA AATTTTTTCCG CATCAAGACG GCTCTGGCTCC 19980
TGCCGGGCTT CTACACCTACT GAGTTGAACTC TGCGGAAGAA CGTCAACATGT ATCTCTGAGA 20040
GCTCCCTCCTG CAACAGACCTG CCCACGGGAGC GGGGCTCCAT CTCTTCACCC AGCATCAACC 20100
TCTACGCACC CCCTTTCTTTT ATGGGCGCAAC AGAGGCCTCT CACGTGCTGAG GGTCTGAGCC 20160
GCACGCGACAC CAACAGACAGGC TCTTCCACAC ACTACCTCTAC GCGGCGCAAC AGGTCTTACC 20220
CCATCCCAGCC CAACGGCACC AACGTGGCACA TCTCCACTCC CGTCCGCAAC CGGGCCGCCCT 20280
TCCGGGCTTG CTCTTTCAAG CAGTCTCAAGA CCAAGGAAGAC GGGCTGGCTG GGCCTGGCGGT 20340
TCCGAAAACTA CTTCTGCTACG TCAGGGCTCCA TCCCTCTACCT CGAACCGCGC TTCTACTAC 20400
ACACACACTTT CAAGAGAGTGC TCCATACCTT TCGACTCTCT CAGTCGCTGG CGGCGCAGCC 20460
ACCGGTCTCTT CAGGCCCCAAC GAGTCTGAAA TCAAGCGCAC GTGCGACGAGG GAGGCTACAA 20520
ACGTGGCGCTA GTGAACATCG ACCAAGACTG GGTTCCGTGT CCAGATGCTG GCCACACTACA 20580
ACCTGCGCTA CCAGGGCTTCT TAGCTGCGCC AGGGCTACAA GCCGACGATG TACTCTCTCT 20640
TCCGGGCTATC AGGCCGCTAG GCAGGCAGGG TGGTGGAGGA GTGTAACATG AAGGACTACG 20700
AGGGCGCTAC TCTGGGCTACT CACGCAACAC ACTGGG GCCCTGCTCT CAGCGCGCCCTG 20760
CCGTCACACC CCAGGGCCAC TACCCCGCCA ACTACCCCTA CGCGGCTCATC GCCAACAGGCC 20820
CGGTACCACTA CGTCATGAGCA AAAAATTTCC TCTGCGACAG GGTCACTGGG CGCATCCCTCT 20880
TCTCCAGCAA CCCTCATGCTC ATGGGCCCGCC TCAGCCGACT CCAGCAGAAC ATGGCTCTATG 20940
CCAATTCCGG CAGCCGCTTA GACATGAAATT TCGAGTGCAG CCCATGGAT GAGTCCACCAC 21000
TTCCTCTATCT TGCTTCGGAA GTCTCCTGAG CTGGCCGAGT GCAACGCCGCA CACCGCGGGG 21060
TCATCGACGC CGTGTCACGGA CCAGCAGGCT CTCGGGGGCC TAAACCCGAC ACTTAAACCT 21120
TTGCTTCTTG CAAGCAGATTG CGCGGGGCTCC CGCGGACAGG GAGTCAAGGG CCACTCATCG 21180
CGACCTGATGC TGGGGCCTGT ACTTCTGTGG CACCTCTGAT AACGCTTCCCG CGGGATTCTAT 21240
GCCGCCGCCA AACGTGATGC CCGCTGGATCA AACAACGCGCC GCGCCGCGAAG CCGGGGGCGG 21300
GCCATGGCGCT GGCTCCGCTC GGAACCCCGGG CTCGAAACAC TGCTACTCTC TCGACCGCGTT 21360
CCGGTTCTCG GAGCAGGCAC TGAGAAGAGAT CATCCAGATG TCAAGGCGGCT GGTCTGTCGG 21420
CCGAGCGCC CTGAGGCAAGG AGGACGGCTG CGTACCCCTG GAAAAATCCA CCCAAGACGT 21480
GCCGCTTGAG CGCTCGGGCC CGTCGCGGCT ATCCGGTCTG AATGTCGCTG AGCGCTTCGTG 21540
GCATTGCGCC GACGGCGCCA TTGAACAGAAA CCCCAATGAG AACTGGCTGA GGGGTGGTCCC 21600
CAACGGCATG CTCACGTCGG CCGAGGCTGA ACCACCTCTG CGCGCGAACG AGGAGGGCTT 21660
CTACCGCTTG CTCGACTCCG ACTCCGCGCTT CCCTCGCTCC CACCGCGCGC GCATCGAGAA 21720
GGCCACCGCC TCGACCGCA TGAATCAAGA CATGTAACC GTGGTGATAT GTTAAATGTC  21780
TTTAAATACG AGCATTCTCA TTGTACACAGT GACATCTGAGA TGTATTTATG AGAAATCGAA  21840
AGGGTTCTCG CCGCTCTCGG CATCGGCCGC GGGCGAGGAC AGGCTGCGGA ACTGATTCTT  21900
GGCCAGGCCA TTGAACCTCGG GGAATACGAGA TTTTGTCCG GGGGTGCGGG GGAAGAGGTC  21960
GTCACACGAC TTCGCGTCTCA GTGGCTAGGCC GCGGCAGCAG TGCGGCCTGG AGATCTTGAA  22020
ATCGAGTCTT GGACCGCGGG TCTGCGCGGG GCAGTATTGG TACAGCAGGT TGCAGCCTTG  22080
GAACACCATC AGGCCCGGTC GGGTCAGGCT GCCACGACC GTGGCTGCTG TGATGCTCTC  22140
CAGCTGAGGC TCTGCTGGGT TGAGGCATCC GAAAGGGGCT GCTTCTGGGG TCTTCTCCTC  22200
CATCGTGGCC AGCAGCCCGG GTTTGTGGCC GCAATCGCC TGCAGGGGG TACGCACTCAT  22260
CTGGGCGTGG TGCGGCCGTC TGCCGGCTTG TACGGCTCTG CAAATGCTC CCAATTCTCT  22320
GAACCCGGCC TGGGCCTTGG CTGGGGCTCG GCAGGAGCC CGCCAGCAGT TCTTAGAGAA  22380
CTGGTGATGG CGCGACCGCC GTCGGTCGAC GCAGCCAGGC GGGTGGTTGC TGGCCACGTC  22440
CACCAAAGCTG CGCCGCAGCC GTTCTGCGGG GATCTTGGCC GGAGCCGGGT TCTTCTCCGC  22500
CGCGCGCTCC CGGTCTCGCC TGCGCACATC CATCTCGATC ATGTGCTCTC TCTGAGCATC  22560
GGGGCCACGC TGGGGGTCGG CTGCACCGGT CTGGGCTCGG TGCCACGGG GCACCCACAC  22620
CGCCGACCGC GTGGACTTTGG GGAGGCAGTG GAATGGCGGT GCAAGAAGGC  22680
CTGGAGGAGG CGGCGCATCA TGTTGCTAGC GGTCTTGGTG CATAGTGAAAG TCGCCGGGAT  22740
GCCGGGCTGC TCTCGTGGCA TGACCTGAGG GCAGATGCCG CGTACACCT CGGGCGCTCC  22800
GGGAGTGCTG CCTGGTGGTG GTGCTTACGC GGGTACGGGT CTGAGCAGCT  22860
AGTCATGATT TCTACACCT TCAGGCGGCG CAGGACAGTG GCAGGCTCAG CAGGGTCTTTT  22920
CACCACTATC TTAGCGCATG CAGCGCGGGC CAGGGGGTGC CTCTGCTGCA GGGTCTTGAA  22980
GCTCGGATTG CCCTGCTTCTGGTCTGAGGG TACCTGAAGG CCAACTGCCC  23040
CAGTCTCCGC TCGCGCTTGG CTGGCTCCTC GCTGTGCTGG CTGACGCCTG GCAGGACCAC  23100
ATGCTGCTGC TTGGCGGGGC TCTTCTGGGG GGAGCCGGGC GGGGGATATG TTGGAGATGG  23160
CGAGGGGGAG CGGGGAGGTCT CGCTCACAGC TACTATCCTC TCTCTTCTCT GTCTCGAGGC  23220
CAGGCCGGGC TAGGTAGTGC TCTGCGGGGG CAGGGCGGGA GGCGCGGGG GGTGGGCTGG  23280
GCCACCTGGG GATGTGCTGG CAGAGCGTGG GGCGCGCTCG GTGGGCTCGG CCCGGCGGGG  23340
CTGTGACTGA CTTCGCCGCC GGGCGCGCAG TGCTTCTCCT TAGGGAGAGA CAACAACGCAT  23400
GGAGACTCAG CCCGCGCAG CTCGGCGGC TGCCGCCCACC GCGGCAGGAG AGCAGCCACG  23460
GCAGAATGAG AGCTTAACGC CCCCCGGCC CAGCCCGCC ACCTCGCGAG CGCCTCGCCC  23520
AGACATGAGA GAGATGGAG AAAGATGCAGA GTAGGACTGG GGGTACCTTG TCACGGTCG  23580
GCACCGGGAG GAGCTGCGAG TGGGCCCTTG ACAAGAAGAG ATACACCAAG AACAGCCAGA  23640
GCAGAAGGA GAGAATGAGC AGAGTCAGGC TGCCGCTGAC TATGACCGGG ACTACCTCCA 23700
CCTGAGCAGG GGGGAGAGGC CGCTCACTCA GCCAATTCGG CCAGCGAGCA CATCATGGGA 23760
GGATGGCCTG CTGACGGCAG CCAGGTTGCC CTCAGCGCGG GAGAGCTACA GCAGCGCTTA 23820
CGAUTGAAC CTCTCTCTGC CGCCCGGGCC CCCCACGCGC CAGCGCAATT GCACCTGGGA 23880
GCCAACCAGG CGCGCTCAACT TCTACCTGGT CTACCCGCGT CCGAGGGCCT TGGCCACCTA 23940
CCACATCTTT TTACAGAACC AAAAAAGATCC GTGTCGCTGC CGGCGCAAAC GCACCGGGGC 24000
CGCACCGCCT TTCCACCTGG GCCTCCGGGG CCGCGTACT CTATGGGAAG TTTGGGAA 24060
GGTTCCCAAG ATCTCCTGAG GTCTGGGCGA CGACAGAAGT CGGCGCGCAG ACGCTCTGCA 24120
AGGAGAAAGG GAGAGCATG AGCCACCAAG CGCCTGGTTC GAGTTGGGAAG GCGAACAGGC 24180
GGGCGCTGGG GTGCTCAAAAC GCAGCGTGCA GCGACACCAT TCGCCCTCCTC TGCCCTGGAA 24240
CCCTCCCCCA AAAGCTGATG CGCCGGTGTAC GGACCAAGGT TCTCATCAAGG CGCGCGTGCC 24300
CATCTCGAG GACGAGGGCA TGCAAGACCT CGAGAGGAGG ACGCCGGTGG TCACCGAGCA 24360
GCAAGCTGGC CCGTGCTGTGA GCTCTGCTAC TTAGCCCTAC AGTTGGAAG AGCGCGCGAA 24420
ACTCATGATG GCCTGGTCTC GTGAGCTGGA GAGCTGGAG TGCTGGCGGC GCTCTCTGCA 24480
CGACGGGGAG ACCCTCGGCA AGTTGCAGGA GAACACTGAC TACCTCCTCC GCCGGCGGTT 24540
CGTGGGCAG GCGCTGCAAGA TCTCCACGAT GGAGCTGACC AACCTGGTCT CCTCATGGG 24600
CATCTTCGAC GAGAACCGCC TGCGGAGAAA CGCTGTCCAC ACCACCTGCG CGCGGGAGGC 24660
CGCGCGCGAC TACATCGGCA ACTGGGTCTA CTTTCTACCT TGCCACACCT GCCGACCGGG 24720
CAAGGGCGTG TGCCAGCAGT GTCTGGGAGA GCAAGAAGTC AAAGACTCCT GCAAGCCTCT 24780
GCAGAGAAGC TCTCAAGGGTC GTGAGCGCGG TTCCGGAGAG CGCCAAACCG CCTCGGAAC 24840
GCCCGACCTC ATTTTCCCGG AGCGCGCTAG GCCGACCGTG GGCAACCGCC TGCCCGACTT 24900
TATGAGCAGA AGGATGTGGA AAAACTCCTGG CTTTTCACTC CTGACGCACT CCTGGACGTC 24960
GCCCGGCACC TGCCCGCGCC TGGCCCTGGGA CTCGGGCTCG CTGAGACCTT GCGCTGCTCC 25020
CCCGCGCTCG TGGGACACTG GCTACCCTGC CGCGCCTGGCC AACTCTCTGG CCTACCACTC 25080
GGAGCTGATC GAGAGCTCA CGCGGGAGGG CCTGCTCGAG TGCCACTGCC GCTGCAACCT 25140
CTGACGGCGG CACCGCTGCA TGCCCTGGCA CCCCAGCTCT TGAGCGAGCA CCCAGATCAT 25200
CCGCCACCTTG CAGTGGCAAG GGCCCGACGA AGGCGAGGCT TGCACCGGCA AGGGGGGCTT 25260
GAAACTCACG CGGGGCGCTG GCACCTGGCG CTAACCCTGG AAGTTGGTGC CGGAGACTA 25320
CAATCCCTTC GAGATCCAGT TCTACCGAGA CCATTCCCAT CGGCCCAAGG CGAGCTGCTG 25380
GGCCTGTGTC ATCAACCAGG GGGCGATCTT GCAGCGCAATTG CAAGCCATCC CGAAATCCCG 25440
CAGAAATTTT TTACGTGAAA AGGGCGCGGG GTTCTACCTC GCAAGCGCGA CGGCTGGAGA 25500
GCTCACCCGG GGCTTCCCCG AGGATCGGCC GAGGAAACAA GAAGCTGAAA GTGGAGCTGC 25560
AATCCATCAA CCAGTCTTTG TTGCTCACCC GGAACGAGAC CGACGCTCCG AGCTCAGTGA
AGGCCCCCAA GAAGTACCTC ACCTGCTGGT TCCACCGGG ATCCATGGCC GTTGTCAAC
ACTCGGACAA CGAGGAGTGT CAGTCAGGCC GCTCTGGCAA CTTACATTTC TCAACCACCA
GAAGCAAGCT CCACGTCTTC CAACCCCTCC TCCCCGGGAC CTATCACTGG GTCCTGGGAC
CTGCGCATCA CACCTTCCAC CTGATCCCCA ATACCACAGG GTGCGTCCCC GCTACTAACA
ACCAAACCTAA CCTCACCACA GCACCACGTC GCAGCATTTCC TGAATCTTAA ACTACCCACC
ACACCCGCTGA TGAGCTCCGA GGTACCACAA CCTCTGGAAT TTACTACGCG CCGTGGGAGG
TGTTTGGTAT AATTACCGTA GGCCTAGTGT CGGGTGATGG TTGGTTTCTC TGCTACTCAT
ACCTCCCTTG CTGTTCGTAC TTCGTGTGTCG GTTTAGAAAG TGGGGAAAGAT
CACCTTAGTG AGCTTCGGTG CGCTGGTGGCC GTGTTGGTCCT TGAGTTGGG GACTGGGGCGG
TGGGCTGATA GTGAAGGGAGA AGGCCGGATCT CGCTCTGCAT TTCAATCCCA ACAATAGCCA
GCTGAGTTTT CAGCCCGATG GCAATCGGGT CCAGTTACTG ATCAAGTGCG GATGGGAATG
CGAGAACGTT AGAATCTCGT ACAATAACAA AACTCGAGAC AATACTCTCG GCTCGGTGTG
CGACGGCGGG GACCCCCGAT GTACACCGGT CTCTGTC CCC GCCGTGAGCG GCTCCCGGGG
CACCGTGAAAT AATACCTTCA TTTTGGCACA CATGTCGAC ACGGTATGGT GAGTGACAA
GCAGTACGAT ATGGGGCCC CACAGGGAAG GAACATCTGT GTCTTTCTCA TGCTTACAG
GCTGTCGACG CCGTATAGCA CGCTACTCGT GTGCTGGCAGG ATCCACATGC TCACTGCAT
TCGGCCAGA AATAATCGGC AAAAGAGAAA ACAGGCATATA CTTTTTTTTT CACCCTTTTT
TCGAGACATTG GCCTCGTTTA AATTTTGCTG TTTAATTCG AGTCCTATGG CGCTCATTCA
TGGAATGAGT AATGAAGAAA TTACTATTTA CACTGCGACT AATCACACAT TGAAGGTGTC
AGAAAGAGCC ACAAGAGTTT CATGTTATAG TTTATTTTAA GAATCAATGT TATCAATCTGA
ACTCTGTGGA ACAACAATCA AAAAATGAGA GAGCATTACT CTCACTAAGT TCCAATGGGG
ATCTGACTTA ACCCTAATTA AACTCATAGT AGACTATGTA GGTATGATT ATGGGACTAC
AGCAGGACATT TGCAATCGGG TTTTTATCA AGTTTTCTGTG TCTGAACCCA CCACGCCTAG
AATGAACACA ACCCAAAAAA CTACACTCTG TACCACTATG AGCTGATGG CCAAAATCAT
TTTGGCATAG CGTCATTAGT TCAACCATAG CACCTCAACC ACCCCACCCA GTGAGGAAT
TCCCACATCC ATGATTGCAC TATTGTGTCG TGTAGTGTTG TGCACTGCTG TCACTGGCTT
TGGCATGTTG TACTATGCC TCTGTCAACG AAGCACAAGA CTGAACGACA AGCTGGGAAC
CGGCTAAGT ATGTTTTTAG AACCAGGAGG ATCCACTGGC TTTAATTTTT
TTCTATCATG ACCCTGGTCG TATGCAATTC GTGACAATAG GACGTTACTG TCGTGTGGCGG
ATCAAATATG ACACGTGAAG GTCCACGGGA GGGTATGCTT TGTGGTATT GCTATTTTGG
ATCTGACACT ACGAGAATCG AATTATGCAC TCTAAGAAAT GGAAAAATTC AAAATCTAA

27540
27600
27660
27720
27780
27840
27900
27960
28020
28080
28140
28200
28260
28320
28380
28440
28500
28560
28620
28680
28740
28800
28860
28920
28980
29040
29100
29160
29220
29280
29340
29400
AAATTAACAT TTATATATGCA ATGGTACTGA TCTGATACCT CTCAATATCA CGAAATCATA 29460
TGCTGGCAGT TACACCTGCC CTGGAGATGA TCTGACAGT ATGATTTTTT ACAAGTAGA 29520
TGTTGTTGAT CCACCACTCT CACCTCACC CACCCACACT ACTCACACCA CACACACAGA 29580
TCAACCACGA GCAGAGAGAC CAGCAAAGTT AGCCTTGACG TCTCAAGACACA GTTTACCCGT 29640
TGCGATTACC CCTACACCTG ATCAGCGGCTG TCCGGGCTTG CTAGTCAGGC GCATGTCTCGG 29700
TGCTTCCTCG GGAATAGACG CATATAATCT AGTCATGATC ATTTTGGCTT GCCGTTATAG 29760
AAGGCTTTAC CGCAAAATGC CAGACCCACT CTTGAACCCTC TATATTTAT ATATTTCCAGA 29820
GCTCATAGAGG CAGTTCAGCG TTACATTGGT TTATATATGGA TTTTGGCATT TTATTTGCAAT 29880
CCTATTCCCTA AAGTTAGCTT TTATTAAGAT GTGATAAGTA CGTGGGGGGG CAAATAGTCA 29940
CTGTAGGGTC TAGAGGGGCC TGAAACACCC ACTTGAGGACA ATATACCCCT AATGAGGCTG 30000
AAAGATATTT CCAATAGGGA TGATTAGCTT TATATATGGA AGAGGTGTTGA TCTACTCATG 30060
GTCAATGCGCA CCTCGACTCA AATGTTGGAA ATCAGAACAGA AAGATGGTACG TGTATCTAAAT 30120
GGGTATTTTA CCCAACATAC TTTTATGTAT GACCTTTAAG TCATACCCACT GCCTACGCCT 30180
AGCCACCTA GCACCTACAC ACAACACAAC CCACTACAC AAGACAACAC AATACAGTACA 30240
TAAATACGG CTACACCCAC TACAGGAGGC GAGGTGTCGA GCCCTGTTCG GTGCGGACTG 30300
GCATTGGGCCC ATCTAGCAGT CCCACTGCTA GTACCAATGGA GCAGACTACT 30360
GAAATTTTGT CCACTGTCGA GACGCAAAAG ACGCTACTCT CCAGTCGCTT CTTAGCACCT 30420
GCCAAATCTCT CCTGGCTTTTC CTCTACACAA ATCAGTCGGG CTACTACTCC TACGCCCCGT 30480
CCCTCTCTCA CTCGCCCTGAA GCACACAGAC GCCGCGGATG AGGAGGAAAGT CACCCCTGCTC 30540
ATTGTGATCG GTGGGTGCTAT CGTGGGGCTG TGCTGTCTAC ACATCTCTCTC CCCGCACATT 30600
CCCAACCGGC ACCGCAAGGC GTTCTACAGG CCACTAATGG CCAGGCGGCTT CAGGCGCGTT 30660
CAGGTGAAG GCCGACTCTAG GAATCTCTCTA TTTCTCTTTTA CAGTATGGTG ATGTAGCTAT 30720
GATTCCTAGA CAATTTCTGA TCTACTTCTT CTATGTCCGC CTCCAGACTCT GTGCCACCTT 30780
CGCTCTGTGG GCACACGCGC GTCCAGAATG TATTTGGGGG TTTCTCCCTC ACGTGTCTCT 30840
TGCTCTCCAC ACCGATCATCT GCTGTGTAG ATAGTCTGC GTGTTTATCA CTTCTTTCGA 30900
GCTGATAGGC TCTGCTGCTG CTAGCTCTGG GCCACACCCCC AGTACGCCTGC CCGCCTCTGT 30960
CCAGGGAGTG GCGCGGCCTGC TCAAGCTCCT CTGATAAGCA CGGGCGCTCT GCATCTTCTC 31020
GCCTCTTCTG CCTATGTCGT CCTCGCGTCCC GTGCAGCCGG CTGCCCCGGC CAGTGCCCCC 31080
GAGGAGTCC CGAAGTCGCA ATTACCAAGG CCGTGAATAC TCTCTTTTAT CCTACGGCAA 31140
AAATCGACA TGGATCCGAG TCAGATCTAG ATCAGTTGGG TCTGAAACAT TCGGCCCTCG 31200
ACCCCTATCT CCTTTGGTAT TTACCCCTGC TTTGACCTTG GTTGGACATC GCCAGGGCGG 31260
CTCTATCTCC GCAGCTGAAC TGACGACCA CCAGAGCAAC CTAGGCGACA CACGACTACCA 31320
CCACTACAGC CTAGGGCACA ATACATGCC ATATTAGACT ATGAGGCGGA GCCACAGCGA 31380
CCCATGCTCC CCGTATTAG TTACTTCAAT CTACCCGGCG GAGATTGACTG ACCCACTGCG 31440
CAACAACCAAC GTCAAGCACC TTCTCCCTTG CATGGACGCC CGGCGTCTGG AGCAGCAGCT 31500
CGCCCAACCTT CGCATTCGCC AGCAGCGAGA GAGACCGGTTC AAGGAGCTGC AGGATGCGGT 31560
GGCCATCCAC CAGTGCAAGA GAGGCACTTT CTGCCTGCGT AAACAGGGCA AGATCTCCTA 31620
CGAGGTCACT CCAAGGGACC ATGCCTCTCC CTGCAGGCTC CTGCAGCAGC GCGAGAGTTT 31680
CACCTGCCTG GTGGGAGGTA ACCCCCATGT CATCACCCAG CAGTCTGGCG ATACCAAAGG 31740
GTGCTACAC TGCTTCTGGG ACTCCCCGGA CTGCGTCCAC ACTCTGATCA AGACCCCTCTG 31800
CGCCTCGCC GACCTCTCCC CCAATGACTA ATCACCCTCT TATCCAGTGA AAAATAAGATC 31860
ATATGTGAGT TGATTTTACA GAAATAAAAA ATATATCAATTT GATTTGAAAT AAAGATACAA 31920
TCATATTGAT GATTTGAGTT TAACAAAAAA AATAAAGATC ACTTACTTTGA AACTCAGTAC 31980
CAGGTCTCTG TCCATGTTTT CTGCCAACAC CACTTCACCT CCCCTCTCCC AGCTCTGGTA 32040
CTGCAGGGCC CGCGGGGGTG CAAACTTCCT CCACAGCTTG AAGGGAGTTG CAAATTCCTC 32100
CTGTCCTCAA ATCTTCATTT TATCTCTCAT CAGATGTCGA AAAAGCGGCGT CCGGGTGAGT 32160
GATGACTTTCG ACCCCGCTCA CCCCCAGTAT GCAGACAAGC CACCCGAGCCT GCCCTTCATC 32220
AACCCCCCCT TGCTCTCTTC AGATGAGTTG CAAGAGAAGC CCGCTCAGGG GTTGGCTCTG 32280
CGACTGCCGC ACCCCGGTAC CACCAAGAAC GGGAAATCA CCCCTCAAGT GGGAGAGGGG 32340
GTGACCCCTG AATCCCTGGG AAAACTCATC TCCAACACGG CCCAAGGCCG CGCGGGCGCT 32400
CTCGATTCCC CCAACAAAC ACATTTTCTC AACATGAGAT ACCCCTTTTA CACTAAAGAT 32460
GAAAATTTAT CCTTACAGTT TTCTCCACCA TTAATATATAC TGAGAAACAAG CATTCTAAAC 32520
ACACTAGCTT TAGGTGTTTG ATCAGGTTTA GGACCCCCTG GCTCCTGGTT GGGACTGAGC 32580
TTGCTCTTTC CACTACATT TGATAGCTGA GGAACATAA AGCTACCTTT AGCAGAGGCT 32640
TTGCAATGTTA CAACAGGAGA TGCAATTGGA AGCAAATATAA GCTGGGCTAA AGGTTAAAAA 32700
TTTGAGATAG GAGCCCATGC AACCAACATT GGAATGGGT TAGAGTTTGG AAGCAGTAGT 32760
ACAGAAACAG GTGTGATGTA TGCTTACCCA ATCCAAGTTA AACCTGGGATC TGCCCTTAGC 32820
TTTGACAGTA CAGGACGGTT AAGGCTCTGG AAAAAGAGA AGCATAAAAC CATTTTGGTG 32880
ACACACCTTG ATCCATCCAA AACCAGCTCA ATACTCGCAG AAAATGATGC AAAACTAACA 32940
CTTTGCTTTG CATAAATGGG TAGTCAAATA CTGGCCACTG TGCTAGCTTT AGTTGTAGGA 33000
AGTGGAAACC TAAACCCCAT TACTGGCACC GTAAGCAGTG CTGAGGTTGT TCTACCTTTT 33060
GATGCAAACG GTGTCTTTTT AACAGAACAT TCTCAACTAA AAAATACAGAG GGGTATAGG 33120
GAGGAGGATA GCATGATAGG CACTCCATAT ACCAATGCTG TAGGATTTAT CCACAAATTTA 33180
AAAGCTTATC CAAAGTCACA AAGTCTACT CTAATAAAAA ATATAGTGGG GCAAGTATAA 33240
TGTTACACTC CTGACCACTC CCCAGATAAT TTTCATTTTT CCAGCCTTGA ATGATTGCAA 35220
CTAGTTCCGTG AGTAAATCCG AACCCAGCAA GATAAAGAG CTGGCCGAGA GCACCCTCCA 35280
CCGGCATTCT TAGGCACACT CTCATAATTCC CAAGATAATG TGCTCTGAGT TCACCTGCAG 35340
CAGATTGACA AGCCGAGATAT CAAAATCTCT GCCCGCAGTC CCAGACCTCT CCCTACGCAA 35400
TAACTGTAAAG TACTCTTTCA TATCTCTCTCC GAAATTTTTA GCCATAGGAC CACCAGGAAT 35460
AAGATTAGGG CAAGCCACAG TACAGATAAAA CCAGAAGTCT CCCCAGTGAG CATTGCCAAA 35520
TGCCAGAAGTGCT CTATAAGCAT GCTGGCTAGA CCCGGTGATA TCTTCCAGAT AACCTGACAG 35580
AAAATCAGCC AGCCAGTAAAA TAAGAAAATC AACAAAAAGAA AAATCTCTCA GCCGACGGTT 35640
TACAGGCTCG GGAACCAAGGA TGAAGTAAAT GCCAGCCTGC TGCTTACGCTA 35700
GCTGATCTGT AGAAAAAAACA AAATAGAACA TTTAACATGTC TAGGCTGTGC GACAGGTGAG 35760
GTAATCTGGT CTCTCCACGA CCAGGCACGG CACGGGCCTC GCGGCGGAC CCCGCTGAAAA 35820
ATTGTCGCTA TGATGGAAA CCAATCACAG AAGAGCTTCC CGTGAGCGCG CCTGAATGAT 35880
TCGACAAGAT GAATAACCCC CCGGAACATT GGGTGCCGG AGTGAAAAAA AGGCCCCGAG 35940
GAAGCAGAAGA GCCACTACA TGCTGACGCTA CAAATTCAGCC AAACCGAGATG CTACGGAGAT 36000
AAGCACAAGA AAAATTGCAGG CCGACAAAT GAAATATTACC CCGTCTGCAGA GAGGACAGCA 36060
AGCCTCCAGCT CCGCTCGGTC ACGACATCAA ACCCTCGCGG TCCGATACGT ATCCAGAGCG 36120
AGCACAACAC AGGCAGCAGA GTCAAGAAAA GCACGATGGTG TAAACCTGCTC ACCCGTCTTC 36180
TGCTCAATAT ATAGCCAGA TCTACACTGA CGTAAGAGGC AAAGTCTAAA AATACCGGCC 36240
AAATAATCAC ACACGCACCG CAACGGCCCA GAAACCGGTG ACACACTCAA AAAAAATACG 36300
GCACTTCCCT AAACGGCCAA AACTGGCCGT ATTTCCGCGT TCCAGCGCTA CTCGATCAAA 36360
ACACGACTTT CAATTCGCTG CGACGCTTAA AAACGTCACC GCGCCCGCGC CTAACGGTGCG 36420
CCGCTCTCTC AGGCAATCAG CGCGCCGCGT CCCCCAATTC AAACACCTCA TTGTCATATT 36480
AAGCGGCACA AAAAGTTTTGA GTGATTATAT TGATGATGG 36519

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8299 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCCCCAATACG CAAACCGGCT CTCCCCCGGC GTCGCGCGGAT TCATTAATGC AGCTGCGCCG 60
TGGCTCCGCTC ACTGAGGCGG CCCGGGACAA GCGCCGGCGT CGCCGACCT TGGCGGCC 120
GGCCCTCAGTG AGCCAGCGAG CCGCGAGAGA GGGAGTGCCCG AACCTCCATCA CTAAGGGGTTC
CTCTAGTTTAA ATGATTTAACC CCCGCTGCTA TCTTCACATA CTAACGATGTA ATTCGAGTTT
GACGTGCCGCA AGGCCGTTAC CTAACCTTAC CCAATGGTG CCGCTGGTCTT ACGGCCCAAC
GACCCCCCGCC CATCGACGC TGAAATAAGCC TAGATTCCCCA TAGTACGCC AATAGGGGACT
TTCCATGGTCA CTTAGGCTAC AAGGCCTATT GACGTCAATG ACGCTGAATG CCGCGCCCTGG
CATATGCCCA AGTACATGAC CTATTGGGCT TTAATCCTTGG TTCCATCTAG ATACGTCACTA
TCATCGCTA TTACCATGGT GATGGGGTTT TGGCTAGTACA TCAATGGGCG TGATAGCAGG
TTGACACTCAC AAGGAGTTCC AACGCTCCAC CCCATGAGCC TCAATGGGAG TTTGTATTGG
CACCAAATAC AGCGGAGCCT TCCCAAATGT CGTAAACACT CGCCTCCATT GACAGCAATG
GGCCGTTAGGC GTGTAGGGGT GGACGCTATG ATAAAGCAGAG CTCGTTTATG GAACCCTGAC
ATCCGCTGGA GAGCGCATCC AGCGTGGTTT GACCTCCATGA GAGAGACCGG GAGCCGATCC
AGCCCTCGGA CTCTAGAGGA TCCTGATTC ACGCCGAGCT CGGACCACT ATGAAGCGCC
GCCAGTGTCC TGGAAATTGT CACTCCAGCG TGGCGCGGTT TGGCGAGTGG TGCTGCTGTG
GCTGTGCGGC CTCAGGGCTAC ACCCTCCTCTT GGCCATCTAC CTAGTTGAGG AGAGGAAACC
GGACCTCTCG AAGCGGCGAG CAGCGCGGAG CCTGCGGTTGG GCCAAGAGCG TGGAGCCTGC
ACAGACACGC CGCAAGACCC TCACTCACTT CCTGGGCGAG GGGATGGGGG TGCTCTAGGT
GACAGCTGCC AGGATCTCAAG AAGGGCAAGA GAGGACAAA CTGGGCGCTG AGATACCCCT
GCCATAGGAC CGCTTCCATAT ATGGTGCTCT TCAGGAGACA TACACATGTG ACAAAACACT
GCAGGCAGACT GAGCCCGACAG CACCCGCTCTA CTCCTGGGCG GTCAAGGGCA ACTCCACAG
CATTGGCCTTAA AGTGCAACGC CCGCGTTTTAA CAAGTGCAAC AGCAGAAGCG GCAAACAGGT
CATCTCCGTTA GATAGTCACG CCAAGAAGGC AGGGAACTCA GTGGGAGTGG TAACACCAC
ACGAGTCGAC CACGCGCCGC CAGCGCGCAC GCTAGCCCCCA CAGCGTGAAC GCAGACTGGA
CTCGGCAGGG CAGGTGCCTGT CTTCCGGCGG CCGAGGAAGG TGCCAGGACA TGCTACGCTA
GCTCATCTCC GACATGGACA TTGTATGGTA CTTAGGTGGCG GGGCGAGAAT ACATGTTCGG
CATGGGAAACC CGACAGCCATG AGTACAGGCA TGACATCACG CGAAAGTGGG CCGGGCTCGG
CGGAAAGAAT CTGTCGCGAG AATGGCTCGGG GAGAGCGCCG GTGGCGCCGT AGTGTGGGA
CCGCAGCGGC GCTAGGCTGGA CCCGCGCTGTA ACCATCACTA TGCTGCTCCT
TGGAGCTGGA GACATGAAAT AGACGGTTCCA CCGAGACTCC AACTGGGACC CTCCTTCTGT
GGAGATGACA GAGGTCGCC TGGCGCTGCT GACGAACGC CCCCGGCCCT TCTCTCTCTT
CGTGAGGAGT GTGCGGCTGC ACCATGGTCA CTGACTAAAGC AGGGTTCCAC GGGACTGCA
TGAGACGATC ATGGTCAGAG AGCGCATTGA GAGGGCGGCG CAGCTCAACA GCGAGGAGGA

180
240
300
360
420
480
540
600
660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
1380
1440
1500
1560
1620
1680
1740
1800
1860
1920
1980
2040
CAGGCTGAGC CTGTCACTG CCGACCACCT CCAGGCTCTCC TCCCTCCGGAG GCTACCCCTT 2100
GGAGGGAGGC TCTTCATCG GGCTGCGCGC TGCCAGAGGG CGGAGACAGA AGGCCTACAC 2160
GGTCCTCTTA ACGGAAAGG GTCCACGCTT TAGTCCTCAGA GACCGGCCC GGCGCGATGT 2220
TACCGAGAGC GAGAGCGAGA CCGCCGAGTA TCGGCAAGAG TAGCAGTGGA CCCTGGACGA 2280
AGACACCCAG GCAGGCGCAG AGCTGGCGGT GTGTCGGCGGC GCCCGCGAGG CCGGACCTGGT 2340
TCAGGGACTG CAGGACAGAG CTTCACTAGC GCAGGCTCAGG GCCTTCCGGC CCTGGCTTGA 2400
GCCTACCCCG GCCGCGACGC TGGCGCCCCC CCGCGCACC ACCGACCGGG CGCAGCCCGG 2460
GGGATTCGTT GTCCCCCCTT CTGGCTACCTC GCTGGCGCGG ACCTGCTGGG CTGGAGCGGC 2520
GGCCACTGCT CCCTGAGTTG CCGCTCCTGC GTGCGCTCTGC TTTCCCATGC GGAGTCTTCT 2580
CTGCTCCCAG CCTCTGCTGG TCTGCGCTGG CCTCTCCGCC GAGTGGCTAGG CGCCGGAGCT 2640
CCCTACAGA AGCTTGGCCA TGGAACCTCTT CCTCCCGCGT CCGCTTGGG GAGCTGGGGC 2700
ATGACACCAA ACCGTCCTCC TGGCTCTCAG GGGAGCTCCT ACCGCAACCG CAGGGACTGC 2760
AGTGTGTGCC CTGGGCTGTT CTCGACCCCC GGAAGGGAGG GGGCTCGGGC CATTGAGCCA 2820
CCACCTACAG CCCAGTGCCG TCACAGCTGG TGTGCTGGCT GAGGATGCG GTGGCGTGA 2880
GGTGGGCGCC GGATGACAGA CATGATAAG CAGTATTGGAG AGTGTGGGCA ACACTACACT 2940
AGAATGTGCC GAAAAATTGT CTTTATTGT GAAAAATGTG ATGTCATGCG TTTATTTGTA 3000
ACCATTATAA GCTGCAATTAA ACAAGTTAAC AACAACAATT GCATCATTT TATGGTTTCA 3060
GTTCAAGGGG AGGTTGTGGGA GTTTTTTCTG GATCCCTCTAG AGTCAGCTCT AGANNNNNNN 3120
NNNNNNNNN NNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN NNNNNNNNNN 3180
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN NNNNNNNNNN 3240
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN NNNNNNNNNN 3300
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN NNNNNNNNNN 3360
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN NNNNNNNNNN 3420
CATTTGGCAT GAGCGTTCAG CCCACCGAGT CTGGATGCTGC TGCGCAGCAT CTGCACAGTA 3480
GGGATCGTCT ACCTCCTTTC GAGACAGAAA CCGCGCTCAC TACACTGCG GACATCCGCC 3540
TGGTCGCCAG AGCTAACACT TTGACAATGC ACAAGTGCGG TTACGCGCGG TCTCTCCTCCT 3600
GCAGTGGTGG ATTACCACTG ATTACAGAT GGGTGTGTCC CTGGGATGAG GTCTTAACGC 3660
GGAGAGGACT GTGAATCTCG AGGAAGGTGA TGCGACGTGG CCTGTGGTCT GCAACAGATT 3720
ATACTAGCAC GAGCATGAGT ATCCATTGTT AGCAGCTCTG GGCTCTCAGA TGTCATTTGT 3780
CCAGTCCCCG TCCCTCGCAAG TCTTACCGG GCCGGCAAGG TTTGCCGACGC TGGTTTACGA 3840
TTGTGGGAGG TTGCTGGTAC TTAAATACAGA ATGGTTATATG GTACCGGGAG GTGGTTGAGT 3900
ACAACATGCC AAAAGAGGTA ATGGTTATGT CCAGGCGTTT TATGAGGGGT GCACACTTAA 3960
TCTACCTGCG CTGTGTGTAT GATGGCCACG TGAGTTCTGT GGTCCCACC ATGAGCTTTG 4020
GATAACGGGC CTTGCACTGT GGGATATTTG ACAATATTGT GGTGTGTGGC TGCAGTTACT 4080
GTGCTGTATTG AAGTGAAGAT AGGGTGCCG CTGGTGCACC GAGGACAGG CGCCTATGCG 4140
TGGCCGGGGGT GCGAATCATT GCTGAGGAGA CCAGCTCCAT GTGTATTCG TGCAAGACGG 4200
AGCGGGGGCC CGAGCAAGTTT ATTCGCGGC CCAGGATAGA CCACCGCCTT ATCCTGATGC 4260
ACGATTAGGA CTCTACCCCC ATGTAAGGAT CCCATCAGT AGTGCAGGCCGG GGGGATCCA 4320
GACATGATAA GATCATTGAG TGAATTGAGA CAACAACACAA G TGAAATTTT CAAGGCGCTAT 4380
TGCTTTATTT GTGAAATGGT TGCTATATTG TTATTTATTG TAACCATTAT AGAAGCTGAAAT 4440
AAACAAGTTT ACACAACAAAC TGTGATTCAT TTATTTTTT AGGTTCTAGG GGGATCTGTTG 4500
GAGGTGTTTG CCGATCTCTT AGAGTGACC TGCAAGCTAG CAGCTGCTAG ATAACTGAGCA 4560
TGGCGGGTTA ATCCACATTG ACAAGGAGCC CCTATCTGATG GATGCTGACCC CTCCTCCTCT 4620
GGCGGGCTGG TCGTCTACGT AGGGCGGGGG ACACAAAAAT GCCAGACGGC GGGGCTTTGC 4680
CCGGGCGGCC TACGAGGGCA AGGACGGGCG CAGCTGCGGT ATAGCCAGAG AGGCCGGCAC 4740
CGATCGCCCT TCCGACACTG CGCCAGCGCT GAATGCCGAA TGAANCTTCC AGACAGATGA 4800
GCGTCAAAAT AGTGGATTTT TTCAAGGGGT TTTTCTGCTG GCAATGCCCT GCGTAAATAT 4860
TGCTTTGATT ATTACCGAGA AGGCGGATAG TTTGATTTCT TCTACTAGGCA CAAGGTGATGT 4920
TATTACAAAT CAAAGAAGTA TTGCACCAAC GTGTATATTT CTGTGATGGA AGACTCTTCTT 4980
ACTCGGTTGG CTCACTACGT ATAAAAACAC TTCTCAAGGT TCTGGCGCTG CTGTCTGGTC 5040
TAAAATCCCT TTACTCGGCC TCTCTGTATT CTCCGCGCTT GATTTAACAG AAGGAAACAGCG 5100
GTATTACGTT CTCGTCAAAG CAACCATGAT AGCGGCGCTG TAGCGGCGGA TTAACGCGGG 5160
CGGTTGTGAT GTTACCGGGC AGCGTGACCG CTACACTTGC CAGCAGCCTA GCGCCCGCTC 5220
CTTCTGGCTT CTCGCGCCCA GTTGCAGCGG CTTCCCGGGT CAAAGCTCTA 5280
ATCGGGGGCG CCTTTTTGAG TCAGGATTTA GTGCTTTATG GACCACTGAC CCCCACACAC 5340
TTGATTAGGG TGATGTGGTC AGTAGGGCCC CTTGGCCCTT AATGACCGGG TTCTGCCCCCT 5400
TGACGGGGA CGAACCCCTT TTTAATAGT GACTCTGTGT CCACACTGTA ACAACACTCA 5460
ACCCATATCT CTGCTATTCT TTGATATTG AAGGATTTT GCGGATTTGC GCCAATTGCT 5520
TAAAGATAGA CACGATTAAA CAAAAAAAA ACAAGGAGAT TAAACAAATA TAAACTTTTA 5580
CAATTAAAT ATTGCTTTAT ACACCTTTTG TGTTTGTGGG GCTTTTCTGA TTATCAACCAG 5640
GAGATCATAT GATTGACAGT CATGTTTTAC GATTACGGTT CATGATCTT CTGGTTTGGCT 5700
CCGACTCTTC AGGCAATGAC CTGATAGCCCT TGTGATAGAC CTCGAAAATA TAGCTACCTT 5760
CTTCGCGCATG ATTTATCAC CTAATTGCGG TGAATATCAT AT GGATTTGTT ATTTGACTGT 5820
CTTCGGCTTT TCTCACCCGCT TGATTCTTTT ACCTAACAAT TACTCAAGGC TTGCAATTAA 5880
AATATATGAG GTTCTAAAAT ATTTTTATCC TTGCCTTGAA ATAAAGGCTT CTCCCGCAAA 5940
AGTATTACAG GGTCTATAATG TTTTTGTTGAC AACCGATTGA GCTTTATGCT CTGAGGCTTT 6000
ATTGGTTAAT TTGGCTATTT CTGCTTCTTG CCTGATGATG TTATTGGATG TTGGAAATTT 6060
CTGATGCGGT ATTTTCTCTC TACGCATCTG TGCTGTATTT CACACCCAG ATGGTGCACT 6120
CTCAGTACAA TCCTGCTCTGA TGCCGCATAG TTTAAGCCAG CCCGACACC GCCAAACCC 6180
GCTGACGCCC CCGACGGGCG TTGTCTGCTC CCGCGATCCG TTCAGACGAA AGCCTGAAAC 6240
GCTCACCAGGA GTACGATGTT CGACAGGTTT TCACCGTCAT CACCGGAAAG CGCAGAAGGA 6300
AAGGGCTCTC TGATAGCTCT ATTTTTATAG GTTAAAGTCA TGATAAATAT GTTTCTTAG 6360
AGCTCAGGTA GCACCTTTGC GGAAATAATG CGCGGAAACC CATTTTTGTT ATTTTTCTAA 6420
ATACATCTAA ATATGATACG GGTACTGACA GAATAACCTT GTAATAAGCTC TCAATTATAT 6480
TGAAAGAGGA AAGATAGTTA ATATACACAT TCGTGTGTCC CCGTTTATCC CTTTTTTGCG 6540
GCATTATGAC TTCCCTTTAT TGCTCACCAA GAAACGCTTG TGAAAGTAAAG AGATGCCTGA 6600
GATCAGTTTG GGTGACGACT GGTTCTACAT GAACTGCGAT CCAAGACGCG TAAAGATCCCT 6660
GAGAGTTTTC GCCCGGAGAG AGCTTTTTCA ATGATAGACA CTTTTAAGTT TGCTGTATGT 6720
GGCGCGTAAA TATCTGAACT TGGCGCGGAG CAGAGGCAAC TCGGTGCGCG CATACACTAT 6780
TCTCAAGAGT CTCCCTGTTG GTCTGCACAA GTCTGACGAAA ACGATCTTAC GGATGCGATG 6840
ACAGTAAAGAG AATTATGACG TGCTGCCATA ACCATGAGTG AATACACTTG GCCAACCTTTA 6900
TCTCTGACAA CGATCGGAAG ACCAGAGGAG CTAACCGCTT TTGTGCACAA CATACTGGAT 6960
CATGTAACCT GCCTGTGACTG TTGGGAACCG GAGCTGAATAG AAGCCATACC AAACAGGAG 7020
CGTGACACCA CGATGCCTTG AGCAATGCGA ACAACGTTTG GCAAGACTATT AACTGGCGAA 7080
CTACTTACTC TAGCTCCTCCG GCAACAATTTA ATAGACTGGG TGGGAGCGGA TAAAGTTGCA 7140
GGACACCTTC TGCCGCTGCG CCTCCCGCGT GCGTTGTTTA TGCTGTGATA ATCTGGAGCC 7200
GGTGAGCGTG TGCTGCGGCG TATCATTTGCA GCACTGGGGG CAGATGGTTA GCCCTCCGC 7260
ATGATAGTTA TCTTACAGAC GGGGATCAG GCAACTATGG ATGAAACGAAA TAGACAGATC 7320
GCTGAGATAC GTGCCCACTG GATTAAGCAT TTGGTACGTT CAGACCAAGT ATCTCATTAT 7380
ATACCTAGTA TTGTATTTAT AACTCATATT TAAATTAAA GGATCTAGGT GAAGATCCCT 7440
TTTATAATTC TCTAGACCAA AATTCCTTTA CGTGAAGTTT GCTTCCACCTG AGCCTGACAG 7500
CCCGTAGAAA AGATAAAGGG ATCTTCTTTG GATCTTTTTT TTCTGCGGCT TATCTGCTGC 7560
TTGCAAACAA AAAAAACACC GCTACGCGCC GTGGTTTGTG TCCGGGATCA AGACTCTACA 7620
ACTCTTTTTC CGAAGGTTAAC TGCGCTCAGC AGAGCGCAGA TACAAATATG TGCTCTTCTA 7680
GTTGAGCGTG AGTTAGGCCA CCACTTCAAG AACTCGTATT CACCGGCTAC ATACCTGCTG 7740
CTGTCTAATC TGTACAGCAG GTGCTGTGCC AGTGGCGATA AGTGTGTCT CACCGGGTTG 7800
GACTCAAGAC GATAAGTACC GGATAAGGCG CAGCGGTCGG GCTGAACGGG GGGTTCTGC  
ACACAGCCCA GCTTGGAGCC AAGGACCTAC ACCGAAGTGA GATAACTACA GCGTGAGCTA  
TGAGAAAGCC CCAGGCTTCC CGAAGGGAAG AAGCCGAGCA GGTATCCCCGTA AAGCGGCAGG  
GTCGGAACAG GAGAGCCAC GAGGGAGCTT CCAGGGGGAA ACGCCTGTTA TCTTTATAGT  
CCTGTGGGT TTGGCCACCT CTGACTTGAG CGTGATTTTT TGTGATGCTC GTCAGGGGGG  
CGAGGCCTAT GGAAAACGCC CAGCAAGCGG GCCTTTTTAC GGGTCTGCCC CTGTGTGCTGG  
CCTTTTGCTC ACGTTTCTT TCCCGGATAA TCCCTCGATT CGTGAGATAA CCCTATTACC  
GCCTTTGAGT GAGCTGATA CGCTGCGCCG AGCGAAGCGA CCGAGGCGAG CGAGTCAGTG  
AGCGAGGAAG CGGAAGAGC  

(2) INFORMATION FOR SEQ ID NO:4:  

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  

(ii) MOLECULE TYPE: other nucleic acid  

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:  

GCAGGTACC CGAGTCGAT CTACAC  26  

(2) INFORMATION FOR SEQ ID NO:5:  

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  

(ii) MOLECULE TYPE: other nucleic acid  

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:  

CTGTCTGAGC TAGAGCTC  18
WHAT IS CLAIMED IS:

1. A vector comprising a chimpanzee adenovirus DNA sequence and a selected heterologous gene operatively linked to regulatory sequences which direct expression of said gene in a heterologous host cell.

2. The vector according to claim 1 wherein said chimpanzee adenovirus sequence comprises at least 5' and 3' cis-elements necessary for replication and virion encapsidation, said cis-elements flanking said selected gene and regulatory sequences.

3. The vector according to claim 1 wherein said chimpanzee adenovirus sequence has a deletion in all or a part of the E1 gene.

4. The vector according to claim 1 wherein said chimpanzee adenovirus sequence comprises the sequence of SEQ ID NO: 1 or a fragment thereof.

5. The vector according to claim 1 wherein said chimpanzee adenovirus sequence comprises the sequence of SEQ ID NO: 2 or a fragment thereof.

6. A host cell transfected with the vector of claim 1.

7. A human cell that expresses a selected gene introduced therein through transduction of the vector of claim 1.

8. A non-simian mammalian cell line that expresses a chimpanzee adenovirus gene.
9. The cell line according to claim 8 wherein said gene is an adenovirus E1 gene or a functional fragment of said E1 gene.

10. The cell line according to claim 8 wherein said chimpanzee adenovirus gene is obtained from the sequence of SEQ ID NO: 1.

11. The cell line according to claim 8 wherein said chimpanzee adenovirus gene is obtained from the sequence of SEQ ID NO: 2.

12. A pharmaceutical composition comprising a recombinant adenovirus vector in a pharmaceutically acceptable carrier, said vector comprising a chimpanzee adenovirus DNA sequence and a selected heterologous gene operatively linked to regulatory sequences which direct expression of said gene in a host cell.

13. A method for delivering a heterologous gene to a mammalian cell comprising introducing into said cell an effective amount of the vector of claim 1.

14. A method for producing a selected gene product comprising infecting a mammalian cell with the vector of claim 1, culturing said cell under suitable conditions and isolating and recovering from said cell culture the expressed gene product.
15. The use of a vector comprising a chimpanzee adenovirus DNA sequence and a selected heterologous gene encoding an antigen of an infective agent operatively linked to regulatory sequences which direct expression of said gene in the production of a medicament for eliciting an immune response in a mammalian host against said infective agent.

16. The use of a vector comprising a chimpanzee adenovirus DNA sequence and a selected heterologous therapeutic gene operatively linked to regulatory sequences which direct expression of said gene in the production of a medicament for treating a patient having an acquired or inherited genetic defect.
<table>
<thead>
<tr>
<th>Protein C1/C68 aa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E1 and pIX Regions</strong></td>
</tr>
<tr>
<td>E1A 6/11K 58/101</td>
</tr>
<tr>
<td>E1A 25K 231/226</td>
</tr>
<tr>
<td>E1A 28K 262/257</td>
</tr>
<tr>
<td>E1B 11K 181/186</td>
</tr>
<tr>
<td>E1B 55K 495/498</td>
</tr>
<tr>
<td>E1B 8.3K 91/102</td>
</tr>
<tr>
<td>pIX 139/143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>E2 and IVa2 Regions</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>E2A 516/513</td>
</tr>
<tr>
<td>E2B pTF 649/628</td>
</tr>
<tr>
<td>E2B pol 1121/1125</td>
</tr>
<tr>
<td>IVa2 448/448</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>E3 Region</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>E3M1 106/106</td>
</tr>
<tr>
<td>E3M2 146/209</td>
</tr>
<tr>
<td>E3M3 172/176</td>
</tr>
<tr>
<td>E3Hy 184/204</td>
</tr>
<tr>
<td>E3Hy 188/204</td>
</tr>
<tr>
<td>E3Hy 103/295</td>
</tr>
<tr>
<td>E3B4 412</td>
</tr>
<tr>
<td>E3B5 134/143</td>
</tr>
<tr>
<td>E3B6 135/135</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>E4 Region</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>E4 123/124</td>
</tr>
<tr>
<td>E4 129/129</td>
</tr>
<tr>
<td>E4 117/117</td>
</tr>
<tr>
<td>E4 124/121</td>
</tr>
<tr>
<td>E4 303/301</td>
</tr>
<tr>
<td>E4 83 64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Late Region</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 16.6K 139/139</td>
</tr>
<tr>
<td>L1 52/55 K</td>
</tr>
<tr>
<td>L2 565/592</td>
</tr>
<tr>
<td>L2 564/534</td>
</tr>
<tr>
<td>L2 192/193</td>
</tr>
<tr>
<td>L2 353/343 M-Core</td>
</tr>
<tr>
<td>L2 76</td>
</tr>
<tr>
<td>L3 250/242</td>
</tr>
<tr>
<td>L3 956/933</td>
</tr>
<tr>
<td>L3 23K 207/206</td>
</tr>
<tr>
<td>L4 126K 828/804</td>
</tr>
<tr>
<td>L4 22K 197/188</td>
</tr>
<tr>
<td>L4 33K 231/222</td>
</tr>
<tr>
<td>L4 227/227</td>
</tr>
<tr>
<td>L5 322/425</td>
</tr>
</tbody>
</table>

Fig. 2

2/11
C1 E1 Expression Plasmid

Fig. 6A

pGPKG
5.5 kb

PGK Promoter

SnaBI
Xbal

EcoRV
Xbal

pGPKG-C1 MU1.3-6.6
Clone 499
7.4 kb

PGK Promoter

Kpnl
KspI

BamHI Partial
Xbal

Fig. 6D

pGPKG-C1 E1-ATG
Clone 620
7.2 kb

PGK Promoter

Fig. 6E

pBS-C1-BamI
Clone 293
4.4 kb

SnaBI
Xbal

pGPKG-C1 E1
Clone 733
8.2 kb

Fig. 6F

Fig. 6G

Fig. 6B

0 1.3 1.6 6.6 Map Unit

SnaBI
E1 ATG
Kpnl
BamHI
BamHI
Xbal

Site

C1 Genome

C1 Genome

Kpnl
SF-40 SF-41

PCR Primers

PCR Product

6/11
GENERATION OF RECOMBINANT C68-CMV-GFP VIRUS BY HOMOLOGOUS RECOMBINATION

Fig. 10A

Fig. 10B

Fig. 10C

Fig. 10D

C-68 Genome
(36519 bp)

C-68-Sapi-A
(35199 bp)

Transfect E1 Cell Line
Agar Overlay
2 Weeks
Isolate plaques

C-68-CMV-GFP
(35511 bp)
CONSTRUCTION OF C1 GENOME WITH UNIQUE NOT-I SITE

Fig. 11A

pNEB-C1-Bam-G
#516
(4642 bp)

Fig. 11B

C1 Genome
(35524 bp)

Not-I Linker
GCGGCCGC
CGCCGCGC

Fig. 11D

Spe-I
Fill-In
Phosphatase

Fig. 11E

pNEB-C1-Ascl-B-Not-I
#955
(10657 bp)

Fig. 11G

C1-Ascl-A
(27587 bp)

Ascl Gel Purify

Fig. 11F

C1 Genome
(35524 bp)

Spe-I

Fig. 11H

C1 - Not-I
(35556 bp)
**INTERNATIONAL SEARCH REPORT**

**International Application No**
PCT/US 97/15694

---

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 6 C12N15/86 C12N5/10 A61K48/00

According to international Patent Classification (IPC) or to both national classification and IPC

---

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K C07K

Documented searched other than minimum documentation to the extent that such documents are included in the fields searched

---

Electronic database consulted during the international search (name of database and, where practical, search terms used)

---

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>WO 94 26914 A (RHÔNE-POULENC RORER S.A.) 24 November 1994 see page 2, line 33 - page 3, line 26</td>
<td>1-3,6-9, 12-16</td>
</tr>
</tbody>
</table>

---

Further documents are listed in the continuation of box C.

---

| Patent family members are listed in annex. |

---

1. Special categories of cited documents:
   - **A** document defining the general state of the art which is not considered to be of particular relevance
   - **E** earlier document but published on or after the international filing date
   - **L** document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
   - **O** document referring to an oral disclosure, use, exhibition or other means
   - **P** document published prior to the international filing date but later than the priority date claimed

2. Later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
3. **X** document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
4. **Y** document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

---

**Date of the actual completion of the international search**

21 January 1998

**Date of mailing of the international search report**

11.02.98

---

**Name and mailing address of the ISA**

European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx: 31 651 epos nl Fax: (+31-70) 340-3016

---

**Authorized officer**

Cupido, M

---

Form PCT/ISA/210 (second sheet) (July 1992)

---

page 1 of 2
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>R. WIGAND ET AL.: &quot;Chimpanzee adenoviruses are related to four subgenera of human adenoviruses&quot; INTERVIROLOGY, vol. 30, no. 1, January 1989 - February 1989, pages 1-9, XP002052837 cited in the application see page 1; table 4</td>
<td>1-16</td>
</tr>
</tbody>
</table>
INTERNATIONAL SEARCH REPORT

Box I  Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☑ Claims Nos., because they relate to subject matter not required to be searched by this Authority, namely:
   see FURTHER INFORMATION sheet PCT/ISA/210

2. ☐ Claims Nos., because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos., because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II  Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims, it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.
Remark: Although claim 13, insofar an in vivo method is concerned, is
directed to a method of treatment of the human or animal body, the search
has been carried out and based on the alleged effects of the vector.
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>W0 9426914 A</td>
<td>24-11-94</td>
<td>FR 2705361 A</td>
<td>25-11-94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 6787894 A</td>
<td>12-12-94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR 9406720 A</td>
<td>06-02-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2163256 A</td>
<td>24-11-94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 1124040 A</td>
<td>05-06-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CZ 9503028 A</td>
<td>14-02-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0698108 A</td>
<td>28-02-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FI 955552 A</td>
<td>27-12-95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HU 73465 A</td>
<td>28-08-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 8510122 T</td>
<td>29-10-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO 954466 A</td>
<td>07-11-95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PL 311660 A</td>
<td>04-03-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SK 144795 A</td>
<td>03-04-96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZA 9403358 A</td>
<td>16-01-95</td>
</tr>
</tbody>
</table>