(54) Titre : ADENOVIRUS AMELIORE ET SES PROCEDES D'UTILISATION

(54) Title: RECOMBINANT ADENOVIRUS AND METHODS OF USE THEREOF

(57) Abrégé/Abstract:
A recombinant adenovirus and a method for producing the virus are provided which utilize a recombinant shuttle vector comprising adenovirus DNA sequence for the 5' and 3' cis-elements necessary for replication and virion encapsidation in the absence of sequence encoding viral genes and a selected minigene linked thereto, and a helper adenovirus comprising sufficient adenovirus gene sequences necessary for a productive viral infection. Desirably the helper gene is crippled by modifications to its 5' packaging sequences, which facilitates purification of the viral particle from the helper virus.
**Title:** RECOMBINANT ADENOVIRUS AND METHODS OF USE THEREOF

**Abstract**

A recombinant adenovirus and a method for producing the virus are provided which utilize a recombinant shuttle vector comprising adenovirus DNA sequence for the 5' and 3' cis-elements necessary for replication and virion encapsidation in the absence of sequence encoding viral genes and a selected minigene linked thereto, and a helper adenovirus comprising sufficient adenovirus gene sequences necessary for a productive viral infection. Desirably the helper gene is crippled by modifications to its 5' packaging sequences, which facilitates purification of the viral particle from the helper virus.
RECOMBINANT ADENOVIRUS AND METHODS OF USE THEREOF

This invention was supported by the National Institute of Health Grant No. P30 DK 47757. 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 thereof.

Background of the Invention

Human gene therapy is an approach to treating 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. 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 that 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 ($10^{13}$ plaque forming units/ml) of recombinant virus can be easily generated in 293 cells (the adenovirus equivalent

Human adenoviruses are comprised 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, *Virology*, 2d edit., ed. B. N. Fields, Raven Press, Ltd., New York (1990)].

The first-generation recombinant, replication-deficient adenoviruses which have been developed for gene therapy contain deletions of the entire E1a and part of the E1b regions. This replication-defective virus is grown on an adenovirus-transformed, complementation human
embryonic kidney cell line containing a functional adenovirus E1a gene which provides a transacting E1a protein, the 293 cell [ATCC CRL1573]. E1-deleted viruses are capable of replicating and producing infectious virus in the 293 cells, which provide E1a and E1b region gene products in trans. The resulting virus is capable of infecting many cell types and can express the introduced gene (providing it carries its own promoter), but cannot replicate in a cell that does not carry the E1 region DNA unless the cell is infected at a very high multiplicity of infection.

However, in vivo studies revealed transgene expression in these E1 deleted vectors was transient and invariably associated with the development of severe inflammation at the site of vector targeting [S. Ishibashi et al, J. Clin. Invest., 93:1885-1893 (1994); J. M. Wilson et al, Proc. Natl. Acad. Sci., USA, 85:4421-4424 (1988); J. M. Wilson et al, Clin. Bio., 3:21-26 (1991); M. Grossman et al, Som. Cell. and Mol. Gen., 17:601-607 (1991)]. One explanation that has been proposed to explain this finding is that first generation recombinant adenoviruses, despite the deletion of E1 genes, express low levels of other viral proteins. This could be due to basal expression from the unstimulated viral promoters or transactivation by cellular factors. Expression of viral proteins leads to cellular immune responses to the genetically modified cells, resulting in their destruction and replacement with nontransgene containing cells.

There yet remains a need in the art for the development of additional adenovirus vector constructs for gene therapy.
Summary of the Invention

In one aspect, the invention provides the components of a novel recombinant adenovirus production system. One component is a shuttle plasmid, pAdΔ, that comprises adenovirus cis-elements necessary for replication and virion encapsidation and is deleted of all viral genes. This vector carries a selected transgene under the control of a selected promoter and other conventional vector/plasmid regulatory components. The other component is a helper adenovirus, which alone or with a packaging cell line, supplies sufficient gene sequences necessary for a productive viral infection. In a preferred embodiment, the helper virus has been altered to contain modifications to the native gene sequences which direct efficient packaging, so as to substantially disable or "cripple" the packaging function of the helper virus or its ability to replicate.

In another aspect, the present invention provides a unique recombinant adenovirus, an AdΔ virus, produced by use of the components above. This recombinant virus comprises an adenovirus capsid, adenovirus cis-elements necessary for replication and virion encapsidation, but is deleted of all viral genes (i.e., all viral open reading frames). This virus particle carries a selected transgene under the control of a selected promoter and other conventional vector regulatory components. This AdΔ recombinant virus is characterized by high titer transgene delivery to a host cell and the ability to stably integrate the transgene into the host cell chromosome. In one embodiment, the virus carries as its transgene a reporter gene. Another embodiment of the recombinant virus contains a therapeutic transgene.

In another aspect, the invention provides a method for producing the above-described recombinant AdΔ virus by co-transfecting a cell line (either a packaging cell
line or a non-packaging cell line) with a shuttle vector or plasmid and a helper adenovirus as described above, wherein the transfected cell generates the AdΔ virus. The AdΔ virus is subsequently isolated and purified therefrom.

In yet a further aspect, the invention provides a method for delivering a selected gene to a host cell for expression in that cell by administering an effective amount of a recombinant AdΔ virus containing a therapeutic transgene to a patient to treat or correct a genetically associated disorder or disease.

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 Figures

Fig. 1A is a schematic representation of the organization of the major functional elements that define the 5' terminus from Ad5 including an inverted terminal repeat (ITR) and a packaging/enhancer domain. The TATA box of the E1 promoter (black box) and E1A transcriptional start site (arrow) are also shown.

Fig. 1B is an expanded schematic of the packaging/enhancer region of Fig. 1A, indicating the five packaging (PAC) domains (A-repeats), I through V. The arrows indicate the location of PCR primers referenced in Figs. 9A and 9B below.

Fig. 2A is a schematic of shuttle vector pAdΔ.CMVlacZ containing 5' ITR from Ad5, followed by a CMV promoter/enhancer, a LacZ gene, a 3' ITR from Ad5, and remaining plasmid sequence from plasmid pSP72 backbone. Restriction endonuclease enzymes are represented by conventional designations in the plasmid constructs.
6

Fig. 2B is a schematic of the shuttle vector digested with EcoRI to release the modified AdΔ genome from the pSP72 plasmid backbone.

Fig. 2C is a schematic depiction of the function of the vector system. In the presence of an E1-deleted helper virus Ad.CBhpAP which encodes a reporter minigene for human placenta alkaline phosphatase (hpAP), the AdΔ.CMVlacZ genome is packaged into preformed virion capsids, distinguishable from the helper virions by the presence of the LacZ gene.

Figs. 3A to 3F [SEQ ID NO: 1] report the top DNA strand of the double-stranded plasmid pAdΔ.CMVlacZ. The complementary sequence may be readily obtained by one of skill in the art. The sequence includes the following components: 3' Ad ITR (nucleotides 607-28 of SEQ ID NO: 1); the 5' Ad ITR (nucleotides 5496-5144 of SEQ ID NO: 1); CMV promoter/enhancer (nucleotides 5117-4524 of SEQ ID NO: 1); SD/SA sequence (nucleotides 4507-4376 of SEQ ID NO: 1); LacZ gene (nucleotides 4320-845 of SEQ ID NO: 1); and a poly A sequence (nucleotides 837-639 of SEQ ID NO: 1).

Fig. 4A is a schematic of shuttle vector pAdΔc.CMVlacZ containing an Ad5 5' ITR and 3' ITR positioned head-to-tail, with a CMV enhancer/promoter-LacZ minigene immediately following the 5' ITR, followed by a plasmid pSP72 (Promega) backbone. Restriction endonuclease enzymes are represented by conventional designations in the plasmid constructs.

Fig. 4B is a schematic depiction of the function of the vector system of Fig. 4A. In the presence of helper virus Ad.CBhpAP, the circular pAdΔc.CMVlacZ shuttle vector sequence is packaged into virion heads, distinguishable from the helper virions by the presence of the LacZ gene.
Figs. 5A to 5F [SEQ ID NO: 2] report the top DNA strand of the double-stranded vector pAdΔc.CMVlacZ. The complementary sequence may be readily obtained by one of skill in the art. The sequence includes the following components: 5' Ad ITR (nucleotides 600-958 of SEQ ID NO: 2); CMV promoter/enhancer (nucleotides 969-1563 of SEQ ID NO: 2); SD/SA sequence (nucleotides 1579-1711); LacZ gene (nucleotides 1762-5236 of SEQ ID NO: 2); poly A sequence (nucleotides 5245-5443 of SEQ ID NO: 2); and 3' Ad ITR (nucleotides 16-596 of SEQ ID NO: 2).

Fig. 6 is a schematic of shuttle vector pAdΔ.CBFCTR containing 5' ITR from Ad5, followed by a chimeric CMV enhancer/β actin promoter enhancer, a CFTR gene, a poly-A sequence, a 3' ITR from Ad5, and remaining plasmid sequence from plasmid pSL1180 (Pharmacia) backbone. Restriction endonuclease enzymes are represented by conventional designations in the plasmid constructs.

Figs. 7A to 7H [SEQ ID NO: 3] report the top DNA strand of the double-stranded plasmid pAdΔ.CBFCTR. The complementary sequence may be readily obtained by one of skill in the art. The sequence includes the following components: 5' Ad ITR (nucleotides 9611-9254 of SEQ ID NO: 3); chimeric CMV enhancer/β actin promoter (nucleotides 9241-8684 of SEQ ID NO: 3); CFTR gene (nucleotides 8622-4065 of SEQ ID NO: 3); poly A sequence (nucleotides 3887-3684 of SEQ ID NO: 3); and 3' Ad ITR (nucleotides 3652-3073 of SEQ ID NO: 3). The remaining plasmid backbone is obtained from pSL1180 (Pharmacia).

Fig. 8A illustrates the generation of 5' adenovirus terminal sequence that contained PAC domains I and II by PCR. See, arrows indicating righthand and lefthand (PAC II) PCR probes in Fig. 1B.
Fig. 8B illustrates the generation of 5' terminal sequence that contained PAC domains I, II, III and IV by PCR. See, arrows indicating righthand and lefthand (PAC IV) PCR probes in Fig. 1B.

Fig. 8C depicts the amplification products subcloned into the multiple cloning site of pAd.Link.1 (IHGT Vector Core) generating pAd.PACII (domains I and II) and pAd.PACIV (domains I, II, III, and IV) resulting in crippled helper viruses, Ad.PACII and Ad.PACIV with modified packaging (PAC) signals.

Fig. 9A is a schematic representation of the subcloning of a human placenta alkaline phosphatase reporter minigene containing the immediate early CMV enhancer/promoter (CMV), human placenta alkaline phosphatase cDNA (hpAP), and SV40 polyadenylation signal (pA) into pAd.PACII to result in crippled helper virus vector pAd.A.PACII.CMVhpAP. Restriction endonuclease enzymes are represented by conventional designations in the plasmid constructs.

Fig. 9B is a schematic representation of the subcloning of the same minigene of Fig. 9A into pAd.PACIV to result in crippled helper virus vector pAd.PACIV.CMV.hpAP.

Fig. 10 is a flow diagram summarizing the synthesis of an adenovirus-based polycation helper virus conjugate and its combination with a pAdA shuttle vector to result in a novel viral particle complex. CsCl band purified helper adenovirus was reacted with the heterobifunctional crosslinker sulfo-SMCC and the capsid protein fiber is labeled with the nucleophilic maleimide moiety. Free sulfhydryls were introduced onto poly-L-lysine using 2-iminothiolane-HCl and mixed with the labelled adenovirus, resulting in the helper virus conjugate Ad-pLys. A unique adenovirus-based particle is generated by purifying the Ad-pLys conjugate over a CsCl gradient to
remove unincorporated poly-L-lysine, followed by extensively dialyzing, adding shuttle plasmid DNAs to Ad-pLys and allowing the complex formed by the shuttle plasmid wrapped around Ad-pLys to develop.

Fig. 11 is a schematic diagram of pCCL-DMD, which is described in detail in Example 9 below.

Fig. 12A - 12P provides the continuous DNA sequence of pAdΔ.CMVMdys [SEQ ID NO:10].

Detailed Description of the Invention

The present invention provides a unique recombinant adenovirus capable of delivering transgenes to target cells, as well as the components for production of the unique virus and methods for the use of the virus to treat a variety of genetic disorders.

The AdΔ virus of this invention is a viral particle containing only the adenovirus cis-elements necessary for replication and virion encapsidation (i.e., ITRs and packaging sequences), but otherwise deleted of all adenovirus genes (i.e., all viral open reading frames). This virus carries a selected transgene under the control of a selected promoter and other conventional regulatory components, such as a poly A signal. The AdΔ virus is characterized by improved persistence of the vector DNA in the host cells, reduced antigenicity/immunogenicity, and hence, improved performance as a delivery vehicle. An additional advantage of this invention is that the AdΔ virus permits the packaging of very large transgenes, such as a full-length dystrophin cDNA for the treatment of the progressive wasting of muscle tissue characteristic of Duchenne Muscular Dystrophy (DMD).

This novel recombinant virus is produced by use of an adenovirus-based vector production system containing two components: 1) a shuttle vector that comprises adenovirus cis-elements necessary for replication and
virion encapsidation and is deleted of all viral genes, which vector carries a reporter or therapeutic minigene and 2) a helper adenovirus which, alone or with a packaging cell line, is capable of providing all of the viral gene products necessary for a productive viral infection when co-transfected with the shuttle vector. Preferably, the helper virus is modified so that it does not package itself efficiently. In this setting, it is desirably used in combination with a packaging cell line that stably expresses adenovirus genes. The methods of producing this viral vector from these components include both a novel means of packaging of an adenoviral/transgene containing vector into a virus, and a novel method for the subsequent separation of the helper virus from the newly formed recombinant virus.

I. The Shuttle Vector

The shuttle vector, referred to as pAda, is composed of adenovirus sequences, and transgene sequences, including vector regulatory control sequences.

A. The Adenovirus Sequences

The adenovirus nucleic acid sequences of the shuttle vector provide the minimum adenovirus sequences which enable a viral particle to be produced with the assistance of a helper virus. These sequences assist in delivery of a recombinant transgene genome to a target cell by the resulting recombinant virus.

The DNA sequences of a number of adenovirus types are available from Genbank, including type Ad5 [Genbank Accession No. M73260]. The adenovirus sequences may be obtained from any known adenovirus serotype, such as serotypes 2, 3, 4, 7, 12 and 40, and further including any of the presently identified 41 human types [see, e.g., Horwitz, cited above]. Similarly adenoviruses known to infect other animals may also be employed in the
vector constructs of this invention. The selection of the adenovirus type is not anticipated to limit the following invention. A variety of adenovirus strains are available from the American Type Culture Collection, Rockville, Maryland, or available by request from a variety of commercial and institutional sources. In the following exemplary embodiment an adenovirus, type 5 (Ad5) is used for convenience.

However, it is desirable to obtain a variety of pAdΔ shuttle vectors based on different human adenovirus serotypes. It is anticipated that a library of such plasmids and the resulting AdΔ viral vectors would be useful in a therapeutic regimen to evade cellular, and possibly humoral, immunity, and lengthen the duration of transgene expression, as well as improve the success of repeat therapeutic treatments. Additionally the use of various serotypes is believed to produce recombinant viruses with different tissue targeting specificities. The absence of adenoviral genes in the AdΔ viral vector is anticipated to reduce or eliminate adverse CTL response which normally causes destruction of recombinant adenoviruses deleted of only the E1 gene.

Specifically, the adenovirus nucleic acid sequences employed in the pAdΔ shuttle vector of this invention are adenovirus genomic sequences from which all viral genes are deleted. More specifically, the adenovirus sequences employed are the cis-acting 5' and 3' inverted terminal repeat (ITR) sequences of an adenovirus (which function as origins of replication) and the native 5' packaging/enhancer domain, that contains sequences necessary for packaging linear Ad genomes and enhancer elements for the E1 promoter. These sequences are the sequences necessary for replication and virion encapsidation. See, e.g., P. Hearing et al, J. Virol., 61(8):2555-2558 (1987); M. Grable and P. Hearing, J.

According to this invention, the entire adenovirus 5' sequence containing the 5' ITR and packaging/enhancer region can be employed as the 5' adenovirus sequence in the pAdA shuttle vector. This left terminal (5') sequence of the Ad5 genome useful in this invention spans bp 1 to about 360 of the conventional adenovirus genome, also referred to as map units 0-1 of the viral genome. This sequence is provided herein as nucleotides 5496-5144 of SEQ ID NO: 1, nucleotides 600-958 of SEQ ID NO: 2; and nucleotides 9611-9254 of SEQ ID NO: 3, and generally is from about 353 to about 360 nucleotides in length. This sequence includes the 5' ITR (bp 1-103 of the adenovirus genome), and the packaging/enhancer domain (bp 194-358 of the adenovirus genome). See, Figs. 1A, 3, 5, and 7.

Preferably, this native adenovirus 5' region is employed in the shuttle vector in unmodified form.

However, some modifications including deletions, substitutions and additions to this sequence which do not adversely effect its biological function may be acceptable. See, e.g., WO 93/24641, published December 9, 1993. The ability to modify these ITR sequences is within the ability of one of skill in the art. See, e.g., texts such as Sambrook et al, "Molecular Cloning. A Laboratory Manual.", 2d edit., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989).

The 3' adenovirus sequences of the shuttle vector include the right terminal (3') ITR sequence of the adenoviral genome spanning about bp 35,353 - end of the adenovirus genome, or map units ~98.4-100. This sequence is provided herein as nucleotides 607-28 of SEQ ID NO: 1, nucleotides 16-596 of SEQ ID NO: 2; and nucleotides 3652-3073 of SEQ ID NO: 3, and generally is
about 580 nucleotides in length. This entire sequence is desirably employed as the 3' sequence of an pAdΔ shuttle vector. Preferably, the native adenovirus 3' region is employed in the shuttle vector in unmodified form.

However, some modifications to this sequence which do not adversely effect its biological function may be acceptable.

An exemplary pAdΔ shuttle vector of this invention, described below and in Fig. 2A, contains only those adenovirus sequences required for packaging adenoviral genomic DNA into a preformed capsid head. The pAdΔ vector contains Ad5 sequences encoding the 5' terminal and 3' terminal sequences (identified in the description of Fig. 3), as well as the transgene sequences described below.

From the foregoing information, it is expected that one of skill in the art may employ other equivalent adenovirus sequences for use in the AdΔ vectors of this invention. These sequences may include other adenovirus strains, or the above mentioned cis-acting sequences with minor modifications.

B. The Transgene

The transgene sequence of the vector and recombinant virus is a nucleic acid sequence or reverse transcript thereof, heterologous to the adenovirus sequence, which encodes a polypeptide or protein of interest. The transgene is operatively linked to regulatory components in a manner which permits transgene transcription.

The composition of the transgene sequence will depend upon the use to which the resulting virus will be put. For example, one type of transgene sequence includes a reporter sequence, which upon expression produces a detectable signal. Such reporter sequences include without limitation an E. coli beta-galactosidase
(LacZ) cDNA, a human placental alkaline phosphatase gene and a green fluorescent protein gene. These sequences, when associated with regulatory elements which drive their expression, provide signals detectable by conventional means, e.g., ultraviolet wavelength absorbance, visible color change, etc.

Another type of transgene sequence includes a therapeutic gene which expresses a desired gene product in a host cell. These therapeutic nucleic acid sequences typically encode products for administration and expression in a patient in vivo or ex vivo to replace or correct an inherited or non-inherited genetic defect or treat an epigenetic disorder or disease. Such therapeutic genes which are desirable for the performance of gene therapy include, without limitation, a normal cystic fibrosis transmembrane regulator (CFTR) gene (see Fig. 7), a low density lipoprotein (LDL) gene [T. Yamamoto et al, Cell, 32:27-28 (November, 1984)], a DMD cDNA sequence [partial sequences available from GenBank, Accession Nos. M36673, M36671, [A. P. Monaco et al, Nature, 323:646-650 (1986)] and L06900, [Roberts et al, Hum. Mutat., 2:293-299 (1993)] (Genbank), and a number of genes which may be readily selected by one of skill in the art. The selection of the transgene is not considered to be a limitation of this invention, as such selection is within the knowledge of the art-skilled.

C. Regulatory Elements

In addition to the major elements identified above for the pAdΔ shuttle vector, i.e., the adenovirus sequences and the transgene, the vector also includes conventional regulatory elements necessary to drive expression of the transgene in a cell transfected with the pAdΔ vector. Thus the vector contains a selected promoter which is linked to the transgene and located,
with the transgene, between the adenovirus sequences of
the vector.

Selection of the promoter is a routine matter
and is not a limitation of the pAdΔ vector itself.

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)]. This
promoter is found at nucleotides 5117-4524 of SEQ ID NO:
1 and nucleotides 969-1563 of SEQ ID NO: 2. Another
promoter is the CMV enhancer/chicken β-actin promoter
(nucleotides 9241-8684 of SEQ ID NO: 3). Another
desirable promoter includes, without limitation, the Rous
sarcoma virus LTR promoter/enhancer. Still other
promoter/enhancer sequences may be selected by one of
skill in the art.

The shuttle vectors will also desirably contain
nucleic acid sequences heterologous to the adenovirus
sequences including sequences providing signals required
for efficient polyadenylation of the transcript and
introns with functional splice donor and acceptor sites
(SD/SA). A common poly-A sequence which is employed in
the exemplary vectors of this invention is that derived
from the papovavirus SV-40 [see, e.g., nucleotides 837-
639 of SEQ ID NO: 1; 5245-5443 of SEQ ID NO: 2; and 3887-
3684 of SEQ ID NO: 3]. The poly-A sequence generally is
inserted in the vector following the transgene sequences
and before the 3' adenovirus sequences. A common intron
sequence is also derived from SV-40, and is referred to
as the SV-40 T intron sequence [see, e.g., nucleotides
4507-4376 of SEQ ID NO: 1 and 1579-1711 of SEQ ID NO: 2].
A pAdΔ shuttle vector 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 and many such sequences are available [see, e.g., Sambrook et al, and references cited therein]. Examples of such regulatory sequences for the above are provided in the plasmid sequences of Figs. 3, 5 and 7.

The combination of the transgene, promoter/enhancer, the other regulatory vector elements are referred to as a "minigene" for ease of reference herein.

The minigene is preferably flanked by the 5' and 3' cis-acting adenovirus sequences described above. Such a minigene may have a size in the range of several hundred base pairs up to about 30 kb due to the absence of adenovirus early and late gene sequences in the vector.

Thus, this AdΔ vector system permits a great deal of latitude in the selection of the various components of the minigene, particularly the selected transgene, with regard to size. Provided with the teachings of this invention, the design of such a minigene can be made by resort to conventional techniques.

II. The Helper Virus

Because of the limited amount of adenovirus sequence present in the AdΔ shuttle vector, a helper adenovirus of this invention must, alone or in concert with a packaging cell line, provide sufficient adenovirus gene sequences necessary for a productive viral infection. Helper viruses useful in this invention thus contain selected adenovirus gene sequences, and optionally a second reporter minigene.

Normally, the production of a recombinant adenovirus which utilizes helper adenovirus containing a full complement of adenoviral genes results in recombinant virus contaminated by excess production of the helper virus. Thus, extensive purification of the viral vector
from the contaminating helper virus is required. However, the present invention provides a way to facilitate purification and reduce contamination by crippling the helper virus.

One preferred embodiment of a helper virus of this invention thus contains three components (A) modifications or deletions of the native adenoviral gene sequences which direct efficient packaging, so as to substantially disable or "cripple" the packaging function of the helper virus or its ability to replicate, (B) selected adenovirus genes and (C) an optional reporter minigene. These "crippled" helper viruses may also be formed into poly-cation conjugates as described below.

The adenovirus sequences forming the helper virus may be obtained from the sources identified above in the discussion of the shuttle vector. Use of different Ad serotypes as helper viruses enables production of recombinant viruses containing the ΔAd (serotype 5) shuttle vector sequences in a capsid formed by the other serotype adenovirus. These recombinant viruses are desirable in targeting different tissues, or evading an immune response to the ΔAd sequences having a serotype 5 capsid. Use of these different Ad serotype helper viruses may also demonstrate advantages in recombinant virus production, stability and better packaging.

A. The Crippling Modifications

A desirable helper virus used in the production of the adenovirus vector of this invention is modified (or crippled) in its 5' ITR packaging/enhancer domain, identified above. As stated above, the packaging/enhancer region contains sequences necessary for packaging linear adenovirus genomes ("PAC" sequences). More specifically, this sequence contains at least seven distinct yet functionally redundant domains.
that are required for efficient encapsidation of replicated viral DNA.

Within a stretch of nucleotide sequence from bp 194-358 of the Ad5 genome, five of these so-called A-repeats or PAC sequences are localized (see, Fig. 1B). PAC I is located at bp 241-248 of the adenovirus genome (on the strand complementary to nucleotides 5259-5246 of SEQ ID NO: 1). PAC II is located at bp 262-269 of the adenovirus genome (on the strand complementary to nucleotides 5238-5225 of SEQ ID NO: 1). PAC III is located at bp 304-311 of the adenovirus genome (on the strand complementary to nucleotides 5196-5183 of SEQ ID NO: 1). PAC IV is located at bp 314-321 of the adenovirus (on the strand complementary to nucleotides 5186-5172 of SEQ ID NO: 1). PAC V is located at bp 339-346 of the adenovirus (on the strand complementary to nucleotides 5171-5147 of SEQ ID NO: 1).

Corresponding sequences can be obtained from SEQ ID NO: 2 and 3. PAC I is located at nucleotides 837-851 of SEQ ID NO: 2; and on the strand complementary to nucleotides 9374-9360 of SEQ ID NO: 3. PAC II is located at nucleotides 859-863 of SEQ ID NO: 2; and on the strand complementary to nucleotides 9353-9340 of SEQ ID NO: 3. PAC III is located at nucleotides 901-916 of SEQ ID NO: 2; and on the strand complementary to nucleotides 9311-9298 of SEQ ID NO: 3. PAC IV is located at nucleotides 911-924 of SEQ ID NO: 2; and on the strand complementary to nucleotides 9301-9288 of SEQ ID NO: 3. PAC V is located at nucleotides 936-949 of SEQ ID NO: 2; and on the strand complementary to nucleotides 9276-9263 of SEQ ID NO: 3.
Table 1 below lists these five native Ad5 sequences and a consensus PAC sequence based on the similarities between an eight nucleic acid stretch within the five sequences. The consensus sequence contains two positions at which the nucleic acid may be A or T (A/T). The conventional single letter designations are used for the nucleic acids, as is known to the art.

<table>
<thead>
<tr>
<th>A-Repeat</th>
<th>Nucleotide sequence</th>
<th>Base Pair Nos. &amp;</th>
<th>[SEQ ID NO:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>241 248 TAG TAAATTTG GCC</td>
<td>262 269 AGT AAGATTTG GCC</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>304 311 AGT GAAATCTG AAT</td>
<td>314 321 GAA TAATTTTGT GTT</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>339 346 CGT AATATTTT GCT</td>
<td>Consensus 5' (A/T)AN(A/T)TTTGT 3'</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>349 350</td>
<td>[SEQ ID NO: 7]</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>369 370</td>
<td>[SEQ ID NO: 8]</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

According to this invention, mutations or deletions may be made to one or more of these PAC sequences to generate desirable crippled helper viruses. A deletion analysis of the packaging domain revealed a positive correlation between encapsidation efficiency and the number of packaging A-repeats that were present at the 5' end of the genome. Modifications of this domain may include 5' adenovirus sequences which contain less than all five of the PAC sequences of Table 1. For example, only two PAC sequences may be present in the crippled virus, e.g., PAC I and PAC II, PAC III and PAC IV, and so on. Deletions of selected PAC sequences may
involve deletion of contiguous or non-contiguous sequences. For example, PAC II and PAC IV may be deleted, leaving PAC I, III and IV in the 5' sequence. Still an alternative modification may be the replacement of one or more of the native PAC sequences with one or more repeats of the consensus sequence of Table 1. Alternatively, this adenovirus region may be modified by deliberately inserted mutations which disrupt one or more of the native PAC sequences. One of skill in the art may further manipulate the PAC sequences to similarly achieve the effect of reducing the helper virus packaging efficiency to a desired level.

Exemplary helper viruses which involve the manipulation of the PAC sequences described above are disclosed in Example 7 below. Briefly, as described in that example, one helper virus contains in place of the native 5' ITR region (adenovirus genome bp 1-360), a 5' adenovirus sequence spanning adenovirus genome bp 1-269, which contains only the 5' ITR and PAC I and PAC II sequences, and deletes the adenovirus region bp 270-360.

Another PAC sequence modified helper virus contains only the 5' Ad5 sequence of the ITR and PAC I through PAC IV (Ad bp 1-321), deleting PAC V and other sequences in the Ad region bp322-360.

These modified helper viruses are characterized by reduced efficiency of helper virus encapsidation. These helper viruses with the specific modifications of the sequences related to packaging efficiency, provide a packaging efficiency high enough for generating production lots of the helper virus, yet low enough that they permit the achievement of higher yields of AdΔ transducing viral particles according to this invention.
B. The Selected Adenovirus Genes

Helper viruses useful in this invention, whether or not they contain the "crippling" modifications described above, contain selected adenovirus gene sequences depending upon the cell line which is transfected by the helper virus and shuttle vector. A preferred helper virus contains a variety of adenovirus genes in addition to the modified sequences described above.

As one example, if the cell line employed to produce the recombinant virus is not a packaging cell line, the helper virus may be a wild type Ad virus. Thus, the helper virus supplies the necessary adenovirus early genes E1, E2, E4 and all remaining late, intermediate, structural and non-structural genes of the adenovirus genome. This helper virus may be a crippled helper virus by incorporating modifications in its native 5' packaging/enhancer domain.

A desirable helper virus is replication defective and lacks all or a sufficient portion of the adenoviral early immediate early gene E1a (which spans mu 1.3 to 4.5) and delayed early gene E1b (which spans mu 4.6 to 11.2) so as to eliminate their normal biological functions. Such replication deficient viruses may also have crippling modifications in the packaging/enhancer domain. Because of the difficulty surrounding the absolute removal of adenovirus from AdΔ preparations that have been enriched by CsCl buoyant density centrifugation, the use of a replication defective adenovirus helper prevents the introduction of infectious adenovirus for in vivo animal studies. This helper virus is employed with a packaging cell line which supplies the deficient E1 proteins, such as the 293 cell line.
Additionally, all or a portion of the adenovirus delayed early gene E3 (which spans nu 76.6 to 86.2) may be eliminated from the adenovirus sequence which forms a part of the helper viruses useful in this invention, without adversely affecting the function of the helper virus because this gene product is not necessary for the formation of a functioning virus.

In the presence of other packaging cell lines which are capable of supplying adenoviral proteins in addition to the E1, the helper virus may accordingly be deleted of the genes encoding these adenoviral proteins. Such additionally deleted helper viruses also desirably contain crippling modifications as described above.

C. A Reporter Minigene

It is also desirable for the helper virus to contain a reporter minigene, in which the reporter gene is desirably different from the reporter transgene contained in the shuttle vector. A number of such reporter genes are known, as referred to above. The presence of a reporter gene on the helper virus which is different from the reporter gene on the pAdΔ, allows both the recombinant AdΔ virus and the helper virus to be independently monitored. For example, the expression of recombinant alkaline phosphatase enables residual quantities of contaminating adenovirus to be monitored independent of recombinant LacZ expressed by an pAdΔ shuttle vector or an AdΔ virus.

D. Helper Virus Polycation Conjugates

Still another method for reducing the contamination of helper virus involves the formation of poly-cation helper virus conjugates, which may be associated with a plasmid containing other adenoviral genes, which are not present in the helper virus. The helper viruses described above may be further modified by resort to adenovirus-polylysine conjugate technology.

Using this technology, a helper virus containing preferably the late adenoviral genes is modified by the addition of a poly-cation sequence distributed around the capsid of the helper virus. Preferably, the poly-cation is poly-lysine, which attaches around the negatively-charged vector to form an external positive charge. A plasmid is then designed to express those adenoviral genes not present in the helper virus, e.g., the E1, E2 and/or E4 genes. The plasmid associates to the helper virus-conjugate through the charges on the poly-lysine sequence. This modification is also desirably made to a crippled helper virus of this invention. This conjugate (also termed a trans-infection particle) permits additional adenovirus genes to be removed from the helper virus and be present on a plasmid which does not become incorporated into the virus during production of the recombinant viral vector. Thus, the impact of contamination is considerably lessened.

III. Assembly of Shuttle Vector, Helper Virus and Production of Recombinant Virus

The material from which the sequences used in the pAdΔ shuttle vector and the helper viruses are derived, as well as the various vector components and sequences employed in the construction of the shuttle vectors, helper viruses, and AdΔ viruses of this invention, are obtained from commercial or academic sources based on previously published and described materials. These materials may also be obtained from an individual patient or generated and selected using standard recombinant molecular cloning techniques known and practiced by those
skilled in the art. Any modification of existing nucleic acid sequences forming the vectors and viruses, including sequence deletions, insertions, and other mutations are also generated using standard techniques.

Assembly of the selected DNA sequences of the adenovirus, and the reporter genes or therapeutic genes and other vector elements into the pAdΔ shuttle vector using conventional techniques is described in Example 1 below. 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 transfection techniques using the HEK 293 cell line. Other conventional methods employed in this invention include homologous recombination of the viral genomes, plaquing of viruses in agar overlay, methods of measuring signal generation, and the like. Assembly of any desired AdΔ vector or helper virus of this invention is within the skill of the art, based on the teachings of this invention.

A. Shuttle Vector

As described in detail in Example 1 below and with resort to Fig. 2A and the DNA sequence of the plasmid reported in Fig. 3, a unique pAdΔ shuttle vector of this invention, pAdΔ.CMVLacZ, is generated. pAdΔ.CMVLacZ contains Ad5 sequences encoding the 5' terminal followed by a CMV promoter/enhancer, a splice donor/splice acceptor sequence, a bacterial beta-galactosidase gene (LacZ), a SV-40 poly A sequence (pA), a 3' ITR from Ad5 and remaining plasmid sequence from plasmid pSP72 (Promega) backbone.
To generate the AdΔ genome which is incorporated in the vector, the plasmid pAdΔ.CMVlacZ must be digested with EcoRI to release the AdΔ.CMVlacZ genome, freeing the adenovirus ITRs and making them available targets for replication. Thus production of the vector is "restriction-dependent", i.e., requires restriction endonuclease rescue of the replication template. See, Fig. 2B.

A second type of pAdΔ plasmid was designed which places the 3' Ad terminal sequence in a head-to-tail arrangement relative to the 5' terminal sequence. As described in Example 1 and Figs. 4A, and with resort to the DNA sequence of the plasmid reported in Fig. 5, a second unique AdΔ vector sequence of this invention, AdΔC.CMVlacZ, is generated from the shuttle plasmid pAdΔ.CMVlacZ, which contains an Ad5 5' ITR sequence and 3' ITR sequence positioned head-to-tail, followed by a CMV enhancer/promoter, SD/SA sequence, LacZ gene and pA sequence in a plasmid pSP72 (Promega) backbone. As described in Example 1B, this "restriction-independent" plasmid permits the AdΔ genome to be replicated and rescued from the plasmid backbone without including an endonuclease treatment (see, Fig. 4B).

B. Helper Virus

As described in detail in Example 2, an exemplary conventional E1 deleted adenovirus helper virus is virus Ad.CBhpAP, which contains a 5' adenovirus sequence from mu 0-1, a reporter minigene containing human placenta alkaline phosphatase (hpAP) under the transcriptional control of the chicken β-actin promoter, followed by a poly-A sequence from SV40, followed by adenovirus sequences from 9.2 to 78.4 and 86 to 100. This helper contained deletions from mu 1.0 to 9.2 and 78.4 to 86, which eliminate substantially the E1 region and the E3 region of the virus. This virus may be
desirably crippled according to this invention by modifications to its packaging enhancer domain.

Exemplary crippled helper viruses of this invention are described using the techniques described in Example 7 and contain the modified 5' PAC sequences, i.e., adenovirus genome bp 1-269; m.u. 0-0.75 or adenovirus genome bp 1-321; m.u. 0-0.89. Briefly, the 5' sequences are modified by PCR and cloned by conventional techniques into a conventional adenovirus based plasmid.

A hpAP minigene is incorporated into the plasmid, which is then altered by homologous recombination with an E3 deleted adenovirus dl7001 to result in the modified vectors so that the reporter minigene is followed on its 3' end with the adenovirus sequences mu 9.6 to 78.3 and 87 to 100.

Generation of a poly-L-lysine conjugate helper virus was demonstrated essentially as described in detail in Example 5 below and Fig. 10 by coupling poly-L-lysine to the Ad.CBhpAP virion capsid. Alternatively, the same procedure may be employed with the PAC sequence modified helper viruses of this invention.

C. Recombinant AdΔ Virus

As stated above, a pAdΔ shuttle vector in the presence of helper virus and/or a packaging cell line permits the adenovirus-transgene sequences in the shuttle vector to be replicated and packaged into virion capsids, resulting in the recombinant AdΔ virus. The current method for producing such AdΔ virus is transfection-based and described in detail in Example 3. Briefly, helper virus is used to infect cells, such as the packaging cell line human HEK 293, which are then subsequently transfected with an pAdΔ shuttle vector containing a selected transgene by conventional methods. About 30 or more hours post-transfection, the cells are harvested, and an extract prepared. The AdΔ viral genome is
packaged into virions that sediment at a lower density than the helper virus in cesium gradients. Thus, the recombinant Ad Δ virus containing a selected transgene is separated from the bulk of the helper virus by purification via buoyant density ultracentrifugation in a CsCl gradient.

The yield of Ad Δ transducing virus is largely dependent on the number of cells that are transfected with the pΔΔ shuttle plasmid, making it desirable to use a transfection protocol with high efficiency. One such method involves use of a poly-L-lysylated helper adenovirus as described above. A pΔΔ shuttle plasmid containing the desired transgene under the control of a suitable promoter, as described above, is then complexed directly to the positively charged helper virus capsid, resulting in the formation of a single transfection particle containing the pΔΔ shuttle vector and the helper functions of the helper virus.

The underlying principle is that the helper adenovirus coated with plasmid pΔΔ DNA will co-transport the attached nucleic acid across the cell membrane and into the cytoplasm according to its normal mechanism of cell entry. Therefore, the poly-L-lysine modified helper adenovirus assumes multiple roles in the context of an Ad Δ-based complex. First, it is the structural foundation upon which plasmid DNA can bind increasing the effective concentration. Second, receptor mediated endocytosis of the virus provides the vehicle for cell uptake of the plasmid DNA. Third, the endosomolytic activity associated with adenoviral infection facilitates the release of internalized plasmid into the cytoplasm. And the adenovirus contributes trans helper functions on which the recombinant Ad Δ virus is dependent for replication and packaging of transducing viral particles. The Ad-based transfection procedure using an pΔΔ shuttle
vector and a polycation-helper conjugate is detailed in Example 6. Additionally, as described previously, the helper virus-plasmid conjugate may be another form of helper virus delivery of the omitted adenovirus genes not present in the pAdΔ vector. Such a structure enables the rest of the required adenovirus genes to be divided between the plasmid and the helper virus, thus reducing the self-replication efficiency of the helper virus.

A presently preferred method of producing the recombinant AdΔ virus of this invention involves performing the above-described transfection with the crippled helper virus or crippled helper virus conjugate, as described above. A "crippled" helper virus of this invention is unable to package itself efficiently, and therefore permits ready separation of the helper virus from the newly packaged AdΔ vector of this invention by use of buoyant density ultracentrifugation in a CsCl gradient, as described in the examples below.

IV. Function of the Recombinant AdΔ Virus

Once the AdΔ virus of this invention is produced by cooperation of the shuttle vector and helper virus, the AdΔ virus can be targeted to, and taken up by, a selected target cell. The selection of the target cell also depends upon the use of the recombinant virus, i.e., whether or not the transgene is to be replicated in vitro or ex vivo for production in a desired cell type for redelivery into a patient, or in vivo for delivery to a particular cell type or tissue. Target cells may be any mammalian cell (preferably a human cell). For example, in in vivo use, the recombinant virus can target to any cell type normally infected by adenovirus, depending upon the route of administration, i.e., it can target, without limitation, neurons, hepatocytes, epithelial cells and
the like. The helper adenovirus sequences supply the sequences necessary to permit uptake of the virus by the AdΔ.

Once the recombinant virus is taken up by a cell, the adenovirus flanked transgene is rescued from the parental adenovirus backbone by the machinery of the infected cell, as with other recombinant adenoviruses. Once uncoupled (rescued) from the genome of the AdΔ virus, the recombinant minigene seeks an integration site in the host chromatin and becomes integrated therein, either transiently or stably, providing expression of the accompanying transgene in the host cell.

V. Use of the AdΔ Viruses in Gene Therapy

The novel recombinant viruses and viral conjugates of this invention provide efficient gene transfer vehicles for somatic gene therapy. These viruses are prepared to contain a therapeutic gene in place of the LacZ reporter transgene illustrated in the exemplary viruses and vectors. By use of the AdΔ viruses containing therapeutic transgenes, these transgenes can be delivered to a patient in vivo or ex vivo to provide for integration of the desired gene into a target cell. Thus, these viruses can be employed to correct genetic deficiencies or defects. An example of the generation of an AdΔ gene transfer vehicle for the treatment of cystic fibrosis is described in Example 4 below. One of skill in the art can generate any number of other gene transfer vehicles by including a selected transgene for the treatment of other disorders.

The recombinant viruses of the present invention 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 recombinant viruses of this invention may be administered in sufficient amounts to transfect the desired cells and provide sufficient levels of integration and expression of the selected transgene to provide a therapeutic benefit without undue adverse effects or with medically acceptable physiological effects which can be determined by those skilled in the medical arts. Conventional and pharmaceutically acceptable parenteral routes of administration include direct delivery to the target organ, tissue or site, intranasal, intravenous, intramuscular, subcutaneous, intradermal and oral administration. Routes of administration may be combined, if desired.

Dosages of the recombinant virus will depend primarily on factors such as the condition being treated, the selected gene, the age, weight and health of the patient, and may thus vary among patients. A therapeutically effective human dosage of the viruses of the present invention is believed to be in the range of from about 20 to about 50 ml of saline solution containing concentrations of from about $1 \times 10^7$ to $1 \times 10^{10}$ pfu/ml virus of the present invention. A preferred human dosage is about 20 ml saline solution at the above concentrations. The dosage will be adjusted to balance the therapeutic benefit against any side effects. The levels of expression of the selected gene can be monitored to determine the selection, adjustment or frequency of dosage administration.
The following examples illustrate the construction of the pAdΔ shuttle vectors, helper viruses and recombinant AdΔ viruses of the present invention and the use thereof in gene therapy. These examples are illustrative only, and do not limit the scope of the present invention.

Example 1 - Production of pAdΔ.CMVlacZ and pAdΔc.CMVlacZ Shuttle Vectors

A. pAdΔ.CMVlacZ

A human adenovirus Ad5 sequence was modified to contain a deletion in the Ela region [map units 1 to 9.2], which immediately follows the Ad 5' region (bp 1-360) (illustrated in Figs. 1A). Thus, the plasmid contains the 5' ITR sequence (bp 1-103), the native packaging/enhancer sequences and the TATA box for the Ela region (bp 104-360). A minigene containing the CMV immediate early enhancer/promoter, an SD/SA sequence, a cytoplasmic lacZ gene, and SV40 poly A (pA), was introduced at the site of the Ela deletion. This construct was further modified so that the minigene is followed by the 3' ITR sequences (bp 35,353-end). The DNA sequences for these components are provided in Fig. 3 and SEQ ID NO: 1 (see, also the brief description of this figure).

This construct was then cloned by conventional techniques into a pSP72 vector (Promega) backbone to make the circular shuttle vector pAdΔCMVLacZ. See the schematic of Fig. 2A. This construct was engineered with EcoRI sites flanking the 5' and 3' Ad5 ITR sequences. pAdΔ.CMVlacZ was then subjected to enzymatic digestion with EcoRI, releasing a linear fragment of the vector spanning the terminal end of the Ad 5'ITR sequence through the terminal end of the 3'ITR sequence from the plasmid backbone. See Fig. 2B.
B. \textit{pAdAc.CMVlacZ}

The shuttle vector \textit{pAdAc.CMVlacZ} (Figs. 4A and 5) was constructed using a pSP72 (Promega) backbone so that the Ad5 5' ITR and 3' ITR were positioned head-to-tail. The organization of the Ad5 ITRs was based on reports that suggest circular Ad genomes that have the terminal ends fused together head-to-tail are infectious to levels comparable to linear Ad genomes. A minigene encoding the CMV enhancer, an SD/SA sequence, the LacZ gene, and the poly A sequence was inserted immediately following the 5' ITR. The DNA sequence of the resulting plasmid and the sequences for the individual components are reported in Fig. 5 and SEQ ID NO: 2 (see also, brief description of Fig. 5). This plasmid does not require enzymatic digestion prior to its use to produce the viral particle (see Example 3). This vector was designed to enable restriction-independent production of LacZ Ad\textalpha vectors.

\textbf{Example 2 - Construction of a Helper Virus}

The Ad.CBhpAP helper virus \cite{Kozarsky1993} is a replication deficient adenovirus containing an alkaline phosphatase minigene. Its construction involved conventional cloning and homologous recombination techniques. The adenovirus DNA substrate was extracted from CsCl purified d17001 virions, an Ad5 (serotype subgroup C) variant that carries a 3 kb deletion between mu 78.4 through 86 in the nonessential E3 region (provided by Dr. William Wold, Washington University, St. Louis, Missouri). Viral DNA was prepared for co-transfection by digestion with ClaI (adenovirus genomic bp position 917) which removes the left arm of the genome encompassing adenovirus map units 0-2.5. See lower diagram of Fig. 1B.
A parental cloning vector, pAd.BglII was designed. It contains two segments of wild-type Ad5 genome (i.e., map units 0-1 and 9-16.1) separated by a unique BglII cloning site for insertion of heterologous sequences. The missing Ad5 sequences between the two domains (adenovirus genome bp 361-3327) results in the deletion of E1a and the majority of E1b following recombination with viral DNA.

A recombinant hpAP minigene was designed and inserted into the BglII site of pAd.BglII to generate the complementing plasmid, pAdCBhpAP. The linear arrangement of this minigene includes:

(a) the chicken cytoplasmic β-actin promoter [nucleotides +1 to +275 as described in T. A. Kost et al, Nucl. Acids Res., 11(23):8287 (1983); nucleotides 9241-8684 of Fig. 7];

(b) an SV40 intron (e.g., nucleotides 1579-1711 of SEQ ID NO: 2),

(c) the sequence for human placental alkaline phosphatase (available from Genbank) and

(d) an SV40 polyadenylation signal (a 237 Bam HI-BclI restriction fragment containing the cleavage/poly-A signals from both the early and late transcription units; e.g., nucleotides 837-639 of SEQ ID NO: 1).

The resulting complementing plasmid, pAdCBhpAP contained a single copy of recombinant hpAP minigene flanked by adenovirus coordinates 0-1 on one side and 9.2-16.1 on the other.

Plasmid DNA was linearized using a unique NheI site immediately 5' to adenovirus map unit zero (0) and the above-identified adenovirus substrate and the complementing plasmid DNAs were transfected to 293 cells [ATCC CRL1573] using a standard calcium phosphate transfection procedure [see, e.g., Sambrook et al, cited above]. The end result of homologous recombination
involving sequences that map to adenovirus map units 9-16.1 is hybrid Ad.CBhpAP helper virus which contains adenovirus map units 0-1 and, in place of the E1a and E1b coding regions from the d17001 adenovirus substrate, is the hpAP minigene from the plasmid, followed by Ad sequences 9 to 100, with a deletion in the E3 (78.4-86 mu) regions.

Example 3 - Production of Recombinant AdΔ Virus

The recombinant AdΔ virus of this invention are generated by co-transfection of a shuttle vector with the helper virus in a selected packaging or non-packaging cell line.

As described in detail below, the linear fragment provided in Example 1A, or the circular AdΔ genome carrying the LacZ of Example 1B, is packaged into the Ad.CBhpAP helper virus (Example 2) using conventional techniques, which provides an empty capsid head, as illustrated in Fig. 2C. Those virus particles which have successfully taken up the pΔd shuttle genome into the capsid head can be distinguished from those containing the hpAP gene by virtue of the differential expression of LacZ and hpAP.

In more detail, 293 cells (4 × 10^7 pfu 293 cells/150 mm dish) were seeded and infected with helper virus Ad.CBhpAP (produced as described in Example 2) at an MOI of 5 in 20 ml DMEM/2% fetal bovine serum (FBS). This helper specific marker is critical for monitoring the level of helper virus contamination in AdΔ preparations before and after purification. The helper virus provides in trans the necessary helper functions for synthesis and packaging of the AdΔCMVLacZ genome.

Two hours post infection, using either the restriction-dependent shuttle vector or the restriction-independent shuttle vector, plasmid pAdΔ.CMVLacZ
(digested with EcoRI) or pAdΔc.CMVlacZ DNA, each carrying a LacZ minigene, was added to the cells by a calcium phosphate precipitate (2.5 ml calcium phosphate transfection cocktail containing 50 µg plasmid DNA).

Thirty to forty hours post-transfection, cells were harvested, suspended in 10 mM Tris-Cl (pH 8.0) (0.5 ml/150 mm plate) and frozen at -80°C. Frozen cell suspensions were subjected to three rounds of freeze (ethanol-dry ice)-thaw (37°C) cycles to release virion capsids. Cell debris was removed by centrifugation (5,000xg for 10 minutes) and the clarified supernatant applied to a CsCl gradients to separate recombinant virus from helper virus as follows.

Supernatants (10 ml) applied to the discontinuous CsCl gradient (composed of equal volumes of CsCl at 1.2 g/ml, 1.36 g/ml, and 1.45 g/ml 10 mM Tris-Cl (pH 8.0)) were centrifuged for 8 hours at 72,128Xg, resulting in separation of infectious helper virus from incompletely formed virions. Fractions were collected from the interfacing zone between the helper and top components and analyzed by Southern blot hybridization or for the presence of LacZ transducing particles. For functional analysis, aliquots (2.0 ml from each sample) from the same fractions were added to monolayers of 293 cells (in 35 mm wells) and expression of recombinant β-galactosidase determined 24 hours later. More specifically, monolayers were harvested, suspended in 0.3 ml 10 mM Tris-Cl (pH 8.0) buffer and an extract prepared by three rounds of freeze-thaw cycles. Cell debris was removed by centrifugation and the supernatant tested for β-galactosidase (LacZ) activity according to the procedure described in J. Price et al, Proc. Natl. Acad. Sci., USA, 84:156-160 (1987). The specific activity (milliunits β-galactosidase/mg protein or reporter
enzymes was measured from indicator cells. For the recombinant virus, specific activity was 116.

Fractions with $\beta$-galactosidase activity from the discontinuous gradient were sedimented through an equilibrium cesium gradient to further enrich the preparation for Ad$\alpha$ virus. A linear gradient was generated in the area of the recombinant virus spanning densities 1.29 to 1.34 g/ml. A sharp peak of the recombinant virus, detected as the appearance of the $\beta$-gal activity in infected 293 cells, eluted between 1.31 and 1.33 g/ml. This peak of recombinant virus was located between two major A260 nm absorbing peaks and in an area of the gradient with the helper virus was precipitously dropping off. The equilibrium sedimentation gradient accomplished another 102 to 103 fold purification of recombinant virus from helper virus. The yield of recombinant Ad$\alpha$.CMVLacZ virus recovered from a 50 plate prep after 2 sedimentations ranged from 107 to 108 transducing particles.

Analysis of lysates of cells transfected with the recombinant vector and infected with helper revealed virions capable of transducing the recombinant minigene contained within the vector. Subjecting aliquots of the fractions to Southern analysis using probes specific to the recombinant virus or helper virus revealed packaging of multiple molecular forms of vector derived sequence. The predominant form of the deleted viral genome was the size (~5.5 kb) of the corresponding double stranded DNA monomer (Ad$\alpha$.CMVLacZ) with less abundant but discrete higher molecular weight species (~10 kb and ~15 kb) also present. Full-length helper virus is 35kb. Importantly, the peak of vector transduction activity corresponds with the highest molecular weight form of the deleted virus. These results confirm the hypothesis that ITRs and contiguous packaging sequence are the only elements
necessary for incorporation into virions. An apparently ordered or preferred rearrangement of the recombinant Ad monomer genome leads to a more biologically active molecule. The fact that larger molecular species of the deleted genome are 2x and 3x ijd larger than the monomer deleted virus genome suggests that the rearrangements may involve sequential duplication of the original genome.

These same procedures may be adapted for production of a recombinant AdΔ virus using a crippled helper virus or helper virus conjugate as described previously.

Example 4 - Recombinant AdΔ Virus Containing a Therapeutic Minigene

To test the versatility of the recombinant AdΔ virus system, the reporter LacZ minigene obtained from pAdΔCMVLacZ was cassette replaced with a therapeutic minigene encoding CFTR.

The minigene contained human CFTR cDNA [Riordan et al, Science, 245:1066-1073 (1989); nucleotides 8622-4065 of SEQ ID NO: 3] under the transcriptional control of a chimeric CMV enhancer/chicken β-actin promoter element (nucleotides +1 to +275 as described in T. A. Kost et al, Nucl. Acids Res., 11(23):8287 (1983); nucleotides 9241-8684 of SEQ ID NO: 3, Fig. 7); and followed by an SV-40 poly-A sequence (nucleotides 3887-3684 of SEQ ID NO: 3, Fig. 7).

The CFTR minigene was inserted into the E1 deletion site of an Ad5 virus (called pAdΔ.E1Δ) which contains a deletion in E1A from μ1-9.2 and a deletion in E3 from μ8.4-86.

The resulting shuttle vector called pAdΔ.CBCFTR (see Figs. 6 and the DNA sequence of Fig. 7 [SEQ ID NO: 3]) used the same Ad ITRs of pAdΔCMVLacZ, but the Ad5 sequences terminated with NheI sites instead of EcoRI.
Therefore release of the minigene from the plasmid was accomplished by digestion with NheI.

The vector production system described in Example 3 was employed, using the helper virus Ad.CBhpAP (Example 2). Monolayers of 293 cells grown to 80-90% confluency in 150 mm culture dishes were infected with the helper virus at an MOI of 5. Infections were done in DMEM supplemented with 2% FBS at 20 ml media/150 mm plate. Two hours post-infection, 50 µg plasmid DNA in 2.5 ml transfection cocktail was added to each plate and evenly distributed.

Delivery of the pAdΔ.CBCFTR plasmid to 293 cells was mediated by formation of a calcium phosphate precipitate and AdΔ.CBCFTR virus resolved from Ad.CBhpAP helper virus by CsCl buoyant density ultracentrifugation as follows:

Cells were left in this condition for 10-14 h, after which the infection/transfection media was replaced with 20 ml fresh DMEM/2% FBS. Approximately 30 h post-transfection, cells were harvested, suspended in 10 mM Tris-Cl (pH 8.0) buffer (0.5 ml/150 mm plate), and stored at -80°C.

Frozen cell suspensions were lysed by three sequential rounds of freeze (ethanol-dry ice)-thaw (37°C). Cell debris was removed by centrifugation (5,000 x g for 10 min) and 10 ml clarified extract layered onto a CsCl step gradient composed of three 9.0 ml tiers with densities 1.45 g/ml, 1.36 g/ml, and 1.20 g/ml CsCl in 10 mM Tris-Cl (pH 8.0) buffer. Centrifugation was performed at 20,000 rpm in a Beckman SW-28 rotor for 8 h at 4°C. Fractions (1.0 ml) were collected from the bottom of the centrifuge tube and analyzed for rAda transducing vectors. Peak fractions were combined and banded to equilibrium. Fractions containing transducing virions were dialyzed against 20 mM HEPES (pH 7.8)/150 mM NaCl
(HBS) and stored frozen at -80°C in the presence of 10% glycerol or as a liquid stock at -20°C (HBS+40% glycerol).

Fractions collected after ultracentrifugation were analyzed for transgene expression and vector DNA. For lacZ ΔAd vectors, 2 μl aliquots were added to 293 cell monolayers seeded in 35 mm culture wells. Twenty-four hours later cells were harvested, suspended in 0.3 ml 10 mM Tris-Cl (pH 8.0) buffer, and lysed by three rounds of freeze-thaw. Cell debris was removed by centrifugation (15,000 x g for 10 min) and assayed for total protein [Bradford, (1976)] and β-galactosidase activity [Sambrook et al, (1989)] using ONPG (o-Nitrophenyl β-D-galactopyranoside) as substrate.

Expression of CFTR protein from the AdΔ.CBCFTR vector was determined by immunofluorescence localization. Aliquots of AdΔ.CBCFTR, enriched by two-rounds of ultracentrifugation and exchanged to HBS storage buffer, were added to primary cultures of airway epithelial cells obtained from the lungs of CF transplant recipients. Twenty-four hours after the addition of vector, cells were harvested and affixed to glass slides using centrifugal force (Cytospin®3, Shandon Scientific Limited). Cells were fixed with freshly prepared 3% paraformaldehyde in PBS (1.4 mM KH₂PO₄, 4.3 mM Na₂HPO₄, 2.7 mM KCl, and 137 mM NaCl) for 15 min at room temperature (RT), washed twice in PBS, and permeabilized with 0.05% NP-40 for 10 min at RT. The immunofluorescence procedure began with a blocking step in 10% goat serum (PBS/GS) for 1 h at RT, followed by binding of the primary monoclonal mouse anti-human CFTR (R-domain specific) antibody (Genzyme) diluted 1:500 in PBS/GS for 2 h at RT. Cells were washed extensively in PBS/GS and incubated for 1 h at RT with a donkey anti-mouse IgG (H+L) FITC conjugated

*Trademark
antibody (Jackson ImmunoResearch Laboratories) diluted 1:100 in PBS/GS.

For Southern analysis of vector DNA, 5 μl aliquots were taken directly from CsCl fractions and incubated with 20 μl capsid digestion buffer (50 mM Tris-Cl, pH 8.0; 1.0 mM EDTA, pH 8.0; 0.5% SDS, and 1.0 mg/ml Proteinase K) at 50°C for 1 h. The reactions were allowed to cool to RT, loading dye was added, and electrophoresed through a 1.2% agarose gel. Resolved DNAs were electroblotted onto a nylon membrane (Hybond-N) and hybridized with a 32-P labeled restriction fragment. Blots were analyzed by autoradiography or scanned on a Phosphorimager 445 SI (Molecular Dynamics).

The results that were obtained from Southern blot analysis of gradient fractions revealed a distinct viral band that migrated faster than the helper Ad.CBhpAP DNA. The highest viral titers mapped to fractions 3 and 4. Quantitation of the bands in fraction 4 indicated the titer of Ad.CBhpAP was approximately 1.5x greater than AdΔCBCFTR. However, if the size difference between the two viruses is factored in (Ad.CBhpAP=35 kb; AdΔCBCFTR=6.2 kb), the viral titer (where 1 particle=1 DNA molecule) of AdΔCB.CFTR is at least 4-fold greater than the viral titer of Ad.CBhpAP.

While Southern blot analysis of gradient fractions was useful for showing the production of AdΔ viral particles, it also demonstrated the utility of ultracentrifugation for purifying AdΔ viruses. Considering the latter of these, both LacZ and CFTR transducing viruses banded in CsCl to an intermediate density between infectious adenovirus helper virions (1.34 g/ml) and incompletely formed capsids (1.31 g/ml). The lighter density relative to helper virus likely results from the smaller genome carried by the AdΔ viruses. This further suggests changes in virus size
influences the density and purification of AdΔ virus. Regardless, the ability to separate AdΔ virus from the helper virus is an important observation and suggests further purification may be achieved by successive rounds of banding through CsCl.

This recombinant virus is useful in gene therapy alone, or preferably, in the form of a conjugate prepared as described herein.

Example 5 - Correction of Genetic Defect in CF airway Epithelial Cells with AdΔCB.CFTR

Treatment of cystic fibrosis, utilizing the recombinant virus provided above, is particularly suited for in vivo, lung-directed, gene therapy. Airway epithelial cells are the most desirable targets for gene transfer because the pulmonary complications of CF are usually its most morbid and life-limiting.

The recombinant AdΔCB.CFTR virus was fractionated on sequential CsCl gradients and fractions containing CFTR sequences, migrating between the adenovirus and top components fractions described above were used to infect primary cultures of human airway epithelial cells derived from the lungs of a CF patient. The cultures were subsequently analyzed for expression of CFTR protein by immunocytochemistry. Immunofluorescent detection with mouse anti-human CFTR (R domain specific) antibody was performed 24 hours after the addition of the recombinant virus. Analysis of mock infected CF cells failed to reveal significant binding to the R domain specific CFTR antibody. Primary airway epithelium cultures exposed to the recombinant virus demonstrated high levels of CFTR protein in 10-20% of the cells.

Thus, the recombinant virus of the invention, containing the CFTR gene, may be delivered directly into the airway, e.g. by formulating the virus above, into a
preparation which can be inhaled. For example, the recombinant virus or conjugate of the invention containing the CFTR gene, is suspended in 0.25 molar sodium chloride. The virus or conjugate is taken up by respiratory airway cells and the gene is expressed.

Alternatively, the virus or conjugates of the invention may be delivered by other suitable means, including site-directed injection of the virus bearing the CFTR gene. In the case of CFTR gene delivery, preferred solutions for bronchial instillation are sterile saline solutions containing in the range of from about $1 \times 10^7$ to $1 \times 10^{10}$ pfu/ml, more particularly, in the range of from about $1 \times 10^8$ to $1 \times 10^9$ pfu/ml of the virus of the present invention.

Other suitable methods for the treatment of cystic fibrosis by use of gene therapy recombinant viruses of this invention may be obtained from the art discussions of other types of gene therapy vectors for CF. See, for example, U. S. Patent No. 5,240,846, incorporated by reference herein.

**Example 6 - Synthesis of Polycation Helper Virus Conjugate**

Another version of the helper virus of this invention is a polylysine conjugate which enables the pAdΔ shuttle plasmid to complex directly with the helper virus capsid. This conjugate permits efficient delivery of shuttle plasmid pAdΔ shuttle vector in tandem with the helper virus, thereby removing the need for a separate transfection step. See, Fig. 10 for a diagrammatic outline of this construction. Alternatively, such a conjugate with a plasmid supplying some Ad genes and the helper supplying the remaining necessary genes for production of the AdΔ viral vector provides a novel way
to reduce contamination of the helper virus, as discussed above.

Purified stocks of a large-scale expansion of Ad.CBhpAP were modified by coupling poly-L-lysine to the virion capsid essentially as described by K. J. Fisher and J. M. Wilson, *Biochem. J.*, 299:49-58 (1994), resulting in an Ad.CBhpAP-(Lys)$_n$ conjugate. The procedure involves three steps.

First, CsCl band purified helper virus Ad.CBhpAP was reacted with the heterobifunctional crosslinker sulfo-SMCC [sulfo-(N-succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate] (Pierce). The conjugation reaction, which contained 0.5 mg (375 nmol) of sulfo-SMCC and 6 x 10$^{12}$ A$_{260}$ helper virus particles in 3.0 ml of HBS, was incubated at 30°C for 45 minutes with constant gentle shaking. This step involved formation of a peptide bond between the active N-hydroxysuccinimide (NHS) ester of sulpho-SMCC and a free amine (e.g. lysine) contributed by an adenovirus protein sequence (capsid protein) in the vector, yielding a maleimide-activated viral particle. The activated adenovirus is shown in Fig. 10 having the capsid protein fiber labeled with the nucleophilic maleimide moiety. In practice, other capsid polypeptides including hexon and penton base are also targeted.

Unincorporated, unreacted cross-linker was removed by gel filtration on a 1 cm x 15 cm Bio-Gel P-6DG (Bio-Rad Laboratories) column equilibrated with 50 mM Tris/HCl buffer, pH 7.0, and 150 mM NaCl. Peak A$_{260}$ fractions containing maleimide-activated helper virus were combined and placed on ice.

Second, poly-L-lysine having a molecular mass of 58 kDa at 10 mg/ml in 50 mM triethanolamine buffer (pH 8.0), 150 mM NaCl and 1 mM EDTA was thiolated with 2-iminothiolane/HCl (Traut's Reagent; Pierce) to a molar
ratio of 2 moles-SH/mole polylysine under N₂; the cyclic thioimidate reacts with the poly(L-lysine) primary amines resulting in a thiolated polycation. After a 45 minute incubation at room temperature the reaction was applied to a 1 cm x 15 cm Bio-Gel P6DG column equilibrated with 50 mM Tris/HCl buffer (pH 7.0), 150 mM NaCl and 2 mM EDTA to remove unincorporated Traut's Reagent.

Quantification of free thiol groups was accomplished with Ellman's reagent [5,5'-dithio-bis-(2-nitrobenzoic acid)], revealing approximately 3-4 mol of -SH/mol of poly(L-lysine). The coupling reaction was initiated by adding 1 x 10^{12} A_{260} particles of maleimide-activated helper virus/mg of thiolated poly(L-lysine) and incubating the mixture on ice at 4°C for 15 hours under argon. 2-mercaptoethylamine was added at the completion of the reaction and incubation carried out at room temperature for 20 minutes to block unreacted maleimide sites.

Virus-polylysine conjugates, Ad.CPAP-p(Lys)_ₙ, were purified away from unconjugated poly(L-lysine) by ultracentrifugation through a CsCl step gradient with an initial composition of equal volumes of 1.45 g/ml (bottom step) and 1.2 g/ml (top step) CsCl in 10 mM Tris/HCl buffer (pH 8.0). Centrifugation was at 90,000 g for 2 hours at 5°C. The final product was dialyzed against 20 mM Hepes buffer (pH 7.8) containing 150 mM NaCl (HBS).

**Example 7 - Formation of AdΔ/helper-pLys Viral Particle**

The formation of Ad.CBhpAP-pLys/pAdΔ.CMVlacZ particle is initiated by adding 20 µg plasmid pAdΔ.CMVlacZ DNAs to 1.2 x 10^{12} A_{260} particles Ad.CBhpAP-pLys in a final volume of 0.2 ml DMEM and allowing the complex to develop at room temperature for between 10-15 minutes. This ratio typically represents the plasmid DNA binding capacity of a standard lot of adenovirus-pLys.
conjugate and gives the highest levels of plasmid transgene expression.

The resulting trans-infection particle is transfected onto 293 cells (4 x 10^7 cells seeded on a 150 mm dish). Thirty hours after transfection, the particles are recovered and subjected to a freeze/thaw technique to obtain an extract. The extract is purified on a CsCl step gradient with gradients at 1.20 g/ml, 1.36 g/ml and 1.45 g/ml. After centrifugation at 90,000 x g for 8 hours, the AdΔ vectors were obtained from a fraction under the top components as identified by the presence of LacZ, and the helper virus was obtained from a smaller, denser fraction, as identified by the presence of hpAP.

Example 8 - Construction of Modified Helper Viruses with Crippled Packaging (PAC) Sequences

This example refers to Figs. 9A through 9C, 10A and 10B in the design of modified helper viruses of this invention.

Ad5 5' terminal sequences that contained PAC domains I and II (Fig. 8A) or PAC domains I, II, III, and IV (Fig. 8B) were generated by PCR from the wild type Ad5 5' genome depicted in Fig. 1B using PCR clones indicated by the arrows in Fig. 1B. The resulting amplification products (Fig. 8A and 8B) sequences differed from the wild-type Ad5 genome in the number of A-repeats carried by the left (5') end.

As depicted in Fig. 8C, these amplification products were subcloned into the multiple cloning site of pAd.Link.1 (IHGT Vector Core). pAd.Link.1 is an adenovirus based plasmid containing adenovirus m.u. 9.6 through 16.1. The insertion of the modified PAC regions into pAd.Link.1 generated two vectors pAd.PACII (containing PAC domains I and II) and pAd.PACIV (containing PAC domains I, II, III, and IV).
Thereafter, as depicted in Figs. 10A and 10B, for each of these plasmids, a human placenta alkaline phosphatase reporter minigene containing the immediate early CMV enhancer/promoter (CMV), human placenta alkaline phosphatase cDNA (hpAP), and SV40 polyadenylation signal (pA), was subcloned into each PAC vector, generating pAd.PACII.CMVhpAP and pAd.PACIV.CMVhpAP, respectively.

These plasmids were then used as substrates for homologous recombination with d17001 virus, described above, by co-transfection into 293 cells. Homologous recombination occurred between the adenovirus map units 9-16 of the plasmid and the crippled Ad5 virus. The results of homologous recombination were helper viruses containing Ad5 5' terminal sequences that contained PAC domains I and II or PAC domains I, II, III, and IV, followed by the minigene, and Ad5 3' sequences 9.6-78.3 and 87-100. Thus, these crippled viruses are deleted of the E1 gene and the E3 gene.

The plaque formation characteristics of the PAC helper viruses gave an immediate indication that the PAC modifications diminished the rate and extent of growth. Specifically, PAC helper virus plaques did not develop until day 14-21 post-transfection, and on maturation remained small. From previous experience, a standard first generation Ad.CBhpAP helper virus with a complete left terminal sequence would begin to develop by day 7 and mature by day 10.

Viral plaques were picked and suspended in 0.5 ml of DMEM media. A small aliquot of the virus stock was used to infect a fresh monolayer of 293 cells and histochemically stained for recombinant alkaline phosphatase activity 24 hours post-infection. Six of eight Ad.PACIV.CMVhpAP (encodes A-repeats I-IV) clones that were screened for transgene expression were
positive, while all three Ad.PACII.CMVhpAP clones that were selected scored positive. The clones have been taken through two rounds of plaque purification and are currently being expanded to generate a working stock.

These crippled helper viruses are useful in the production of the AdΔ virus particles according to the procedures described in Example 3. They are characterized by containing sufficient adenovirus genes to permit the packaging of the shuttle vector genome, but their crippled PAC sequences reduce their efficiency for self-encapsidation. Thus less helper viruses are produced in favor of more AdΔ recombinant viruses. Purification of AdΔ virus particles from helper viruses is facilitated in the CsCl gradient, which is based on the weight of the respective viral particles. This facility in purification is a decided advantage of the AdΔ vectors of this invention in contrast to adenovirus vectors having only E1 or smaller deletions. The AdΔ vectors even with minigenes of up to about 15 kb are significantly different in weight than wild type or other adenovirus helpers containing many adenovirus genes.

Example 9 - AdΔ Vector Containing a full-length dystrophin transgene

Duchenne muscular dystrophy (DMD) is a common x-linked genetic disease caused by the absence of dystrophin, a 427K protein encoded by a 14 kilobase transcript. Lack of this important sarcolemmal protein leads to progressive muscle wasting, weakness, and death. One current approach for treating this lethal disease is to transfer a functional copy of the dystrophin gene into the affected muscles. For skeletal muscle, a replication-defective adenovirus represents an efficient delivery system.
According to the present invention, a recombinant plasmid pAdΔ.CMVmdys was created which contains only the Ad5 cis-elements (i.e., ITRs and contiguous packaging sequences) and harbors the full-length murine dystrophin gene driven by the CMV promoter. This plasmid was generated as follows.

pSL1180 [Pharmacia Biotech] was cut with Not I, filled in by Klenow, and religated thus ablating the Not I site in the plasmid. The resulting plasmid is termed pSL1180NN and carries a bacterial ori and Amp resistance gene.

pAdΔ.CMVLacZ of Example 1 was cut with EcoRI, klenow, and ligated with the ApaI-cut pSL1180NN to form pAdΔ.CMVLacZ (ApaI).

The 14 kb mouse dystrophin cDNA [sequences provided in C. C. Lee et al, Nature, 349:334-336 (1991)] was cloned in two large fragments using a lambda ZAP cloning vector (Stratagene) and subsequently cloned into the bluescript vector pSK- giving rise to the plasmid pCCL-DMD. A schematic diagram of this vector is provided in Fig. 11, which illustrates the restriction enzyme sites.

pAdΔ.CMVLacZ (ApaI) was cut with NotI and the large fragment gel isolated away from the lacZ cDNA. pCCL-DMD was also cut with NotI, gel isolated and subsequently ligated to the large NotI fragment of NotI digested pAdΔ.CMVLacZ (ApaI). The sequences of resulting vector, pAdΔ.CMVmdys, are provided in Fig. 12A-12P [SEQ ID NO:10].

This plasmid contains sequences form the left-end of the Ad5 encompassing bp 1-360 (5' ITR), a mouse dystrophin minigene under the control of the CMV promoter, and sequence from the right end of Ad5 spanning
bp 35353 to the end of the genome (3' ITR). The minigene is followed by an SV-40 poly-A sequence similar to that described for the plasmids described above.

The vector production system described herein is employed. Ten 150mm 293 plates are infected at about 90% confluency with a reporter recombinant E1-deleted virus Ad.CBhpAP at an MOI of 5 for 60 minutes at 37°C. These cells are transfected with pAdΔ.CMVdys by calcium phosphate co-precipitation using 50 μg linearized DNA/dish for about 12-16 hours at 37°C. Media is replaced with DMEM + 10% fetal bovine serum.

Full cytopathic effect is observed and a cell lysate is made by subjecting the cell pellet to freeze-thaw procedures three times. The cells are subjected to an SW41 three tier CsCl gradient for 2 hours and a band migrating between the helper adenovirus and incomplete virus is detected.

Fractions are assayed on a 6 well plate containing 293 cells infected with 5λ of fraction for 16-20 hours in DMEM + 2% PBS. Cells are collected, washed with phosphate buffered saline, and resuspended in 2 ml PBS. 200λ of the 2ml cell fractions is cytopsased onto a slide.

The cells were subjected to immunofluorescence for dystrophin as follows. Cells were fixed in 10N MeOH at -20°C. The cells were exposed to a monoclonal antibody specific for the carboxy terminus of human dystrophin [NCL-DYS2; Novocastra Laboratories Ltd., UK]. Cells were then washed three times and exposed to a secondary antibody, i.e. 1:200 goat anti-mouse IgG in FITC.

The titer/fraction for seven fractions revealed in the immunofluorescent stains were calculated by the following formula and reported in Table 2 below.

DFU/field = (DFU/200λ cells) x 10 = DFU/10^6 cells = (DFU/5λ viral fraction) x 20 = DFU/100λ fraction.
<table>
<thead>
<tr>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

A virus capable of transducing the dystrophin minigene is detected as a "positive" (i.e., green fluorescent) cell. The results of the IF illustrate that heat-treated fractions do not show positive immunofluorescence. Southern blot data suggest one species on the same size as the input DNA, with helper virus contamination.

The recombinant virus can be subsequently separated from the majority of helper virus by sedimentation through cesium gradients. Initial studies demonstrate that the functional AdCMVΔmDys virions are produced, but are contaminated with helper virus. Successful purification would render AdΔ virions that are incapable of encoding viral proteins but are capable of transducing murine skeletal muscle.

**Example 10 - Pseudotyping**

The following experiment provides a method for preparing a recombinant AdΔ according to the invention, utilizing helper viruses from serotypes which differ from that of the pAdΔ in the transfection/infection protocol. It is unexpected that the ITRs and packaging sequence of
51

Ad5 could be incorporated into a virion of another serotype.

A. Protocol

The basic approach is to transf ect the AdΔ.CMVlacZ recombinant virus (Ad5) into 293 cells and subsequently infect the cell with the helper virus derived from a variety of Ad serotypes (2, 3, 4, 5, 7, 8, 12, and 40). When CPE is achieved, the lysate is harvested and banded through two cesium gradients.

More particularly, the Ad5-based plasmid pAdΔ.CMVlacZ of Example 1 was linearized with EcoRI. The linearized plasmids were then transfected into ten 150 mm dishes of 293 cells using calcium phosphate co-precipitation. At 10-15 hours post transfection, wild type adenoviruses (of one of the following serotypes: 2, 3, 4, 5, 7, 12, 40) were used to infect cells at an MOI of 5. The cells were then harvested at full CPE and lysed by three rounds of freeze-thawing. Pellet is resuspended in 4 mL Tris-HCl. Cell debris was removed by centrifugation and partial purification of AdΔ.CMVlacZ from helper virus was achieved with 2 rounds of CsCl gradient centrifugation (SW41 column, 35,000 rpm, 2 hours). Fractions were collected from the bottom of the tube (fraction #1) and analysed for lacZ transducing viruses on 293 target cells by histochemical staining (at 20h PI). Contaminating helper viruses were quantitated by plaque assay.

Except for adenovirus type 3, infection with Ad serotypes 2, 4, 5, 7, 12 and 40 were able to produce lacZ transducing viruses. The peak of β-galactosidase activity was detected between the two major A260 absorbing peaks, where most of the helper viruses banded (data not shown). The quantity of lacZ virus recovered from 10 plates ranged from 10^4 to 10^8 transducing particles depending on the serotype of the helper. As
expected Ad2 and Ad5 produced the highest titer of lacZ transducing viruses (Table 3). Wild type contamination was in general $10^2$-$10^3$ log higher than corresponding lacZ titer except in the case of Ad40.

B. Results

Table 3 summarizes the growth characteristics of the wild type adenoviruses as evaluated on propagation in 293 cells. This demonstrated the feasibility of utilizing these helper viruses to infect the cell line which has been transfected with the Ad5 deleted virus.

<table>
<thead>
<tr>
<th>Adenovirus serotypes</th>
<th>p/ml</th>
<th>pfu/ml</th>
<th>p:pfu</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$5 \times 10^{12}$</td>
<td>$2.5 \times 10^{11}$</td>
<td>20:01</td>
</tr>
<tr>
<td>3</td>
<td>$1 \times 10^{12}$</td>
<td>$6.25 \times 10^9$</td>
<td>160:1</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$3 \times 10^{12}$</td>
<td>$2 \times 10^9$</td>
<td>150:1</td>
</tr>
<tr>
<td>5</td>
<td>$1 \times 10^{12}$</td>
<td>$5 \times 10^{10}$</td>
<td>20:01</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>$5 \times 10^{12}$</td>
<td>$1 \times 10^{11}$</td>
<td>50:1</td>
</tr>
<tr>
<td>12</td>
<td>$6 \times 10^{11}$</td>
<td>$4 \times 10^9$</td>
<td>150:1</td>
</tr>
<tr>
<td>35</td>
<td>$1.2 \times 10^{12}$</td>
<td>$4.4 \times 10^8$</td>
<td>5000:1</td>
</tr>
<tr>
<td>30</td>
<td>$2.2 \times 10^{12}$</td>
<td>$4.4 \times 10^8$</td>
<td>5000:1</td>
</tr>
</tbody>
</table>

Table 4 summarizes the results of the final purified fractions. The middle column, labeled LFU/μl quantifies the production of lacZ forming units, which is a direct measure of the packaging and propagation of pseudotyped recombinant AdΔ virus. The pfu/μl titer is an estimate of the contaminating wild type virus. AdΔ virus pseudotyped with all adenoviral strains was generated except for Ad3. The titers range between $10^7$ - $10^4$. 
Table 4

<table>
<thead>
<tr>
<th>Serotypes</th>
<th>LFU/ml</th>
<th>PFU/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>4.6 x 10⁷</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>6.7 x 10⁶</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6.3 x 10⁷</td>
</tr>
<tr>
<td></td>
<td>7a</td>
<td>3 x 10⁶</td>
</tr>
<tr>
<td>15</td>
<td>12</td>
<td>1.2 x 10⁵</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>9.5 x 10⁴</td>
</tr>
</tbody>
</table>

Table 5A-5D represents a more detailed analysis of the fractions from the second purification for each of the experiments summarized in Table 4. Again, LFU/µl is the recovery of the AdΔ viruses, whereas pfu/µl represents recovery of the helper virus.

Table 5A

<table>
<thead>
<tr>
<th>Ad2 Fraction #</th>
<th>VOLUME/µl</th>
<th>LFU/µl</th>
<th>PFU/µl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120</td>
<td>9532</td>
<td>8 x 10⁶</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>5.8 x 10⁴</td>
<td>3 x 10⁶</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>8.24 x 10⁴</td>
<td>6 x 10⁵</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>9.47 x 10⁴</td>
<td>1.2 x 10⁵</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>6 x 10⁴</td>
<td>8 x 10⁴</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>2 x 10⁴</td>
<td>6 x 10⁴</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>5434</td>
<td>5 x 10⁴</td>
</tr>
<tr>
<td>Total/10 pH</td>
<td></td>
<td>3.32 x 10⁷</td>
<td>1.35 x 10⁹</td>
</tr>
</tbody>
</table>
### Table 5B

<table>
<thead>
<tr>
<th>Ad4 Fraction #</th>
<th>Volume/ul</th>
<th>LFU/ul</th>
<th>PFU/ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>1000</td>
<td>1.75 x 10⁵</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>1.79 x 10⁴</td>
<td>2.8 x 10⁵</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>1.8 x 10⁴</td>
<td>5.5 x 10⁴</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>2909</td>
<td>1.25 x 10⁴</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>920</td>
<td>4 x 10⁴</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>153</td>
<td>3 x 10³</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Total/10 pH</td>
<td>4 x 10⁶</td>
<td>5.6 x 10⁷</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ad5 Fraction #</th>
<th>Volume/ul</th>
<th>LFU/ul</th>
<th>PFU/ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120</td>
<td>1.98 x 10⁴</td>
<td>6 x 10⁶</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>5.8 x 10⁴</td>
<td>3 x 10⁶</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>1.2 x 10⁵</td>
<td>1.5 x 10⁶</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>1 x 10⁵</td>
<td>1.4 x 10⁵</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>7.96 x 10⁴</td>
<td>8 x 10⁴</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>6860</td>
<td>6 x 10⁴</td>
</tr>
<tr>
<td>25</td>
<td>Total/10 pH</td>
<td>3.88 x 10⁷</td>
<td>1.2 x 10⁹</td>
</tr>
</tbody>
</table>
### Table 5C

<table>
<thead>
<tr>
<th>Ad7 Fraction #</th>
<th>VOLUME/ul</th>
<th>LFU/ul</th>
<th>PFU/ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>1225</td>
<td>$5 \times 10^5$</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>5550</td>
<td>$4 \times 10^5$</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>4938</td>
<td>$2 \times 10^5$</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>3866</td>
<td>$8 \times 10^4$</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>4134</td>
<td>$6 \times 10^4$</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>995</td>
<td>$7 \times 10^4$</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>230</td>
<td>$6 \times 10^3$</td>
</tr>
<tr>
<td>Total/10 pH</td>
<td></td>
<td>2.09 x $10^6$</td>
<td>1.3 x $10^8$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ad12 Fraction #</th>
<th>VOLUME/ul</th>
<th>LFU/ul</th>
<th>PFU/ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>31</td>
<td>$5 \times 10^5$</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>169</td>
<td>$8.5 \times 10^5$</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>245</td>
<td>$1.8 \times 10^5$</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>161</td>
<td>$1.1 \times 10^5$</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>62</td>
<td>$7 \times 10^3$</td>
</tr>
<tr>
<td>Total/10 pH</td>
<td></td>
<td>6.14 x $10^4$</td>
<td>1.65 x $10^8$</td>
</tr>
</tbody>
</table>
Table 5D

<table>
<thead>
<tr>
<th>Ad40 Fraction #</th>
<th>VOLUME/ul</th>
<th>LFU/ul</th>
<th>PFU/ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>80</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>80</td>
<td>184</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>80</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>80</td>
<td>168</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>80</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>100</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>100</td>
<td>32</td>
</tr>
<tr>
<td>20</td>
<td>Total/10 pH</td>
<td>$6.65 \times 10^4$</td>
<td>$1.1 \times 10^3$</td>
</tr>
</tbody>
</table>

C. **Characterization of the Structure of Packaged Viruses**

Aliquots of serial fractions were analysed by Southern blots using lacZ as a probe. In the case of Ad2 and 5, not only the linearized monomer was packaged but multiple forms of recombinant virus with distinct sizes were found. These forms correlated well with the sizes of dimers, trimers and other higher molecular weight concatamers. The linearized monomers peaked closer to the top of tube (the defective adenovirus band) than other forms. When these forms were correlated with lacZ activity, a better correlation was found between the higher molecular weight forms than the monomers. With pseudotyping of Ad4 and Ad7, no linearized monomers were packaged and only higher molecular weight forms were found.

These data definitively demonstrate the production and characterization of the Δ virus and the different pseudotypes. This example illustrates a very simple way of generating pseudotype viruses.
Example 11 - AdΔ Vector Containing a FH Gene

Familial hypercholesterolemia (FH) is an autosomal dominant disorder caused by abnormalities (deficiencies) in the function or expression of LDL receptors [M.S. Brown and J.L. Goldstein, *Science*, 232(4746):34-37 (1986); J.L. Goldstein and M.S. Brown, "Familial hypercholesterolemia" in *Metabolic Basis of Inherited Disease*, ed. C.R. Scriver et al, McGraw Hill, New York, pp1215-1250 (1989).] Patients who inherit one abnormal allele have moderate elevations in plasma LDL and suffer premature life-threatening coronary artery disease (CAD). Homozygous patients have severe hypercholesterolemia and life-threatening CAD in childhood. An FH-containing vector of the invention is constructed by replacing the lacZ minigene in the pAdΔ.CMVlacZ vector with a minigene containing the LDL receptor gene [T. Yamamoto et al, *Cell*, 39:27-38 (1984)] using known techniques and as described analogously for the dystrophin gene and CFTR in the preceding examples. Vectors bearing the LDL receptor gene can be readily constructed according to this invention. The resulting plasmid is termed pAdΔ.CMV-LDL.

This plasmid is useful in gene therapy of FH alone, or preferably, in the form of a conjugate prepared as described herein to substitute a normal LDL gene for the abnormal allele responsible for the gene.

A. *Ex Vivo Gene Therapy*

*Ex vivo* gene therapy can be performed by harvesting and establishing a primary culture of hepatocytes from a patient. Known techniques may be used to isolate and transduce the hepatocytes with the above vector(s) bearing the LDL receptor gene(s). For example, techniques of collagenase perfusion developed for rabbit liver can be adapted for human tissue and used in transduction. Following transduction, the hepatocytes
are removed from the tissue culture plates and rein infused into the patient using known techniques, e.g. via a catheter placed into the inferior mesenteric vein.

B. **In Vivo Gene Therapy**

Desirably, the in vivo approach to gene therapy, e.g. liver-directed, involves the use of the vectors and vector conjugates described above. A preferred treatment involves infusing a vector LDL conjugate of this invention into the peripheral circulation of the patient. The patient is then evaluated for change in serum lipids and liver tissues.

The virus or conjugate can be used to infect hepatocytes in vivo by direct injection into a peripheral or portal vein (10^7-10^8 pfu/kg) or retrograde into the biliary tract (same dose). This effects gene transfer into the majority of hepatocytes.

Treatments are repeated as necessary, e.g. weekly. Administration of a dose of virus equivalent to an MOI of approximately 20 (i.e. 20 pfu/hepatocyte) is anticipated to lead to high level gene expression in the majority of hepatocytes.

Numerous modifications and variations of the present invention are included in the above-identified specification and are expected to be obvious to one of skill in the art. Such modifications and alternations to the compositions and processes of the present invention, such as various modifications to the PAC sequences or the shuttle vectors, or to other sequences of the vector, helper virus and minigene components, are believed to be encompassed in the scope of the claims appended hereto.
SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Trustees of the University of Pennsylvania
Wilson, James M.
Fisher, Krishna J.
Chen, Shu-Jen
Weitzman, Matthew

(ii) TITLE OF INVENTION: Improved Adenovirus and Methods
    of Use Thereof

(iii) NUMBER OF SEQUENCES: 10

(iv) CORRESPONDENCE ADDRESS:
    (A) ADDRESSEE: Howson and Howson
    (B) STREET: Spring House Corporate Cntr, PO Box 457
    (C) CITY: Spring House
    (D) STATE: Pennsylvania
    (E) COUNTRY: USA
    (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:
    (B) FILING DATE:
    (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
    (A) APPLICATION NUMBER: US 08/331,381
    (B) FILING DATE: 28-OCT-1994

(viii) ATTORNEY/AGENT INFORMATION:
    (A) NAME: Bak, Mary E.
    (B) REGISTRATION NUMBER: 31,215
    (C) REFERENCE/DOCKET NUMBER: GNPVN.008PCT

(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: 7897 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

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

GAACCTGAGC AGCTGAAGCT TGAATTCCAT CATCAATAAT ATACCTTATT 50
TTGGATGGAA GCCAATATGA TAATGAGGCG GTGAGTTGCT TGACGTGGCG 100
CGGGCGCTGG GAACGGGGCC GTGACCTAGT GTTTCTAGGC GGAAGTACTT 150
GTATGTTGTT GGAATTCTAG TTTTCTTTAA ATGGGAAGTT ACGTAACGTT 200
GGAAAAACGA AGTACGAGTT TGAAGAAGTT GTGGGTTTTG TGGCCTTTCGT 250
TTCTGGGCGT AGTTTCGGCT GCGTCTTTCT GGGGTTTTTG TGGGACTTTT 300
AACCGTTCAG TCACTTTTTA GTCTATATAT TACTGGCTCT GCACTTGGCC 350
CTTTTTTACA CTGTGACTGA TTAGGCTGCT GGGTGTGCA GTGGGTTTTT 400
TTAATAGGT TTTTTTTTT ACTGGGTAAGG CTGACTGTCA GCTGCGCGCT 450
GTGAAGCGCT GTATGTTGTT CTGGAGCAGG AGGGTGCTAT TTTGCCCTAGG 500
CAGGAGGGTT TTTCAAGGTG TTATGTTGTT TTCTTCTCCTA TTAATTGTTT 550
TATACCTCCT ATGGGGGGCTG TAATGTTGTC TCTACGCTCT CGGTATGTA 600
TTCCCCCAAC GCTGGCATGC CTCAAGTCGCG ACTCTAGAGG ATCCGAAAGA 650
ACCTCCCACA CCTCCCCCTG AACCTGAAC ATAAAATGAA TGCAATTGTT 700
GTTGTTAACG TGGTATGTC ATCCGTTGAAT GGAAGCCTAA TAAAGAATAG 750
CATCACAATA TTCAAAATA AAGCACTTCT TCTACGCTAT TTGTCCTATG 800
GTTGTTGGGA ACTCATCAAT GTACATTGAC ATGTCTGGAT CCCCAGGGCC 850
GCCTAAGGTC GAGGGCAGTC TGGTCAGAAA GCAGAACCAG AAGGAGTTGG 900
AATAATAGCG AGAACAGAGA AATACGGGCA AAAATAATAC CCGTATCAGT 950
TTTGCTGATA TGGTGGATTG CATCTAGCCA AATCGGAAA AACGGAAGT 1000
AGGCTCCCAT GTAATAAAGAA AGAATAAACC GAACATCCAA 1050
AAGTTTTGTGT TTTTTAAATA GTACATAATG GATTTCTTTA CGCGAAATAC 1100
GGGCAGACAT GGGCTGCCG GGTTATTTAT TTTTTGACAC CAGACCAACT 1150
GGTAACTTTGA GCGACCGGCG CTACGCTTGA A_TCCGCCGA TACTGACGGG 1200
CTCCAGGAGT CGTGCGCCACC AATCACCATA TGGAAACGT CGATATTCCAG 1250
CCATGTGCGC TCTTCGCGGT GCAGGAGATG GCGATGGCCT CTTTTCAATCA 1300
GTTGCTGTGG ACTGTAACGG CTGATAGTTA ACTGGAAGTC GCCGGCCAC 1350
TGGTGTGGCC CATAAATTCA TCGCGCGGTC CCGCAGCGCA GACCGTTTTT 1400
GCTCGGAAG AGCTACCGGG TATACAGTGC TGACAATGGA AGATCCCCG 1450
GGTCAAAACA GCCGCGACTG AGCGGTTCGG GATAGTTTTT TTGCGGCCCT 1500
AAATCCCGAC AGTATTCCCG CTCTGCTACC TGGCCGAGCT GCCAGTTTAG 1550
GCAATCCGC GCCGAGATCG GTGTACGCTT CCCACCTTGA ACATCAACGG 1600
TAATCGCCAT TCGCACTACT CATCAATCCC GGTAGTTTCC CGGCCTGATA 1650
AAATAAGGGTT TCCCATGTATG CTGCCACCGG TGAGCGGTCA TAACTCAGC 1700
CGCATCGACA AGTGTATCTG CGTGCACTG CAACAAAGCT GCTTCCGGCT 1750
GGTAATGGCC CGCAGGCCTTC CAGCGGTTCGA CCCAGGCTTT AGGTCAATGG 1800
CGGGTCGGCT CATCAGGCC AATGTGTTA TCGAGCGGTC CACGGGTGAA 1850
CTGATCGGCAC AGGGGCGTCA GCAGTGTGTTC TTTTAGGCCA ATCCACATCT 1900
GTGAAGGAGGCC CCGTACCTGG CGGGTAATT GCCAAGCTTT ATTACACG 1950
TCGTAGCCAA AAATCCATTTG CTCTGCGGTTA AGATGGCGGA TGCGTGGGAA 2000
CGCGCGCGGG AGCGTCACAC TGAGGTTCCT CGCCAGAGCG CACTGCTGCCG 2050
AGGGCGCTGAT GTCGCGCGCT TCTGACCATG CGGTCGGCTG CGTTGCACTG 2100
ACGGCTACTG TGAGCCAGAT TTGCAGCAGG CTCGCAGGCCT GCAGTAAGTT 2150
AGGCAGTTCA ATCACTGTGT TACCTGTGGA AGCGACATCC AGAGGCACCT 2200
CACCAGCTTC AGCCGGCTTA CCATCCACGC CCGACCACCA GTGCAGGAAGC 2250
TCGGTATCCG TATGACGGAAG CAGGTATTCG CTGGTCACCTT CGATGGTTTG 2300
TGATCTTCCA GATAACTGCC GTCACTCCAA CGCAGCACCA TCACCGCGAG 3650
CCGGTTTTCT CGGGCGCGTA AAAATGCCCT CAGGTCAAAT TCAGACGGCA 3700
AACGACTGTC CTGGCCGTAA CGGACCCGAC GGGCGGTGCA CCACAGATGA 3750
AACGCGGAGT TAACGCGCCTC AAAATAATT CTGCTGCGGC CTTCCCTGAT 3800
CCAGCTTTCG TCAACATTAA ATGATAGCGA GTAACAACCC GTGGGATTCT 3850
CCGTTGGGAAC AAACCGCGGAA TTGACCGTAA TGGGATAGGT TACGTTGGTG 3900
TAGATGGGCG CATCGTAAACC GTGCTATCGC CAGTTTGAGG GGAGCAGGAC 3950
AGTATCGGCCT TCAGGAAGAT CGCAGCTCAG CCAGCTTTCC GCCACCCGCTT 4000
CTGGTGCCGG AAACCGGCCG AAGCGCCATT CGCCATCCAG CTTGGCACAAC 4050
TGGTGGGAAG GGCATGTGGT GCCGCCCTCT TCGCTATTAC GCCAGCTGGC 4100
CAAAGGGGGA TGGTGCAGAA GGGGGATTAG TTGGTAACGC CAGAGTTTT    4150
CCCAGTCAG ACCTTGTAAG ACGACGGGAT CGCGTCTGAG CAGCTCTTGG    4200
CTGGTGTCGCA GACCAATGCC TCCAGACCGC GCAACGAAAA TCACGGTTCT   4250
GGTTGGTCAA GTAAACGACA TGGTACTTC TTTTTTGATT TAGCAGCACCT   4300
TTTGGATCCC CGGGAATGGC GGCAGCGGGT ACAATTCGGC AGCTTTTGA     4350
GGAGATGAA CACCTCGTGT CAGCGCTCAGA AGTAAAGCGA ACATTTACTG    4400
AGGAGCGATT CTTTGATTTG CACCAACACC GGATCCGGGA CCTGAAATAA    4450
AAGACAACAA AGCATAACTT ACCAGTAAAC TTCTGGGTTT TTCAAGTTCT    4500
CGAGTACCGG ATCCTCTAGA GTCGGGAGGC TGGATCCGGTC CGGTCTCTTT   4550
CTATGGAGGT CAAAACAGGC TGGATCCGGC TCCAGGGGGA TCTGAGGTTT    4600
CAGTAAACGA GCTCTCCTTA TATAGACCTC CACGCTACA GCCTACGGC     4650
CCATTTTGCGT CAATGGGGCG GAGTTGTATG GACATTTTG AAATGTCGCG    4700
TGATTTTGTG GCCAAACAAA ACTCCCCATGG ACCTGATGGG GTGGGAGACT   4750
TGGAAATCCC CTTGATTCAA ACAGCGCTA CCAGCCATTTG ATGACTGGCC    4800
AAACCCGCGT CCCATGTGTA ATAGCGATGA CTAATACTGA GATGACTGCG    4850
CAAGTAGGAA AGTCCCCATAA GGTCTATGAC TGGGCATAAT GCCAGGCAGG    4900
64

CCATTTACGG TCATTGACGT CAATAGGGGG CGTACTTGGC ATATGATAACA 4950
CTTGATGTAC TGCCAAGTGG GCAGTTTACC GTAAATACTC CACCCATTGA 5000
CGTCAATGGA AAGTCCCCTT TGCGGTACT ATGGGAAACAT ACGTCATTAT 5050
TGACGTCAAT GGGCAGGGGT CTGGTGCCCG TCACGCCAGC GGGCCATTTA 5100
CCGTAAGTTA TGTAACGACC TGCCGCGTGA CTCTGAGGAA TCTCCCTAGA 5150
CAAATATTAC GCGCTATGAC TAACACAAAA TTATTCAGAT TTCACTTCTC 5200
CTTATTACGT TTTCCCGCGA AAATGGCCCA ATCTTTACTCG GTTACGCCCA 5250
AAATTACTAC AACATCCGGC TAAAACCCGC CGAAAAATTGT CACTTTCTGT 5300
GTACACCGGC GCACACAAAA AACGTCACTT TTGCCACATC CGTCGCTTAC 5350
ATGTGTTCCG CCACACTTGAC AACATACACAT TTCCGGCACA CTACTACGT 5400
ACCCGGCCCG TTCCACCGCC CGCGGCGACG TCACAACACTC CACCCCTCTA 5450
TTATCATATT GGCTTCAATC CAAAATAAGG TATATATTTG ATGATGCTAG 5500
CGAATTACATC GATGATATCA GATCTGCGCG TCTCCTTTAT GAGAGTGCTA 5550
TTAATTTCGA TAAGCCAGGT TAACCTGAGT TAATGAATCG GCAACCGCGC 5600
GGGGAGAGGC GGTTCGGGCTA TTGGGCCTTC TTCCGGCTCC TCGCTCAGTG 5650
ACTCGCTGGC CCTGCGTGGTT CGGCGGCGGC GAGGGTATGC AGCTGACTCA 5700
AAGCGGTAAC TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAGAAA 5750
CATGTGACCA AAAAGCAGAC AAAAGGCAGC GAACCCTAAA AAGGCCCGGT 5800
TGCTGGCGTT TTTCATAGCT CTCCGCBBBB CTGACCAGCA TCACAAAAAT 5850
CGACGCCTCA GTCAGAGGTC GCAGAACCCG ACAGAGCATAT AAAGATACCA 5900
GGCGTTTCC CCTGGAAGCT CCTCTGCGG CTCTCTCTGT CAGCCGCTCG 5950
CGCTTACGGG ATACCTGTCG GCCTTTTCCTT CTCCGGGAAG CGTGGCGCGT 6000
TCTCAATGCT CAGCCTGTAG TGATCTCAGT TCGGTTGAGG TCGTTCGCTC 6050
CAAGCTGGGC TGTTGACCCG AACCCCCCGT TCAAGCCGAC CGCTGCGCCT 6100
TATCCCGTAA CTATCGTCTT GAGTCCAAA CGGTAAAGACA CGACTTATCG 6150
CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA GGTATGTAGG 6200
CGGTGCTACA GAGTTCTTGA AGTGTTGGCC TAACTACGCG TACACTAGA 6250
GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTGGGAAAA 6300
AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG CTAGCGGTGG 6350
TTTTTTTTGT TGCAAGCAGC AGATTACGCG ChGAIAAAAAA GGAATCTCAAG 6400
AAGATCCCTT GATCTTTTCT ACGGGCTTCT AGCGTCTTAG GAAAGAIAAC 6450
TCACGGTTAA AGGATTTTTGAT CATGAGATT TCAAAAAGGA TCTTCCAATA 6500
GATCCTTTTA AATTAAAAAT GAAATTTTAA ATCAATCTAA AGTATATATG 6550
AGTAACCTTG GTCTGACGTA TACCAATGCT TAACTAGTGA GCCACCTATC 6600
TCAGCGATCT GTCTATTTTG TTATACCTATA GTTGCTTGAC TCCCCCGTGT 6650
GTAGATAACT AGCATAGGCG AGGGCTTTAC ATCTGGCCCC AGTGGCTGAA 6700
TGATACCAGG AGAGCCAGCC TCACCGGCTC CAGATTTATC AGCAATAAAC 6750
CAGCCAGCCG GAAGGGGCGA GCAGAGAAAGT GTGTTGTCAA CTGGATTCGC 6800
CTCCATCCAG TCTATTAAAG TTGCGCGAGA AGCTAGAGTA AGTATGCCCG 6850
CAGTTAATAG TTTGCAGCAAC GTTGTTTGCA TTGCTACAGG CATCGTGGTG 6900
TCACGGTCTG CGTTTTGTAT GGGCTCATTGC AGCTCGGTTT CCCAAGGATC 6950
AAGGCGAGTT ACATGAGCCC CCAATTTGGA CAAAAAGCG GTTAGCTCCT 7000
TCGCTTCCTC GATCGTTGTC AGAAGTAGTA TGGCAGCGAT GTTATCACC 7050
ATGGTTATGG CAGCAGTCCA TAATTTCTTT ACTGTCAATG CATCGTGAAG 7100
ATGCTTCTCT GTGACTGGTG AGTACTTAAC CAAGTCTACT TGAGAATAGT 7150
GTATGCGCGG ACCAGGGTGCC TCTTGGCGCG CGTCAATACG GGATAATAACC 7200
GCGCACAATA GCAAGACCTTT AAAAGTGGCTC ATCATTTGAA AACGTCTTTC 7250
GGGGCGAAAA CTCTCAAAGA TCTTTCCGTG TGAGATCC AGTCTCGATGT 7300
AACCACCTCG TGCAACCAAC TGACCTCTAC CATCTTTTCAT TTCCACCAGC 7350
GTTTCTGGGT GAGGAAAAAC AGGAAAGCGA AATGCCGCAA AAAAAAGGAAAT 7400
AAGGCGGACA CGGAATTTGA GAATACTCAT ACTCTCTCTT TTTCAATATT 7450
ATTGAAGCAT TTATCAGGGT TATTGTCTCA TGAGGCGATA CATATTTGAA 7500
TGTATTTAGA AAAATAACA AATAGGGTTT CGCGCAGCAG ATCCCCGAAA 7550
AGTGGCACCT GACGCTTAAAG AAACATTAT TATCATGACA TTAACCTATA 7600
AAAATAAGGC GATCAGGAG CCCTTTGCTC TCGCCGCTTT CGGTGATGAC 7650
GGTGAAAACC TCTGACACAT GCAGCTCCCG GAAGGCTCA CAGCTTGTCT 7700
GTAAGGGGAT GCGGGGAGCA GACAAGCCCG TCAGGGGCGG TCAGCGGGTG 7750
TTGGGGGTTG TCCGGGGCTGG CGTAACTATG CGCCGCTGAC GCAGATTGTA 7800
CTGAGAGTGC ACCATATGGA CATATTGTGG TTAGAACCGG GCTACAATTAT 7850
ATACATAACC TTATGTATCA TACACATACG ATTTAGGTGA CACTATA 7897

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7852 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
GAATTCGCTA GCTAGGGGGA GAATACATAC CGCCAGGGGT AGAGACAAAC 50
TTAACGCCCT CATAGGAAGGATAAAATG TAATAGGAGA GAAGGACACA 100
TAAACACCTG AAAAACCCCTC CTGCGCTGGG AAAATAGCAC CTCCTGGGCTC 150
CAGAACAACA TACAGGCGTT CACAGCGGCA GCCTAACAGT CAGCCCTACC 200
AGTAAAAAAG AAAACCTAAT AAAAAGACAC CACTCGACAC GGCACCAGCT 250
CAATCAGTCA CAGTGAAAAG AAGGGCCAAG TCAGGAGGCA GTATATATAG 300
GACTAAAAAA TGACGGTAACGT GTAAAACTCC ACAAAAAACA CCCAGAAAAC 350
CGCAAGCGAA CCTAGGCCCC GAAAGCAGAG CCCAAAAACAC CACAACTTCCC 400
TCAAATCAGTC ACTCCGGTTT TCCACGTTA CGTAACCTCC CATTTTAAGA 450
AAACTACAAAT TCCCAACACA TACAAGTTAC TCCGCCCATAA AACCTACGTC 500
ACCCGCCCCG TTCCACGCGC CGCGCCAGCG TCACAACAGC CACCCCCCTCA 550
CGTTGCCGCT TGCAAGGGCA TTGGTCGGGA CACCAGCAAG GACTGCTACA
AGCGCGATCC GTGCTTTTTA CAAAGTGGTG ACTGGGAAA CCCCTGGGTT
ACCCACACTT ATCGCCCTTG ACAGACATCC CCTTTGCGCA CGTGGCGTA
TAGCAGAGAG GCCGCGACCG ATCGCCCTTC CGAACAGTTG CGCAGCCCTGA
ATGGGCAATG GCGCTTTCGC TGGTTTCGGC CACCAGAGC GGTGCGCCGA
AGCTGGAAGC AGTGCGATCT TCCTGAGGCC GATACTGTCG TGTTCCCCTC
AAACCTGGCAG ATGACAGGTT AGATGCGGCC CATCTACACC AACGTAACCT
ATCCCATTAC GGTCATCCG CCCGTTGGTG CCACGGAGAA TCGGACGGGT
TGTTACTCGC TCACATTTAA TGGTTGATGA AGCTGGCTAC AGGAAGGCAA
GACGCCAATT ATTTTGGAG GGGTTAATCG GGGTTTCAT CTCTGGTGCA
ACGCGCGCTG GGTCGTTAC GGGGAGGCA GTGCTTTGGC GTCTGAAATTG
GACTGAGCG CATTCTACG CGCGGAGGA AACGCCTCG CGGGTGACTT
GCTGCGTGGG AGTGAGGGCA GTTATCTGGG AGATCAGGAT ATGGCGCCGA
TGAGCGGCAT TTCCGGTGAC GTCTCGTTGC TGCAAAAACC GACTACACAA
ATCAGCAGTT TCCAGTGGCC CACTGCGTTT AATGATGATT TCAGCGCGGC
TGTACTGGAG GCTGAAGGTTG AGATGTCGCG CGAGTTGCGT GACTACCTAC
GGTAAACAGT TCTTTTATGG CAGGTTGAAA GCGAGGTCGC CAGCGGCACC
GCGCCTTTCG GGGTTGAAAT TATGATGAG GTCTGGTGGT ATGCGCATCG
CGTCACACTA CTTGCTAACG TGCAAAACC GAAACTGTTG ACGCGCCGAA
TCCGCAATCT CTAATCGTGCG TGTTGGTGAA TCACACCCG CGACCAGCAG
CTGATTGAG CAGAAAGCCTG CGATGTCGTT TTCCGGAGG TGCGGATTTGA
AAATGACTCTG CTTGGTGCTGA ACGGCAAGCC GTTGCTGATT CAGGGCGTTA
ACCGTCACGA GCATCATCCT CTGCAATGTC AGTGCATGGA TGAGCAGACC
ATGGTGCGAG AATACCTGCT GATGAACAG AAWAACTTTA ACGCGTGCGG
CTGTTCGCATT TATCGGAACT ATCCGCCTTG GTCACAGGGT TCGGACCCGT
ACGGCCTGTA TGGTGTTGGAT GAAGCCAATA TGGAACCCA CGCAGTGCTG

1900
1950
2000
2050
2100
2150
2200
2250
2300
2350
2400
2450
2500
2550
2600
2650
2700
2750
2800
2850
2900
2950
3000
3050
3100
3150
CCAATGAATC GTCTGACCGA TGATCCCGCC TGGCTACCGG CGATGAGCGA  3200
ACGCGTAAACG CGAATGGTGC AGGCCGATCG TAATCACCCG AGTGATGATCA  3250
TCTGCTCGCT GGGGAATGAA TCAGGCCAGC CGCCTAATCA CGACGCCTGT  3300
TATCGCTGGGA TCAATACCTG CATCCTTTCC CCCCAGGCATG AGTATGAAGG  3350
CGGCCGAGCC GACACCACCG CCACGGATAT TTATTGCCGG ATGTACGCCG  3400
GCGTGGATGA AGACACCGCC TCCCCGCTGT TGCGGAAATG GTCCATCAAA  3450
AAATGGCTTT GCCTACCTGG AGAGACGCGC CCCGCTGATCC TTGGCAATA  3500
GCACCCGGCC ATGGGTAACA GTCTGGCGCG TTTTGCTTTA TACTCGAG  3550
CGTGTTCGTC GTATCCCCGT TTACAGGGGC GCTTCGTCGT GAATGCGGTG  3600
GATCACTCGC TGATTAATAA TGATGAAACG GCCACCGGTT GTCCGCTTTA  3650
CGGCCGATAT TTTGGCGATA CGCGCAACGA TCGCCAGTTT TGATGACG  3700
GTCGGTGCTT TGCGGACCAGC ACGCCGATTC CAGCGCTGAC GAAAGCAAAA  3750
CACCACGCAC AGTCTCTTTCA GTTCGGTTTA TCCGGCAAA CCATCGAGT  3800
GACCAGCCAA TACGCTTTCC GTCATAGCGA TAACGAGCTC CTGCACTGGA  3850
TTGTCGGGCTT GGGATGTAAC CGGCGGGCAG GCCTGAGATTC GCCTCCGAT  3900
GTCGGTCCAC AAGTAAAACA GTGATGAGGA CTGCCTGAA TCACGCAAGC  3950
GGAGAGCGCC GGGCAGCTG GCCCGTACAGT ACGCGATGTC CAACGAAAGG  4000
CGACCGCATG TGCAGAGGCC GGGCAGCTCA GCCCGCTGCA GCATGCGGT  4050
CTGGCGGAAA ACCTCATGCT GACGCGTCGG CCGCGTCCCC ACGCCGATCC  4100
GCATCTGACC ACCAGGGAA TGAGATTTGG CATCGAGCTG GGTAAATAAGC  4150
TTGGCAATT TAAACCGCAAG TCAGGCTTTT TTTCAGAGAT GTGAGTAGGC  4200
GATAAAAAAC AACTGTCGAC CGCCGCTGCC GATCATGTTC CCCGTCACC  4250
GCTGGTAAAC GCACGTTGGG TAAATGAGGC ACCCGCATTT GACCTAAAGG  4300
CCTGGGGCGA CGCCTGGAAG CGCGCGGGCC ATTACCAGGC CAAAGCAGCG  4350
TTGTGGTGAT GCACCGCGAG TACACTGTCC GTACGCGGTGC TGTATTACGAG  4400
CGTCACCGCG TGGTACGACTC AGGGGAAACC TTATATTATC AGCGGAAAAA  4450
CCTACCGGAT TGATGTTAGT GGTCAATAGG CGATTACCGT TGATGTGGAA
GTGGCGAGGC ATACACGGCA TCCGGCGCGG ATTTGCTGTA ACTGCCAGCT
GGCGACGGTA CGAGACGGGG TAAACTGGCT CGGATAGGG CCGCAAGAAA
ACTATCCCCA CGCCTTTACT GCAGCGCTTT TTAACCCTGG TGATCTGCCA
TGTGCAGACA TGTATACCCC GTACAGTCTTC CCGAGGAAAC ACGGTCTGCG
CTGGCGGAGC CGCGAATTGA ATTATGGCCC ACACCAGTGG CGCGCGGACT
CTCAGTCCAA CATCAGCCGC TACAGTCAAC AGCAACTGAT GGAACCGAGC
CATCAGCATC TGTCGACAGC GGAAGAGGGA ACATGCGTGA ATATCGACGG
TTTCCCATATG GGGATTGTTG CGCACGACTC CTGGAGGGCC TCAGTATCGG
CGGAATTACA GCTGAGCCGC GGTGCGTACC ATTACAGTT GGCTGCGTGT
AAAATAAT AATAACCGGG CAGGCCATGG CTGCCCCGTAT TTCCGCTAAG
GAATCCATT ATGTACTATT TAAAACAC AAACCTTTGGG ATGTCGGTT
TATTCTTTTT CTTTTACTTT TTTATCATGG GAGCCTACTT CCCGTTTTTC
CGATTTGGGC TACATGACAT CAACCATATC AGCAAAGTG ATACGCGTAT
TATTTTTGCC GCTATTTCTC TGTTCTGCTT ATATTCCAA CGCGTGGTTG
GTCTCTTCTC TGCAAAACTC GCCTCGACT CTAGGCGGCG GCAGGGATCC
AGACATGATA AGATACTATT ATGAGTTTGG ACAAAACCACA ACATGAATGC
AGTGAAGAAA ATGCTTTATT TGTAAGATTG TGATGCTAT TGCTTTATT
GTAACCATTA TAAGCTGCAA TAAAAGAGTT AAACAAACAA ATAGCATTCA
TTTTATGTTT CAGGTCAGG GGGAGGTGTT GGAGGTTTTT TCAGGATCTC
TAGAGTCGAC GACGCCAGGC TGGATGGCCT TCCCCATTAT GATTCTTTTC
GCTCTCGCGG GCATCGGGAT GCCCGCGTTG CAGGCCATGC TGCTCAGGCA
GGTAGATGAC GACCATCGG CAGACCTTCA AGGATCGCTC GCGGCTCTTA
CCAGCTAAC TTTGATCAGT GGACCCTGTA TGTCAGGCG GATTATTGCC
GCTCGCCGCA GCACATGGAA CGGTTGGCA TGGATTGTAG GCCGCGCCCT
ATACCTTGTG CTGCTCCCCG CGTTCGCTCG CGTGACATTG AGCCGGGCCA
71

CCTCGACCTG AATGGAAGCC GGGGCGCACCT CGCTAAGCGGA TTCACCACCT  5800
CAAGAATTGG AGCCCAATCAA TTCTTGCGGA GAACGTGAAA TGCAGAACC  5850
AACCCTTGGC AGAACATATC CATCGGCTGC GCCATCTCCA GCAAGCAGAC  5900
CGGCGGCATC TGCGGCGAGG TTGGGTCTCTG GACGACGGTTG CGCATAGCTC  5950
TGCTCTCTGC GTTGAGGACC CGCTTAGGCT GGGGGGTTTG CTTTACGTGT  6000
TAGCAGAATG AACACCCGAT ACCGAAGCGA AGCTGAAGCG ACTGCTGCTG  6050
CAAAACGTCT GGGACCGTAG CAACAAACTG AATGTCCTTC GGTTTCCGTG  6100
TTTCGTAAGG TCTGGAAAGC CGGAAGTCAG CGCCTGCAAC CATTATGTCC  6150
CGGATCTGCA TCAGCAGTAG CTGCTGTGCTA CCGCTGGAAC CACCTACATC  6200
TGATTTAACG AGCCCTTTCT CAATGCTCAG GCTGTAGGTA TCTCAGTTGG  6250
GTGTTAGGTGC TCTGTCCCAA GCTGGGCTGT GTCGACGACG CCCCCGGTTA  6300
GCCCCAGCC CGCGCCTTAT CGGTAACTA TCCTCTGAG TCCAACCGGG  6350
TAAGACACGA CCTATCGGCA CTGGCGCAAG CCACCTGTTA CAGGATTAGGC  6400
AGAGCGGAGT ATGTACGCCG TGCTAGCAGG TTCTTGAAAGT GGGGCCTAA  6450
CTACGGCTAC ACTAGAAGGA CAATATTTTG TATCTGCAGT CTGCTGAGGC  6500
CAGTTACCTT CGGAAAAAGA GTGGTGACCT CTGATCGGG CAACAAACCC  6550
ACGGCTTGTA GGCGTGTTTT TTTTTTTTGA AACGCACATA TTACGCGCAG  6600
AAAAAAAAAG TCTCAAGAG ATCCTTTTGAG CTTTTCTACG GGGTCTGACG  6650
CTCAGTGGA ACGAAACTCA CGTTAGGGGA TTGTTGTCAT GAGATTATCA  6700
AAAGGATCT TCACCTGATC CTTTTAATAAT TAAAAATGAA GTTTTAAATC  6750
AATCTAAAGT ATATATGAGT AAACCTTGGTC TGACAGTTAC CAATGTTCAA  6800
TCAGTGAGGC ACCTATCTCA GCGATCTGTC TATTTGCTTC ATCATAGTTT  6850
GCCTGACTCC GGTCGCAGTA GATAACTACG ATACGGGAGG GCTTACCATC  6900
TGCCCCAGCT GCTGCAATGA TACCGCGAGA CCCAGCTCA CGGCGCTCAG  6950
ATTTATCACG AATAAACACG CCGCAGGGA GGGCGAGCG CAGAAGTTGG  7000
CCTGCAACTT TATCCGCCCC TCATCACTCT ATTAATTGTT GCGGGAAGCG  7050
(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9972 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

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

TCTTCCGCTT CCTGCTAC TGACTCGCTG CGCTCGCTG TTCGGCTCGG 50
GCGAGCGGTA TCAGCTCAGT CAAAGCGCT AATACCGTTA TCCACAGAAAT 100
CAGGGGATAA  CGCAGGAALG  AACATGTGAG  CAAAAGGCCA  GCAAAAGGCC  150
AGGAACCGTA  AAAAGGCCGC  GTTGCTGGCG  TTTTTCCATA  GGCTCCGCCC  200
CCTGTACGAG  CATCACAAA  ATCGACGCTC  AAGTCAAGGG  TGCCGAACCC  250
GACAGGACT  ATAAAGATAC  CGGCGTTTCG  C.CCTGGAAG  CTCCCTCCTG  300
CGCTCTCCTG  TTTCCGACCTC  GCCGCTACCT  GGATACCTGT  CCGCTTTTCT  350
CCCTTCCGGA  AGCGTGGCCGC  TTTCTCATAG  CTCAGCGCTG  AGGTACTCTA  400
GTTCGGTGTA  GGTCTGTTCCG  TCCAAAGCTGG  GCTGTGTCGA  CGAACCACCC  450
GTTCAGCCCG  ACCGCTGCGC  TTATCCCGGT  AACTATCGTC  TTGAGTCCCA  500
CCCGTAAGAG  CACGACTTAT  GCCCAGTGCC  AGCCGCCACT  GGTAAACAGGA  550
TTAGCAAGAC  GAGGTATGTA  GCCGCGTGCTA  CAGAGTCTTT  GAAGTGTTGG  600
CCTAACTACG  GCTACACTAG  AAGAACAGTA  TTTGGTATCTG  GCCTGCTGCT  650
GAAGCCAGTT  ACCTCCGAA  AAAGAGTGGG  TAGCTCTTGA  TCCGGCAACA  700
AAACCAACGC  TGTTAGGCGT  GTTTTTTTTG  TTTGCAAGCA  GCAGATTACGC  750
CGCAAAAAA  AAGGATCTCA  AGAAGATCTC  TTGATCTTTT  CTACGGGGTC  800
TGACGTCAG  TGGACGCAA  ACTCACGTTA  AGGGATTTTG  GTCATGAGAT  850
TATCAAAAAAG  GATCITTCACC  TAGATCCTTT  TATAATTAAA  ATGAAGTTTT  900
AAATCAATCT  AAAGTAATTATA  TGAGTAAACT  TTGCTGACA  GTTACCAATG  950
CTTAATCAGT  GAGGCACCTA  TCTCAGGCTG  CTGTCTATTTC  GTTCCTACCA  1000
TAGTTGCTTG  ACTCCCGGTC  GTTAGATAAA  CTACGATAGC  GGAGGGCTTA  1050
CCATCTGGCC  CCAAGTGCTGC  AATGATAACCG  CCAGACCCAC  GCTCAACCGGC  1100
TCCAGATTTA  TACGCAATAA  ACCAGCCAGC  CGGAAGGGCC  GAGCGCAGAA  1150
GTGGTCTCTG  AACTTTATCC  GGCCTCATCC  AGTCTATAA  TTGTGTCGGG  1200
GAAGCTAGAG  TAAGTAGGCTG  GCCAGTTAAT  ATTTGGCGCA  AGCTTTGTCG  1250
CATTGCTACA  GGCATCGTGG  TGTCAGCGCTC  GTGCTTTGGT  ATGGCTTCTAT  1300
TCAGCTCCGC  TTCCCAACGA  TCAAGGCGAG  TTAAGTAGTC  CCCCATGTTG  1350
TGCAAAAAAG  CGGTTAGCTC  TTTCCGCTCT  CCGATCGTTG  TCAAAAGTAA  1400
74

GTTGGCCGCA GTTTATCAC TCATGTTTAT GGCGAAGCTG CATATTCTC 1450
TTACTGTGAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACCTA 1500
ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCAGGTT GCTCTTGGCC 1550
GGGCGTCAATA CGGGATAATA CCGGCCACA TGACGAACT TTAAGTGC 1600
TCATCATTTG AAAACGGTCTC TCGGGGCAGA AACCTTCAAG GATCTTACC 1650
CTGTGGAGAT CAGTGGATG TGAACCCACT CGTGCACCC ACGATCTTC 1700
AGCATCCTTT ACTTTACACCA GCGTATCTTG GTGACGAAAAC AGAGAAAGGC 1750
AAAATGCCGC AAAAAGGGGA ATAGGGCGA CACGGAATG TGAATACCT 1800
ATACTCTTCC TTTTTCAATA TATTGAGGCT AATTTACGAG GTTATTGTCT 1850
CATGAGCAGA TACATATTG ATGATTTTA GAAAAATAAA AAAATTGGGG 1900
TTCCGCGCAC ATTTTCACCA AAAATGCCAC CGACGTCTA AGAAACATT 1950
ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCAGA GGGCTTTTGC 2000
TCTCGCGCGT TCGGTGTATG AGGCTGAAAAT CCTCTGCACG ATGCAGCTCC 2050
CGGAGACGGT CACAGCTTGT TGGTAAAGCG ATGCGGGGAG CAGCAGGCC 2100
CGTCAGGGCG CGTCAGCGGG TGGGGCGGCT TGGCGGGCTG GCTAAACTA 2150
TGCGGCATCA GAGCAGTGG TACGAGAGAT GCCACATAAA AGTAAACG 2200
TTAAATTTTT GTTAAATTCC GCGTAAATT TTTGTTAAAT CAGCTCATTT 2250
TTTAACCAAT AGGCCAAAT TGCAAAGATAT CTTTAAATG AAAAGATA 2300
GCCGAGATA GGTTGAGTG TTGTTCCAGT TTGGACGAGA AGTCACCTAT 2350
TAAAGAACGT GGAATCCAC GCACGAGGAC GAAAAAGCGT CATACTGGGC 2400
GATGGCCCCC TACGTGAACT ACTACCAAAT TCAAGTTTTT TGGGTCAGG 2450
GTGCCGTAAG GACACCCAAT GAAACCCATA AAGGAGGCC CGATTAGAG 2500
CTTGACCGGG AAAAGCCGCG AACCCTGGCGA GAAAGAAGGA GAAAGAAAGGC 2550
AAAGGAGCGG GCCCCGAGGC GCCTGGCAAGT GTAGCGGTCA CGCTGCAGGT 2600
AACACCACA CCCGCGCGCG TTAATGCGCC GCTACAGGCG CGTCATATGG 2650
GTTGCTTTGA CGTATGGCGGT GTGAAATACCC GCACAGATGC GTAGGAGA 2700
AATACCGCAT CAGGCCCTT TCGCCATTTA GGCCTGGCAAT CTGTTGGGAA 2750
GGGCGATCGG TGGGGGCTTC TTAGCTATTCA GGCAGGCTTC GCAGAGGGGG 2800
ATGTGCTGCA AGGCAGTTAA GTGGGTAAC GCCAGGTTTT TCCACGTCAC 2850
GACGCCGTCG TACGAGCCGC ATTGCCAAAGC T2AAGGTGCA CGGCAACGCT 2900
GGCCACTAGT ACTCTCGAG ATCCTGCTATC GTCCGGCGTCG GCGACGTAAG 2950
CGTATCGATG CGGCCAGCTG CAGCCGGCGG CCAATATGAT CTTAAGGCTA 3000
TTAATATTTCC GGAGATATCG TGTGCGCTA AGCTTAACAA CGGTACCTC 3050
TAGAACATTA GCTAGCCAAT TCTACATCAATA AATATACTTTATTTTGGAGA 3100
TTAAGGCCAA TATGATAATG AGGGGGTGGA TTTTGAGACG TGGCGCGGGG 3150
CGGGGAAAGG GCCGGGGGGT GCTAGTTTTT AGGGCGGAGT AACATGTATG 3200
TGTTGGGAAAT TGTAGTTTTC TTAATATGGG AAGTACGTA ACTGGGAAA 3250
ACGGAAGTGA CGATTGAGG AGTTCTGCG TTTTGGGCT TCGTTTTTCTC 3300
GGCGTACGTT GCCGTGCGGT TTTCGGGCTT TTTTTTTGAG ACTTTAACC 3350
TTACGTCAATT TTATGCTCCT ATATATAACT GCTGCTCACT TGGCCCTTTT 3400
TTACGTGTG ACTGATTGAG CTGTTGCGGT GTGAGTGGT GTTTTTTTAA 3450
TAGGTTTCTC TTTTTCTGAG TAAAGGTGAC TGTGCTGTAG CGCGTGTAAG 3500
GGCGTGATAG TGTTCCTGAG GGGGGAGGCT GCTATTTTGC CATGGCGAGA 3550
GGTTTTTCTA GGTGTTTATG TTTTTTCTC TCTGTATATTTT CTTTTTTATAC 3600
CTCTATGGG GGCTGTAATT TGTTCTCTAC GCCCTCGGGGT AGTATTTCCC 3650
CCCAAGCTTG CATGCCCTGA GGTGCTAAGT AGAGGATCGG AAAAAACCTC 3700
CCACACCTCC CCTGAACCT GAGAACATAA AATGAGGCAA TGTTTTGTGG 3750
TAACCTTGGTT GTTGCACGTTG ATAATGAGTA CAATAAAAGC ATAGCAACTA 3800
CAAATTTCA CAAATAAGCA TTTTTTCTAC TGCAATCTTAG TGGTTGTTTG 3850
TCAAAGCTCA TCAAATGATC TCAATTGAGT TGGATCCCTG TACCTTGCCA 3900
AACCCTACGG TGGGGTCTCT CAITCCCCC TTTTTCTGGA GACTAAAATCA 3950
AATCTTTTTAT TTATATCTATG GCTGCTACTC TATAGCTTTC AGCTGGTGAT 4000
ATTGTGAGT CAAACTAGA GCCTGGACCA CTGATATCCT GTCTTTAACAA 4050
AATTTGGACTA ATCGCGGGAT CAGCCAATTC CATGAGTTAA TGTTCCATGT 4100
CAACATTATG TGTGCCTCTCT AAAGCTCTTG ATCTTGGATC TCCTTTCTCTG 4150
TCTCCTCTTT CAGAGCAGCA ATCTGGGCGT TAT3ACTTGCA CTTGCTTGAG 4200
TTCCGGTGGG GAAAGAGCCTT CCACCTGTCG GAGGGGGGCA TGGGTTGCCG 4250
GAAGAGGCTTC TCTCTGGTCA GCAGTTTCCTG GATGGAAATC TACTGCCCGCA 4300
CTTGGTTCCT TCTATGACC AAAAATTGTT GCATTCCAG CATGCTTCTT 4350
ATCCTGTGTT CACAGAGAAT TACTGTGCCA TCACCAATG CTGTTTTTAG 4400
AGTCTCTCTA ATTATTGTT ATGTTACTGG ATCCAAATGA GCACCTGAGGTT 4450
CATCAAGCAG CAGATCTTC GCCTTACTGA GAACAGATCT AGCCGAGCAC 4500
ATCAACTGCT GTGGCCATAG CCTAGGACA CAGCCCCCAT CCACAGGAC 4550
AAAGTCAAGC TTCCAGGAA ACTGTTCTAT CACAGATCTG AGCCCAACCT 4600
CATCTGCAAC TTCCATATT TCTTTGATCAC TCCACTGTTC ATAGGGATCC 4650
AAGTTTTTTC TAAATGTTCC AGAAAAATA AATACTTCTT GTGGATACAC 4700
TCCAAGAGCT TTCTCCACTG GTTGCAAAGT TATTGAAATCC CAAGACACAC 4750
CATCGATTCTG GATTCTCTCT TCAGTGTTCA GTACTCTCAA AAAAGCTGAT 4800
AACAAAGTAC TCTTCCCTGA TCCAGTTCTT CCCAGGCCC CCACCTCTCG 4850
GCCAGGACTT ATTTGAAGGG AAATGTCTCT TAAATGACCA TTTCACCCTT 4900
CTGTGTATTT TGCTGTGAGA TCTTTGACAG TCAATTGGCC CCGTGAGGCG 4950
CAGATTTCAT TTCTTCACAT GTTGAAATTC TCAATAATCA TAACTTTTGA 5000
GAGTTGGCCA TTCTTTATAG GTTGGTGGTA CTGGTGGAGT TACCTTTGCTG 5050
TTGGCATGTC AATGAACTTA AAGACTGCGC TCACAGATCG CATCAAGCTA 5100
TCCACATCTA TGCTGGAGTT TACAGCCAC TGCAATTAC TCATGATATT 5150
CATGGCTAAA GTCAGGATAA TACCAACTCT TCCTTTCTCT TCTCTGTTG 5200
TTAAAAATGGA AATGAAGGTA ACAGCAATGA AGAAGATGAC AAAAATCATT 5250
TCTATTCTGCA TTTGGAAACCA GCACGATGTT GACAGGTACA AGAAACAGGTT 5300
GGCAGTATGT AAATTCAGAG CTGTTTGGA A CAGAGTTC A AAGTAAAGCT 5350
GCCGTCCGAA GCCCAAGAAG GTTCCATAGTC CTTTAAAGCT TGAACAAGA 5400
TGATGTGAATA TTGGAATCCT GCCTTCAGAT TCCAGTGGTT TGATTGTCTG 5450
TGAGGTGTTG AGGAAATATG CTCTCACAAT AAATAAAGCC ACTATCACTG 5500
GCACTGTGGC AACAAAGATG TAGGGTTGTA A AACTGCGAC AACTGCTATA 5550
GTCCTCAAATCA CAATTAATAA CAACTCGATG AAGTCAATA TGGTAAAGGG 5600
CAGAAGGTCA TCCAAAATTG CTATATCTTT GGGAATCTA TTAAGAATCC 5650
CACCTGCTTT CAACGGTTTG AGGGTTGACA TTAGGTCTTG AAGAACAGAA 5700
TGTAACATTT TGTGTTGTA AAATTTTCAGAC ACTGTGATTA GAGTATGCAC 5750
CAGTGTTAGA CCTCTGAAAG ATCCCCATAG C AAGCAAGATG TCGGCTACTC 5800
CCACGGTAAT GTAAACAGCA TAATAGCAAC TGGTGTGTG GATAATCACT 5850
GCATAGTCCT TATTTTACT TATGACTATA TTTCCCTTGT CTGGAAGG 5900
AGGTGTTCCTA AGGAGCCACA GCACAAACAA AGAAGCAGCC ACCTCTGCCA 5950
GAAAAATTAC TAAGCAACAA ATTACGCA AAATAAGCT CTTGTTGACA 6000
GTAATATATC GAAGTGATATG GTTCCATGTA GTCACTGTG TGTATGCTCTC 6050
CATATCATCA AAAAGCACT CCTTTAAGTC TTCTCGTATA ATTCTTTAC 6100
TTATTTACA GCCAGTTTCT TGAGATAACC TTCTTTGAATA TATATCCAGT 6150
TCAGTCAAGT TTGGCTGGAG GGCAGTGAC ACTTTTCTGT TGAGTGTCTG 6200
TGCTTTTCCG TGAATGTCTT GACCTTGATT AACTGAGTGT GTCACTCAGGT 6250
TCAGGACAGA CTGCTCCTTT CGTGCTGAA GCGTGCGGCG AGTGCTGATC 6300
ACGCTGATGC GAGCGATAT CGGCTCTCCC TGCTCAGAAT CTGTTACTAA 6350
GGACAGCCCT CTCTCTAAAG GTCTCATCAG TATCTCTTCTG ATGCATTCA 6400
TTTGTAAGGG AGTCGTTTGC ACAATGGAAA ATTTTCTAT TAGAGTGGATT 6450
GGATGGAGAA TAGAATTCTT CCTTTTTCC CCAACTCTG CAGTCTTGGTT 6500
AAAAGATTGG TTTTTTTTGT GTGTCAGAG AAGAGGGAC TCTCCTTCTA 6550
ATGAGAAACG GTGTAAGGTC TCACTTAA GAAGTTTCT CTTTTCTGCA 6600
CTAATTGTT CTGAAGAATC ACATCCCAGT AGTTTTTGAG TAAAGTCTGG 6650
CTGTAGATTT TGGAGTTCTG AAAATGCTCC ATAAAAATAG CTGCTACCTT 6700
CATGCAAAAT TAATATTTTG TCAAGTTTCT TAAATGTTCC CATTTTGAAG 6750
GTGACCAAAG TCCTAGTTTT GTTAGCCAATC AGTTTACAAG CACAGCCTTC 6800
AAATATTTCT TTTTCTGTAA AAACATCTAG GTATCCAAAA GGAGAGCTTA 6850
ATAAATACAA ATCAGCATTCT TTTGATATCT CTTGTGCTAA AGAAATTTCT 6900
GTCGCTTGAC CTCCACTCAG TGTGATTCCA CTTTCTCCAA GAACATTATTT 6950
GTCTTTTCTCT GCAAACCTTGG AGATGTCTTC TTCTATTTGG CATGCTTTGA 7000
TGACGCTTCT GTATCTATAT TCATCATAGG AAAACCAAAG GATGATATTTT 7050
TCTTTAATGG TGCCAGGCAAT AATCCAGAAG AAGATGAAAT 7100
TCTTCCACTG TGCTTATATT TACCCCTGTA AGGCTCCAGT TCTCCATAA 7150
TCATCATTAG GAGTGAGGTC TTTGGCTGCTC CAGTGAGATT AGCAACCGCC 7200
AACAACCTGTC CTCTTCTTAT CTTGAAATTA ATATCTTTTCA GAGAGAGGT 7250
ACCAAGAAATG GAAATTAC TGAAAGAAGG GCTGTATCGA CCAATTGAAG 7300
TTTTTCTATT GTTATTTGTT TGTTTTTGCT TCTCAAAATA TTCCCCAAAT 7350
CCCTCCTCCC AGAAGGCTGT TACATTCTCC ATCACTACTT CTGATGTCG T 7400
TAAGTTATAT TCCAAATGCT TATATTCTTG CTTTTGTAAG AAATCCTGTA 7450
TTTTGTITAT TGCTCAGAAG GAGTCATACC ATGTTTGTAC AGCAGCAGGA 7500
AATGGCCGAG TGACGCCTAG GCGCAGAACA ATGCAGAATG AGATGGGTGT 7550
GAATATTTTC CGGAGGATGA TTTCTTTGAT TAGTGATAG GGAAGCAGAG 7600
ATAAAACAC ACACAAAGAAG CCTGAGAAAG AGAAGGCTGA GCTATTGAAG 7650
TATCCTCAGT AGGCTGCTTT CCGAGTCAGT TCTGTTTCTG TTTGCTTTAA 7700
GTTTTCAATC ATTTTTTCCA TTTCTCTTTC CCCAGCGATG GCCCTCAACAG 7750
ATGAGATGTT TTGAGTTTCA TTTGGGATAA TCACAAGCTT TTCACTGATC 7800
TTCCAGCCTC TGATCTCTCT GTACTTCATC ATCAATCTCC CTAGCCCCGCC 7850
CTGAAAAAGG GCAAGGACTA TCAGAAACC AAGTCCACAG AAGGACAGCG 7900
80

GGCGGTCAGC CAGGCGGCCC ATTTACCGTA AGTTATGTAA CGACCTGCAG 9250
GCTGATCTCC CTAGACAAT ATTACGCGCT ATGAGTAACA CAAAATTATT 9300
CAGATTTCAC TTCCTCTTAT TGAGTTTACC CGCGAAATCG GCAAATCTTT 9350
ACTCGGGTAC GCCAAATATT ACTACAACAT CCCGTTAAA CCGCGCAGA 9400
ATTGTCACTT CCTGTGTACA CCGGCCACA CCAAAAACTT CACTTTTGCC 9450
ACATCCGCTG CTTCATGTG TTGGCCCACT ATGGAACACT CACAATTCGG 9500
CCACACTACT ACGTACCCCG GCCGTTTTCC AGGCCCGCGC CCAGGCACA 9550
AACTCCACCCT CCTCATTATC ATATGGGCTT CAATCCAAAA TAAGGTATAT 9600
TATTGATGAT GCTAGCATGC GCAAATTTAA AGCCGTGATA TCGATCGGCG 9650
GCAGATCTGT CATGATGATC ATGGCAATTG GATCCATATA TAGGGCCGCG 9700
GTTATAATTAT CCTCAGGTTC GCAGTCATGC GCCATTCGAA TTGGTAATCA 9750
TGCTCATAGC TGTTTCTGT GTGAAATTGT TATCCGCTCA CAAATCCACA 9800
CAACATACGA GCCGGAAGCA TAAAGGTGAA AGCCGGGGGT GCCTAATGAG 9850
TGAGCTAACT CACATTAATT GCCTGCGGCT GACTGCCCAG TTTCCATCGC 9900
GGAACGCTGT CAGTGCCAGCT ATCGGAATTG ACGGCAACGC GCGCGGGGAG 9950
AGGCCTTTTG CGTATGGGCC GC

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

TAGTAAATTT GGGC
(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
AGTGAAATCT GAAT

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
GAATAATTTT GTGT

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown
(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGTAATATTT GTCT

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

WANWTTTG

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCAATTCAT CATCAAATAAT ATACCTTATT TTGGATGAA GCCAATATGA
TAATGAGGGG GTGGAGTTTG TGACGTGGCG CGGGCCGTGG GAACGGGCG
GGTGACGTAG GTTTTAGGGC GGAGTAACCT GTATGTTTG GGAATTTGAT
TTTTCTTAAA ATGGGAAGTT ACGTAACCTG GGAAACCGGA AGTGACGATT
TGAAGGAAGTT GTGGGTCTTT TTGCTTTTGAT TTCTGGCGGT AGGTTCGCCGT
GGGTTCCTCT GGTTTTTTTT TTGATGATTTT ATCCCGTTACG TCATTTTTTTA
GTCTATATATA TACTCGCTCT GCACTTGGCC CTTTTTTACA CTGTGACTGA
TTGAGCTGGT GCCGTGTCGA GTGGTCTTTTT TTTAATAGGT TTCTTTTTTT
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Start Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCTGGCTAGG CTTGACTGTTA</td>
<td>450</td>
</tr>
<tr>
<td>CGCTGCGGCT GTGAAGCGCT</td>
<td></td>
</tr>
<tr>
<td>GTATGTGGTT CTGAGGCCTTTAT</td>
<td>500</td>
</tr>
<tr>
<td>TTTTGCAGATT CAGGAGGTT</td>
<td></td>
</tr>
<tr>
<td>TTTCCCAGGT CTAGCTCTCTA</td>
<td>550</td>
</tr>
<tr>
<td>TATATCGCTTT CAGATGGTG GA</td>
<td></td>
</tr>
<tr>
<td>CAGGTAGCTA AATCCCGGCCC</td>
<td>600</td>
</tr>
<tr>
<td>GTCTAGACAT TAAGCCCAGA</td>
<td></td>
</tr>
<tr>
<td>CTCCTGATGA CATCAGACAT</td>
<td>650</td>
</tr>
<tr>
<td>AGCTGGATT TTTTCTACTT</td>
<td></td>
</tr>
<tr>
<td>TTTTGAAGCT ACTTTCT AGTCCGCGC</td>
<td>700</td>
</tr>
<tr>
<td>GTCTAGACT GTGTTTACACT</td>
<td></td>
</tr>
<tr>
<td>GGTGGGTGGC TAACGAGT</td>
<td>750</td>
</tr>
<tr>
<td>ACTTACTTTA AAGCGAAATG</td>
<td></td>
</tr>
<tr>
<td>CATCATATAC CGAGAATT GGA</td>
<td>800</td>
</tr>
<tr>
<td>GCCATATA TGGGTAAAT CAGTATTTA</td>
<td></td>
</tr>
<tr>
<td>TGATAGTTGT AATATGCA TTA</td>
<td>850</td>
</tr>
<tr>
<td>ACTGAAGCTG TAACTACCA AAGAATGC</td>
<td></td>
</tr>
<tr>
<td>TTATTTACAA AAAAACTGAA AGTTTAAG</td>
<td>900</td>
</tr>
<tr>
<td>CAGATCAGT AATCCGCTC</td>
<td></td>
</tr>
<tr>
<td>ACCTGGGATG AATGATG ATGAGGCTG</td>
<td>950</td>
</tr>
<tr>
<td>GATCGAGT TGTGTTTT TAAACTTG</td>
<td></td>
</tr>
<tr>
<td>GAGATCTT GAAATCGCA TTAACACA</td>
<td>1000</td>
</tr>
<tr>
<td>ACA CAGTTA TACAAGTA GGTGTA</td>
<td></td>
</tr>
<tr>
<td>TGGTAAATC ATCTGTGTA CTTAGCTA</td>
<td>1050</td>
</tr>
<tr>
<td>CATGAT CACATGCTA CATGCTAGT</td>
<td></td>
</tr>
<tr>
<td>CTGGGAGCTA AGCAGGATG TACAGGA</td>
<td>1100</td>
</tr>
<tr>
<td>AGTTTCGAGT GAAATGCTA AAATAGCA</td>
<td></td>
</tr>
<tr>
<td>TTTTCTGTTA GGAATGGCA AGCACCC</td>
<td>1150</td>
</tr>
<tr>
<td>AGTCTTGA GCTT CTC GTAGATGA</td>
<td></td>
</tr>
<tr>
<td>TAACAAACTA TGCCATAGT AATGTTATG</td>
<td>1200</td>
</tr>
<tr>
<td>CAGGT GTCCCA ATTATGCTT AATC</td>
<td></td>
</tr>
<tr>
<td>GAACATGAA TGGTTTAA AATCAGCA</td>
<td>1250</td>
</tr>
<tr>
<td>TAAGAG AAGTAA AATAAAGCTCA</td>
<td></td>
</tr>
<tr>
<td>TTTTGG CTTTGA TT TGGTAA AAG</td>
<td>1300</td>
</tr>
<tr>
<td>AGATCAAGG AAGAAGCTA TACGCTTCA</td>
<td></td>
</tr>
<tr>
<td>AGTCTTAAG GAAATGCTG TTTGGAA</td>
<td>1350</td>
</tr>
<tr>
<td>ATTAATG TGTGTTTT TCTTTGAAAA</td>
<td></td>
</tr>
<tr>
<td>AGGAGGAC AGGAAGACTG AAGAGAC</td>
<td>1400</td>
</tr>
<tr>
<td>TGA TGCTA TTTTGG GTG ATAGTTA</td>
<td></td>
</tr>
<tr>
<td>GAAGCTT GTCCTCATGTTA AGAATGCAA</td>
<td></td>
</tr>
<tr>
<td>ACTGAGTTAA TTTCTTTA AAGCGATTTACAT CACGATTTTG</td>
<td>1450</td>
</tr>
<tr>
<td>GAAATGCTA AAGCAAGG AAACAGGCT AGCTGTCCTCA</td>
<td>1500</td>
</tr>
<tr>
<td>AGATGGTGA AATGTTGAA GAAGATGGT AACA CACATGAGTCA</td>
<td>1550</td>
</tr>
<tr>
<td>AAAGCTGTT TGCAGTAA TAAGTGGT AATGGGCAA ACCA CACTGC</td>
<td>1600</td>
</tr>
<tr>
<td>TAGAAGT AATTTA TTACACTT GA CGT AAGGTTT GA AAAATGATA</td>
<td>1650</td>
</tr>
</tbody>
</table>
GGACAGTGAT TTCTATTTGAC AGAATGCTCT TTAATGCTA AGAAGAGAA
ACTGCGATGA GAGAGTTAAA GCTCTCTCTG GAGGTCTTTA GCTTTTCTGG
GCACTTTTTTC TCGGGTTGCA ATGACCTGGA TTTGTTTAGA CATTCCAGCC
CGTCACCTAG ACCAGAGATG TGGGAGAGCC T1_7TCTCTC AAAACCTTCC
AACCACGTG TCTTCTCACC CAAATCCTTG TGGGAGTTA TTTGCAGGGA
AACCAGTAGCA AAGGAGACAA ATGCACTTCA TGGGCTTCTG GACTGATATT
CACCAGGGTC ACAATGTGAT TGGGGTACTT TCTTAACAGT ATGCTCTAAGT
CTTGCAAGCAT TAAAAAATATA AATCATCACA ATGAAGAAAA AAAAAACCAA
AAAATCTAAA ATCTAAAATT CATCATCATC ATCAACACA ACAACAAACA
CAACAACAAA ACCACCCACT TCGGGTGGAT TATATGAAGA GGGCGAACA
ATTTAGTGTG AATATAGAG ATGTGTTATAT GTATAGTTG AATATCCTAT
CCATTCCTTT ACAGAGTTGT TGCTCCCCCC ATATAATTG ACTGGAGGAC
CGCAACCTTT AGCTCTTACC ATCTTCTCC TACTGTCTGG GAGTTAAAAA
TGTCACTCTGA TTATCTATTG CAGAAACATC ATTAATATA ACCCAACAGT
AGGAAGTGGAA ATATATCAGC CAACAAATTA CTATGATAGT AAGTCCTGTG
TATTCATTCG CATGTCTCTT GAAAAATATG AATCTCTTAG CTCTCAGTGG
AAAGTTTTAA ACTAGAAACA TCTGAGGCC TAGACAATAT TTTATGTGG
CGGTAGTCTC CGTGGTTTGG GCTCCAGGGA AAATTCACTC TTGGCCCAAGC
AGATAAGCCC AGATGACCTAG AAGCAAATTC CATTAGGAAG TGGAAGAAGAC
ATTTGAAGAA GATAATCTCAT ATCTATTTAT CATATACCT ATAGTATTAA
TAT/AIDSGTA GACATATAGA TGTATAAAAT GAAAGCCCAT AGCCAGGCCC
ACTCATGCAA CAATTCTCAA AAGAGCAATA TGAAGACGTC ATTTGCTGGG
GTTCTATGC AAGAAATAAA AAAAACTCAA TGAATCTCCAT ATGAATACCA
CGCTAAAGTA AGTCAAAACA ATGTGCTGCC TCAATGTGTG ATGTGCTGTA
TGTTGTTGAT GTGGGTTGCT GCATGTATTG GTGCGTGTTG ATGTGCTGTA
TGTTGTTGAT GTGTGCTGTC GTGTGCTGTTT TTTTATAAAC
AACTTTTTTT ATAAAGCACA\ CTTAGTTTA CAATCCTCTT TTATAACTGT 3050
TATAAAATTTT TAAACAACC AAAAAAGGTT CCATATAAAG AAATGCAAG 3100
TTATTTAGCT ATCAAGATT TACATGTGTT TTTTAAACTT TTTGTTACAA 3150
TTGCATTAGAC GTGATAAAAC TGCCATTGTT AAJAAACAAC TAACAGACTT 3200
AGAAACTACT GAAATCCTCA GTATAGTACC ACTAACCCTTC ACAAATAAT 3250
AGATTTTATT TCTTGTAAAAC TCTTACTGTC TAATCCTCTT TGGTGTACGA 3300
ATATATAAAA ACCATGCGG GAACAGGAGG TGTAAACAAC TTTATCTGC 3350
TCCTTCTTCA TCTGTCATGA CTGAAACTAA GGAATCCACG CTCTGCCCCA 3400
AATCATCTGC CATGTGGAAA AGGCTCTTCA CATGGTGTCG TCTCTCATTG 3450
GCTTTCCGGA GGACATTTCTT CTCCTTGAAC TAGGAAGGAGA TGGTGTTGAGT 3500
TGCTCCATCA CTCTTCTAA CCCGTGTGTT GTGTCTCTGG GAGGACTCAG 3550
AGATCTTCC TACCCCATAG TTCTCGATT TGACTGCCA ACCACTCGGA 3600
GCAGGATAGG CTGACTGCTA TCTGACCTCT GCAGAGAGGT GGAAGGAGAG 3650
GACACCGTGGA TGCCATTCA CTTAGCTCTCA GCCTGGGCTG CCTCAGGGAG 3700
CTGTCTCAGT CTAAGTAACT GAACTCCAG CTGTATATTG TGGTCTTCCA 3750
GGATTTGCAT CCTGGCTTCC AGGGGCTTTT GTGTGTGGCG CAGTAGCTTA 3800
GCCCTAGCAA TGAGCTCAGC ATCCCTGAGG CTCTGAGGAG AGGTGGGCAT 3850
CATCTCACGA GGAGATGGCA GTGGAGACAG GCCCTTATGC TCATGCTGCT 3900
GCTTCAAGCG ATCATATTCT GCTTGCGAGAT TCTCCTGTTTTC TCCCTCAAGA 3950
TCTGCTAGGA TTCTCCTAG GTCCCCCTTT TTACTCACTCT TAAAGGAAT 4000
CAAGATCTGG GCAGGACTAC GAGGCTGGCT CAGGAGGGAG TCCTGGTTCA 4050
AACTTTGGCA GATAGCTGG ATAAACAAAT GTTCATCAGC TATGCTCTCA 4100
TTAGGAGAGA TGCTATCATT TAGATAAGAT CCAATTGCTGT TTTCCATTTC 4150
TGCTAGCCTG CTAGCATAAT GTTTAATGCTG TGAATGAGTA TCATCGTGTG 4200
AAAGCTGGGG GGACAGGAGC GGCCTCAGAT CTACTGGCCA GAGGTTGATC 4250
AGAGTTACGG GAGTTTCCAT GTTGTCCTCC TCTAACACAG TCTGACCTGG 4300
CAGGTAGCCC ATTCGGGGAT GCTTCGCAA A ATACCTTTTG GTPC GAATTT
TGT TTTTTTAG TACCTTGGCG AAGTCGCGAA CATCTTCTCC GAGATGT AGTC
GGAGTGCAAT ACTCTCACCT GGGTAGTGC ATTTATG GC CTTTGGCAAC
TCGGCCAGAA AAAAAGCAAC TTGGGCAGAT GTATAATTA AAATGCTTTA
GGCTTTCTGA CCTGAATCCA ATGATTG GACTCCTTACA GATGTTACAC
TTGGCTTGAT GCTTGGCAGT TCCAGCGACA GCCACTCTGT GCAA GACCGG
CAGCCACACC ATAGACTGGG GTTCCAGCGC C ATCCAG TAC AGGAAGAGAG
CAGCTTCAAT CTCAGGTTTA TTATGGCAAA ATGGGAAGCA GCTC C TGAACA
CTCGGCTCAA TGTTACTGCC CCCAAGAGA GCAACTTCAC CCAACTGTCT
TGGAATTTGA ATAGAATCAT GCAGAAGAAG ACCAGGCTA CGCTGGTCAC
AAAAGCCAGT TGAACCTGCC ACCTGTCTTA AAAGGTATCT GTACTTTGCT
TCCAAGTGTG CTCTCACAG AGAAATGATG CCAGTTTTAA AAGACAGGAC
ACGATCCCTC CCTGTTCTGC CCGTATCATA AACATGAGA AGCCAGT GTA
GACACATATC CACACAGAGA GGGACATTGA CCAGATGT TTGCTCTTG C
TCCAGACGAT CATAAATTGT AGTCAACAG T TAAATTATCT GCAGGATATC
CATGGGCTGG TCATT TGGCCTGAGT TGGGT GGTG CAGT CAGTCAT
CAGCTGACAG GCTCAAGAGA TCCAAGCAA GGGGCTTCTG GAGCCTCTG
AGCTTCATGG CAGT CCTATA CGCGGAGAAC CTGACATTAT TCAGGTCA CG
TAAAGACTGG TAGAGCTCTG TCATTGTGGG GTGTTCCA AAAGTTGTTT
GGTCTGGTG GTTAGATAG T AGGGC A CTT G T G TGTTGTTG GATGGCTCTC
TCCCAGGAC C C T G A C T G A G T G A A A G G A G T C T G G G A ATCCAGGACC
AAAGTCCCTG TGG GCTT CAT GC A G CTG T C T G A G C A G T C C T C T C
TAAGGTGAGA GCTGAATGCC CAGTGTTGTC AGCTGATGTG CAAGGTCATT
GACAGGATTG ACATTCTCTT TAAAGAGTGTC AATTTCCTCCC CGAAGTG CCT
TGAC T T T T T C C A A G T G ATC T GCAGAGAGT GATGGAGGAG ATCC C CCA C T
GGCTGCAGG ATCCCTTGATCACCTCAGCT TGGCGGAACCT GAGGTCCAG 5650
TTCATCGGCA GCTTCCTGAA GTTCCCTGGAG TCTTTCAAGA GCTTCATCTA 5700
TTTTTCTCTG CCAATCGACT GAGCGCAAGT TCAATTTGTC CCAATTCACG 5750
TGACCTCTTT CAGCCCTCCT TCGTAGGAGC CAGAGCAT TCTGAGCTCT 5800
TTCTTCAGGA GGCAGTTCCT TGAGCCTCGT GTAGGTTTC TCTAGTCCTT 5850
CAGAAGGCCTG CTCTGTCAAG AATATTTCTCA CAGTCTCCAG AGTACTCATG 5900
ATTACAGGTT CTTTATGTTT CAAATCTCCT TCTGAGGCCC TATGTATATC 5950
ATTTGCTTT CAGACCTGCT TTAGAGTGGT GGAATACCC ACCGATGGGT GCCTGACGCC 6000
TCAGTTTCATC ATCTTTTCAGC TGTAGGCCAA CAAAGAAGTC CTGAAGAGAA 6050
AGATGCACAC ACTTCCTCGT CTAAGAAGCT CTTGCAAGAT GGGCACTTAT 6100
GTTGAGAGAC TTTTTCTGAA GTTGAATTTCT CTTGAATTC ATGTTATCCA 6150
AACGTCTTTG TAACAGGGCT GCTTCATCCG AACCTTCACG GGATCTCAGG 6200
ATTTTTTGCG CATTTTTCATC AAGATTTGTA TAGATATCTG TGTGAGTTTC 6250
AATTTCTCTT TGGAGATCTT GCCATGGTTT CATCAGCTCT CTAATCCC 6300
TGGAGTCTTG TAGGAGCTCC TCTTACGGG AAGCTTCTTG TAGGACATTG 6350
GCAGTTTGGT CTGGCTCGGG CAACTCAGGA AGAAACTCTG CCAGTCTCAG 6400
AGGGACTGC TGCAATGATT TATGAATTTG TTCCAAGCA GCCTCTTGGT 6450
CACTACTCTT TTTATGAATG TTTCCCAAG AAGTATTGAT ATTCTCTGTT 6500
ATCATGTTGA CTTTTCTGGT ATCATCAGCA GAATATCCC GAAGAGTTT 6550
CAGTGCCGAA TCATTGGCAA GTCTCAGACT TATCTGCGGT TGACGGAGGT 6600
CTTTGGCCCA GTGCTTGGTT TCTGTGATCT TCTTTTGGAT TGCATCTACT 6650
GTGTGAGAGAC TTGTTTCCCA GAGTCAAGG TCTCTTGTA CCTGTCCTAT 6700
GACCTGTCG CTCTCTCCTC TAGCTCAGAG CCATTGTTT GAATCTTTTA 6750
ACATTTACATT CAACGTGGTT CGCTCTGTCC GAGCTGTTTC TCTGAGCCTCA 6800
TCCACTGAA TCTGAAATTCT TTCAATTGGCA TCGTAATGA TCTGTCTAGC 6850
TTCTTGGATTG CTGGTTTTCT TTTTCAATTT CTGGGCAGCA GTAATGAGTT 6900
CTTCCAATTTG GGGGCCTCCTC TTGTCCCAATT CGTGCAGTGT TGCTTTCTGTT
TTGATGTACG TTTCATTCAG GTCTTCCAGA TCCACCCCA CACTCTCTGT
TGATTTTATA ACTGAGTCAA GCAGGACAGC CCAGTGCATA AGGTCTGTCC
AAGCTGCGGTT GAAGTCTGCC AGTGACGAGT CCGCAGCAGA CAAAGAAGAT
GGCAATTCTA GGTGGGAGAT GACAGTTTCC TTAGTAACCA CGATTTGTGT
CAGCTAGAGTA ACAGTCTGAC TGGCAGAGGC TCCAGTAGTG CTCAGTCCAG
GGGCACGGTC AGGCTGCTTTT GTCTCAGCTT CCGAAGTAA ATGGTTTACA
GCCTCCCACT CAGACCTCAG ATCTTCTAAAC TTCTCTTCTAG CTGGTGGACT
GCTTGGTTTT TCTTTTATACA AATGCTGCCA TTTGCACAAA AGCCTTTCCA
CATCCGGCTTG TTACCGTGTA ACTGTATCTT CAATCTCTTT TATGTCAAAT
GGTCCTGCTT GAATGGTTTG TTTATAATT TTCAACGTTG TTCTAATAGG
AGAGACCCAC AGAAGCAGGT GATCCACGCT CTCTTCAAGC TGCTAAAAAT
CTTCTAAGTG AACCTCAAGC TCTCTTGTGG TCTCAGTTAA AGCTCTGGAAG
ACCTTCATCC ACTGAGATTG TGCTCTGTGG AGCTTTTTTT CAAGTTTATC
TTGGCTCTTCG GCCCTATAGG GAGCACTTAC AAGTACTGCT CCTCTTTTTT
CATTTATAAG TGTTAGAATT CCGGCGGCA GGGCAACTC TTCTGCGAGT
AACTTGACTT GTTCAAGTGG TTCTTTTACG TGCTGCTCAT CTCCAAGTGG
AGTAATAGCA ATGTTATCTG CTTCTTCAG CCACAAAAC AATTCATTTTA
AATCTTCGGT AATAATCTGAC AAGACATTCT TTTGGTCTCC AATACCTTTT
CTCGTTCTTG CCAAGCTCTC GCAGATGTGC TGCCACCGCA GACTCAAGCT
TCCTAATTGG TCTTGTAAGA TATGGCATAC TGTTTTGGA GACTGGTGAA
TTAATTTCTC CCCAGTGGCA TCCAGTGTCC TGACAACGAG TTACGCGCTGC
CCAATGCCCC CCGGAGTTC CTTTATGACAT CATTAGTATT TAGCATGTTTC
CAGGTTTAAT GAATTTCTGT TCTTTTGGA AATCTGTTCA ACTTCTACAG
GCCATGGATT AAATACCTTC ATATCATATA GAAAGTGTCC CCAATTTTCA
ACTGATCTGT CGATCCGCCC TTGCGCTTCC TTGATCATTT CTTAAGAGTTT
TTCCCCCTGG AAATCCATCT GTGCCACGCC TTCTGTACT TTCACCTTTT 8250
CCATGGAGGT GGCACTTTGC AAGGCTGCTG TCTTCTCTTT GTGAATATAA 8300
TCAATCCGAC CTGAGATTTC TTGCAATTTG TCTTTTATAT TCTTAAAAGA 8350
CTCCTCTTTGC TTTAAAAAGAT CTTCAAAATTC T1PAGCAGC AGTTCAAGAG 8400
TATTTAGAAG ATGATCAACT TCTGAAGAGAG CTGTTAAGAT ATGACTGACTC 8450
TCGGTCAAAAT AAGTTGAAAGG CACATAAGAA ACATCCAAAG GCATATCTTC 8500
AGTCTGTCAT ACCATAGTTT CTTCATGGAG AGTGTGAAATT TGTGCAAAGT 8550
TGAGTCTCTCG AAATCGAGCA AAATGCTCTT CAATTTGCCG CCAGCGCTTG 8600
CTGAGCTGGA TCTGAGTTTG CTCACCTGCC ATTGCGGCC CATTTCTCAG 8650
CAAGCCTCTCA GCTGGCCTGC GCACTGCAAT CAGCTCCTCT TTCTTTCTTC 8700
GCAAATTCAG ATCAATTTC TTTAATTTTC TTTCATCTCT TGTCTCCAGT 8750
AGGCTGGCTA AAAAAAAATT AATTTCTACC AAGCATTTCA GGAATCATC 8800
AGCCTGCTTC TGTTACTGAT ACCACTGGTG AGAAATTCTT AGGGCCTTTT 8850
TTCTTCTTTG AGACCCTCAA TCCCTGAGAG CATATGTTTT TGTCTGTAAC 8900
AGCTGCTGTT TTATCTTTAT TTCCCTTCGC TTTTCTCTAT CTGTGATTCT 8950
TTGTTGTAAG TTGCTCTTCTC TTGCAACAAA TTTATTTACA GTACTCTCAT 9000
TGTTTCAACT CATATCTTTA TGAAGTTTCT CCTCTTTCAG ATTCACCCCC 9050
TGCTGAATTG CGCCTCCAG TGGTTCAAGC AATTTTGTGA TATCTGAGTT 9100
AAACTGCTTC AATTCCTTCA AAGGAATGGA GGCCTTTCCA GTTTAATTTC 9150
TGATGAAATG AGTGCAAAAT CGACGGTGTG GCTCAGAGAT TTGGGCTCTC 9200
ACTACTTTCC TGCAGTGGTC ACCGCGGTTT GCCATCAATT TTGCTGCTTG 9250
GTCAGGTGTG GAGTCCACCT TTGGGCGCAT GTCAATTCACT TCACGCCCTTA 9300
AACGCTTAAAG AATGTCTTTCC TTTTGTGGTG TTTCTTCTCT TTTGACACTCA 9350
TCTAAAAGTT CATCTGCATG AATGATCCAC TTTGTTGGTTT GTTCTATGGT 9400
CTGATCAAAAG GTTTCCATGT GTTTCTGGTA TTCAACAAAA AGATTTTAGC 9450
ATTCTTCTAC TCTGGAGGTG ACAGCTATCC AGTTACTGTG CAGAAGACTC 9500
AGTTTATCTT TCTACCAAGGT TTCTTTCTTG CCAACACAAC TTTTCAAGA
CTCTCCCTAT TCTGTAACAC TCTTCAAGTG ACGCTTCCTGT TTCTCAATCT
CTTTTTTGGT AGCGTCCTCC CAGGCAACTT CAGAAATCAA ATTACTTGGC
ATTCCTTCAGTCTGCTGCTC TTGGTCACAT C1JATCTGT TTGGTGGCCAG
CCATCTCTTTT AAGACATTTCA TTTTCTCTCT CTTTACCAAGC ACACCAACTCA
AGCATTTTCCT CAACCTGGTC TTTTCTCTCG TTACCTTGG CACCAACTCAG
TTGTGATAATGCA ATTTCAACAG TGGATCTCGT TCAATCAAGCT CTTTGGATT
TTCTGTCTGC TTTTTCTGTA CAATTTGAGC TCCGGTTTTA ATCACCATT
CCACTTCAGA CTTGACTTTCA TCTGCTCTTT TATACAATTT CACACAAATGA
CTTAGTGGTG ACTGAATTAC TTTCTGTGCA ACACTTTGG TTTCCATGCG
AGGCAAAATGC ATCTTGAATT CATCTAAAAT CATCTTACTT TTCTCTAGAC
GTGGTTCAGA ATGGCTGCT GTTGGGATA ATCGGAAATT CATGGAGACA
TCTTTGATAAT TTTTCTGTGC AACATCAATT TGAGAAGAACA CCTTTAGGGT
GGCATCCTTC CCGTGTTAT GTTTCTCTCAT TTCTCTAAA CTTATCTCAT
GACTTGCAGA ATCTGATTGG ATTTCTCTGG CTTCTGAGG CATTGGAGCT
GCATCCACCT TGCAATGTGT AATAGCTGCC AACCTGTGAG CAATGAATTC
AAGCACTCC TGAATTAAGT GCAAGACCTT TCCAATTCTCG TGGGCAAGCT
GGATACTCTG TTCAAGCAAC TTTTGTTCCT TCACAGCCTC TCTAGTGTAG
TCACCTCAAC GAGATATATA CTTCTCAAGC TCCTCATGCA TCAGGCATAC
CAGACTTCTC CACATTGAA GTAGCTGGTG CAAATAGCGA ATCTGATTGT
GTTCTCTTCC TGAATGATGC ATCAGATTTC CAAGATATTC TAGCATTCAA
GTGATTCTCT CAGTCTGCTC AGGAACATT TCCATGATTTC TAAGTTCAG
TTACCTTACA TTTGAGCCACT TGGTGTACTGTT CTTCTAAATAT GACAATAC
CATGCCAACA TGGCCAACACT TCTTCACAAAG TTTTGACTTT TCCATTACGC
CTGTTGCACAT CCAATTGTA GTGGTGGTCG AGAGTTGCAA TTTCCTTATT
TAAGGCCCTCT TGAGCTGAGG GGTGGACCTG AGCTATTACA CTAATTTACAG
TCTCAGTAAG GAGTTTCAC TTTAGTTTCTT TTTGTAAGTG CTCCTTTCTTA 10850
GCTCTCTGCA TTTCTCTCAC AGCATGCTGT AAATCATCTG GAGTTTATA 10900
TTCAAAATCT GTCTGTAGAT ATTCTTCTCTT AGCTTGATTC ACACCTCATT 10950
GCATCTCTGA TAGATCCTTT TGGAGGCTTA CGJTTTTATCT CAAACCTGCG 11000
TTAAGGCTTT CTTTTCTGTA GTAGACCTGG CGGCATATGT GATCCACTGT 11050
AGTGTAAAGG CTCTCTTGAAG GTGTCTCCAG CTCTGGATGC AAATCAAGTT 11100
CAGCTTCATG CTTTATCTTC TCCCACCTTC CTTAACACTT ATTTAAACTG 11150
GGCTGAATTC TTTGAATATC ACCAATCTAA AGTCTGCACT TTTTGAGCTG 11200
TTTTTTCAAG ATTTCAAGCT CCCCAAGGCG AGGCATTCC TCTTTGAGGA 11250
AAACATCAAC TTCAAGCGATC CATTTTCTGA AGGGTTTTAT GTGATTTCTG 11300
AATTTTCTAA GTTTATCTAC ATGTTCCTCT ACCTTTGGCG AGCCCTTCAC 11350
CAACCTGGAG GAAAGTTTCT TCCAGTGGCC CTCAATCTCT TCAAATTCTG 11400
ACAGATATTT CTGAAGGCTT CCTTTCCTGG TATCTTATGG CATCTTCTTC 11450
ACAGTGTCAC TCAGATATGG GAAGCCTTTT TGGATCTTTT TCCAAAAGCT 11500
TTGCAGAGCG TGTTATATTT CGAGTCTCTT TCACCTATT TCAATCTCTG 11550
TAACACTAAC ATGAGTTGCA GAGATTTGCG TTTCTGACTG CGGAGTCAC 11600
GTCTGTATGC TACTCATTTG CTCTGTAAG CGGATTGCTG CAAAGTGTC 11650
AAAAATGGTC TTTGACCTTT TCTTCTGGCC CCTACACCA TCAAAGATGT 11700
GGTTAAAATG ATTAGTAAG GCCACAAAGT GTGCACTCCAG AAACATTGGC 11750
CCCTGTCCCT TTTTTCTTCAG TTTGACACT GTGAAATTTC ATTGTCTCAAT 11800
TTGAGGCTGA AGAGCTGACA ATCTGTTGAC TTACCTCCTA CAACATTTATA 11850
ACTGGCTTTT AATTGCTTGT GGCTGATAA GGGTGTTGTA CGGATTTTTC 11900
AACAAGTTTT CCACGATAGT TGTCATCTGT TCCAAATTGT GTAGCTGATT 11950
ATAAAAGGTA ATGATCTTGG TTTGATACTC TAGCCATTTA ACTCTCTCAC 12000
TCAGCAGATTG GCAGAATTTCT GTCACCCGCG GTTCAGTTG TTCTGAAGCT 12050
TGCTGTATAC TTTCGACATT AACACCTCAG TTTGCCATCT GTTCCACCAG 12100
GAAAGTACTT CTTCTAAAGGCT GATTTTGTTAA CATATCCAGAT TTACTTTCCGTT 13450
CTCCATCAAT GAACGTGTCAAGTGACTTTGTCTG TCTGGGAGCT TCCAAATGCT 13500
GTGAAAGGATA GGGGCTCTGTTG GAATATCAGT AGGTGGAACCC ATAAGCACGCC 13550
TGTTGAGAGG CATACACTTTT GAATCAGGCC T1.1GGAATATG AAGAAGTTTG 13600
TTCATAGCCC TGTTGCTAGAC TGACTGCTGAT CTGTTGAGAGC TAAATGATCT 13650
GGTGATGTAA TTGAAATATG TCTTCTCTAG TTAATTTGGA AGATGTCCTG 13700
GGCAACATTTT ACCATTCTTTG GATGGCTTCA ATGGCTCATT GTTGTGGCAA 13750
AATTTGAGAGT GTGATGTGTA TGTACATATAA GATGGAATCC TGTTCTGGAT 13800
AATGTGCCAGA AACATCTTTCA GATCAAGAA AATTTTCTAT GCCAATCTGG 13850
CATTTTTGCAAT TGGGAGGCT TGGTGCCAGT GTTTGGTGAG CAGATGTGCTG 13900
TGAAACCACA CATTCCCAAT CAAAGGACTC GGGCCTGTGA CTATGGATAA 13950
GAGCATTCAA AGGCAACCCT TGGGATCAGC TAGGATGCAA GTTGTAGACG 14000
TTAACCTGTTG GATAATTACGG TGGTACTGCT CGAACCCAGCG TCAGAAGAAT 14050
CTTTTCAGTG TGTTTGTGCT GCAATCCAGC CATGATAGTT TTCATCATCAAT 14100
TTTTGACTTG CCAAGGAGG CTATATTATCC AAATCAAACC AAGATGAGT 14150
TTGTGACTTC CATCCACTAT TGGCAATGCTT CCTATATATCA CTAAATCAATAC 14200
ATTTTTTTTC TGGTGGGACC GCAGTGCTTT GTGAGATCTT TTCAGGCGAT 14250
GAGGCTTTTTG AGATCCCTTT TCTTTTGGCC GATAATGCCC TGTAAGGCCTT 14300
TCCAGAGGTT CTAGGAAGGGG TTTTCCCCATG TGGCAGTCAG TGAAGAGGTT 14350
GCTTGATGTG TGCTTTTCAAA ACTTGGAAAA TGTTGCATTG ATCCATTTTTG 14400
TGAAATTTTTT CTTTTGAAA CTTTCTCTTT CATCAAGATC CTCTACTTCTT 14450
TCCCCAAAAC CATTGGGAAA GAAAAAGTAT ATATCAAGGC AGGGATAAAA 14500
ATTTTGCTAA AAGTTTTCTCC CAGTTTTATT GCTCCAGGAG GCTTATGTGAC 14550
GATGGAGGAG CAAATACCTT CAGCAGCTTT GACAAAAAAA AAAAAAAAAA 14600
TAGGACTTCA AGCCTTCCAA TCTGTTTTTT CTATAAAGCT ATGGCTTTCA 14650
AGAGCGGAAT TCCTGCAGCC CGGGGATCAG ACTAGTCTAA GACCCGCCC 14700
GGGTACAATT CCGCAGCTTT TAGAGCAGAA GTAACACTTC CGTACAGGCC 14750
TAGAAGTAAA GGAACACATCC ACTGAGGAGC AGTTCTTTGA TTTGACCCAC 14800
CACCAGATCC GGGACCTGAA ATAAAAGACA AAAAGACTAA ACTTACCAGT 14850
TAACCTTCTG GTTTTTCCAGT TCCTCGAGTA CACGATATCTC TAGATGCTCG 14900
AGGCTGGATC GGGCCCGGTG TCTTCTATGG AGGTCAAAC AGCGGGATG 14950
GGGTCTCCAG GGGATCTGAC GGTTCACTAA AGCGACTCTG TTATATATAGA 15000
CCCTCCACCG TACACGGCTA CGGCCCATT GGGTCAATGG GGGCGAGTTG 15050
TTACGGACATT TTGGAAAGTC CGGTGTGATT TGGTGCCAAA ACAACTCCC 15100
ATTGACGTCA ATGGGGTGGA GACTTGGAAA TCCCCGTGAG TCAAACCGCT 15150
ATCCACGCC ATCGATGTAC TGCCAAAACC GCATCACCAT GTGAAATAGCG 15200
ATGACTAATA CGTAGATGTG CTGCAAGTGA GGAAGTCCC ATAAGGCTAT 15250
GTACTGGGCA TAATGCCAG CCGGCCATT GACGTCACTG ACCTCATTG 15300
GGGGCGTACT TGGCATAATGA TACACTTGAT GTACTGCCAA GTGGCGAGTT 15350
TACCCTAAT ACTCCACCCA TTGGACGTAA TGGAAGTCCC ATATTGGCGT 15400
TACTATGGGA ACATACGTCA TTATGGACGT CAAATGGGG GGGTCGTGGG 15450
GGGTCAGCC AGGCGGGCCA TTTACGGTAA GTTATGTAAC GACCTGCAAG 15500
TCGACTCTAG AGGATCCTCC TAGACAAATA TTACGGCTTA TGAGTAACAC 15550
AAAATTATTC AGATTTCCAT TCTCTCTATT CAGTTTTCCCG CGGAAAAATGG 15600
CCAAATCTTA CTGGGGTACG CCAAAATTGA CTACACATC CGCCTAAAAC 15650
CGGGCGAAAA TTGTCACTTC CTGTTGACAC CGGGCGACAC CAAAAAGGTC 15700
ACTTTTGCCA CATCCGTTCG TTACATGTGTT TCGGCCACAC TTGCAAACAC 15750
ACACTTCGCG CACACTACTA GTCACCCGC CCGTTCCCA CGCCCGGGGC 15800
CACGTCACAA ACTCCACCCCC TCTATTATCA TATTGGCTTC AATCCAAAT 15850
AAGGTATATT ATTTGATGATG CTACGGGGGC CCAATTATGA GATCCAATTG 15900
CAATGATCAT CATGACAGAT CTGGCGCGGA TCGATATCG CGCTTTAAAT 15950
TTGCGCATGC TAGCTATAGT TCTAGAGGTA CGGTTGGTTA ACGTTAGCCG 16000
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTACGTATA CTCCGGAATTA TTAATAGGCC TAGGATGCAT ATGGCGGCCG</td>
<td>16050</td>
</tr>
<tr>
<td>GCCGCCTGCA GCTGGCGCCA TCGATACCGG TACGTGGCGA CGCCGGACAT</td>
<td>16100</td>
</tr>
<tr>
<td>GTACAGAGCT CGAGAAGTAC TAGTGCCAC GTGGGCGCTG CACCTAAAGC</td>
<td>16150</td>
</tr>
<tr>
<td>TTGGCCTGTT CGGGCTGTTTT ACAACGTCGT GATGGGAAA ACCCTGGCCTG</td>
<td>16200</td>
</tr>
<tr>
<td>TACCCLACTTT AATCGCTTGG CAGCACATCC CCCCCGCCC AGCTGGCGTA</td>
<td>16250</td>
</tr>
<tr>
<td>ATAGCGAAGA GGCCCGCACC GATCG GCCCTT CCCCAACAGT GCCGAGCCTG</td>
<td>16300</td>
</tr>
<tr>
<td>AATGGCGAAT GGCGCCCTGA CGCGTATTTTT CTCTTACCG AATCTGTGGG</td>
<td>16350</td>
</tr>
<tr>
<td>TATTTCACAC CGCATAGCTC AAAGCAACCA TAGTCGCGG CCTGTAGCGG</td>
<td>16400</td>
</tr>
<tr>
<td>CGCATTAAGC GGCGGCGTGG TGTTGCTTAC CGCGAGCGTG ACCGCTACAC</td>
<td>16450</td>
</tr>
<tr>
<td>TGGCCAGGCG CCTAGCGCC GCTCCCTTCC CTTCTCTCCT TCTCTTCCTC</td>
<td>16500</td>
</tr>
<tr>
<td>GCCACGTTCC CGGCTTCCCG CGCTACAGCT CTAAATCGGG GGTCCCCTTT</td>
<td>16550</td>
</tr>
<tr>
<td>AGGTTCGGGA TTTAGCTTCT TAGGGCGACCT GACACCAAA AAATTTGATT</td>
<td>16600</td>
</tr>
<tr>
<td>TGGGGTATGG TTACAGCTAGT GGGCCATCGC CCGTATAGAC GCTTTTCCGC</td>
<td>16650</td>
</tr>
<tr>
<td>CTTTTGACGT TGAGCTCCAC GTCTTCTTAT AGTGGACCTCT TGTTCCAAAC</td>
<td>16700</td>
</tr>
<tr>
<td>TGGAACCAAC CTCAACCCCTA TCTCGGCTTA TTCTTTTGAT TTATAAGGGA</td>
<td>16750</td>
</tr>
<tr>
<td>TTTTGCGCAT TTTGGCGCTAT TGTTAAAAA ATGAGCTGAT TAAACAAAAA</td>
<td>16800</td>
</tr>
<tr>
<td>TTTAACGCGA ATTTAACA AAATAAAGC TTTACAATTT TATGGTGACAC</td>
<td>16850</td>
</tr>
<tr>
<td>TCTCAGTACA ATCTGCTCTG ATGCCGCATA GTAAGCCAG CCGGACACC</td>
<td>16900</td>
</tr>
<tr>
<td>CGCCACCAACC CGCTGACGGG CCGTGACGGG CTTGTCTCTG CCCGGCATCC</td>
<td>16950</td>
</tr>
<tr>
<td>GCTTACGACG AAGCTGTCG CGTCTCCGAG AGCTGACATG GTCAAGGTTT</td>
<td>17000</td>
</tr>
<tr>
<td>TTCAACCGTC TAACCCGACAC GCGCGAGACG AAGGGCCTCT GCGTACGCGC</td>
<td>17050</td>
</tr>
<tr>
<td>TATTTTTATA GTTAAATGTC ATGATAATAA TGGTTCTTTA GACGTCAGGT</td>
<td>17100</td>
</tr>
<tr>
<td>GCCACTTTTC GGGGAAATGT GCGGGCAACC CCTATTTGTT TATTTTCTA</td>
<td>17150</td>
</tr>
<tr>
<td>AATACATTCA AATATGATGC CGTCATGAG ACAATAACC TGATAATGTC</td>
<td>17200</td>
</tr>
<tr>
<td>TTCAATAATA TTGAAAAAGG AGAGTATGAG GTATTAACCA TTCTCCGTGC</td>
<td>17250</td>
</tr>
<tr>
<td>GCCCTTATTC CTTTTTCTGC GGCGATTTGC CTTCTGTGTT TGCTCACCC</td>
<td>17300</td>
</tr>
</tbody>
</table>
AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG 17350
TGGGTTACAT CGAACTGGAT CTCAACAGCG GTAAGATCTT TGAGAGTTTT 17400
CGCCCCGAAAG AAGCTTTTCC AATGATGAGC ACTTTTAAAG TTCTGCTATG 17450
TGGCGCGGTAA TTATCCCGTA TTGACCGCAG GCCAGAGCAA CTGGGTCCGAC 17500
GCATACACTA TTCTCAGAAT GACTTGGTGT AGTACTCACC AGTCACAGAA 17550
AAGCATCTTA CGGATGGGCAT GACAGTAAGA GAATTATGCA GTCGTGCCAT 17600
AACCATGAGT GATAACACTG CGGCCAATTT ACTTTCTGACAGACGATCGGAG 17650
GACCGAAGGA GCTAACCCTT TTTTGGACAGA CATGAGGGGATCAGTAACATG 17700
CGCCTTGATC GTTGGGAAACC GGGCGCTAAG GAAACCCACGCAAACGACGA 17750
GCGTGAAACC ACGATGCTGG TAGCAATTGG AAACACGGTTGGCAAACTAT 17800
TAATGGGCA ACTACTACTT CTAGCTTTCC GGCAACAATT AATAGACTGGG 17850
ATGGAGCCGG ATAAATTTGC AGGCCAATCT CTGCCCTCGG CCCCCTCCCGC 17900
TGATTCGATT ATTCTGTGATA AACTGCGGAC CGGTGAGCGTG GGTCCTCCCGG 17950
GTATCATTGC AGCACTGGGG CCAGATGGTA GCCTTTCCCACGGTAATAGTTT 18000
ATCTACAGCA CGGGAGTGCA GCACAATTGA GATGAAGGGAATACGACAGAT 18050
CGCTGAGATA GTGCTCTCAG TGATTAAGCCT TGTCCTACTGGCAGACCAAG 18100
TTTACTCATAT TATCTTTTAG ATGATTGTTA AACTCAATT TTAATTTTTAAA 18150
AGGATCTAGG TGAAGATCCTT TTTTGTGGAAT CTCTGAGCCA AACTCCTTTA 18200
ACCTGGAGTT TGCTGCTTAC GAGCCGCA CCCCCTAGGA AAAAGTCAGAG 18250
GATCTTCTTGA GATCCTTTTT TTTCTGCGGG TAATCTGCTGG TGGCAAAACA 18300
AAAAAACCAC CGCTACCCAG GTGTTTGTGT TTGCGCCGATAC AAGAGCTTACC 18350
AATCTTTTTT CGGAAGGTTAA CGGGCCCTCAG CAGAGCCGCAATACAAATA 18400
CTGTTCTTCTG AGTGTTAAGG TCGTACTGCG ACCACTCTCAA GAACACTCTGTA 18450
GCACCGCTTA CATACCTGGC TCCTGTAAATC CTGGTACCSG TTGCTGCTGCG 18500
CAGTGCGGAT AGATCGTGCTA TTACCGGTGTT GGAACATCAGA CGATAGTTAC 18550
CGGATAAGGC GCAGCCGGTCG GGCTGAACCAG GGGTTCGCTG CACACAGCCC 18600
AGCTGGGAGC GAACGACCTA. CACCGAAGCT AGATACCTAC AGCGTGAGCT 18650
ATGAGAAAGC GCCACGCTTC CCGAAGGGAG AAAGCCGAC AGGTATCCGG 18700
TAAGCGGCAG GGTCCGAACA GGAGAGCGCA CGAGGAGCT TCCAGGGGGA 18750
AACGCTCTGG ATCTTTTATAG TCCTGTCCGG TTACGCCACC TEGACTTGA 18800
GCGTGCATTT TTGTGATGCT CGTCAGGGGG GCGGAGCGTA TGGAAAAACG 18850
CCAGCAACGC GGCTTTTTTA CGGTTCCTGG CTTTTTGCTG GCCTTTTGCT 18900
CACATGTTCT TTCTTCGGTT ATCCCGCTAT TCTGTGGATA ACCGTATTAC 18950
CGCCTTTTGA TGAGCTGATA CCGCTCGCGG CAGCGGAAGC ACCGAGCGA 19000
GCGAGTCAGT GAGGAGGAAG GCGGAAGAGC GCCAATACG AAACCGCCT 19050
CTCCCCGCAG GTTGGCCGAT TCATTAATGC AGCTGGCACG ACAGGTCTCC 19100
CGACTGGAAA GCCGGCAGTG AGCGCAACGC AAATATGCTG AGTTAGCTCA 19150
CTCATAGGCC ACCCCAGGCT TTACACTTTA TGCTTCCCGC TCGTATGTTG 19200
TGTGGAATGT TGAGCGGATA ACAATTCAC ACAGGAACA GCTATGACCA 19250
TGATTACGAA TTGGAAGGG CATGGGAGGT CGACCTGAGG TAATTATAAC 19300
CCGGGGCC 19307
CLAIMS:

1. A method for producing an AdΔ virus which comprises the steps of:
   (a) introducing into a selected host cell:
       (i) a recombinant shuttle vector which comprises adenovirus nucleic acid sequences and a minigene, wherein said adenovirus sequences are of a first serotype and consist of the adenovirus 5' and 3' cis-elements necessary for replication and virion encapsidation; and wherein said minigene comprises a selected gene operatively linked to regulatory sequences which direct expression of said selected gene in a target cell, said adenovirus cis-elements flanking said minigene; and
       (ii) a helper virus comprising adenovirus gene sequences necessary for adenoviral infection and encoding an adenoviral capsid of a second serotype which differs from said first serotype;
   (b) culturing said host cell containing the shuttle vector and the helper virus so as to permit the formation of an AdΔ virus comprising the adenovirus sequences and minigene of the shuttle vector (i) in an adenoviral capsid of a second serotype which differs from said first serotype.

2. The method according to claim 1, wherein the helper virus lacks all or a sufficient portion of the adenovirus E1 and/or the E3 genes so as to eliminate their biological function.

3. The method according to claim 1 or 2, wherein the first and second serotypes are individually selected from an adenovirus serotype of the group consisting of 2, 4, 5, 7, 12 and 40.

4. The method according to any one of claims 1 to 3, further comprising the steps of isolating and purifying the AdΔ virus from said host cell or cell culture.
5. A method for producing an AdΔ virus which comprises the steps of:
   (a) introducing into a selected host cell
       (i) a recombinant shuttle vector which comprises adenovirus nucleic acid sequences and a minigene, wherein said adenovirus sequences are of a first serotype and consist of the adenovirus 5' and 3' cis-elements necessary for replication and virion encapsidation; and wherein said minigene comprises a selected gene operatively linked to regulatory sequences which direct expression of said selected gene in a target cell, said adenovirus cis-elements flanking said minigene; and
       (ii) a helper virus comprising adenovirus gene sequences of a second serotype which differs from the serotype of the adenovirus sequences of (i);
   (b) culturing said host cell containing the shuttle vector and the helper virus, wherein the helper virus and the host cell provide the adenovirus genes which are necessary for adenoviral infection and which permit the formation of an AdΔ virus comprising the adenovirus sequences and minigene of the shuttle vector (i) in an adenoviral capsid of a second serotype which differs from said first serotype.

6. The method according to claim 5, wherein the first and second serotypes are individually selected from an adenovirus serotype of the group consisting of 2, 4, 5, 7, 12 and 40.

7. The method according to claim 5 or 6, further comprising the steps of isolating and purifying the AdΔ virus from said host cell or cell culture.

8. An AdΔ virus comprising:
   (a) adenovirus nucleic acid sequences and a minigene, wherein said adenovirus nucleic acid sequences comprise the adenovirus 5' and 3' cis-elements necessary for replication and virion encapsidation; wherein said minigene comprises a selected gene operatively linked to regulatory sequences which direct expression of said
selected gene in a target cell, and wherein said adenovirus cis-elements are of a first serotype and flank said minigene; and

(b) a capsid of a second adenovirus serotype which differs from said first serotype and which encapsidates said adenovirus nucleic acid sequences and minigene of (a).

9. An AdΔ virus according to claim 8, wherein the first and second serotypes are independently selected from an adenovirus serotype of the group consisting of 2, 4, 5, 7, 12 and 40.

10. An AdΔ virus according to claim 8 or 9, wherein said adenovirus 5' cis-elements comprise native adenovirus 5' inverted terminal repeats (ITRs) and packaging sequences.

11. An AdΔ virus according to any one of claims 8 to 10, wherein said 3' cis-elements comprise the native adenovirus 3' inverted terminal repeats (ITRs).

12. An AdΔ virus according to any one of claims 8 to 11, wherein said selected gene is a reporter gene selected from the group consisting of the genes encoding β-galactosidase, alkaline phosphatase and green fluorescent protein.

13. An AdΔ virus according to any one of claims 8 to 12, wherein said selected gene is a therapeutic gene selected from the group consisting of a normal CFTR gene, a DMD Becker allele and a normal LDL receptor gene.

14. A pharmaceutical composition comprising an AdΔ virus according to any one of claims 8 to 13 and a pharmaceutically acceptable carrier.
15. The use of an AdΔ virus according to any one of claims 8 to 13 for the manufacture of a pharmaceutical composition suitable for delivering and integrating a selected gene into the chromosome of a target cell.
<table>
<thead>
<tr>
<th>Sequence</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TGGTGTGGGC</td>
<td>CATAATTCAA</td>
<td>TTCCGCGGTCC</td>
<td>CCGCAGCCGCA</td>
<td>GACGTTTTTC</td>
<td>1400</td>
</tr>
<tr>
<td>GCTCGGGAAAG</td>
<td>ACGTACCGGGG</td>
<td>TATACATGTC</td>
<td>TGACAATGGGC</td>
<td>AGATCCCGACG</td>
<td>1450</td>
</tr>
<tr>
<td>GGTCAAAACA</td>
<td>GGGCGCGTGA</td>
<td>AGGCGGTCGG</td>
<td>GATAGTTTTC</td>
<td>TTTCCGCGCCT</td>
<td>1500</td>
</tr>
<tr>
<td>AATCCGACGC</td>
<td>AGTTTACCCG</td>
<td>CTCTGCTACC</td>
<td>TGCCGCGACTG</td>
<td>GCCAGTTTCA</td>
<td>1550</td>
</tr>
<tr>
<td>GCCAATCCGCG</td>
<td>GCCGGATGCGG</td>
<td>GTGCTATCGCT</td>
<td>CCCTACCTTC</td>
<td>ACATCAACGG</td>
<td>1600</td>
</tr>
<tr>
<td>TAATGCACCTA</td>
<td>TTTGACCAACTA</td>
<td>CCAATGCACTC</td>
<td>GGTAGTTTTC</td>
<td>CCGGCTGATA</td>
<td>1650</td>
</tr>
<tr>
<td>AATAAGGTTT</td>
<td>TCCCTGATG</td>
<td>CTGCCAGGCG</td>
<td>TAGCCGCTCG</td>
<td>TAATCACGAC</td>
<td>1700</td>
</tr>
<tr>
<td>CGCATCAGCA</td>
<td>AGTGTATCTG</td>
<td>CCGTGACTCG</td>
<td>CAACAGCAGC</td>
<td>GTCTCCGCGCT</td>
<td>1750</td>
</tr>
<tr>
<td>GTGAATGGCC</td>
<td>CCAGGCTCTTC</td>
<td>CAGCGTTGCA</td>
<td>CCAGGGCGT</td>
<td>AGGTCAATG</td>
<td>1800</td>
</tr>
<tr>
<td>CGGCTCGCTT</td>
<td>CACTTACGCC</td>
<td>AATGTCTGTA</td>
<td>TCCGCCGCTG</td>
<td>CACCGGTGAA</td>
<td>1850</td>
</tr>
<tr>
<td>CTGATCGGCC</td>
<td>AGCCCGGCTCA</td>
<td>GCAGTTTGGT</td>
<td>TTTATCCGCA</td>
<td>ATCCACATCT</td>
<td>1900</td>
</tr>
<tr>
<td>GTGAAGAAAG</td>
<td>GCCTGACTGG</td>
<td>CGTTGAAATT</td>
<td>GCAAAGCTTC</td>
<td>ATACCCAGC</td>
<td>1950</td>
</tr>
<tr>
<td>TCGATCGAA</td>
<td>AATCCATTTTC</td>
<td>GCTGGTGTTG</td>
<td>AGATGGCGGA</td>
<td>TGGCGGGGAG</td>
<td>2000</td>
</tr>
<tr>
<td>CCGGCGGGGG</td>
<td>AGGCTCAGAC</td>
<td>TGAGGCTTTC</td>
<td>CCAGACGCGC</td>
<td>CACTGCTGCC</td>
<td>2050</td>
</tr>
<tr>
<td>AGGCGCTGAT</td>
<td>GTGCCGGCTG</td>
<td>TCTGCCGACT</td>
<td>CGGTGCGGT</td>
<td>CGGTAACCT</td>
<td>2100</td>
</tr>
<tr>
<td>AGCCGTACTG</td>
<td>TGAGCCAGAG</td>
<td>TTCCGGCGGG</td>
<td>CTCTCCGCTG</td>
<td>GCGGTAGTTC</td>
<td>2150</td>
</tr>
<tr>
<td>AGGCAGTTCA</td>
<td>ATCAACTGTG</td>
<td>TACCTTTG</td>
<td>AGGACATCC</td>
<td>AGAGCGACT</td>
<td>2200</td>
</tr>
<tr>
<td>CACCACCGTC</td>
<td>CAGAGCCTTA</td>
<td>CCATCCAGCG</td>
<td>CCACCATCCA</td>
<td>GTCCAGGAGC</td>
<td>2250</td>
</tr>
<tr>
<td>TCGTTATCGC</td>
<td>TATGACGGAA</td>
<td>CAGGTATCCTG</td>
<td>CTGGTACTTT</td>
<td>CGATGGTTTG</td>
<td>2300</td>
</tr>
<tr>
<td>CCCGATAAAG</td>
<td>CGGAACTTGG</td>
<td>AAAACTGCTG</td>
<td>CTGGTGTTTTC</td>
<td>GCTTCGCTCA</td>
<td>2350</td>
</tr>
<tr>
<td>GGGCTGGAGT</td>
<td>CCGCGGCTCGG</td>
<td>TCGGCAAGAG</td>
<td>CCAGACGCT</td>
<td>CATACGAGC</td>
<td>2400</td>
</tr>
<tr>
<td>TGGCGATCTG</td>
<td>TGCGCTATC</td>
<td>GCAAAATCA</td>
<td>CCGCGTAAAG</td>
<td>CGGACCACGG</td>
<td>2450</td>
</tr>
<tr>
<td>GTTGGCGCTTT</td>
<td>TCATCATATT</td>
<td>TATCACGCGA</td>
<td>CTGATCCACC</td>
<td>CAGTCCAGA</td>
<td>2500</td>
</tr>
<tr>
<td>CGAAGCCGCC</td>
<td>CTGTAAAGCG</td>
<td>GGAATCTGAC</td>
<td>GAAAGCCTG</td>
<td>CGATGATTTA</td>
<td>2550</td>
</tr>
<tr>
<td>CGGAAACCGC</td>
<td>CAAGACTGTG</td>
<td>ACCCATCGCG</td>
<td>TGGGCTATTT</td>
<td>CGGAAGGAT</td>
<td>2600</td>
</tr>
<tr>
<td>CAGCGGGGCGC</td>
<td>GTCTCCGCTG</td>
<td>TAGGCGAAGG</td>
<td>CCATTGTTTG</td>
<td>ATGGCAAATT</td>
<td>2650</td>
</tr>
</tbody>
</table>
TCGGCACAGC CGGGAAGGGC TGGTCTTCAT CCACGCAGGC GTACATCCGG
2700
CAAATAATAT CGGTGGCGGT GTGGTGCGGT CCGCGGGGCTT CATACTGCAC
2750
CGGGCGGGAAG CAGATGGACAG ATTTGATCCA GCGATACGC GCGTCGTGAT
2800
TAGCGCGCTG GCCTGATTTCA TTCCCCAGGC ACCAGATGAT CACACTCGGG
2850
TGATATGCAG CAGGCTTCGAC CATTCGGTCTT ACGCGTTCGCC TCGACGCAGG
2900
TAGCAGCGGG GAGTATCCG TCAGACGGATT CATTGGCACG ATGCGGTGGG
2950
TTTCAATATT GGGCTTCATCC ACCACATACA GCGGTAGCC GTCGACAGC
3000
GTGTACCACA CGGATGCGTT CGGATAATGC GAACAGCGCA CGGGCCTAAA
3050
GTGTTTCTGC TTCACTGCAA GGTATCTTGG CACCATCGTC TGCTCATCCA
3100
TGACCTGACC AGTCGAGGA TGATGCTGCT GACGGTTAAC GCCTCGAATC
3150
AGCAACGGCT TGCCGTCGCA CAGCAGCAGA CCATTCTCAA TCCGCACTTC
3200
GGCAAAACCG ACATCGCAGG CTTCTGCCTTC AATCAGCGTG CCGTCGGCGG
3250
TGTCAGGTTT ACCACCAGCA CGATAGAGAT TGCGGATTTT GGGCCTCCAC
3300
AGTGTTCGHT TCTCAGCTGTT CAGACGTAAT GTGAGCGGAT GGGCATAACC
3350
ACCACGCTCA TGATATATTT CACCGCCGAA AGGCCCGCTG CCGGTGGCGA
3400
CTTCCGTTCG ACCCTGCGAT GAAGAAACTG TTACCCGTAG GTAGTCACGC
3450
AACTCGCCGC ACATCTGAAC TCTACGCTTCA AGTACAGCGC GGCTGAAATC
3500
ATCATTAAG CGAGTGGCAA CATGAAATC GCTGTATTTTT GTGACGCGTT
3550
TATGCACGAA CGAACGTGCA CCGAAATGCA CGCTCAGCTG CCACATATCC
3600
TGACCTTCCA GTAACTGCC GCTACTCAGA GCAGACCGA TCACCGCGAG
3650
GGGTTTTCT CCGCCCAGTA AAAATGGCT CAGGTCAATG TCAGCCGGCA
3700
AACGACTGTC TCGCCGTAAA CGCACCGAG GCCCGTGGCA CCACAGATGA
3750
AACCGCGAGT TAAGCCTATC AAAATAATT CCGGTCTGCG CTTCTGTAG
3800
CCACGGTTCA TCAACATTAA ATGGTAGCGA GTAACAAACC GTGGGATTCT
3850
CGTGGGAAAC AAACGGCGGA TTGACGTAAC TGATAGAGT TACGTGGGTG
3900
TAGATGGCGG CATCGTAACC GTGCATCTGC CAGTTGAGG AGACGACGAC
3950
FIGURE 3D

AGTATCGGCC TCAGGAAGAT CGCACTCCAG CCAGCTTTCC GGCACCGCTT 4000
CTGGTGCCGG AAACCGGCA AAGGCCATT CGCCATTCA GCTGCCCAAC 4050
TGTTGGGAAG GCAGATCGTT GGCGGCCCTCT TCCTATTACC GCCAGCTGGC 4100
CAAAAGGGGA TGTTGCAGAA GGGATTTAAC TTGGGTAAAC CCAGGGTTTT 4150
CCAGCTCAGC ACGTGTAAAA ACGACGGGAT GCAGCTTGAG CAGCTCTTTA 4200
CTGGTGCTCA AGCAATGCC TCCCAGACCC GCAAGAAATA TCACGTCTTT 4250
GTTGGTCAAA GTAAACGACA TGTAGACTTC TTTTTTGCTT TAGCAGGCTC 4300
TTTCGATCC CGGGAATTGC GCAGCCGGGT ACAATTCCGC AGCTTTTAGA 4350
GCAGAAGTAA CACTCCGTA CAGCCCTAGA AGTAAGGCA AATCCACCTG 4400
AGGAGCAGTT CTTTGTATTTG CACCCACACC GGATCCGGGA CTCGAAATAA 4450
AAGAACAAAA GACTAAACTT ACCAGTTAAC TTTCCTGGTT TTCACTCTCT 4500
CGAGTACCGG ATCCCTCTAGA GTCCGGAGGC TGATCCGGTC CCGGTCTCTT 4550
CTATGGAGGT CAAAACACGG TGGATGGGCT CTCCAGGGGA TCTGACGTTT 4600
CACTAAACGA GCTCTGCTTA TATAGACCTC CCACCGTACA GCCTTACGC 4650
CCATTGGGCT CAATGGGCG GAGTTGTTAC GACATTTTGG AAAGTCCCGT 4700
TGATTGTGGT GCACAAAAAC ACTCCCATAG ACGTCAATGG GCTGGAGACT 4750
TGGAATCCCC CGTAGTCAA ACGCTATCCC ACGCCATTG ATGTACTGCC 4800
AAAACCCGAT CACCATGTA ATACGCGATA CTAATACGTG GATGTACTGC 4850
CAAGTAGAAA AGTCCCATAA GGTATGTAC TGGGCATAAT GCCAGGGGGA 4900
CCATTACCG TCATTAGCGT CAATAGGGG CCTACGCTTC ATATGATACA 4950
CCTGATGTAC TGCCCAAGTG GCAGTTAAC GTAATATCTC CACCCATGGA 5000
CGTCAATGGA AAGTCCCTAT TGCCGGTACT ATGGGAACAT ACGTCAATTAT 5050
TGACGTCATA GCCGCGCGGT CGTGGGGCCG TCAGCCAGGC GGGCCATTAT 5100
CGCGAAGTTA TGTTACCGAC GCCAGGCTCA CTTTAGGGA TCTCCCTAGA 5150
CAAAATTCG GCGCTATGAG TAAACAAAAA TTATCCAGAT TTACCTTTCT 5200
CTTATTTCGTT TTTCCCGGGA AAATGGGCAA ATCTTACTCG GTTAGGCCCA 5250

SUBSTITUTE SHEET (RULE 26)
FIGURE 3F

AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCAACTATC 6600
TCAGCGATCT GTCTATTTTCG TTTCATCCATA GTTGCCCTGAC TCCCGTCGCT 6650
GTGATAAAGT AGGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA 6700
TGATACCAGG AGACCCAGGC TCAAGGGCTC CP ATTTATTGC AGCAATAAAC 6750
CAGCCAGCGG GAAAGGCCGA GCGCAGAGTT GGTCTTGCAA CTTTATCCGC 6800
CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA AGTAGTCGCT 6850
CAGTTAATAG TTTGCGCAAC GTTGTGTGCA TTGCTACAGG CATCGTGGTG 6900
TCAGGCTCGT GTTTGTTGAT GGCCTTATTC AGCTCCGTTT CCAACGATC 6950
AAGGCGAGGT ACATGATCCC CCATGTTTGT CAAAAGACCG GTTAGCTCCT 7000
TCGGTCTCCC GATCGTGTGC AGAAGTAAGT TGGCGCAGCT GTATATCTC 7050
ATGTTTATGG CAGCAGTCCA TAATCTCTTT ACTGTCTAGC CATCCGTAAG 7100
ATGCCTTTCT GTGACTGGTG AGTACTCAAC CAAGTCTTTG TGAGATAGT 7150
GTATCGGCCG ACCAGTTGCG TCTTGCCGGG GCTCAATACG GAATAATACC 7200
GGGCCAGATA CGAGAACTTTT AAAAGTGCTC ATCATGGGAA AACGTCTTC 7250
GGGGCGAAAA CTCTCAAGGA TCTTACCGCT GTGAGATCC AGTTGCAAGT 7300
AACCCACTCG TGCAACCAAAC TGATCTTCAG CATCTTTTAC TTTCAACCAG 7350
GTCTTGTGGT GAGCAAAAAAC AGGAAGCCAA AATGCCGCAA AAAAGGGAAAT 7400
AAGGGCAGCA CGGAATAAGTT GAATACTCAT ACTCTTCTTT TTTCATATTT 7450
ATTGAAGCAT TTATCAGGGT TATGTGCTCA TGAGCGGATA CATATTGGA 7500
TGATATTGAA AAAATAAACA AAATAGGGTT CGGCGACAT TTCCCAGGAA 7550
AGTGCCACTG GAGCTTAAG AAACCATTAT TATCATGACA TTAACCTATA 7600
AAAATAGGGC TATCAGGAGG CCCCCGTCG TCGCGCGTTT CGGTGATGAC 7650
GGTGAAAGC TCTGACACAT GCAGCTCCCC GAGACGGTCA CAGCTTGTCT 7700
GTAAGCGGAT GCGGGGACGA GACAAGCCCG TCAAGGCGCG TCGCGGCGTG 7750
TTGGGCGGTG TCGGGGTCTT CTTAATTAG CGGCATCAGA CGAGATTGTA 7800
CTGAGAGTGG ACCATAGGGA CATATTGCG TTGAAAGCGG GCTACAATTA 7850
ATACATAACC TTATGATCA TACACATACG ATTTAGGTGA CACTATA 7897
FIG. 4A

pAdΔc.CMVLacZ

HEAD-TO-TAIL JUNCTION

SD/SA

CMV

5'ITR

3'ITR

EcoRI(1)

pAT153

XbaI (5449)

pA

XbaI (1567)

ClaI(2722)

NotI(1761)

NotI(5235)

LacZ

FIG. 4B

PACKAGE INTO EMPTY CAPSID HEAD PROVIDED BY HELPER VIRUS

Ad.CBhpAP

HELPER VIRUS

hpAP

hpAP

hpAP

hpAP

Ad.CBhpAP

LacZ

LacZ

LacZ

LacZ

AdΔ.CMVLacZ

SUBSTITUTE SHEET (RULE 28)
FIGURE 5A

GAATTCGCTA GCTAGCGGGG GAATACATAC CCGCAGGCCT AGAGACAACA  50
TTACAGCCCC CATAGGAGGT ATAACAAATATA TAGAGAGA GAAAAACACA  100
TAAACACCTG AAAAACCTTC CTGCTTAGGC AAAATAGCAC CCTCCCGCTC  150
CAGAAACAACA TACAGGGCTT CACAGCGGCA GCCTAACAGT CAGGCTTACC 200
AGTAAAAAG AAAACCTATT AAAAAGACAC CACTCGACAC GGCACCAGCT  250
CAATCAGTCA CAGTGTTAAA AAGGGCAAG TGCAAGAGCA GTATATATAG  300
GACTAAAAAA TGACGTAAAC GTTAAGGTCC AAAAAAACA CCCAGAAAAC  350
GCACGCCGAA CCTAGGCCC AAGACGAAAG CAAAACACC CACACCTTCC  400
TCAAATCGTC ACTTCCGTTT TCCACGTTA CGTAACTTCC CATTCTAAGA  450
AAACTACAAT TCCCAACACA TACAAGTTAC TCCGCCCCTAA AACTACGTG  500
ACCCGCCCGG TTCCACCGCC CCGGCCACG TCACAAACTC CACCCCCCTCA  550
TTATCATATT GGCTTCAATC CAAATAAGG TATATTATTC ATGATGCTAG  600
CATCATCAAT AATATACCTT ATTTTGAGTT GAAGCCAATA TGATAATGAG  650
GGGGTGAGGT TTGTCAGCGT GGGGCGGCGG TGGAACGAG GGCGGTAGCG  700
TAGTACTTGT GGCGGACTGT GATTTGCAA GTGTGCGGGA ACAGACTAA  750
GCGACGGATG TGGCAAAAGT GACGTTTTTG GTGTGGCGCG GTGTACACAG  800
GAACTGACAA TTTTCCGCGG GTTTTAGGCC GATGTTGTAG TAAATTTGGG  850
CGTAAACGAG TAAAGTTTGG CCATTTTCC CAGGAAAACG AATTAGAGGA  900
AGTGAATATCT GAATAATTTT GTGTACTCA TAGCGGTTAA TATTTGTCTA  950
GGGAGATCAG CCTGCAAGTC GTTACATAAC TTACGGTTAA TGGCGGCGCT 1000
GGGTGACGGG CCAACGGCCC CCGCCCATTG ACCTCAATAA TGACGTTATG 1050
TCCCCATAGTA AGCGCAATAG GGACTTTCGA TTGACGGTAA TGGGTTGAGT 1100
ATTACGTTA AACTGCCAC TTGGCAGTAC ATCAAGTGTAA TCAATATCCA 1150
AGTACGCCG CTAATGACGT AAATGCGCGG CCTGGCATTA 1200
TGGGGAGTAC ATGAGCTTAT GGGACTTTC TACTGGCAG TACATCTACG 1250
TATTAGTCAT CGCTATTACC ATGCTAGTGCG GGTCTTGGCA GTACATCAAAT 1300
14/45

FIGURE 5E

AGACATGATA AGATACATTG ATGAGTTTGG ACAACCACA ACTGAATGC 5300
AGTGAAAAA ATGCTTTATT TGTGAAATTT GTGATGCTAT TGCTTTATT 5350
GTAACCATT A TAAGCTGCAA TAAACAGT T AAACAACA A ATGCA TTA 5400
TTTATGTGTT CAGGTCAGG GGGAGGTGTG GGAGGT TTTT TCGGATCTTC 5450
TAGAGTCGAC GACGGCGGTC TGGATGGCCT TCCCCATTAT GATTTCTCTC 5500
GCCTCCGC G CCATCGGAT GCCCGCGGTT G AGGCGATGC TGTCAGGGCA 5550
GGTAGATGAC GACCACCGG GACAGCTTCA AGGATCGCTC CCGCCTCTTA 5600
CCAGCCTAAC TTGCATCAGT GGACCGCTGA TCGTCAGCGC GATTTATGCC 5650
GCCTGGCGGA GAACATGGAA CGGGTGGGCA TGGATTGTAG GGGCGCCCT 5700
ATACCTTGTC TGCTCCCCCG GTTGGCTGCG CGGTGCATGG AGCCGGGCCA 5750
CCTCGACCTG AATGGAAGCC GGCGGACACT CGTACAGGA TCCACCACTC 5800
CAAGAATGGG AGCCAATCAA TTCTTGCGGA GAACTGTGAA TGGCGCAACC 5850
AACCCCTTGGG AGAGCATATC CATCGCTTCC GCACCTGCCA GCAGCGCCAC 5900
GGGCGCATC TCACGGGACC TTGGGTCTGT GCGACGGGT GCATGATCG 5950
TGCTCCTGTG GTGGAGGACG CCGCTAGGGT GGGGCGGTG CCTACTGGGT 6000
TAGCAGAAATG AATCACCGGAT AGCGGAGCGA AGCTGAAGCG ACTGCTGCTG 6050
CAAAACGTCT CCGCAGCTGAG CAACCAACTG AATGCGTTC TGGTTCCGTC 6100
TTGCTGAAAG TCTGGAACCG CGGAAATCGA CGCCCTGCA CATTATGTTC 6150
CGGATCTGCA TCCGAGGATG CGTCTGGCTA CCCGTGGCAA CACCTACATC 6200
TGTATTAAGG AAGCCCTTCT CAATGCTCA GTGCGTGGTA TCTCAGTTCG 6250
GTGTAGGTCG TTGCTCACA GCTGCGGTGT GTCAGCGAAC CCCGCGTTCA 6300
GCCCGACCGC TGGCGCCTAT CCGTAACTA TGGCTTGGAG TCCAACCCCAG 6350
TAAGACACGCA CTTATGCAGA CTGGCGACAG CCCTGGTCAA CAGGATTAGC 6400
AGAGCGGCTG ATGTAATGGG TGTCATAGGC TTCTGAACT GGTCGGCCTAA 6450
CTACGGCTAG CTAGAAGaga CAGTATTTGG TATCTGGCTT CTGCTGAAAGC 6500
CAGTTACCTT CGGAAAAAGA GTGTCAGCT CTTGATCCCG CAAACAAACC 6550
ACCGCTGGTA GCGTTGTTTT TTTTTGGTGAG AAAAGCACGA AAGACGCAGA TTACGCCGCA
AAAAGAAAAGA TCTCAAGAG ATCCCTTGAT CTTTTTCATG GGGTCTGACG 6650
CTCAGTGAAA CGAAAACCTCA GTTAAAGGA TTTTGGTCAT GAGATTATCA
AAAAGGATCT TCACCTAGAT CCTTTTTAAAT TA\`AAATGAA GTTTTAAATC 6750
AATCTAAGCT ATATATGAGT AAACCTGGCT TGCAGTTTAC CAATGCTTAA
TCAGTGAGGC ACCTATCTCA GCGATCGTCA TATTTGGTTC ATCCATAGTT
GCCTGACTCC CCGTCTGTA GATAACTCCG ATACGCGAAG GCCGCTCCAT
TGGCCCGAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA CCGGCTCCAG
ATTATACGC AATAAACACG CCAAGCCGGAA GGGCGAGCC CAGAACTGCT
CCTGCAACTT TATCCCGCTC CATCCAGTTT ATTAAATGTT GCCGCGAAGC
TAGTAAGTT AGTTCGGCAG TTAATAGTTT GGCAACGGT GTTGCCATTT
CTGCAGGGCT CGTGAGTGA TGGTCTGCG TTGTATGAGG TTCAATTCAG
TCCGGTTCCC AACGATAAAG GCGGTTTACA TCATCCCAA TGTTGTCGAA
AAAACGGGTT AGCTCCTTCCG GTCTCCGAG ATGGTGCAGA AGTAAGTGGG
CCGCAGTGGT ATCACTCATG GCTATGCGCAG CACTGCATAA TTCTCTTACT
GCTATGCTCC CGTAAGATG CTTTCTTGGT ACTGCTGGATG ACTCAAACCA
GCTATTTCTCA GAATAGGTTA TGCGGGAGCC GAGTTGCTCT TGCCGGGCCG
CAACAGGGAA TAACTGCGGC CCACATAGCA CAACCTTTAA AGTGCTCAGC
ATTGAAACA GTTCTTCCGG GCGAAAACCTC TCAAGAGATCT TACCGCTGCTT
GAGATCCTAGT TCAGATGTTAC CCACTCTGGC ACCCAACTGA TCTTCAGCAG
CTTTTACTTT CACCAGGCTT TCTGGGTGAG CAAAACAGG AAGGCAAAAT
GCCGCAAAAA AGGGAAATAAG GGGAGACCGG AAATGGTGAAG TACTCATACT
CTTCCCTTTT CAATTTATT GAAGCATTTA TCAGGGTTAT TGCTCATGAT
GCGATACTAT ATTTGAAATGT ATTTGAAGAA AATAAACAAT AGGGGTTCGC
CGCAGCTTTC CCGAAAGGCT GCCACCTGAG GTCTAAGAAA CCATTATAT
CATGACATTT ACCATAAAA ATAGGCGTAT CAGGAGCCCC TTTCGTCTTC
AA 7852
FIGURE 7B

TCAGCTCCGC TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG 1350
TGCAAAAAAG CGGTAGCTC CTTGGGTCTT CCGATCGTTG TCAGAAGTAA 1400
GGTGGCCGCA GTGGTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC 1450
TTACTGTGAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA 1500
ACCAAGTCAT TCTGACAATA GTGTAGCGGG CGACCGAGTT GCTCTGCCC 1550
GGCGTCAATA CGGATAATA CGCGGCAACA TAGCAGAACT TTAAAAGTGCC 1600
TCATCTTGG AAAACGTCTT CTGGGGCGAA AACTCTCAAG GATCTTACCAG 1650
CTGTTGAGAT CCAGTTCGAT GTAACCCACT CTGTCACCCA ACTGATCTTC 1700
AGCATCTTTT ACTTTCACCA CGGTTCTGG GTGAGCAAAA ACAGGAGGCG 1750
AAAATGCCG AAAAGGGAATA AATAGGGCGA CACGGAAATG TTGAATACTC 1800
ATACTCTTCC TTTTTCAATA TTATGGAGCT ATTTATCGAG GTTATTGTCT 1850
CATGAGCCGA TACATATTG AATGTATTTA GAAAAATAAA CAAATAGGGG 1900
TTCCGCGCACT ATTTCCCCGA AAAGTGGCACC CTGACGTCTA AGAAACCAIT 1950
ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCAGCA GGCCCTTTCG 2000
TCTCGCAGGT TTGCTGATAG ACGGTGAAAA CCTCTGACAC ATGCAGCTCC 2050
CGGAGACGGT CACAGCTTGT CTGTAAGCGG ATGCCGGGAG CAGACAGGCC 2100
CGTCAGGGCG CGTACAGGGG TGTGAGGGGG GTGCGGGCTT GGCTTAACCTA 2150
TGCGGCATCA GAGCAGATTG TACTGAGAGT GCACCTATAC AAAATGTTAGG 2200
TTAATATTTT GTTTAAATTC GCACCCAAA TTTTATCAGCA AACCTATT 2250
TTTAAACAAAT AGGCGGAAAT CGGAAAAATC CTTTATAAAT AAAAGAATA 2300
GGCCCGAGATA GGTTGAGTG TTGTTCCAGT TTGGAACAGG AGTCCACTAT 2350
TAAAGAACGT GGACTCCAAC GTCAAGGCG GAAACCGCT CTATCGGGC 2400
GATGGCCAC TACGTGAACC ATCAACAAA TCAAGTTTTT TGGGGTGCAG 2450
GTGCCGTAAA GCACTAAATC GGAACCTAA AGGGAGCCCC CGATTAGAG 2500
CTTGACGGGG AAAGCAGGGC AACTGCGGGA GAAGAGACCG GAGAAAGC 2550
AAAGGAGGCG GCGCTAGGGC GTTGGCAGGT GTAGCGGTCA CGCTCCGGTG 2600

SUBSTITUTE SHEET (RULE 26)
Figure 7G

TTCCCAGCTC TCTGATCTCT GTACTTCATC ATCATTTCCC CTAGCCAGC 7850
CTGAAAAAGG GCAAGGACTA TCAGGAAACC AAGTCCACAG AAGGCAGACG 7900
CCTGTAACAA CTCAGGAGATT AGCCCCATGA GGAGTCCAC TGGAAGAGGA 7950
GGGATCCACA CGAAATGTGC CAATGCAGT CTTCCATCAA ATTTGTTCAG 8000
GTTGTTGGAA AGGAGACTAA CAAGTGTGTC AATACTTATT TTATCTAGAA 8050
CACGCGTTGA CAGCTTTAAA GTCTTCTTAT AAATCAAACG AACAATAGCT 8100
ATTCTCATCT GCATTCCAAT GTGATGAAGG CCAAATAGG CTGGTGTGATG 8150
GGGACGTGTGTC TCGACAATAA AGAGAAGGCA TAAGCTATAG CCTAGATAAA 8200
TCGCGATAAG ATCGTTCCCT TTGTTATCCG GGTCATAGGA AGCTATGATT 8250
CTTCCAGTA AGAGAGGCTG TACTGTTTTG GTGAATCCC CTAAATATAA 8300
AAAGATTCCA TAGAACATAA ATCTCCAGAA AAAACATGCG CGAAGGCGAT 8350
TAATGAGTATT AGATTTTTTC TTCTGAGCCA GCTCTCTATC CACTTCTCTT 8400
TCCAATTTTT CAGATAGATT GTACACGAA TCAACAGAAG GATTTTGGTA 8450
TAATGCTGAC AATTCGACGC GCTGTCTGTA TCTTTCCTC AAAATTGGTC 8500
TGGTCCAGGT GAAAAAAGGT TGGAAGACAA CGCTGGCCTT TTCCAGAGGC 8550
GACCTCTGCA TGGTCTCTCG GGCGCTGGGG TCCCTGTAG GGGCTCTGGG 8600
GCTCAAGCTC CTAATGCAAA AGGATTTCTT GCAGCAGGGG GGATCATACTA 8650
GTTCTAGAGC GGCGCCCACCC GGCGGGTGTG ATCCGGCTCC CGCGCGCGGC 8700
GCGCTTCGCT TTTTATAAGG CGCGGCACCG CGCGCCCTCG CCATAAAGG 8750
AAACCTTGCG AGCCGCCGCC TCTGATGCC TGCGCGCCGA CCTCTCAGCC 8800
TCGGCCGCCG CGGGACTCAG CCCGGCTGGG CCCGGCTGGG CGCCGGCGGC 8850
CCGCCCCCGG CGCCGGCGG CGCTGCACAA AATAATTAAA AAATAAAATAA 8900
ATACAAATTT GGGGGTGGGG AGGGGGGGGA GATGGGGAGA GTGAAACAGA 8950
AGTGCGCCTC GAGTAGATGT ACTGCAAGT GGAAAGTCC CATAGGTCA 9000
TGACTGGGGA ATGATGCCG GCAGGGCAAT TACGCACATT GACGTCAATA 9050
GGGGGGGTAC TTGGCATATG ATACACTTGA TGTAATCGCA AGTGGGCGATG 9100

Substitute Sheet (Rule 26)
TTACCGTAAA TACTCCACCC ATGAGCTCA ATGAAAGTC CCTATTGGCG 9150
TTACCTATGGG AACATACGTCC ATTTATGACG TGATTTGGCG GCCGTCGTGG 9200
GCGGTCAGC CAGCGGGGCG ATTTACCTGA AGTTATGTA CGAACCCTGCAG 9250
GCTGATCTCC CTAGCAAAATG ATTAGCGCGCT ATGAGTAAAC CAAAATATT 9300
CAGATTTTCAT TTCTCTTTAT TGATTTTTCC CCGGAAATGT GCGAAATCTT 9350
ACTCGGTTAC GCCCAATTTT ACTACAACAT CGGCCCTAAA CCAGCGCGAAA 9400
ATTGTCACTT CCTGTGTAAT ACGCGCGACAT CAAAACCTGG GTAATTTGCCC 9450
ACATCCGTCG CTTACATGTG TTCCGCCACA CTTGCAACAG CACACTTCCG 9500
CCACACTACT ACCTCACCAG TCGGGTCGTCC ACGCCCCGCC CACCGTCACA 9550
AATCCTCCAC CCCTTATATTAT ATATTGCGTT CAATCCAAAA TAAGGTATAT 9600
TATTGATGAT GCTGCAATGC GCCTATTTAA AGCGCATGTC TCGATCGCGC 9650
GCAGATCTGT CATGATGATG ATGCCAAATTG GATCCATATA TAGGCCCCCGG 9700
GTTATAATTA CCTCAGGCTCG ACCTCCATGT GCCATTCGAA TTCCGATATCA 9750
TGGTCATAGG TTGTTGCTGT GTGAAATTGT TATCCGCTCA CATTTCCACA 9800
CAACATACGA GCCGGAGCA TAAAGTGTAA AGCCGCGGTC GCCTAAATGAG 9850
TGAGCTAAGT CAGGATATT GCGTTGCGCT CACTGCCCGC TTTCCAGTGCG 9900
GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAAC GCCGGGGGAG 9950
AGCGGTTTTG CGTATTGGGC GC 9972
FIG. 8A

FIG. 8B

FIG. 8C
FIG. 10

MALEIMIDE ACTIVATION

THIOLATION

GEL FILTRATION

GEL FILTRATION

20 h @ 4°C

CsCl STEP GRADIENT

Ad-pLys CONJUGATE

PLASMID DNA

AD-BASED COMPLEX
Figure 12B

AGATAAAACG AAAAGATTTA AACAAAACTA TGCACCTCAT CTACCTTGG 1300
TTTTGAAATG AAACCTTAAAT CTTCTTAGTA GGAAGGAAC CTCCTTTTTT 1350
AAATCTTTGGT GAAACACAAT CCTTGGAATA AGAATAATGCA CAGTGCCACA 1400
TAAGGAGAGG AGAGAGAGAA AGAACAAGCC AGAAACCAAT TTTCAATTTGT 1450
TATCTTATAG CGTTGGGTGT TCTTTGGGAA ATTATAATG AAAAAGGAA 1500
ACTGGTGTC CCACCAACAGA CAAGTGTGGA AGTTTGGAAT TTAGGTTGTC 1550
ACCAATTACAA GAACACACCC AAAACCAAGG TGAAGTGAGA ATAGCATGAG 1600
AAGCTGTGTT TGATGTTAAAT TACAATAT TAATGGACAAA ACCCACTGCA 1650
TAGAAGTTAAT TTACACTTGAA CGTTAGAGGT AACAGATTTG CAAAATGATA 1700
GGACAGTGAT TTCTATTGAG AGAATGCTCT TTAATGCTTA AGAAGGAAGA 1750
ACTGGCATGA GAGGAGTTAA GCTCTTCTTA GCAGTCTTTA GCTTTCTGTT 1800
GCACCTTTTC TCCTGGGTTCA ATGACTTGCA TTTTTTGGAA CATTTTCAGCC 1850
CGTCAAATAG ACCAGAGAGT TGGGAGACGC TTTTGGCTCCT AAAACCTTCC 1900
AACCACTGTG CCTTCTCACC CACAAACTCTG TGTTGGATTA CTTGCAGGGA 1950
AACCAATGCA AGGAGACAA ATGCAGTTCA TGCGTTCTTG GACTGTATT 2000
CACCAAGGTCA ACAATGTGAT TGGGTCTATTT TCTTAAACGT AATCCTAAGT 2050
CTTGAGGCTAC TAAAAAAAAT AATCATCACA ATGAAGAAAA AAAAACCCAA 2100
AAAATCTAAA ATCTAAATTT CATCATCATC ATCAACAACA ACAAACAACA 2150
CAACAAACAA ACCAAACACT TCAGGGTGA TTTATGAAAG GGGCAGAAC 2200
ATTAGTTGAT AATTATAGAG ATGTGTATAT GTATAGTTGT AAATATTCA 2250
CCATTTTTTC ACAGAGTTGT TGCTCCCCCT ATATAATATTG ACTGGGAGGC 2300
CGAACACTTT GCAGCCTCTCC ATCTTCCCTC TACGTGCTGG GAGTTAAAAA 2350
TGTCACTCAGA TGTTCTATTG CAGAAACATC ATTAAATATA ACCCAACAGT 2400
AGGAAGTTGA ATATATCAGC CAACAAATTA CTATGATAGT AAGTCTCTGTG 2450
TATTTATCG CATGTTCTTT GAAAAAATG AATCCCTTAG CTCTCAGTGG 2500
FIGURE 12E

CATGGGCTGG TCATTTCTCT TGAGGTTCTG CTGATCCAGG GCATCACATG 5100
CAGCTGACAG GCTCAAGAGA TCCAAGCAAA GGGCTCTTCTG GAGCCTTCTG 5150
AGCTTCATGG CAGTCTCTAT CCGGAGAAC CTGACATTAT TCAGGCTAGC 5200
TAAGACGTGG TAGAGCTCTG TCTATTTTGG GTGTCCTCAA CAAGTGCTTTT 5250
GGGTCTGAGT GTGATATTAG TAGGGCACTT TGTTTGCTGA GATGGCTCTC 5300
TCCCAGGGAC CCGGAATCTGA ATGGAAGAGG AATGCTGTTG ATGGAAGACC 5350
AAAGTCCCTG TGGCTCTCAT GCAGCTGCTT GACAGCGTCC TCCACAGCCA 5400
CCTCTAGAGA CCTCCATCTG GTATTCAGAT CTTCCAAAGT GCTGAGGGTA 5450
TAAGGAGA GCTGAAATCC CAGTGTGGTC AGCTGATGTT CAAGGTCATT 5500
GACAGATTTA ACATTCTTCT TAAAGGTTGC AATTTCTCCC CGAAGTGCCCT 5550
TGACTTTTTCT AAGGGTAGCT TGGCGGAGAT CAATGAGGAG ATCCCCACT 5600
GGCTGCCAGG ATCCCCGT CATCCTCACCT GGCGCAACT TGAGGTCAGG 5650
TTTCTCCGCA GCTTCTCGAA GTTCTCTGAG TTTTCAAGA GCTTCTCCTCTA 5700
TTTTTCTCTT CCAATCAGCT GAGCGCAGGT TCAATTTGTC CCATTCAGCG 5750
TGACCTCTTT CAGCCTGGTT TGCTAGGAGG CGAGTGACAT TCTGAGGCTT 5800
TTTTAGAGGA GCGAGTCTCC TGGCGCTCTT GTAGAGGTTTC TCTAGGCTCTT 5850
CCAAAGGCGT CTCTGTCGAA AATATTCTCA CAGTCTCCCC AGTACTCATG 5900
ATTACAGGTT CTTTTAGTTCT CAATTCCTTC GCTGAGGCCC TATGTTATATC 5950
ATTCTGCTTC TGAACCTGCTG GGAATCCACC ACCGATGGGT GCTGAGCGGC 6000
TCAGTCCCAT ATCTTCACAG TGGAGCCTTTA CAAGAGTTTC CTGAGAGAAA 6050
AGATGCAACA GCTTCCACTG GTCAAGACTT GCTTCCAAAT GGGACCTAAT 6100
GTTGAGAGAC TTTTTCTGCAA GCTTCACTCCA CTTGAAATGCT ATGTTATCCA 6150
AACGTCTTTG TAACAGGCGT GCTTCATCCG AACCTCCAG CGATTCGAGG 6200
ATTTTTTGCA CTTTTTTCCAT CAGATTCTGA TAGATATCTG TGTTGAGTTCT 6250
AAATTCTCTCT TGGAAGCTTT GCCATGTGTTT CATGAGCTCT CTTGACTCCCT 6300
TGAGTCTCCTG TAGAGCTCTC TCTTTAGGG AAGCGTCTCTG TAGGACATTTG 6350
FIGURE 12G

TTGCTCTTCT GGCCTTATGG GAGCCTTAC AAGTACTGCT CCTCCTGTTT 7650
CATTTAATTG TTTTGAATTT CCTGCTGCA GGGGCAAATC TTCTGCTATT 7700
AACTTGACC GTGCAAGTTG TTCTTTAAGC TGTCTGCTAT CTCCAAAGTG 7750
AGTAATAGCA ATGGTATCTG CTTCTTCCAG CCACAAACA CAATTCAATT 7800
AATCTCTTTG AAATTCTGAC AAGACATTCT TTTGTTCTTC AATCCCTCTT 7850
CTCCTTTCTG CCAGCTCTTT GCAGATGTCC TGCCACCGCA GACTCAAGCT 7900
TCCTAAATTG TCTTTGTTAA TATTGACATC TGTGTTTGA GACTGTGAA 7950
TTATTTCTTC CCCAGTGCA TTTAGTGTC TGCAAACAGC TTGAGCTGTC 8000
CCAATGCGAT CTTGAGATTC TTTATGTTAT TAGCATGTTC 8050
CCAGTTTTCA GGATTTTTTG TCTTTTTGAA AAACGTTGCA ACTTCACTCA 8100
GCCATTGAAT AAATACCTTC ATATCATAA GAAAGTGTCG CCATTTTTTCA 8150
ACTGATCTGT CGAATCGCCC TTGCTGTTCC TTGACATTC TATGAAGTTT 8200
TTCCCCCTGG AAATCCATCT GTGCCACGGC TTCTGACTT TTACACCTTT 8250
CCATGGAGGT GGCACTTGGC AGGGCTGCTG TCTTCTTCTT GTGAAATAATA 8300
TCAATCCGAC CTGAGATTTG TTGCAAATGG TCTTTATAT TCTTAAAGAGA 8350
CTCCTCTTCG TTTAAAAGAT CTTCAGAATC TTGACACAG AGTTCAGGAG 8400
TATTTGAAAG ATGATCAACT TCTGAAAGAG CTTGTAAGAT ATGACTGATC 8450
TCGGTCAAAT AAGTAGAAGG CACATAAGAA ACATCAGAAG GCCATATCTTC 8500
AGTCGTCATC ACCATAGTTT CTTCATGGAG AGTGTGAATT TGTGAAAGAT 8550
TGAGTCTTGG AAATCGGACA AAATTGCTCT CAATTTGCGC CCAGCGCTTG 8600
CTGAGCTGGA TCTGAGTTGG CTCCAGTGCC ATTGGCGCCC CATTCTCAGA 8650
CAAGCCCTCA GCTTGCCTGC GCACGCTATT CAGCTCTCTT TTCTTTCTT 8700
GCAATTTCAG ATCATTCCCC TTTAAATTTT TTTATCTCTT GGGTGCGGTT 8750
AGGCTGCGTA ATTTTTTTTC AATTCATTTT AAGCAATTCA GGAGACTCATC 8800
AGCCCTGCCTC TTGTACTGAT ACCACTGGTG AGAAAATTCT AGGGCCTTTT 8850
FIGURE 12H

TTCTTTTTTG AGACCTCAAA TCCCTGAGAG CATTATTTTT TGCTCTGAAC 8900
AGCTGCTGTT TTTATTATAT TCCCTCCTGC TTTCTCTCAT CTGTGATTCT 8950
TTTTGTAAG TTGTCTTACT TGGCAACAAA TCTTATTTAC GTACCTCAT 9000
TGCTCTCAGT CATATTTTAA TTAGGTCCTT CTTTTTCAAG ATTCACCCCC 9050
TGCTGAATTT CAGGCTCCAG TGGTTCAAGC AATTTTTGTA TATCTGAGTT 9100
AAACTGCTCC AATTCCTTTA AAGGAATGGA GGCCTTCTCA GCTTTAATTC 9150
TGAGAGAAAT AGCTGCAAAAT CGACGGTTGA GCTGAGAGAT TGGGGGCTCT 9200
ACTACTTTCC TGGCAGTGTTG ACCGGGTGGT GCACTCAATT TGCTGCTTTG 9250
GTCACGTGGT GAGTCCACCT TGGGGCCGAT GCTATCCATT TCGGCTTTTA 9300
AACGTTAAG AATGCTTCTT TTTTTTGTGT GTTTTTCTTT TTGAGACTCA 9350
TCTAAAAGTT CATCGCATGT AATGATCCAC TTTGCTATTT TTTCTATGTT 9400
CTGATCAAAG GTTTCCAGTT GTTTCTGGTA TTCCAACAAG AGATTTAGCC 9450
ATTCTTTCTAC TCTGGAGGTT ACAGCTATCC AGTTACTTGG CAGAAGACTC 9500
AGTTTTATCTT CTACCAAGGT TTCTTCTTGT CCCACACCCA TTTTCAAGAG 9550
CTCTCCTAAT TCTGGAACAC TCTTCAAGTG AGCCTTCTGT TTCTCAAATCT 9600
CTTTTTGAGT AGCCTTCTCC CAGGCAAATT CAGAAATCCCA ATTACTTGGC 9650
ATTCTTTCAA CTGTCTGATCT CCTGCTCAAT TCTGTATCTG TGCTGCCAG 9700
CCATTTCTGT AAGACATTCA TTTCTTTCTT CATCTTACGG GACAATTCCA 9750
AGCATTTCTC CAACTGTTGC TTTCTCTCTG TAACCCAGCC ACCAATCTCA 9800
TTGTAATGCA ATTTCAAGGC TGTTACTCTG TCACTAAGCT TTTTGGGATT 9850
TTTGCTCTGC TTTTTCTGTA CAAATTGAG GCCGGTTTTA ATCACCATT 9900
CCACTTGACA CTTGGACTCT CAGGCGTTTT TATACAGTTT CACACAATGA 9950
CTTAGTTGTC ACCTGAAATTAC TTCCCTTTCA AGACTCTTGG TTTCAATGGT 10000
AGGCAATGAC ATCTTGACTT CATCTAAAT CATTCATTCT TCCTCTGAC 10050
TTGGTTCAA AATGGCCTTTG TTTTTGGAATA ATCGAAATTG ATGGGAGACA 10100
TTGTTAATT TTTTCTGTGC AACATCAATT TGTAAGAGAA CCCCTTGGTT 10150
ACAGATATTT CTGGCATATT TCTGAAGGTG CTTTCTTGGC CATCCTCTTC 11450
ACAGTGTCAC TCAGATAGTT GAAGCCCATTT TGGTGCTTCT TCAAAGAACT 11500
TTGCAGAGCC TGTAATTCCC CGAGTCTCTC TTCCATTATT TCTATTCAG 11550
TAACACTAAG ATAAAGTACA GAGAGTTTGC TTCCTGACTG CTGGATCCAC 11600
GTCCTGATGC TACTCATTTG CTCTCTGATAG CGCATTGGTG GTAAAGTGTC 11650
AAAAATTGTC TGTTACTCTTT TCTCTTGGGC CCTCACAACCT CAAAGATGT 11700
GGTTAAAAATG ATTAGTAAAG GCCACAAGTT CTGCAATTTA AACAATTCCC 11750
CCCTGTCCCT TTTCCTTCAG TTGTAGACGC TGAATTTTTA ATGTCTTAAT 11800
TTGAGGCTGA AGAGCTGACA ATCTGTGAGC TTCACTTTTA CAAATTTTTA 11850
ACTGGCTTTT AATTGCTGTG GGCTGTGAGG GGCTGGTAGA CTGGGTGTTT 11900
AACAAAGTTTT CGGCAGTATG TGTCATCTGT TCCAATTTG GCTAGCTGATT 11950
ATAAAAGGTA ATGATGTGGG TTTGATACTC TAGCCAGTTA ACTCTTCAC 12000
TCAGCAATTG GCCAATTCT GTCCACCGGC TTGTCACTTT TTCTGAAGCT 12050
TGTCAGTTAC TTTCGACTT AACCACCTCA TTTGCCATCT GTTCCACCA 12100
GGCTCTAGCT CATCTCTGCG CATCTTGCAAG TTTCCTGAAC TTCTCTGCTT 12150
TTTCTCGTGC TATGGCATTG ACTTTTCTTT GCAAGTCTGA GATGTGGCCT 12200
TCTTTTGAGT AGACTGCAAA TTTCAGAACCT TGAATACAG CTTCTGAACC 12250
AGTAATCCAA CTGTCAGAGTT CAGTTATATC GACATCAAAC CTTTCTCAG 12300
GTTCAGATTC CAGATTATC TGCCCTTCTT TGGAGGGAGG TTGGTTGTTA 12350
AGTTCCTCTT GGGCATGGTT TACCATGATT GTTCCCTTTT TGGTACCAC 12400
AGTTACGGTT TCCATTACAG TTGTCTGTTG TAGGGATGGT TGAAGTGGTTG 12450
TGACAGCCGT GAAATTTGT GCTGAACTCT TTCTCAAGTTT TTGGGTTAAA 12500
TTGTCCCAAC GTTGTCACAA GTTTTCCCAT CAGATTTCCA TCTTTTGAGT 12550
CAGTGACTTA TTTTCTGTAAG CCGAAAGTAG ATCTTGATTT ACGAATTTTA 12600
GTTTTCCCAT GGGTGGCTTT TTCTTTCTTA GATCTATTTT TAAAGTAGAT 12650
FIGURE 12K

ATTTTGTAAC GACTTGACAT CATTTCATTT TGATCTTTCAA AGCCACTTG T 12700
CTGAATGTTC TCTCTTGCAT CTCTTTTTTC TGAAGGCCAT GTACTAAAA 12750
GGCAGCTTTT CTCAGTAAAA TGCTGCTATT TTAGAAGAAT ATCTTGTAAT 12800
ACAATCCAGC GGTCTTTGAG CCATCTGACG ATATTTGCCA ATGAGCCTCC 12850
CAGTACTTTA AGTTGGTTCTT CCAAGGCAGC TGTTGACGAT TCACCGCTTG 12900
ATTCATCAAC CACTACCCAT ATGTGAGTGA GCCAGTTGAC CCTGACTCTC 12950
TCCTGTCTTA GATCTTTCTTG AAGCACCCTT TGTTGGTATC GTPGGAATTT 13000
TAGATCTTCA AGATCGAGTC CAAAGGGCTC TTCCTCCATT TTCTTAGTTC 13050
TCCTCTGAGT TTTGTATAC GATCTCATTA GTTCTTTTACA TTTCTGATTG 13100
TGGAGATCCA TTAGAATTCT GTGTAATTGT CTTTGGTTTT CCATGCCAGC 13150
TACCCGAGA CATTCCATG TTAGATTTAG GAGATTATTG TGTTCTGACA 13200
CTCCAGCTTC TCTCTCTCCT GATAAATTTCC CTTTTCACAC TAGTTGACTT 13250
CTTTGCTGTT GTAAGATTCC AACAGTCTT TCTGAGATGT TCAGATCCTC 13300
CATGAATCCC TCATGGACAT GAAACTGTTT TTACACTTCT TCAACATCAT 13350
TGGAATCTTC TCTTTGTGCT CCAAGTAGT TCTCGGCAGA AAGAAGCCAT 13400
GAAAGTACTCT TTCTCAAAGC ATTTGGTAA CTTACCAGAT TTACTTCCGT 13450
CTCCATCAAT GAATCTGCAA GTGACTTTGT CTTGGGAGCT TCAAATGCT 13500
TGGAAGGATA GGGCTCTGT GTGGAAATCG AGTGGCAAC ATAAACGCC 13550
TGTTGAAGG GATACTCTTG GAATCGAGGC TTAGGAGATG AAGAAGTTTG 13600
TTCTAGCCA TTGTCTAGAC TGAATGTGCT TCTTGGGAGG TAATGCATCT 13650
GGTGATGAAA GTGGAATACTG TCTTTCTCAG TACTGTTTGA AGATGTCCGT 13700
GGCAAGATTT CCACTTCTTG AATGGCTTCA ATGCTCATTG GTTTGGCAAA 13750
AATTTGAAG AGTGATGTTG TGTACATTA GATGGACTTC TTTGCTGAG 13800
AAGTGGTAGC AACATCTTCA GGAATCAAGA GTTTTTCTAT GCCAACTGG 13850
CATTTCGAAA TGTTGAAAGGC ATGTCCACGT CTGTTGCTGG CTGAGTC 13900
FIGURE 12M

ATGACTAATA CGTAGATGT A CTGCCAAGTA GGAAGTCCC A TAAAGTTCAT 15250
GTACTGCGCA TAATGCCAGG GCGGCCATTT ACCGTATTG AGCTCAATAG 15300
GGGGCGTACT TGGCATATGA TACACTTTGT GTACTGCAAA GTGGGCAGTT 15350
TACCGTAAAT ACTCCACCCA TTTAGCTCAA TGGAAGTCCC C TATTGGCGT 15400
TACTATGGGA ACATACGTCA TTTATGAGCT CAATGGGCGG GGGTCTGTTG 15450
GCGTGCAGCC AGGGCGGGCA TTTACCGTAA GGTATGTAAC GCACCTGAGG 15500
TCAAATCTAG AGGATCTCCC TGACAAATAA TTACGGCCTA TGAATGAACAC 15550
AAAATTATTC AGATTTCCTC TCTCTTTATT CAGTTTTTCC GCCAAATTGG 15600
CCCAATCTTTA CTCGGTGTAC CCGCAATTTTA CTACAAATCT CCGCTAAAC 15650
CGGCAGAAA TGTCACCTTC CTTGTTGACAC CGGCGCAAC CCAAAACGTC 15700
ACTTTTGCCCA CATCCGTCGC TTACCTAGTT TCCGCACAC CGGAACACATC 15750
ACACTTGGCC CAGACTCTA CTCACCCGCC CCGGTTCCTC CGCCCGCCGC 15800
CAGTCACAA ACTCCACCCC TGCAAGTTCA ATATTTGCTTTC AATCCAAAAT 15850
AAGGTATATT ATGAGCTGAT CAGCCGGGGT CCAAGATATCG GATCCAAATTG 15900
CAATGATCAT CATGACAGAT CTGGCGCGCA TGATATCGA CGTATTTAAAT 15950
TTGCCATAGC TAGCTATAGT TTAGAGGCTA CGGTTGTTTA ACATTAGGGG 16000
GCTACGTATA CTCGGGAATA TTAAGGCTG TGGATGACAT ATGGGGGGCG 16050
GCGCCCTGCA GCTGGCGCCA TCGATACGCG TAGCTGGCGA CCACGACCAT 16100
GTACAGAGCT CGAGAAGTAC TAGTGCCAGC GTGGGGCGGTC CACCTTAAGC 16150
TTTGCCATGGG CGTGTGTTTTT ACAAAGTCTG TACTGGAAAA ACCCTGCGGT 16200
TACACCACCTT AACGGCCTTG CAGCACACTCC CAGTTTGCGC AGCTGGCGTA 16250
ATAGCAGAGA GGCCCGCACC GATCGCCTCTT CCAACAGATT GGCAGCCTG 16300
AATGGCGGAAT GGGCGCTGAT GGGTTTTTTT CCTGCACGTC ATCTGGCGG 16350
TATTTTACAC CGCATACTGC AAAGCAACCA TAGTGCGCCG CCTGAGCGG 16400
CGCATTAAGC CGGGCGGTTGT TGGTGTTTAC GGCAGCGGTG ACCGCTACAC 16450
FIGURE 120

CGCCTTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATA CAAACGACGA 17750
GCGTGACACC ACGATGCTGT TAGCAATGGC AACAAGTTG CGAAYATAT 17800
TAACCTGGCA AACTTTAAT CTTAGTCCCTT CCGAACAATT ATAGACTGG 17850
ATGGAGCCGG ATAAAGGTTG AGGACCACCT TCTGCGTCGG CCTTCCGGC 17900
TGGCTGATTT AGTGCACTGATA AACCTGGACG CGGTGAGGCT GGGCTCGCG 17950
GTATCATTGC AGCAATGGG CGCAGTGGTA AGCCCTCCGG TATCGTAGTT 18000
ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAAGCAA ATAGAGCAGAT 18050
CGCTGACACTA GCTGAATGCTA TTTTGTAAT TCATGACCA AAATCCCTTA 18100
TTTACTCATA TATACATTTA ATTGATTTAA AACTTCATTT TTAATTTAAAA 18150
AGGATCATGC TGAAGATCAT TTTGATAAA CTCATGACCA AAATCCCTTA 18200
ACGTGACTTT CTGTTGCTACT GAGCTCCAG CCCCCTAGAA AAGATCAAAAG 18250
GATCTTTCCTT AGATCTTTTT TTTCTCGCGG TAACTTCCTG CTGCAAACA 18300
AAAACGCAC CGCTACCAGC GGTGGTTTGT TTGCGGCCATC AAGAGCTACC 18350
AACTCTTTTT CGGAGGATCA CTGGTCTTCGG CAGAGGCGAC ATACAAATA 18400
CTGTTCTCTT AGTGAAGCCG TAGTTAGGCC ACCACTTCAA GAACTCTGTA 18450
GCACCGCCTG CATTACCCTG TCTGCTATAC CTCGGTACCAG TGGCTGCCTG 18500
CAGTGGCAT AAGCTCTTCT TTACCGGGTT GGAATCAAGA CGATAGTTAC 18550
CGGATAAGGC GCAGCGGCTG GGCTGAAAGG GGGTTCTGTT CACACAGCAGC 18600
AGCTTGGACG CAAGGACTCT CACCGAACCTT ATACACCTAC AGCTTACGT 18650
ATGAGAAAGC GCCACGCTTC CGGAAGGGG AAAGCGGAG AGGTATCCGG 18700
TAAGCGCCAG GTGTGGAACA GGAAGGGCG AGAGGGAGCT TCCAGGCGGA 18750
AACGCCCTGT ATCTTTTATAG TCTTGTGCGG TTTCCGCACC TCTGACTGGA 18800
GCGTGGATTG TTTGACTGCT CGTCTGGGGG CGCGAGGCTA TGGAAAAACG 18850
CCAGAAGGCC GGCCTTTTTA CGTTCTCGGT CCTTCTGCTG GCCTTTTTTCT 18900
CACATGTGCT TTCTCGCGGT ATCCCGTAT CTTGTTCA ACGTATTAC 18950
Ad.CBhpAP

HELPERS VIRUS

hpAP

LacZ

AdΔ.CMVLacZ

PACKAGE INTO EMPTY CAPSID HEAD PROVIDED BY HELPER VIRUS