Production of host cells containing multiple integrating vectors by serial transduction

C07K 14/76 (2006.01) C12N 15/82 (2006.01)
C07K 14/02 (2006.01) C12N 15/85 (2006.01)
C07K 16/12 (2006.01) C12N 15/86 (2006.01)
C07K 16/30 (2006.01) C12N 15/864 (2006.01)
C12N 5/10 (2006.01) C12N 15/867 (2006.01)
C12N 7/04 (2006.01) C12P 21/02 (2006.01)
C12N 15/09 (2006.01) C12Q 1/02 (2006.01)

C07K 16/12 (2006.01) C12N 16/14 (2006.01)
C12N 7/04 (2006.01) C12P 21/02 (2006.01)
C12N 15/09 (2006.01) C12Q 1/02 (2006.01)

Related Art
Title: PRODUCTION OF HOST CELLS CONTAINING MULTIPLE INTEGRATING VECTORS BY SERIAL TRANSDUCTION

Abstract: The present invention relates to the production of proteins in host cells, and more particularly to host cells containing multiple integrated copies of an integrating vector comprising an exogenous gene and methods of making such host cells by serial transduction or transfection. The present invention further provides methods of expressing increased levels of protein in host cells using such vectors.
PRODUCTION OF HOST CELLS CONTAINING MULTIPLE INTEGRATING VECTORS BY SERIAL TRANSDUCTION

This application in a continuation-in-part of U.S. Patent Application serial number 10/397,079, filed 3/26/03, which is a continuation in part of U.S. Patent Application serial number 09/897,511, filed 6/29/01, which claims priority to provisional application 60/215,925, filed 7/3/00.

FIELD OF THE INVENTION

The present invention relates to the production of proteins in host cells, and more particularly to host cells containing multiple integrated copies of an integrating vector comprising an exogenous gene and methods of making such host cells by serial transduction or transfection.

BACKGROUND OF THE INVENTION

The pharmaceutical biotechnology industry is based on the production of recombinant proteins in mammalian cells. These proteins are essential to the therapeutic treatment of many diseases and conditions. In many cases, the market for these proteins exceeds a billion dollars a year. Examples of proteins produced recombinantly in mammalian cells include erythropoietin, factor VIII, factor IX, and insulin. For many of these proteins, expression in mammalian cells is preferred over expression in prokaryotic cells because of the need for correct post-translational modification (e.g., glycosylation or silation; see, e.g., U.S. Pat. No. 5,721,121, incorporated herein by reference).

Several methods are known for creating host cells that express recombinant proteins. In the most basic methods, a nucleic acid construct containing a gene encoding a heterologous protein and appropriate regulatory regions is introduced into the host cell and allowed to integrate. Methods of introduction include calcium phosphate precipitation, microinjection, lipofection, and electroporation. In other methods, a selection scheme is used to amplify the introduced nucleic acid construct. In these methods, the cells are co-transfected with a gene encoding an amplifiable selection marker and a gene encoding a heterologous protein (See, e.g., Schroder and Friedl, Biotech. Bioeng. 53(6):547-59 [1997]). After selection of the initial
transformants, the transfected genes are amplified by the stepwise increase of the selective agent (e.g., dihydrofolate reductase) in the culture medium. In some cases, the exogenous gene may be amplified several hundred-fold by these procedures. Other methods of recombinant protein expression in mammalian cells utilize transfection with episomal vectors (e.g., plasmids).

Current methods for creating mammalian cell lines for expression of recombinant proteins suffer from several drawbacks. (See, e.g., Mielke et al., Biochem. 35:2239-52 [1996]). Episomal systems allow for high expression levels of the recombinant protein, but are frequently only stable for a short time period (See, e.g., Klehr and Bode, Mol. Genet. (Life Sci. Adv.) 7:47-52 [1988]). Mammalian cell lines containing integrated exogenous genes are somewhat more stable, but there is increasing evidence that stability depends on the presence of only a few copies or even a single copy of the exogenous gene.

Standard transfection techniques favor the introduction of multiple copies of the transgene into the genome of the host cell. Multiple integration of the transgene has, in many cases, proven to be intrinsically unstable. This intrinsic instability may be due to the characteristic head-to-tail mode of integration which promotes the loss of coding sequences by homologous recombination (See, e.g., Weidle et al., Gene 66:193-203 [1988]) especially when the transgenes are transcribed (See, e.g., McBurney et al., Somatic Cell Molec. Genet. 20:529-40 [1994]). Host cells also have epigenetic defense mechanisms directed against multiple copy integration events. In plants, this mechanism has been termed “cosuppression.” (See, e.g., Allen et al., Plant Cell 5:603-13 [1993]). Indeed, it is not uncommon that the level of expression is inversely related to copy number. These observations are consistent with findings that multiple copies of exogenous genes become inactivated by methylation (See, e.g., Mehtali et al., Gene 91:179-84 [1990]) and subsequent mutagenesis (See, e.g., Kricker et al., Proc. Natl. Acad. Sci. 89:1075-79 [1992]) or silenced by heterochromatin formation (See, e.g., Dorer and Henikoff, Cell 77:993-1002 [1994]).

Accordingly, what is needed in the art are improved methods for making host cells that express recombinant proteins. Preferably, the host cells will be stable over extended periods of time and express the protein encoded by a transgene at high levels.

SUMMARY OF THE INVENTION
In a first aspect, the invention provides a method for transducing host cells comprising:

a) providing:
   i) at least one host cell comprising a genome, and
   ii) a plurality of retroviral vectors encoding a gene of interest operably linked to a promoter internal to the retroviral 5' and 3' long terminal repeats and a segment encoding a secretion signal sequence;

b) contacting said at least one host cell with said plurality of retroviral-vectors under conditions such that said host cells are transduced to produce transduced host cells, wherein said conditions comprise contacting said host cell at a multiplicity of infection of from about 10 to 1000;

c) repeating steps a) and b) a plurality of times to provide host cells comprising from about 20 to about 100 integrated retroviral vectors, wherein said host cell does not produce infectious retroviral vectors.

In a second aspect, the invention provides a transduced host cell produced in accordance with the method of the first aspect, wherein said host cell is transduced with at least two different vectors encoding different genes of interest.

In a third aspect, the invention provides a method for transducing host cells comprising:

a) providing:
   i) at least one host cell comprising a genome, and
   ii) a plurality of retroviral vectors encoding a gene of interest; and

b) contacting said at least one host cell with said plurality of integrating vectors under conditions such that said host cells are transduced to produce transduced host cells;

c) repeating steps a) and b) a plurality of times to provide host cells comprising greater than about 20 integrated retroviral vectors;

d) clonally selecting a host cell expressing said gene of interest; and

e) purifying a protein of interest encoded by said gene of interest.

In a fourth aspect, the invention provides a transduced host cell produced in accordance with the third aspect, wherein said host cell comprises greater than about 20 integrated retroviral vectors.

The present invention relates to the production of proteins in host cells, and more particularly to host cells containing multiple integrated copies of an integrating vector comprising an exogenous gene and methods of making such host cells by serial transduction or transfection.
Accordingly, in some embodiments, the present invention provides a host cell comprising a genome, the genome comprising at least one integrated integrating vector, wherein the integrating vector comprises at least one exogenous gene operably linked to a promoter, and wherein the integrating vectors lacks a gene encoding a selectable marker. In some embodiments, the integrating vector further comprises a secretion signal sequence operably linked to the exogenous gene. In some embodiments, the integrating vector further comprises an RNA stabilizing element operably linked to the exogenous gene. In some embodiments, the integrating vector is a retroviral vector. In some embodiments, the retroviral vector is a pseudotyped retroviral vector. In certain embodiments, the pseudotyped retroviral vector comprises a G glycoprotein selected from the group including, but not limited to, vesicular stomatitis virus, Piry virus, Chandipura virus, Spring viremia of carp virus and Mokola virus G glycoproteins. In further embodiments, the retroviral vector comprises long terminal repeats selected from the group including, but not limited to, MoMLV, MoMuSV, and MMTV long terminal repeats. In some embodiments, the host cell is clonally derived. In other embodiments, the host cell is non-clonally derived. In some preferred embodiments, the genome is stable for greater than 10 passages, and preferably, stable for greater than 100 passages. In certain particularly preferred embodiments, the integrated exogenous gene is stable in the absence of selection. In some embodiments, the at least one exogenous gene is selected from the group consisting of genes encoding antigen binding proteins, pharmaceutical proteins, kinases, phosphatases, nucleic acid binding proteins, membrane receptor proteins, signal transduction proteins, ion channel proteins, and oncoproteins. In some embodiments, the genome comprises at least 5, and preferably, at least 100 integrated vectors. In some preferred embodiments, the host cell expresses greater than about 3, and preferably, greater than about 10 picograms of the exogenous protein per day.

The present invention also provides a method for transfecting host cells comprising providing a plurality of host cells comprising a genome, and a plurality of integrating vectors, wherein the integrating vectors comprise at least one exogenous gene, and wherein the
integrating vectors lack a gene encoding a selectable marker; contacting the host cell with the plurality of integrating vectors to generate transfected host cells comprising at least one integrated copy of the integrating vector; and clonally selecting the transfected host cells. In some preferred embodiments, the integrated exogenous gene is stable in the absence of selection.

In some embodiments, the host is contacted with the integrating vectors at a multiplicity of infection of greater than 10. In some embodiments, the host cells are contacted with the plurality of integrating vectors under conditions such that at least 2, preferably 5, and even more preferably 10, integrating vectors integrate into the genome of the host cell. In some embodiments, the clonally selecting comprises detecting nucleic acid of the exogenous gene. In some embodiments, detecting nucleic acid of the exogenous gene comprises a detection assay selected from the group consisting of a PCR assay and a hybridization assay. In other embodiments, the clonally selecting comprises detecting protein expressed by the exogenous gene. In some embodiments, detecting protein expressed by the exogenous gene comprises a detection assay selected from the group consisting of an immunoassay and a biochemical assay.

In some embodiments, the immunoassay is selected from the group consisting of ELISA and Western blot. In some embodiments, the integrating vector is a retroviral vector. In some preferred embodiments, the host cells synthesize greater than about 1, preferably greater than about 10, and even more preferably, greater than about 50 picograms per cell per day of protein from the exogenous gene of interest.

The present invention further provides a method of producing a protein of interest comprising providing a host cell comprising a genome, the genome comprising at least one integrated copy of at least one integrating vector comprising an exogenous gene operably linked to a promoter, wherein the integrating vector lacks a gene encoding a selectable marker, and wherein the exogenous gene encodes a protein of interest, and culturing the host cells under conditions such that the protein of interest is produced. In some preferred embodiments, the integrated exogenous gene is stable in the absence of selection. In some embodiments, the integrating vector further comprises a secretion signal sequence operably linked to the exogenous gene. In some embodiments, the method further comprises the step of isolating the protein of interest. In some embodiments, the method further comprises the step of clonally selecting at least 10 colonies. In some embodiments, the clonally selecting comprising detecting the protein expressed by the exogenous gene. In some embodiments, detecting protein expressed by the
exogenous gene comprises a detection assay selected from the group consisting of an
immunoassay and a biochemical assay. In some embodiments, the immunoassay is selected
from the group consisting of ELISA and Western blot. In some embodiments, the genome of the
host cell comprises greater than 5, and preferably greater than 10, integrated copies of the
integrating vector. In some embodiments, the integrating vector is a retroviral vector. In some
embodiments, the host cells synthesize greater than about 1, preferably greater than 10, and even
more preferably, greater than 50 picograms per cell per day of the protein of interest.

The present invention also provides a retroviral vector comprising a gene construct
comprising an exogenous promoter operably linked to an exogenous gene, wherein the vector
lacks a gene encoding a selectable marker. In some embodiments, the retroviral vector is a
pseudotyped retroviral vector. In some embodiments, the pseudotyped retroviral vector
comprises a G glycoprotein selected from the group including, but not limited to, vesicular
stomatitis virus, Piry virus, Chandipura virus, Spring viremia of carp virus and Mokola virus G
glycoproteins. In some embodiments, the retroviral vector comprises long terminal repeats
selected from the group including, but not limited to, MoMLV, MoMuSV, and MMTV long
terminal repeats.

In some embodiments, the present invention provides methods for transducing host
cells comprising: a) providing: i) at least one host cell comprising a genome, and ii) a
plurality of retroviral vectors encoding a gene of interest; and b) contacting said at least one
host cell with said plurality of integrating vectors under conditions such that said host cells
are transduced to produce transduced host cells; and c) repeating steps a) and b) a plurality of
times to provide host cells comprising multiple integrated retroviral vectors. The present
invention is not limited to repeating steps a) and b) any particular number of times. Indeed,
in some embodiments, the steps a) and b) may be repeated at least 3, 4, 5, 6, 7, 8, 10 or
between about 3 and 20 times. The present invention is not limited to the integration of any
particular number of vectors. In some embodiments, about 10 to about 100 retroviral vectors
are integrated. The present invention is not limited to retroviral vectors produced by any
particular method. In some embodiments, the retroviral vectors utilized in steps 1 and 2 are
produced from packaging cells transfected with an envelope plasmid and a vector plasmid.

In some embodiments, the methods of the present invention further comprise step d)
transducing said host cells comprising multiple integrated retroviral vectors produced by
steps 1 and 2 with vectors produced from packaging cells produced by transducing said packaging cells with a retroviral vector encoding said gene of interest and transfecting said packaging cell with a plasmid expressing an envelope protein. In some embodiments, the packaging cells express retroviral gag and pol proteins. In some preferred embodiments, the packaging cells are 293-GP cells. In some embodiments, the envelope plasmid encodes a G protein. In some preferred embodiments, the G protein is VSV-G protein. In still other embodiments, the retroviral vector comprises MoMLV elements.

In some embodiments of the present invention, the conditions comprise contacting said host at a multiplicity of infection of from about 10 to 1000. In some embodiments, the gene of interest is operably linked to an exogenous promoter. In further embodiments, the gene of interest is operably linked to a signal sequence. In still further embodiments, the retroviral vector encodes at least two genes of interest. In some embodiments, the at least two genes of interest are arranged in a polycistronic sequence. In some preferred embodiments, the at least two genes of interest comprise immunoglobulin heavy and light chains. In still other embodiments, the retroviral vector is a lentiviral vector. In some embodiments, the host cell is selected from Chinese hamster ovary cells, baby hamster kidney cells, human 293 cells, and bovine mammary epithelial cells.

In some embodiments, the methods further comprise the step of clonally selecting said transduced host cells. In further embodiments, the methods comprise the step of culturing said clonally selected host cells under conditions such that a protein of interest encoded by said gene of interest is produced. In some embodiments, the retroviral vector further comprises a secretion signal sequence operably linked to said exogenous gene. In still further embodiments, the methods comprise the step of isolating said protein of interest. In some preferred embodiments, the culture conditions are selected from the group consisting of roller bottle cultures, perfusion cultures, batch fed cultures, and petri dish cultures. In some embodiments, the host cells synthesize greater than about 1 picograms per cell per day of said protein of interest. In some embodiments, the host cells synthesize greater than about 10 picograms per cell per day of said protein of interest. In some embodiments, the host cells synthesize greater than about 50 picograms per cell per day of said protein of interest.

In some embodiments, the retroviral vector further encodes an amplifiable marker. In some preferred embodiments, the amplifiable marker is selected from the group consisting of
DHFR and glutamine synthetase. In further embodiments, the methods comprise the step of culturing said transduced host cells under conditions that allow for amplification of the integrated retroviral vectors. In some embodiments, the conditions comprise culturing said transduced host cells in the presence of a selection agent selected from the group consisting of methotrexate, phosphinothricin and methionine sulfoxime. In some embodiments, the immunoglobulins are selected from the group consisting of IgG, IgA, IgM, IgD, Ibe and sIg.

In other embodiments, the host cell is transduced with at least two different vectors encoding different genes of interest.

In still further embodiments, the present invention provides host cells produced by the foregoing methods.

DESCRIPTION OF THE FIGURES

Figure 1 is a western blot of a 15% SDS-PAGE gel run under denaturing conditions and probed with anti-human IgG (Fc) and anti-human IgG (Kappa).

Figure 2 is a graph of MN14 expression over time.

Figure 3 is a Western blot of a 15% PAGE run under non-denaturing conditions and probed with anti-human IgG (Fc) and anti-human IgG (Kappa).

Figure 4 provides the sequence for the hybrid human-bovine alpha-lactalbumin promoter (SEQ ID NO:1).

Figure 5 provides the sequence for the mutated PPE sequence (SEQ ID NO:2).

Figure 6 provides the sequence for the IRES-Signal peptide sequence (SEQ ID NO:3).

Figures 7a and 7b provide the sequence for CMV MN14 vector (SEQ ID NO:4).

Figures 8a and 8b provide the sequence for the CMV LL2 vector (SEQ ID NO:5).

Figures 9a-c provide the sequence for the MMTV MN14 vector (SEQ ID NO:6).

Figures 10a-d provide the sequence for the alpha-lactalbumin MN14 Vector (SEQ ID NO:7).

Figures 11a-c provide the sequence for the alpha-lactalbumin Bot vector (SEQ ID NO:8).

Figures 12a-b provide the sequence for the LSRNL vector (SEQ ID NO:9).

Figures 13a-b provide the sequence for the alpha-lactalbumin cc49IL2 vector (SEQ ID NO:10).
Figures 14a-c provide the sequence for the alpha-lactalbumin YP vector (SEQ ID NO:11).

Figure 15 provides the sequence for the IRES-Casein signal peptide sequence (SEQ ID NO:12).

Figures 16a-c provide the sequence for the LNBOTDC vector (SEQ ID NO:13).

Figure 17 provides a graph depicting the INVADER Assay gene ratio in CMV promoter cell lines.

Figure 18 provides a graph depicting the INVADER Assay gene ratio in α-lactalbumin promoter cell lines.

Figures 19a-d provides the sequence of a retroviral vector that expresses a G-Protein coupled receptor and antibody light chain.

Figure 20 shows a graph demonstrating increased expression of a gene of interest in the absence of a selectable marker.

Figure 21 provides SEQ ID NO:37, the coding sequence for a vector encoding an IgM.

Figure 22 provides SEQ ID NO:38, the coding sequence for one vector of a two vector system for producing an IgM.

Figure 23 provides SEQ ID NO:39, the coding sequence for one vector of a two vector system for producing an IgM.

Figure 24 provides SEQ ID NO:40, the coding sequence for a retroviral vector comprising an amplifiable marker (dhfr).

Figure 25 provides SEQ ID NO:41, the coding sequence for a retroviral vector comprising an amplifiable marker (gs).

DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

As used herein, the term "host cell" refers to any eukaryotic cell (e.g., mammalian cells, avian cells, amphibian cells, plant cells, fish cells, and insect cells), whether located in vitro or in vivo.

As used herein, the term "cell culture" refers to any in vitro culture of cells. Included within this term are continuous cell lines (e.g., with an immortal phenotype), primary cell
cultures, finite cell lines (e.g., non-transformed cells), and any other cell population maintained in vitro, including oocytes and embryos.

As used herein, the term "vector" refers to any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, virus, virion, etc., which is capable of replication when associated with the proper control elements and which can transfer gene sequences between cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

As used herein, the term "integrating vector" refers to a vector whose integration or insertion into a nucleic acid (e.g., a chromosome) is accomplished via an integrase. Examples of "integrating vectors" include, but are not limited to, retroviral vectors, transposons, and adenovirus vectors.

As used herein, the term "integrated" refers to a vector that is stably inserted into the genome (i.e., into a chromosome) of a host cell.

As used herein, the term "multiplicity of infection" or "MOI" refers to the ratio of integrating vectors:host cells used during transfection or transduction of host cells. For example, if 1,000,000 vectors are used to transduce 100,000 host cells, the multiplicity of infection is 10. The use of this term is not limited to events involving transduction, but instead encompasses introduction of a vector into a host by methods such as lipofection, microinjection, calcium phosphate precipitation, and electroporation.

As used herein, the term "genome" refers to the genetic material (e.g., chromosomes) of an organism.

The term "nucleotide sequence of interest" refers to any nucleotide sequence (e.g., RNA or DNA), the manipulation of which may be deemed desirable for any reason (e.g., treat disease, confer improved qualities, expression of a protein of interest in a host cell, expression of a ribozyme, etc.), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (e.g., reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, etc.), and non-coding regulatory sequences which do not encode an mRNA or protein product (e.g., promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, etc.).

As used herein, the term "protein of interest" refers to a protein encoded by a nucleic acid of interest.
As used herein, the term "signal protein" refers to a protein that is co-expressed with a protein of interest and which, when detected by a suitable assay, provides indirect evidence of expression of the protein of interest. Examples of signal protein useful in the present invention include, but are not limited to, immunoglobulin heavy and light chains, beta-galactosidase, beta-lactamase, green fluorescent protein, and luciferase.

As used herein, the term "exogenous gene" refers to a gene that is not naturally present in a host organism or cell, or is artificially introduced into a host organism or cell.

The term "gene" refers to a nucleic acid (e.g., DNA or RNA) sequence that comprises coding sequences necessary for the production of a polypeptide or precursor (e.g., proinsulin). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. The sequences that are located 5' of the coding region and which are present on the mRNA are referred to as 5' untranslated sequences. The sequences that are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' untranslated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

As used herein, the term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (e.g., mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (i.e., via the enzymatic action of an RNA polymerase), and for protein encoding genes, into protein through "translation" of mRNA. Gene expression can be regulated at many stages in the process. "Up-regulation" or "activation" refers to regulation that increases the production of gene expression products (i.e., RNA or protein), while "down-regulation" or...
"repression" refers to regulation that decrease production. Molecules (e.g., transcription factors) that are involved in up-regulation or down-regulation are often called \"activators\" and \"repressors,\" respectively.

Where \"amino acid sequence\" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, \"amino acid sequence\" and like terms, such as \"polypeptide\" or \"protein\" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

As used herein, the terms \"nucleic acid molecule encoding,\" \"DNA sequence encoding,\" \"DNA encoding,\" \"RNA sequence encoding,\" and \"RNA encoding\" refer to the order or sequence of deoxyribonucleotides or ribonucleotides along a strand of deoxyribonucleic acid or ribonucleic acid. The order of these deoxyribonucleotides or ribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA or RNA sequence thus codes for the amino acid sequence.

As used herein, the term \"variant,\" when used in reference to proteins, refers to proteins encoded by partially homologous nucleic acids so that the amino acid sequence of the proteins varies. As used herein, the term \"variant\" encompasses proteins encoded by homologous genes having both conservative and nonconservative amino acid substitutions that do not result in a change in protein function, as well as proteins encoded by homologous genes having amino acid substitutions that cause decreased (e.g., null mutations) protein function or increased protein function.

As used herein, the terms \"complementary\" or \"complementarity\" are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence \"5'-A-G-T-3',\" is complementary to the sequence \"3'-T-C-A-5'.\" Complementarity may be \"partial,\" in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be \"complete\" or \"total\" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

The terms \"homology\" and \"percent identity\" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (i.e., partial identity) or
complete homology (i.e., complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (i.e., an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (i.e., it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.
As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the Tm of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

As used herein, the term "Tm" is used in reference to the "melting temperature" of a nucleic acid. The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the Tm of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the Tm value may be calculated by the equation: 

\[ T_m = 81.5 + 0.41(\% G + C) \]

when a nucleic acid is in aqueous solution at 1 M NaCl (See e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization [1985]). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of Tm.

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "weak" or "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄•H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄•H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH),
0.5% SDS, 5X Denhardt's reagent and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄•H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharamcia), 5 g BSA (Fraction V; Sigma)] and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

A gene may produce multiple RNA species that are generated by differential splicing of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example, representing the presence of exon "A" on cDNA 1 wherein cDNA 2 contains exon "B" instead).

Because the two cDNAs contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

The terms "in operable combination," "in operable order," and "operably linked" as used herein refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

As used herein, the term "selectable marker" refers to a gene that encodes an enzymatic activity that confers the ability to grow in medium lacking what would otherwise be an essential nutrient (e.g. the HIS3 gene in yeast cells); in addition, a selectable marker may confer resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed. Selectable markers may be "dominant"; a dominant selectable marker encodes an enzymatic activity that can be detected in any eukaryotic cell line. Examples of dominant selectable markers include the bacterial aminoglycoside 3' phosphotransferase gene (also referred to as the neo gene) that confers resistance to the drug G418 in mammalian cells, the bacterial hygromycin G phosphotransferase (hyg) gene that confers resistance to the antibiotic hygromycin and the
bacterial xanthine-guanine phosphoribosyl transferase gene (also referred to as the gpt gene) that confers the ability to grow in the presence of mycophenolic acid. Other selectable markers are not dominant in that their use must be in conjunction with a cell line that lacks the relevant enzyme activity. Examples of non-dominant selectable markers include the thymidine kinase (tk) gene that is used in conjunction with tk\textsuperscript{-} cell lines, the CAD gene, which is used in conjunction with CAD-deficient cells, and the mammalian hypoxanthine-guanine phosphoribosyl transferase (hprt) gene, which is used in conjunction with hprt\textsuperscript{-} cell lines. A review of the use of selectable markers in mammalian cell lines is provided in Sambrook, J. \textit{et al.}, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York (1989) pp.16.9-16.15.

As used herein, the term "lacking a selectable marker" as in integrating vectors that lack a gene encoding a selectable marker refers to integrating vectors that do not contain a gene encoding a selectable marker.

As used herein, the term "selection free growth" refers to growth in the absence of selective conditions required for a given selectable marker (\textit{e.g.}, antibiotics or the deficiency of a nutrient of enzymatic activity). In some preferred embodiments, host cells comprising integrating vectors that "lack a selectable marker" are also subjected to selection free growth.

As used herein, the term "regulatory element" refers to a genetic element that controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, RNA export elements, internal ribosome entry sites, etc. (defined infra).

Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis \textit{et al.}, Science 236:1237 [1987]). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect and mammalian cells, and viruses (analogous control elements, \textit{i.e.}, promoters, are also found in prokaryotes). The selection of a particular promoter and enhancer depends on what cell type is to be used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types (for review see, Voss \textit{et al.}, Trends Biochem. Sci., 11:287 [1986];
and Maniatis et al., supra). For example, the SV40 early gene enhancer is very active in a wide variety of cell types from many mammalian species and has been widely used for the expression of proteins in mammalian cells (Dijkema et al., EMBO J. 4:761 [1985]). Two other examples of promoter/enhancer elements active in a broad range of mammalian cell types are those from the human elongation factor 1α gene (Uetsuki et al., J. Biol. Chem., 264:5791 [1989]; Kim et al., Gene 91:217 [1990]; and Mizushima and Nagata, Nuc. Acids. Res., 18:5322 [1990]) and the long terminal repeats of the Rous sarcoma virus (Gorman et al., Proc. Natl. Acad. Sci. USA 79:6777 [1982]) and the human cytomegalovirus (Boshart et al., Cell 41:521 [1985]).

As used herein, the term "promoter/enhancer" denotes a segment of DNA which contains sequences capable of providing both promoter and enhancer functions (i.e., the functions provided by a promoter element and an enhancer element, see above for a discussion of these functions). For example, the long terminal repeats of retroviruses contain both promoter and enhancer functions. The enhancer/promoter may be "endogenous" or "exogenous" or "heterologous." An "endogenous" enhancer/promoter is one that is naturally linked with a given gene in the genome. An "exogenous" or "heterologous" enhancer/promoter is one that is placed in juxtaposition to a gene by means of genetic manipulation (i.e., molecular biological techniques such as cloning and recombination) such that transcription of that gene is directed by the linked enhancer/promoter.

Regulatory elements may be tissue specific or cell specific. The term "tissue specific" as it applies to a regulatory element refers to a regulatory element that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (e.g., liver) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (e.g., lung).

Tissue specificity of a regulatory element may be evaluated by, for example, operably linking a reporter gene to a promoter sequence (which is not tissue-specific) and to the regulatory element to generate a reporter construct, introducing the reporter construct into the genome of an animal such that the reporter construct is integrated into every tissue of the resulting transgenic animal, and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic animal. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the regulatory element
is "specific" for the tissues in which greater levels of expression are detected. Thus, the term "tissue-specific" (e.g., liver-specific) as used herein is a relative term that does not require absolute specificity of expression. In other words, the term "tissue-specific" does not require that one tissue have extremely high levels of expression and another tissue have no expression. It is sufficient that expression is greater in one tissue than another. By contrast, "strict" or "absolute" tissue-specific expression is meant to indicate expression in a single tissue type (e.g., liver) with no detectable expression in other tissues.

The term "cell type specific" as applied to a regulatory element refers to a regulatory element that is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a regulatory element also means a regulatory element capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue.

Cell type specificity of a regulatory element may be assessed using methods well known in the art (e.g., immunohistochemical staining and/or Northern blot analysis). Briefly, for immunohistochemical staining, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is regulated by the regulatory element. A labeled (e.g., peroxidase conjugated) secondary antibody specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy. Briefly, for Northern blot analysis, RNA is isolated from cells and electrophoresed on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support (e.g., nitrocellulose or a nylon membrane). The immobilized RNA is then probed with a labeled oligo-deoxyribonucleotide probe or DNA probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists.

The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (i.e., upstream) of a nucleotide sequence of interest.
whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (e.g., heat shock, chemicals, etc.). In contrast, a "regulatable" promoter is one that is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (e.g., heat shock, chemicals, etc.), which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The presence of "splicing signals" on an expression vector often results in higher levels of expression of the recombinant transcript. Splicing signals mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York [1989], pp. 16.7-16.8). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length. The term "poly A site" or "poly A sequence" as used herein denotes a DNA sequence that directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable as transcripts lacking a poly A tail are unstable and are rapidly degraded. The poly A signal utilized in an expression vector may be "heterologous" or "endogenous." An endogenous poly A signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly A signal is one that is isolated from one gene and placed 3' of another gene. A commonly used heterologous poly A signal is the SV40 poly A signal. The SV40 poly A signal is contained on a 237 bp BamHI/BclI restriction fragment and directs both termination and polyadenylation (Sambrook, supra, at 16.6-16.7).

Eukaryotic expression vectors may also contain "viral replicons" or "viral origins of replication." Viral replicons are viral DNA sequences that allow for the extrachromosomal replication of a vector in a host cell expressing the appropriate replication factors. Vectors that
contain either the SV40 or polyoma virus origin of replication replicate to high "copy number" (up to $10^4$ copies/cell) in cells that express the appropriate viral T antigen. Vectors that contain the replicons from bovine papillomavirus or Epstein-Barr virus replicate extrachromosomally at "low copy number" (~100 copies/cell). However, it is not intended that expression vectors be limited to any particular viral origin of replication.

As used herein, the term "long terminal repeat" or "LTR" refers to transcriptional control elements located in or isolated from the U3 region 5' and 3' of a retroviral genome. As is known in the art, long terminal repeats may be used as control elements in retroviral vectors, or isolated from the retroviral genome and used to control expression from other types of vectors.

As used herein, the term "secretion signal" refers to any DNA sequence which, when operably linked to a recombinant DNA sequence, encodes a signal peptide which is capable of causing the secretion of the recombinant polypeptide. In general, the signal peptides comprise a series of about 15 to 30 hydrophobic amino acid residues (See, e.g., Zwizinski et al., J. Biol. Chem. 255(16): 7973-77 [1980], Gray et al., Gene 39(2): 247-54 [1985], and Martial et al., Science 205: 602-607 [1979]). Such secretion signal sequences are preferably derived from genes encoding polypeptides secreted from the cell type targeted for tissue-specific expression (e.g., secreted milk proteins for expression in and secretion from mammary secretory cells). Secretory DNA sequences, however, are not limited to such sequences. Secretory DNA sequences from proteins secreted from many cell types and organisms may also be used (e.g., the secretion signals for t-PA, serum albumin, lactoferrin, and growth hormone, and secretion signals from microbial genes encoding secreted polypeptides such as from yeast, filamentous fungi, and bacteria).

As used herein, the terms "RNA export element" or "Pre-mRNA Processing Enhancer (PPE)" refer to 3' and 5' cis-acting post-transcriptional regulatory elements that enhance export of RNA from the nucleus. "PPE" elements include, but are not limited to Mertz sequences (described in U.S. Pat. Nos. 5,914,267 and 5,686,120, all of which are incorporated herein by reference) and woodchuck mRNA processing enhancer (WPRE; WO99/14310 and U.S. Pat. No. 6,136,597, each of which is incorporated herein by reference).

As used herein, the term "polycistronic" refers to an mRNA encoding more than polypeptide chain (See, e.g., WO 93/03143, WO 88/05486, and European Pat. No. 117058, all of which are incorporated herein by reference). Likewise, the term "arranged in polycistronic
sequence" refers to the arrangement of genes encoding two different polypeptide chains in a single mRNA.

As used herein, the term "internal ribosome entry site" or "IRES" refers to a sequence located between polycistronic genes that permits the production of the expression product originating from the second gene by internal initiation of the translation of the dicistronic mRNA. Examples of internal ribosome entry sites include, but are not limited to, those derived from foot and mouth disease virus (FDV), encephalomyocarditis virus, poliovirus and RDV (Scheper et al., Biochem. 76: 801-809 [1994]; Meyer et al., J. Virol. 69: 2819-2824 [1995]; Jang et al., 1988, J. Virol. 62: 2636-2643 [1998]; Haller et al., J. Virol. 66: 5075-5086 [1995]).

Vectors incorporating IRES's may be assembled as is known in the art. For example, a retroviral vector containing a polycistronic sequence may contain the following elements in operable association: nucleotide polylinker, gene of interest, an internal ribosome entry site and a mammalian selectable marker or another gene of interest. The polycistronic cassette is situated within the retroviral vector between the 5' LTR and the 3' LTR at a position such that transcription from the 5' LTR promoter transcribes the polycistronic message cassette. The transcription of the polycistronic message cassette may also be driven by an internal promoter (e.g., cytomegalovirus promoter) or an inducible promoter, which may be preferable depending on the use. The polycistronic message cassette can further comprise a cDNA or genomic DNA (gDNA) sequence operatively associated within the polylinker. Any mammalian selectable marker can be utilized as the polycistronic message cassette mammalian selectable marker. Such mammalian selectable markers are well known to those of skill in the art and can include, but are not limited to, kanamycin/G418, hygromycin B or mycophenolic acid resistance markers.

As used herein, the term "retrovirus" refers to a retroviral particle which is capable of entering a cell (i.e., the particle contains a membrane-associated protein such as an envelope protein or a viral G glycoprotein which can bind to the host cell surface and facilitate entry of the viral particle into the cytoplasm of the host cell) and integrating the retroviral genome (as a double-stranded provirus) into the genome of the host cell. The term "retrovirus" encompasses Oncorvirinae (e.g., Moloney murine leukemia virus (MoMOLV), Moloney murine sarcoma virus (MoMSV), and Mouse mammary tumor virus (MMTV), Spumavirinae, and Lentivirinae (e.g., Human immunodeficiency virus, Simian immunodeficiency virus, Equine infection anemia
virus, and Caprine arthritis-encephalitis virus; See, e.g., U.S. Pat. Nos. 5,994,136 and 6,013,516, both of which are incorporated herein by reference).

As used herein, the term "retroviral vector" refers to a retrovirus that has been modified to express a gene of interest. Retroviral vectors can be used to transfer genes efficiently into host cells by exploiting the viral infectious process. Foreign or heterologous genes cloned (i.e., inserted using molecular biological techniques) into the retroviral genome can be delivered efficiently to host cells that are susceptible to infection by the retrovirus. Through well known genetic manipulations, the replicative capacity of the retroviral genome can be destroyed. The resulting replication-defective vectors can be used to introduce new genetic material to a cell but they are unable to replicate. A helper virus or packaging cell line can be used to permit vector particle assembly and egress from the cell. Such retroviral vectors comprise a replication-deficient retroviral genome containing a nucleic acid sequence encoding at least one gene of interest (i.e., a polycistronic nucleic acid sequence can encode more than one gene of interest), a 5' retroviral long terminal repeat (5' LTR); and a 3' retroviral long terminal repeat (3' LTR).

The term "pseudotyped retroviral vector" refers to a retroviral vector containing a heterologous membrane protein. The term "membrane-associated protein" refers to a protein (e.g., a viral envelope glycoprotein or the G proteins of viruses in the Rhabdoviridae family such as VSV, Piry, Chandipura and Mokola) that are associated with the membrane surrounding a viral particle; these membrane-associated proteins mediate the entry of the viral particle into the host cell. The membrane associated protein may bind to specific cell surface protein receptors, as is the case for retroviral envelope proteins or the membrane-associated protein may interact with a phospholipid component of the plasma membrane of the host cell, as is the case for the G proteins derived from members of the Rhabdoviridae family.

The term "heterologous membrane-associated protein" refers to a membrane-associated protein which is derived from a virus that is not a member of the same viral class or family as that from which the nucleocapsid protein of the vector particle is derived. "Viral class or family" refers to the taxonomic rank of class or family, as assigned by the International Committee on Taxonomy of Viruses.

The term "Rhabdoviridae" refers to a family of enveloped RNA viruses that infect animals, including humans, and plants. The Rhabdoviridae family encompasses the genus Vesiculovirus that includes vesicular stomatitis virus (VSV), Cocal virus, Piry virus, Chandipura
virus, and Spring viremia of carp virus (sequences encoding the Spring viremia of carp virus are available under GenBank accession number U18101). The G proteins of viruses in the Vesiculovirus genera are virally-encoded integral membrane proteins that form externally projecting homotrimeric spike glycoproteins complexes that are required for receptor binding and membrane fusion. The G proteins of viruses in the Vesiculovirus genera have a covalently bound palmitic acid (C_{16}) moiety. The amino acid sequences of the G proteins from the Vesiculoviruses are fairly well conserved. For example, the Piry virus G protein share about 38% identity and about 55% similarity with the VSV G proteins (several strains of VSV are known, e.g., Indiana, New Jersey, Orsay, San Juan, etc., and their G proteins are highly homologous). The Chandipura virus G protein and the VSV G proteins share about 37% identity and 52% similarity. Given the high degree of conservation (amino acid sequence) and the related functional characteristics (e.g., binding of the virus to the host cell and fusion of membranes, including syncytia formation) of the G proteins of the Vesiculoviruses, the G proteins from non-VSV Vesiculoviruses may be used in place of the VSV G protein for the pseudotyping of viral particles. The G proteins of the Lyssa viruses (another genera within the Rhabdoviridae family) also share a fair degree of conservation with the VSV G proteins and function in a similar manner (e.g., mediate fusion of membranes) and therefore may be used in place of the VSV G protein for the pseudotyping of viral particles. The Lyssa viruses include the Mokola virus and the Rabies viruses (several strains of Rabies virus are known and their G proteins have been cloned and sequenced). The Mokola virus G protein shares stretches of homology (particularly over the extracellular and transmembrane domains) with the VSV G proteins which show about 31% identity and 48% similarity with the VSV G proteins. Preferred G proteins share at least 25% identity, preferably at least 30% identity and most preferably at least 35% identity with the VSV G proteins. The VSV G protein from which New Jersey strain (the sequence of this G protein is provided in GenBank accession numbers M27165 and M21557) is employed as the reference VSV G protein.

As used herein, the term “lentivirus vector” refers to retroviral vectors derived from the Lentiviridae family (e.g., human immunodeficiency virus, simian immunodeficiency virus, equine infectious anemia virus, and caprine arthritis-encephalitis virus) that are capable of integrating into non-dividing cells (See, e.g., U.S. Pat. Nos. 5,994,136 and 6,013,516, both of which are incorporated herein by reference).
The term “pseudotyped lentivirus vector” refers to lentivirus vector containing a heterologous membrane protein (e.g., a viral envelope glycoprotein or the G proteins of viruses in the Rhabdoviridae family such as VSV, Piry, Chandipura and Mokola).

As used herein, the term “transposon” refers to transposable elements (e.g., Tn5, Tn7, and Tn10) that can move or transpose from one position to another in a genome. In general, the transposition is controlled by a transposase. The term “transposon vector,” as used herein, refers to a vector encoding a nucleic acid of interest flanked by the terminal ends of transposon. Examples of transposon vectors include, but are not limited to, those described in U.S. Pat. Nos. 6,027,722; 5,958,775; 5,968,785; 5,965,443; and 5,719,055, all of which are incorporated herein by reference.

As used herein, the term “adeno-associated virus (AAV) vector” refers to a vector derived from an adeno-associated virus serotype, including without limitation, AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAVX7, etc. AAV vectors can have one or more of the AAV wild-type genes deleted in whole or part, preferably the rep and/or cap genes, but retain functional flanking ITR sequences.

AAV vectors can be constructed using recombinant techniques that are known in the art to include one or more heterologous nucleotide sequences flanked on both ends (5' and 3') with functional AAV ITRs. In the practice of the invention, an AAV vector can include at least one AAV ITR and a suitable promoter sequence positioned upstream of the heterologous nucleotide sequence and at least one AAV ITR positioned downstream of the heterologous sequence. A "recombinant AAV vector plasmid" refers to one type of recombinant AAV vector wherein the vector comprises a plasmid. As with AAV vectors in general, 5' and 3' ITRs flank the selected heterologous nucleotide sequence.

AAV vectors can also include transcription sequences such as polyadenylation sites, as well as selectable markers or reporter genes, enhancer sequences, and other control elements that allow for the induction of transcription. Such control elements are described above.

As used herein, the term "AAV virion" refers to a complete virus particle. An AAV virion may be a wild type AAV virus particle (comprising a linear, single-stranded AAV nucleic acid genome associated with an AAV capsid, i.e., a protein coat), or a recombinant AAV virus particle (described below). In this regard, single-stranded AAV nucleic acid molecules (either the sense/coding strand or the antisense/anticoding strand as those terms are generally defined)
can be packaged into an AAV virion; both the sense and the antisense strands are equally infectious.


Suitable nucleotide sequences for use in AAV vectors (and, indeed, any of the vectors described herein) include any functionally relevant nucleotide sequence. Thus, the AAV vectors of the present invention can comprise any desired gene that encodes a protein that is defective or missing from a target cell genome or that encodes a non-native protein having a desired biological or therapeutic effect (e.g., an antiviral function), or the sequence can correspond to a molecule having an antisense or ribozyme function. Suitable genes include those used for the treatment of inflammatory diseases, autoimmune, chronic and infectious diseases, including such disorders as AIDS, cancer, neurological diseases, cardiovascular disease, hypercholesteremia; various blood disorders including various anemias, thalassemias and hemophilia; genetic defects such as cystic fibrosis, Gaucher's Disease, adenosine deaminase (ADA) deficiency, emphysema, etc. A number of antisense oligonucleotides (e.g., short oligonucleotides complementary to sequences around the translational initiation site (AUG codon) of an mRNA) that are useful in antisense therapy for cancer and for viral diseases have been described in the art. (See, e.g., Han et al., Proc. Natl. Acad. Sci. USA 88:4313-4317 [1991]; Uhlmann et al., Chem. Rev. 90:543-584 [1990]; Helene et al., Biochim. Biophys. Acta. 1049:99-125 [1990]; Agarwal et al., Proc. Natl. Acad. Sci. USA 85:7079-7083 [1989]; and Heikkila et al., Nature 328:445-449 [1987]). For a discussion of suitable ribozymes, see, e.g., Cech et al. (1992) J. Biol. Chem. 267:17479-17482 and U.S. Patent No. 5,225,347, incorporated herein by reference.
By "adeno-associated virus inverted terminal repeats" or "AAV ITRs" is meant the art-recognized palindromic regions found at each end of the AAV genome which function together in cis as origins of DNA replication and as packaging signals for the virus. For use with the present invention, flanking AAV ITRs are positioned 5' and 3' of one or more selected heterologous nucleotide sequences and, together with the rep coding region or the Rep expression product, provide for the integration of the selected sequences into the genome of a target cell.

The nucleotide sequences of AAV ITR regions are known (See, e.g., Kotin, Human Gene Therapy 5:793-801 [1994]; Berns, K.I. "Parvoviridae and their Replication" in Fundamental Virology, 2nd Edition, (B.N. Fields and D.M. Knipe, eds.) for the AAV-2 sequence. As used herein, an "AAV ITR" need not have the wild-type nucleotide sequence depicted, but may be altered, e.g., by the insertion, deletion or substitution of nucleotides. Additionally, the AAV ITR may be derived from any of several AAV serotypes, including without limitation, AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAVX7, etc. The 5' and 3' ITRs which flank a selected heterologous nucleotide sequence need not necessarily be identical or derived from the same AAV serotype or isolate, so long as they function as intended, i.e., to allow for the integration of the associated heterologous sequence into the target cell genome when the rep gene is present (either on the same or on a different vector), or when the Rep expression product is present in the target cell.

As used herein the term, the term "in vitro" refers to an artificial environment and to processes or reactions that occur within an artificial environment. In vitro environments can consist of, but are not limited to, test tubes and cell cultures. The term "in vivo" refers to the natural environment (e.g., an animal or a cell) and to processes or reaction that occur within a natural environment.

As used herein, the term "clonally derived" refers to a cell line that it derived from a single cell.

As used herein, the term "clonally selecting" refers to selecting (e.g., selecting for the presence of a integrated vector) cell lines derived from a single cell.

As used herein, the term "non-clonally derived" refers to a cell line that is derived from more than one cell.
As used herein, the term “passage” refers to the process of diluting a culture of cells that has grown to a particular density or confluency (e.g., 70% or 80% confluent), and then allowing the diluted cells to regrow to the particular density or confluency desired (e.g., by replating the cells or establishing a new roller bottle culture with the cells).

As used herein, the term “stable,” when used in reference to genome, refers to the stable maintenance of the information content of the genome from one generation to the next, or, in the particular case of a cell line, from one passage to the next. Accordingly, a genome is considered to be stable if no gross changes occur in the genome (e.g., a gene is deleted or a chromosomal translocation occurs). The term "stable" does not exclude subtle changes that may occur to the genome such as point mutations.

As used herein, the term “response,” when used in reference to an assay, refers to the generation of a detectable signal (e.g., accumulation of reporter protein, increase in ion concentration, accumulation of a detectable chemical product).

As used herein, the term “membrane receptor protein” refers to membrane spanning proteins that bind a ligand (e.g., a hormone or neurotransmitter). As is known in the art, protein phosphorylation is a common regulatory mechanism used by cells to selectively modify proteins carrying regulatory signals from outside the cell to the nucleus. The proteins that execute these biochemical modifications are a group of enzymes known as protein kinases. They may further be defined by the substrate residue that they target for phosphorylation. One group of protein kinases are the tyrosine kinases (TKs) which selectively phosphorylate a target protein on its tyrosine residues. Some tyrosine kinases are membrane-bound receptors (RTKs), and, upon activation by a ligand, can autophosphorylate as well as modify substrates. The initiation of sequential phosphorylation by ligand stimulation is a paradigm that underlies the action of such effectors as, for example, epidermal growth factor (EGF), insulin, platelet-derived growth factor (PDGF), and fibroblast growth factor (FGF). The receptors for these ligands are tyrosine kinases and provide the interface between the binding of a ligand (hormone, growth factor) to a target cell and the transmission of a signal into the cell by the activation of one or more biochemical pathways. Ligand binding to a receptor tyrosine kinase activates its intrinsic enzymatic activity (See, e.g., Ullrich and Schlessinger, Cell 61:203-212 [1990]). Tyrosine kinases can also be cytoplasmic, non-receptor-type enzymes and act as a downstream component of a signal transduction pathway.
As used herein, the term "signal transduction protein" refers to a protein that is activated or otherwise affected by ligand binding to a membrane receptor protein or some other stimulus. Examples of signal transduction protein include adenyl cyclase, phospholipase C, and G-proteins. Many membrane receptor proteins are coupled to G-proteins (i.e., G-protein coupled receptors (GPCRs); for a review, see Neer, 1995, Cell 80:249-257 [1995]). Typically, GPCRs contain seven transmembrane domains. Putative GPCRs can be identified on the basis of sequence homology to known GPCRs.

GPCRs mediate signal transduction across a cell membrane upon the binding of a ligand to an extracellular portion of a GPCR. The intracellular portion of a GPCR interacts with a G-protein to modulate signal transduction from outside to inside a cell. A GPCR is therefore said to be "coupled" to a G-protein. G-proteins are composed of three polypeptide subunits: an α subunit, which binds and hydrolyses GTP, and a dimeric βγ subunit. In the basal, inactive state, the G-protein exists as a heterotrimer of the α and βγ subunits. When the G-protein is inactive, guanosine diphosphate (GDP) is associated with the α subunit of the G-protein. When a GPCR is bound and activated by a ligand, the GPCR binds to the G-protein heterotrimer and decreases the affinity of the Ga subunit for GDP. In its active state, the G subunit exchanges GDP for guanine triphosphate (GTP) and active Ga subunit disassociates from both the receptor and the dimeric βγ subunit. The disassociated, active Ga subunit transduces signals to effectors that are "downstream" in the G-protein signalling pathway within the cell. Eventually, the G-protein's endogenous GTPase activity returns active G subunit to its inactive state, in which it is associated with GDP and the dimeric βγ subunit.

Numerous members of the heterotrimeric G-protein family have been cloned, including more than 20 genes encoding various Ga subunits. The various G subunits have been categorized into four families, on the basis of amino acid sequences and functional homology. These four families are termed Ga13, Ga14, Gaq, and Gai2. Functionally, these four families differ with respect to the intracellular signaling pathways that they activate and the GPCR to which they couple.

For example, certain GPCRs normally couple with Ga13 and, through Ga13, these GPCRs stimulate adenylyl cyclase activity. Other GPCRs normally couple with Gai13 and through Gai4, these GPCRs can activate phospholipase C (PLC), such as the β isoform of phospholipase C (i.e., PLCβ), Sternweis and Smrcka, Trends in Biochem. Sci. 17:502-506 [1992]).
As used herein, the term “nucleic acid binding protein” refers to proteins that bind to nucleic acid, and in particular to proteins that cause increased (i.e., activators or transcription factors) or decreased (i.e., inhibitors) transcription from a gene.

As used herein, the term “ion channel protein” refers to proteins that control the ingress or egress of ions across cell membranes. Examples of ion channel proteins include, but are not limited to, the Na^+-K^+ ATPase pump, the Ca^{2+} pump, and the K^+ leak channel.

As used herein, the term “protein kinase” refers to proteins that catalyze the addition of a phosphate group from a nucleoside triphosphate to an amino acid side chain in a protein. Kinases comprise the largest known enzyme superfamily and vary widely in their target proteins. Kinases may be categorized as protein tyrosine kinases (PTKs), which phosphorylate tyrosine residues, and protein serine/threonine kinases (STKs), which phosphorylate serine and/or threonine residues. Some kinases have dual specificity for both serine/threonine and tyrosine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain. This domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure that binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a serine, threonine, or tyrosine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. STKs and PTKs also contain distinct sequence motifs in subdomains VI and VIII, which may confer hydroxyamino acid specificity. Some STKs and PTKs possess structural characteristics of both families. In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain.

Non-transmembrane PTKs form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that signal through non-transmembrane PTKs include cytokine, hormone, and antigen-specific lymphocytic receptors. Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine...
phosphorylation activity (See, e.g., Carbonneau, H. and Tonks, Annu. Rev. Cell Biol. 8:463-93 [1992]). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Examples of protein kinases include, but are not limited to, cAMP-dependent protein kinase, protein kinase C, and cyclin-dependent protein kinases (See, e.g., U.S. Pat. Nos. 6,034,228; 6,030,822; 6,030,788; 6,020,306; 6,013,455; 6,013,464; and 6,015,807, all of which are incorporated herein by reference).

As used herein, the term “protein phosphatase” refers to proteins that remove a phosphate group from a protein. Protein phosphatases are generally divided into two groups, receptor and non-receptor type proteins. Most receptor-type protein tyrosine phosphatases contain two conserved catalytic domains, each of which encompasses a segment of 240 amino acid residues (See, e.g., Saito et al., Cell Growth and Diff. 2:59-65 [1991]). Receptor protein tyrosine phosphatases can be subclassified further based upon the amino acid sequence diversity of their extracellular domains (See, e.g., Krueger et al., Proc. Natl. Acad. Sci. USA 89:7417-7421 [1992]). Examples of protein phosphatases include, but are not limited to, cdc25 a, b, and c, PTP20, PTP1D, and PTPL (See, e.g., U.S. Pat. Nos. 5,976,853; 5,994,074; 6,004,791; 5,981,251; 5,976,852; 5,958,719; 5,955,592; and 5,952,212, all of which are incorporated herein by reference).

As used herein, the term “protein encoded by an oncogene” refers to proteins that cause, either directly or indirectly, the neoplastic transformation of a host cell. Examples of oncogenes include, but are not limited to, the following genes: src, fps, fes, fgr, ros, H-ras, abl, ski, erbA, erbB, fns, fos, mos, sis, myc, myb, rel, kit, ras, K-ras, and ets.

As used herein, the term "immunoglobulin" refers to proteins that bind a specific antigen. Immunoglobulins include, but are not limited to, polyclonal, monoclonal, chimeric, and humanized antibodies, Fab fragments, F(ab')2 fragments, and includes immunoglobulins of the following classes: IgG, IgA, IgM, IgD, Ibe, and secreted immunoglobulins (sIg). Immunoglobulins generally comprise two identical heavy chains (y, a, d, or e) and two light chains (k or l).

As used herein, the term "antigen binding protein" refers to proteins that bind to a specific antigen. "Antigen binding proteins" include, but are not limited to, immunoglobulins, including polyclonal, monoclonal, chimeric, and humanized antibodies; Fab fragments, F(ab')2
fragments, and Fab expression libraries; and single chain antibodies. Various procedures known in the art are used for the production of polyclonal antibodies. For the production of an antibody, various host animals can be immunized by injection with the peptide corresponding to the desired epitope including but not limited to rabbits, mice, rats, sheep, goats, etc. In a preferred embodiment, the peptide is conjugated to an immunogenic carrier (e.g., diphtheria toxoid, bovine serum albumin (BSA), or keyhole limpet hemocyanin (KLH)). Various adjuvants are used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (Bacille Calmette-Guerin) and Corynebacterium parvum.

For preparation of monoclonal antibodies, any technique that provides for the production of antibody molecules by continuous cell lines in culture may be used (See, e.g., Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). These include, but are not limited to, the hybridoma technique originally developed by Köhler and Milstein (Köhler and Milstein, Nature 256:495-497 [1975]), as well as the trioma technique, the human B-cell hybridoma technique (See e.g., Kozbor et al. Immunol. Today 4:72 [1983]), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96 [1985]).

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; herein incorporated by reference) can be adapted to produce specific single chain antibodies as desired. An additional embodiment of the invention utilizes the techniques known in the art for the construction of Fab expression libraries (Huse et al., Science 246:1275-1281 [1989]) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibody fragments that contain the idiotype (antigen binding region) of the antibody molecule can be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')2 fragment that can be produced by pepsin digestion of an antibody molecule; the Fab' fragments that can be generated by reducing the disulfide bridges of an
F(ab')2 fragment, and the Fab fragments that can be generated by treating an antibody molecule with papain and a reducing agent.

Genes encoding antigen binding proteins can be isolated by methods known in the art. In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art (e.g., radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, in situ immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), Western Blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.) etc.

As used herein, the term "reporter gene" refers to a gene encoding a protein that may be assayed. Examples of reporter genes include, but are not limited to, luciferase (See, e.g., deWet et al., Mol. Cell. Biol. 7:725 [1987] and U.S. Pat Nos. 6,074,859; 5,976,796; 5,674,713; and 5,618,682; all of which are incorporated herein by reference), green fluorescent protein (e.g., GenBank Accession Number U43284; a number of GFP variants are commercially available from CLONTECH Laboratories, Palo Alto, CA), chloramphenicol acetyltransferase, β-galactosidase, alkaline phosphatase, and horse radish peroxidase.

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An "isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated.

The term "test compound" refers to any chemical entity, pharmaceutical, drug, and the like contemplated to be useful in the treatment and/or prevention of a disease, illness, sickness, or disorder of bodily function, or otherwise alter the physiological or cellular status of a sample. Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the screening methods of the present invention. A "known therapeutic compound" refers to a therapeutic compound that has been shown (e.g., through animal trials or prior experience with administration to humans) to be effective in such treatment or prevention.
DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to the production of proteins in host cells, and more particularly to host cells containing multiple integrated copies of an integrating vector. The present invention utilizes integrating vectors (i.e., vectors that integrate via an integrase or transposase) to create cell lines containing a high copy number of a nucleic acid encoding a gene of interest. The transfected genomes of the high copy number cells are stable through repeated passages (e.g., at least 10 passages, preferably at least 50 passages, and most preferably at least 100 passages). Furthermore, the host cells of the present invention are capable of producing high levels of protein (e.g., more than 1 pg/cell/day, preferably more than 10 pg/cell/day, more preferably more than 50 pg/cell/day, and most preferably more than 100 pg/cell/day).

The genomic stability and high expression levels of the host cells of the present invention provide distinct advantages over previously described methods of cell culture. For example, mammalian cell lines containing multiple copies of genes are known in the art to be intrinsically unstable. Indeed, this instability is a recognized problem facing researchers desiring to use mammalian cell lines for various purposes, including high throughput screening assays (See, e.g., Sittampalam et al., Curr. Opin. Chem. Biol. 1(3):384-91 [1997]).

It is not intended that the present invention be limited to particular mechanism of action. Indeed, an understanding of the mechanism is not necessary to make and use the present invention. However, the high genomic stability and protein expression levels of the host cells of the present invention are thought to be due to unique properties of the integrating vectors (e.g., retroviral vectors). For example, it is known that retroviruses are inherited elements in the germ line of many organisms. Indeed, as much as 5-10% of the mammalian genome may consist of elements contributed by reverse transcription, indicating a high degree of stability. Likewise, many of these types of vectors target active (e.g., DNase I hypersensitive sites) transcriptional sites in the genome.

Many investigations have focused on the deleterious effects of retroviral and transposon integration. The property of targeting active regions of the genome has led to the use of retroviral vectors and transposon vectors in promoter trap schemes and for saturation mutagenesis (See, e.g., U.S. Pat. Nos. 5,627,058 and 5,922,601, all of which are herein incorporated by reference). In promoter trap schemes, the cells are infected with a promoterless reporter vector. If the promoterless vector integrates downstream of a promoter (i.e., into a
gene), the reporter gene encoded by the vector is activated. The promoter can then be cloned and further characterized.

As can be seen, these schemes rely on the disruption of an endogenous gene. Therefore, it is surprising that the methods of the present invention, which utilize integrating vectors at high multiplicities of infection that would normally be thought to lead to gene disruption, led to the development of stable cell lines that express high quantities of a protein of interest. The development of these cell lines is described more fully below. The description is divided into the following sections: I) Host Cells; II) Vectors and Methods of Transfection; and III) Uses of Transfected Host Cells.

I. Host Cells

The present invention contemplates the transfection of a variety of host cells with integrating vectors. A number of mammalian host cell lines are known in the art. In general, these host cells are capable of growth and survival when placed in either monolayer culture or in suspension culture in a medium containing the appropriate nutrients and growth factors, as is described in more detail below. Typically, the cells are capable of expressing and secreting large quantities of a particular protein of interest into the culture medium. Examples of suitable mammalian host cells include, but are not limited to Chinese hamster ovary cells (CHO-K1, ATCC CCL-61); bovine mammary epithelial cells (ATCC CRL 10274; bovine mammary epithelial cells); monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture; see, e.g., Graham et al., J. Gen Virol., 36:59 [1977]); baby hamster kidney cells (BHK, ATCC CCL 10); mouse sertoli cells (TM4, Mather, Biol. Reprod. 23:243-251 [1980]); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather et al., Annals N.Y. Acad. Sci., 383:44-68 [1982]); MRC 5 cells; FS4 cells; rat fibroblasts (208F cells); MDBK cells (bovine kidney cells); and a human hepatoma line (Hep G2).
In addition to mammalian cell lines, the present invention also contemplates the transfection of plant protoplasts with integrating vectors at a low or high multiplicity of infection. For example, the present invention contemplates a plant cell or whole plant comprising at least one integrated integrating vector, preferably a retroviral vector, and most preferably a pseudotyped retroviral vector. All plants that can be produced by regeneration from protoplasts can also be transfected using the process according to the invention (e.g., cultivated plants of the genera Solanum, Nicotiana, Brassica, Beta, Pisum, Phaseolus, Glycine, Helianthus, Allium, Avena, Hordeum, Oryzae, Setaria, Secale, Sorghum, Triticum, Zea, Musa, Cocos, Cydonia, Pyrus, Malus, Phoenix, Elaeis, Rubus, Fragaria, Prunus, Arachis, Panicum, Saccharum, Coffea, Camellia, Ananas, Vitis or Citrus). In general, protoplasts are produced in accordance with conventional methods (See, e.g., U.S. Pat. Nos. 4,743,548; 4,677,066, 5,149,645; and 5,508,184; all of which are incorporated herein by reference). Plant tissue may be dispersed in an appropriate medium having an appropriate osmotic potential (e.g., 3 to 8 wt. % of a sugar polyol) and one or more polysaccharide hydrolases (e.g., pectinase, cellulase, etc.), and the cell wall degradation allowed to proceed for a sufficient time to provide protoplasts. After filtration the protoplasts may be isolated by centrifugation and may then be resuspended for subsequent treatment or use. Regeneration of protoplasts kept in culture to whole plants is performed by methods known in the art (See, e.g., Evans et al., Handbook of Plant Cell Culture, 1: 124-176, MacMillan Publishing Co., New York [1983]; Binding, Plant Protoplasts, p. 21-37, CRC Press, Boca Raton [1985],) and Potrykus and Shillito, Methods in Enzymology, Vol. 118, Plant Molecular Biology, A. and H. Weissbach eds., Academic Press, Orlando [1986]).

The present invention also contemplates the use of amphibian and insect host cell lines. Examples of suitable insect host cell lines include, but are not limited to, mosquito cell lines (e.g., ATCC CRL-1660). Examples of suitable amphibian host cell lines include, but are not limited to, toad cell lines (e.g., ATCC CCL-102).

II. Vectors and Methods for Transfection

According to the present invention, host cells such as those described above are transduced or transfected with integrating vectors. Examples of integrating vectors include, but are not limited to, retroviral vectors, lentiviral vectors, adeno-associated viral vectors, and
transposon vectors. The design, production, and use of these vectors in the present invention is described below.

A. Retroviral Vectors

Retroviruses (family Retroviridae) are divided into three groups: the spumaviruses (e.g., human foamy virus); the lentiviruses (e.g., human immunodeficiency virus and sheep visna virus) and the oncoviruses (e.g., MLV, Rous sarcoma virus).

Retroviruses are enveloped (i.e., surrounded by a host cell-derived lipid bilayer membrane) single-stranded RNA viruses which infect animal cells. When a retrovirus infects a cell, its RNA genome is converted into a double-stranded linear DNA form (i.e., it is reverse transcribed). The DNA form of the virus is then integrated into the host cell genome as a provirus. The provirus serves as a template for the production of additional viral genomes and viral mRNAs. Mature viral particles containing two copies of genomic RNA bud from the surface of the infected cell. The viral particle comprises the genomic RNA, reverse transcriptase and other pol gene products inside the viral capsid (which contains the viral gag gene products), which is surrounded by a lipid bilayer membrane derived from the host cell containing the viral envelope glycoproteins (also referred to as membrane-associated proteins).

The organization of the genomes of numerous retroviruses is well known to the art and this has allowed the adaptation of the retroviral genome to produce retroviral vectors. The production of a recombinant retroviral vector carrying a gene of interest is typically achieved in two stages.

First, the gene of interest is inserted into a retroviral vector which contains the sequences necessary for the efficient expression of the gene of interest (including promoter and/or enhancer elements which may be provided by the viral long terminal repeats (LTRs) or by an internal promoter/enhancer and relevant splicing signals), sequences required for the efficient packaging of the viral RNA into infectious virions (e.g., the packaging signal (Psi), the tRNA primer binding site (-PBS), the 3' regulatory sequences required for reverse transcription (+PBS)) and the viral LTRs. The LTRs contain sequences required for the association of viral genomic RNA, reverse transcriptase and integrase functions, and sequences involved in directing the expression of the genomic RNA to be packaged in viral particles. For safety reasons, many recombinant retroviral vectors lack functional copies of the genes that are essential for viral replication (these
essential genes are either deleted or disabled); therefore, the resulting virus is said to be replication defective.

Second, following the construction of the recombinant vector, the vector DNA is introduced into a packaging cell line. Packaging cell lines provide proteins required in trans for the packaging of the viral genomic RNA into viral particles having the desired host range (i.e., the viral-encoded gag, pol and env proteins). The host range is controlled, in part, by the type of envelope gene product expressed on the surface of the viral particle. Packaging cell lines may express ecotropic, amphotropic or xenotropic envelope gene products. Alternatively, the packaging cell line may lack sequences encoding a viral envelope (env) protein. In this case the packaging cell line will package the viral genome into particles that lack a membrane-associated protein (e.g., an env protein). In order to produce viral particles containing a membrane associated protein that will permit entry of the virus into a cell, the packaging cell line containing the retroviral sequences is transfected with sequences encoding a membrane-associated protein (e.g., the G protein of vesicular stomatitis virus (VSV)). The transfected packaging cell will then produce viral particles, which contain the membrane-associated protein expressed by the transfected packaging cell line; these viral particles, which contain viral genomic RNA derived from one virus encapsidated by the envelope proteins of another virus are said to be pseudotyped virus particles.

The retroviral vectors of the present invention can be further modified to include additional regulatory sequences. As described above, the retroviral vectors of the present invention include the following elements in operable association: a) a 5' LTR; b) a packaging signal; c) a 3' LTR and d) a nucleic acid encoding a protein of interest located between the 5' and 3' LTRs. In some embodiments of the present invention, the nucleic acid of interest may be arranged in opposite orientation to the 5' LTR when transcription from an internal promoter is desired. Suitable internal promoters include, but are not limited to, the alpha-lactalbumin promoter, the CMV promoter (human or ape), and the thymidine kinase promoter.

In other embodiments of the present invention, where secretion of the protein of interest is desired, the vectors are modified by including a signal peptide sequence in operable association with the protein of interest. The sequences of several suitable signal peptides are known to those in the art, including, but not limited to, those derived from tissue plasminogen activator, human growth hormone, lactoferrin, alpha-casein, and alpha-lactalbumin.
In other embodiments of the present invention, the vectors are modified by incorporating an RNA export element (See, e.g., U.S. Pat. Nos. 5,914,267; 6,136,597; and 5,686,120; and WO99/14310, all of which are incorporated herein by reference) either 3' or 5' to the nucleic acid sequence encoding the protein of interest. It is contemplated that the use of RNA export elements allows high levels of expression of the protein of interest without incorporating splice signals or introns in the nucleic acid sequence encoding the protein of interest.

In still other embodiments, the vector further comprises at least one internal ribosome entry site (IRES) sequence. The sequences of several suitable IRES's are available, including, but not limited to, those derived from foot and mouth disease virus (FDV), encephalomyocarditis virus, and poliovirus. The IRES sequence can be interposed between two transcriptional units (e.g., nucleic acids encoding different proteins of interest or subunits of a multisubunit protein such as an antibody) to form a polycistronic sequence so that the two transcriptional units are transcribed from the same promoter.

The retroviral vectors of the present invention may also further comprise a selectable marker allowing selection of transformed cells. A number of selectable markers find use in the present invention, including, but not limited to the bacterial aminoglycoside 3' phosphotransferase gene (also referred to as the neo gene) that confers resistance to the drug G418 in mammalian cells, the bacterial hygromycin G phosphotransferase (hyg) gene that confers resistance to the antibiotic hygromycin and the bacterial xanthine-guanine phosphoribosyl transferase gene (also referred to as the gpt gene) that confers the ability to grow in the presence of mycophenolic acid. In some embodiments, the selectable marker gene is provided as part of polycistronic sequence that also encodes the protein of interest.

In some embodiments, the retroviral vectors further comprise an amplifiable marker. Suitable amplifiable markers include, but are not limited to, the genes encoding dihydrofolate reductase (DHFR) and glutamine synthetase (GS). These genes are described in U.S. Pat. Nos. 5,770,359; 5,827,739; 4,399,216; 4,634,665; 5,149,636; and 6,455,275; all of which are incorporated herein by reference. In some embodiments, these genes replace the neo or hyg gene in the vectors disclosed herein in the examples and figures (See Figures 24 and 25 for sequences of vectors comprising DHFR (SEQ ID NO:40) and GS (SEQ ID NO:41) selectable markers, respectively). In embodiments, where amplifiable markers are utilized, it is contemplated that culture of transduced host cells in a medium comprising an inhibitor of the gene. Suitable
inhibitors include, but are not limited to methotrexate for inhibition of DHFR and methionine sulfoximine (Msx) or phosphinothricin for inhibition of GS. It is contemplated that as concentrations of these inhibitors are increased in a cell culture system, cells with higher copy numbers of the amplifiable marker (and thus the genes or genes of interest) are selected. Thus, the genes are amplified.

In some embodiments, the amplifiable marker system is used in conjunction with the introduction of multiple retroviral vectors through transduction at a high multiplicity of infection and/or by serial transductions. In some of these embodiments, cells that are transduced are cultured in amounts of inhibitor that allow for selection of cells with multiple integrated retroviral vectors. Thus, the present invention provides methods for selecting cells in which multiple copies of a vector have integrated in the substantial absence of amplification of the integrated provirus by duplication of regions of the chromosome(s) containing the provirus. In other embodiments, the integrated proviruses are amplified by selection in increasing concentrations of the inhibitor. The present invention is not limited to any particular mechanism of action. Indeed, an understanding of the mechanism of action is not necessary to practice the present invention. Nevertheless, as previously described, when vectors such as plasmids are used to create cell lines, they are often inserted into a chromosome as a series of head to tail repeats. It is believed that this multiple repeat segment is inherently unstable, and that when this region is amplified (e.g., in a DHFR or GS selection system) the resulting amplified segments are inherently unstable. The present invention solves this problem by using retroviral vectors to introduce the amplifiable genes into the host cells. When this introduction is performed at high multiplicities of infections and/or in a serial manner, multiple copies of the retroviral vector are introduced into multiple chromosomes in a stable manner. Thus, when these stable regions are amplified, the resulting cell line is stable.

In still other embodiments of the present invention, the retroviral vectors may comprise recombination elements recognized by a recombination system (e.g., the cre/loxP or flip recombinase systems, see, e.g., Hoess et al., Nucleic Acids Res. 14:2287-2300 [1986], O'Gorman et al., Science 251:1351-55 [1991], van Deursen et al., Proc. Natl. Acad. Sci. USA 92:7376-80 [1995], and U.S. Pat. No. 6,025,192, herein incorporated by reference). After integration of the vectors into the genome of the host cell, the host cell can be transiently transfected (e.g., by electroporation, lipofection, or microinjection) with either a recombinase
enzyme (e.g., Cre recombinase) or a nucleic acid sequence encoding the recombinase enzyme and one or more nucleic acid sequences encoding a protein of interest flanked by sequences recognized by the recombination enzyme so that the nucleic acid sequence is inserted into the integrated vector.

Viral vectors, including recombinant retroviral vectors, provide a more efficient means of transferring genes into cells as compared to other techniques such as calcium phosphate-DNA co-precipitation or DEAE-dextran-mediated transfection, electroporation or microinjection of nucleic acids. It is believed that the efficiency of viral transfer is due in part to the fact that the transfer of nucleic acid is a receptor-mediated process (i.e., the virus binds to a specific receptor protein on the surface of the cell to be infected). In addition, the virally transferred nucleic acid once inside a cell integrates in a controlled manner in contrast to the integration of nucleic acids which are not virally transferred; nucleic acids transferred by other means such as calcium phosphate-DNA co-precipitation are subject to rearrangement and degradation.

The most commonly used recombinant retroviral vectors are derived from the amphotropic Moloney murine leukemia virus (MoMLV) (See e.g., Miller and Baltimore Mol. Cell. Biol. 6:2895 [1986]). The MoMLV system has several advantages: 1) this specific retrovirus can infect many different cell types, 2) established packaging cell lines are available for the production of recombinant MoMLV viral particles and 3) the transferred genes are permanently integrated into the target cell chromosome. The established MoMLV vector systems comprise a DNA vector containing a small portion of the retroviral sequence (e.g., the viral long terminal repeat or "LTR" and the packaging or "psi" signal) and a packaging cell line. The gene to be transferred is inserted into the DNA vector. The viral sequences present on the DNA vector provide the signals necessary for the insertion or packaging of the vector RNA into the viral particle and for the expression of the inserted gene. The packaging cell line provides the proteins required for particle assembly (Markowitz et al., J. Virol. 62:1120 [1988]).

Despite these advantages, existing retroviral vectors based upon MoMLV are limited by several intrinsic problems: 1) they do not infect non-dividing cells (Miller et al., Mol. Cell. Biol. 10:4239 [1990]), except, perhaps, oocytes; 2) they produce low titers of the recombinant virus (Miller and Rosman, BioTechniques 7: 980 [1980] and Miller, Nature 357: 455 [1990]); and 3) they infect certain cell types (e.g., human lymphocytes) with low efficiency (Adams et al., Proc. Natl. Acad. Sci. USA 89:8981 [1992]). The low titers associated with MoMLV-based vectors
have been attributed, at least in part, to the instability of the virus-encoded envelope protein. Concentration of retrovirus stocks by physical means (e.g., ultracentrifugation and ultrafiltration) leads to a severe loss of infectious virus.

The low titer and inefficient infection of certain cell types by MoMLV-based vectors has been overcome by the use of pseudotyped retroviral vectors, which contain the G protein of VSV as the membrane associated protein. Unlike retroviral envelope proteins that bind to a specific cell surface protein receptor to gain entry into a cell, the VSV G protein interacts with a phospholipid component of the plasma membrane (Mastromarino et al., J. Gen. Virol. 68:2359 [1977]). Because entry of VSV into a cell is not dependent upon the presence of specific protein receptors, VSV has an extremely broad host range. Pseudotyped retroviral vectors bearing the VSV G protein have an altered host range characteristic of VSV (i.e., they can infect almost all species of vertebrate, invertebrate and insect cells). Importantly, VSV G-pseudotyped retroviral vectors can be concentrated 2000-fold or more by ultracentrifugation without significant loss of infectivity (Burns et al. Proc. Natl. Acad. Sci. USA 90:8033 [1993]).

The present invention is not limited to the use of the VSV G protein when a viral G protein is employed as the heterologous membrane-associated protein within a viral particle (See, e.g., U.S. Pat. No. 5,512,421, which is incorporated herein by reference). The G proteins of viruses in the Vesiculovirus genera other than VSV, such as the Piry and Chandipura viruses, that are highly homologous to the VSV G protein and, like the VSV G protein, contain covalently linked palmitic acid (Brun et al. Intervirol. 38:274 [1995] and Masters et al., Virol. 171:285 (1990)). Thus, the G protein of the Piry and Chandipura viruses can be used in place of the VSV G protein for the pseudotyping of viral particles. In addition, the VSV G proteins of viruses within the Lyssa virus genera such as Rabies and Mokola viruses show a high degree of conservation (amino acid sequence as well as functional conservation) with the VSV G proteins.

For example, the Mokola virus G protein has been shown to function in a manner similar to the VSV G protein (i.e., to mediate membrane fusion) and therefore may be used in place of the VSV G protein for the pseudotyping of viral particles (Mebatsion et al., J. Virol. 69:1444 [1995]). Viral particles may be pseudotyped using either the Piry, Chandipura or Mokola G protein as described in Example 2, with the exception that a plasmid containing sequences encoding either the Piry, Chandipura or Mokola G protein under the transcriptional control of a suitable promoter element (e.g., the CMV intermediate-early promoter; numerous expression
vectors containing the CMV IE promoter are available, such as the pcDNA3.1 vectors (Invitrogen) is used in place of pHCMV-G. Sequences encoding other G proteins derived from other members of the Rhabdoviridae family may be used; sequences encoding numerous rhabdoviral G proteins are available from the GenBank database.

The majority of retroviruses can transfer or integrate a double-stranded linear form of the virus (the provirus) into the genome of the recipient cell only if the recipient cell is cycling (i.e., dividing) at the time of infection. Retroviruses that have been shown to infect dividing cells exclusively, or more efficiently, include MLV, spleen necrosis virus, Rous sarcoma virus and human immunodeficiency virus (HIV; while HIV infects dividing cells more efficiently, HIV can infect non-dividing cells).

It has been shown that the integration of MLV virus DNA depends upon the host cell's progression through mitosis and it has been postulated that the dependence upon mitosis reflects a requirement for the breakdown of the nuclear envelope in order for the viral integration complex to gain entry into the nucleus (Roe et al., EMBO J. 12:2099 [1993]). However, as integration does not occur in cells arrested in metaphase, the breakdown of the nuclear envelope alone may not be sufficient to permit viral integration; there may be additional requirements such as the state of condensation of the genomic DNA (Roe et al., supra).

**B. Lentiviral Vectors**

The present invention also contemplates the use of lentiviral vectors to generate high copy number cell lines. The lentiviruses (e.g., equine infectious anemia virus, caprine arthritis-encephalitis virus, human immunodeficiency virus) are a subfamily of retroviruses that are able to integrate into non-dividing cells. The lentiviral genome and the proviral DNA have the three genes found in all retroviruses: gag, pol, and env, which are flanked by two LTR sequences. The gag gene encodes the internal structural proteins (e.g., matrix, capsid, and nucleocapsid proteins); the pol gene encodes the reverse transcriptase, protease, and integrase proteins; and the pol gene encodes the viral envelope glycoproteins. The 5' and 3' LTRs control transcription and polyadenylation of the viral RNAs. Additional genes in the lentiviral genome include the vif, vpr, tat, rev, vpu, nef, and vpx genes.

A variety of lentiviral vectors and packaging cell lines are known in the art and find use in the present invention (See, e.g., U.S. Pat. Nos. 5,994,136 and 6,013,516, both of which are
herein incorporated by reference). Furthermore, the VSV G protein has also been used to pseudotype retroviral vectors based upon the human immunodeficiency virus (HIV) (Naldini et al., Science 272:263 [1996]). Thus, the VSV G protein may be used to generate a variety of pseudotyped retroviral vectors and is not limited to vectors based on MoMLV. The lentiviral vectors may also be modified as described above to contain various regulatory sequences (e.g., signal peptide sequences, RNA export elements, and IRES's). After the lentiviral vectors are produced, they may be used to transfect host cells as described above for retroviral vectors.

C. Adeno-Associated Viral Vectors

The present invention also contemplates the use of adeno-associated virus (AAV) vectors to generate high copy number cell lines. AAV is a human DNA parvovirus, which belongs to the genus Dependovirus. The AAV genome is composed of a linear, single-stranded DNA molecule that contains approximately 4680 bases. The genome includes inverted terminal repeats (ITRs) at each end that function in cis as origins of DNA replication and as packaging signals for the virus. The internal nonrepeated portion of the genome includes two large open reading frames, known as the AAV rep and cap regions, respectively. These regions code for the viral proteins involved in replication and packaging of the virion. A family of at least four viral proteins are synthesized from the AAV rep region, Rep 78, Rep 68, Rep 52 and Rep 40, named according to their apparent molecular weight. The AAV cap region encodes at least three proteins, VP1, VP2 and VP3 (for a detailed description of the AAV genome, see e.g., Muzyczka, Current Topics Microbiol. Immunol. 158:97-129 [1992]; Kotin, Human Gene Therapy 5:793-801 [1994]).

AAV requires coinfection with an unrelated helper virus, such as adenovirus, a herpesvirus or vaccinia, in order for a productive infection to occur. In the absence of such coinfection, AAV establishes a latent state by insertion of its genome into a host cell chromosome. Subsequent infection by a helper virus rescues the integrated copy, which can then replicate to produce infectious viral progeny. Unlike the non-pseudotyped retroviruses, AAV has a wide host range and is able to replicate in cells from any species so long as there is coinfection with a helper virus that will also multiply in that species. Thus, for example, human AAV will replicate in canine cells coinfected with a canine adenovirus. Furthermore, unlike the retroviruses, AAV is not associated with any human or animal disease, does not appear to alter
the biological properties of the host cell upon integration and is able to integrate into nondividing cells. It has also recently been found that AAV is capable of site-specific integration into a host cell genome.


Recombinant AAV virions can be produced in a suitable host cell that has been transfected with both an AAV helper plasmid and an AAV vector. An AAV helper plasmid generally includes AAV rep and cap coding regions, but lacks AAV ITRs. Accordingly, the helper plasmid can neither replicate nor package itself. An AAV vector generally includes a selected gene of interest bounded by AAV ITRs that provide for viral replication and packaging functions. Both the helper plasmid and the AAV vector bearing the selected gene are introduced into a suitable host cell by transient transfection. The transfected cell is then infected with a helper virus, such as an adenovirus, which transactivates the AAV promoters present on the helper plasmid that direct the transcription and translation of AAV rep and cap regions.

Recombinant AAV virions harboring the selected gene are formed and can be purified from the preparation. Once the AAV vectors are produced, they may be used to transfect (See, e.g., U.S. Pat. 5,843,742, herein incorporated by reference) host cells at the desired multiplicity of infection to produce high copy number host cells. As will be understood by those skilled in the art, the AAV vectors may also be modified as described above to contain various regulatory sequences (e.g., signal peptide sequences, RNA export elements, and IRES's).

D. Transposon Vectors

The present invention also contemplates the use of transposon vectors to generate high copy number cell lines. Transposons are mobile genetic elements that can move or transpose from one location another in the genome. Transposition within the genome is controlled by a transposase enzyme that is encoded by the transposon. Many examples of transposons are
known in the art, including, but not limited to, Tn5 (See e.g., de la Cruz et al., J. Bact. 175: 6932-38 [1993], Tn7 (See e.g., Craig, Curr. Topics Microbiol. Immunol. 204: 27-48 [1996]), and Tn10 (See e.g., Morisato and Kleckner, Cell 51:101-111 [1987]). The ability of transposons to integrate into genomes has been utilized to create transposon vectors (See, e.g., U.S. Pat. Nos. 5,719,055; 5,968,785; 5,958,775; and 6,027,722; all of which are incorporated herein by reference.) Because transposons are not infectious, transposon vectors are introduced into host cells via methods known in the art (e.g., electroporation, lipofection, or microinjection). Therefore, the ratio of transposon vectors to host cells may be adjusted to provide the desired multiplicity of infection to produce the high copy number host cells of the present invention.

Transposon vectors suitable for use in the present invention generally comprise a nucleic acid encoding a protein of interest interposed between two transposon insertion sequences. Some vectors also comprise a nucleic acid sequence encoding a transposase enzyme. In these vectors, the one of the insertion sequences is positioned between the transposase enzyme and the nucleic acid encoding the protein of interest so that it is not incorporated into the genome of the host cell during recombination. Alternatively, the transposase enzyme may be provided by a suitable method (e.g., lipofection or microinjection). As will be understood by those skilled in the art, the transposon vectors may also be modified as described above to contain various regulatory sequences (e.g., signal peptide sequences, RNA export elements, and IRES’s).

E. Transfection at High Multiplicities of Infection

Once integrating vectors (e.g., retroviral vectors) encoding a protein of interest have been produced, they may be used to transfect or transduce host cells (examples of which are described above in Section I). Preferably, host cells are transfected or transduced with integrating vectors at a multiplicity of infection sufficient to result in the integration of at least 1, and preferably at least 2 or more retroviral vectors. In some embodiments, multiplicities of infection of from 10 to 1,000,000 may be utilized, so that the genomes of the infected host cells contain from 2 to 100 copies of the integrated vectors, and preferably from 5 to 50 copies of the integrated vectors. In other embodiments, a multiplicity of infection of from 10 to 10,000 is utilized. When non-pseudotyped retroviral vectors are utilized for infection, the host cells are incubated with the culture medium from the retroviral producers cells containing the desired titer (i.e., colony forming units, CFUs) of infectious vectors. When pseudotyped retroviral vectors are utilized, the
vectors are concentrated to the appropriate titer by ultracentrifugation and then added to the host cell culture. Alternatively, the concentrated vectors can be diluted in a culture medium appropriate for the cell type. Additionally, when expression of more than one protein of interest by the host cell is desired, the host cells can be transfected with multiple vectors each containing a nucleic acid encoding a different protein of interest.

In each case, the host cells are exposed to medium containing the infectious retroviral vectors for a sufficient period of time to allow infection and subsequent integration of the vectors. In general, the amount of medium used to overlay the cells should be kept to as small a volume as possible so as to encourage the maximum amount of integration events per cell. As a general guideline, the number of colony forming units (cfu) per milliliter should be about $10^5$ to $10^7$ cfu/ml, depending upon the number of integration events desired.

The present invention is not limited to any particular mechanism of action. Indeed, an understanding of the mechanism of action is not necessary for practicing the present invention. However, the diffusion rate of the vectors is known to be very limited (See, e.g., U.S. Pat. No. 5,866,400, herein incorporated by reference, for a discussion of diffusion rates). Therefore, it is expected that the actual integration rate will be lower (and in some cases much lower) than the multiplicity of infection. Applying the equations from U.S Pat. No. 5,866,400, a titer of $10^6$ cfu/ml has an average vector-vector spacing of 1 micron. The diffusion time of a MMLV vector across 100 microns is approximately 20 minutes. Accordingly, the vector can travel approximately 300 microns in one hour. If 1000 cells are plated in a T25 flask, the cells are spaced 2.5 mm apart on average. Using these values, the only 56 viral particles would be expected to contact a given cell within an hour. The Table below provides the expected contact rate for a given number of cells in a T25 flask with a particular vector titer. However, as shown below in the examples, the actual number of integrations obtained is much lower than may be predicted by these equations.

<table>
<thead>
<tr>
<th>Vector Titer</th>
<th>Cells/T25 Flask</th>
<th>MOI</th>
<th>Contacts/Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^6$</td>
<td>1000</td>
<td>1,000</td>
<td>56</td>
</tr>
<tr>
<td>$10^5$</td>
<td>100</td>
<td>10,000</td>
<td>&lt;56</td>
</tr>
<tr>
<td>$10^3$</td>
<td>1000</td>
<td>100</td>
<td>5.6</td>
</tr>
</tbody>
</table>
Accordingly, it is contemplated that the actual integration rate is dependent not only on the multiplicity of infection, but also on the contact time (i.e., the length of time the host cells are exposed to infectious vector), the confluency or geometry of the host cells being transfected, and the volume of media that the vectors are contained in. It is contemplated that these conditions can be varied as taught herein to produce host cell lines containing multiple integrated copies of integrating vectors. As demonstrated in Examples 8 and 9, MOI can be varied by either holding the number of cells constant and varying CFU's (Example 9), or by holding CFU's constant and varying cell number (Example 8).

In some embodiments, after transfection or transduction, the cells are allowed to multiply, and are then trypsinized and replated. Individual colonies are then selected to provide clonally selected cell lines. In still further embodiments, the clonally selected cell lines are screened by Southern blotting or INVADER assay to verify that the desired number of integration events has occurred. It is also contemplated that clonal selection allows the identification of superior protein producing cell lines. In other embodiments, the cells are not clonally selected following transfection.

In some embodiments, the host cells are transfected with vectors encoding different proteins of interest. The vectors encoding different proteins of interest can be used to transfect the cells at the same time (e.g., the host cells are exposed to a solution containing vectors encoding different proteins of interest) or the transfection can be serial (e.g., the host cells are first transfected with a vector encoding a first protein of interest, a period of time is allowed to pass, and the host cells are then transfected with a vector encoding a second protein of interest). In some preferred embodiments, the host cells are transfected with an integrating vector encoding a first protein of interest, high expressing cell lines containing multiple integrated copies of the integrating vector are selected (e.g., clonally selected), and the selected cell line is transfected with an integrating vector encoding a second protein of interest. This process may be repeated to introduce multiple proteins of interest. In some embodiments, the multiplicities of infection may be manipulated (e.g., increased or decreased) to increase or decrease the expression of the protein of interest. Likewise, the different promoters may be utilized to vary the expression of the proteins of interest. It is contemplated that these transfection methods can
be used to construct host cell lines containing an entire exogenous metabolic pathway or to provide host cells with an increased capability to process proteins (e.g., the host cells can be provided with enzymes necessary for post-translational modification).

In still further embodiments, cell lines are serially transfected with vectors encoding the same gene. In some preferred embodiments, the host cells are transfected (e.g., at an MOI of about 10 to 100,000, preferably 100 to 10,000) with an integrating vector encoding a protein of interest, cell lines containing single or multiple integrated copies of the integrating vector or expressing high levels of the desired protein are selected (e.g., clonally selected), and the selected cell line is retransfected with the vector (e.g., at an MOI of about 10 to 100,000, preferably 100 to 10,000). In some embodiments, cell lines comprising at least two integrated copies of the vector are identified and selected. This process may be repeated multiple times until the desired level of protein expression is obtained and may also be repeated to introduce vectors encoding multiple proteins of interest. Unexpectedly, serial transfection with the same gene results in increases in protein production from the resulting cells that are not merely additive.

The present invention contemplates a variety of serial transfection procedures. In some embodiments, where retroviral vectors are utilized, serial transduction procedures are provided. In preferred embodiments, serial transduction is carried out on a pool of cells. In these embodiments, an initial pool of host cells is contacted with retroviral vectors, preferably at a multiplicity of infection ranging from about 0.5 to about 1000 vectors/host cell. The cells are then cultured for several days in an appropriate medium (e.g., with a selection agent such as neomycin). An aliquot of the cells is then taken to determine the number of integrated vectors and to freeze for future possible use. The remaining cells are then recontacted with retroviral vectors, again preferably at a multiplicity of infection ranging from about 0.5 to about 1000 vectors/host cell. This process is repeated until cells with a desired number if integrated vectors are obtained. For example, the process can be repeated up to 10 to 20 or more times. In some embodiments, cells can be clonally selected after any particular transduction step if so desired, however, utilizing a pool of cells in the absence of transduction results in a decreased time to the desired integrated vector copy number.

In some embodiments, the retroviral vectors are produced by the Vector Initial Production process. In this process, cells containing gag and pol genes (e.g., 293GP cells) are
co-transfected with a vector or vectors encoding a retroviral backbone comprising the gene or genes of interest and an envelope protein (e.g., VSV-G protein). These cells produce vector that can be optionally concentrated and then used to transduce host cells. In alternative embodiments, vectors are produced by transducing a cell line comprising retroviral gag and pol genes (e.g., 293-GP cells) with a retroviral vector comprising the gene of interest. This cell line is then transfected with a plasmid encoding the desired env protein (e.g., VSV-G protein). Combinations of these two approaches may also be used.

Following the serial transduction process, cell lines are clonally selected and analyzed for integrated vector copy number and protein production characteristics. Superior cell lines are chosen and stored in a master cell bank.

F. Transfection in the Absence of Selectable Markers

In some embodiments, the present invention provides methods of transfecting host cells with integrating vectors lacking selectable markers. Experiments conducted during the course of development of the present invention (Example 26) demonstrated that vectors lacking selectable markers and grown in selection-free media resulted in higher levels of protein expression at the same vector copy number than vectors comprising selectable markers. In some embodiments, host cells comprising integrated vectors comprising an exogenous gene and lacking a selectable marker express at least 20%, preferably at least 30%, even more preferably, at least 50%, and still more preferably, at least 60% more protein than a host cell with the same number of integrated vectors than contain selectable markers.

In some embodiments, host cell lines derived from integrating vectors comprising an exogenous gene and lacking a selectable marker are clonally selected for the presence of the exogenous gene of interest. In preferred embodiments, selection is performed via clonal analysis of individual cells. In preferred embodiments, expression of a protein of interest is detected directly. For example, in some embodiments, selection is performed via an immunoassay (e.g., an ELISA assay) with an antibody specific for the protein of interest. In other embodiments (e.g., those where the protein of interest is an enzyme) proteins are detected via a biochemical assay (e.g., via the altering of the substrate of an enzyme).

In other embodiments, nucleic acid encoding the protein of interest is detected. For example, in some embodiments, a PCR assay is performed using primers specific for the protein
of interest. In other embodiments, nucleic acid is detected via a hybridization assay (e.g.,
including, but not limited to, Southern Blot, Northern Blot, INVADER Assay (Third Wave
Technologies, Madison, WI), TaqMan assay (Applied Biosystems, Foster City, CA), and SNP-IT
primer extension assay (Orchid Biosciences, Princeton, NJ).

III. Uses of Transfected Host Cells

The host cells transfected at a high multiplicity of infection can be used for a variety of
purposes. First, the host cells find use in the production of proteins for pharmaceutical,
industrial, diagnostic, and other purposes. Second, host cells expressing a particular protein or
proteins find use in screening assays (e.g., high throughput screening). Third, the host cells find
use in the production of multiple variants of proteins, followed by analysis of the activity of the
protein variants. Each of these uses is explained in more detail below.

A. Production of Proteins

It is contemplated that the host cells of the present invention find use in the production of
proteins for pharmaceutical, industrial, diagnostic, and other uses. The present invention is not
limited to the production of any particular protein. Indeed, the production of a wide variety of
proteins is contemplated, including, but not limited to, erythropoietin, alpha-interferon, alpha-1
proteinase inhibitor, angiogenin, antithrombin III, beta-acid decarboxylase, human growth
hormone, bovine growth hormone, porcine growth hormone, human serum albumin, beta-
interferon, calf intestine alkaline phosphatase, cystic fibrosis transmembrane regulator, Factor
VIII, Factor IX, Factor X, insulin, lactoferrin, tissue plasminogen activator, myelin basic protein,
insulin, proinsulin, prolactin, hepatitis B antigen, immunoglobulins, monoclonal antibody
CTLA4 Ig, Tag 72 monoclonal antibody, Tag 72 single chain antigen binding protein, protein C,
cytokines and their receptors, including, for instance tumor necrosis factors alpha and beta, their
receptors and their derivatives; renin; growth hormone releasing factor; parathyroid hormone;
thyroid stimulating hormone; lipoproteins; alpha-1-antitrypsin; follicle stimulating hormone;
calcitonin; luteinizing hormone; glucagon; von Willebrands factor; atrial natriuretic factor; lung
surfactant; urokinase; bombesin; thrombin; hemopoietic growth factor; enkephalinase; human
macrophage inflammatory protein (MIP-1-alpha); a serum albumin such Mullerian-inhibiting
substance; relaxin A-chain; relaxin B-chain; prorelaxin; mouse gonadotropin-associated peptide;
beta-lactamase; DNase; inhibin; activin; vascular endothelial growth factor (VEGF); receptors for hormones or growth factors; integrin; protein A or D; rheumatoid factors; a neurotrophic factor such as bone-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), or a nerve growth factor such as NGF-beta; platelet-derived growth factor (PDGF); fibroblast growth factor such as aFGF and bFGF; epidermal growth factor (EGF); transforming growth factor (TGF) such as TGF-alpha and TGF-beta, including TGF-β1, TGF-β2, TGF-β3, TGF-β4, or TGF-β5; insulin-like growth factor-I and -II (IGF-I and IGF-II); des(1-3)-IGF-I (brain IGF-I), insulinslike growth factor binding proteins; CD proteins such as CD-3, CD-4, CD-8, and CD-19; osteoinductive factors; immunotoxins; a bone morphogenetic protein (BMP); an interferon such as interferon-alpha, -beta, and -gamma; colony stimulating factors (CSFs), e.g., M-CSF, GM-CSF, and G-CSF; interleukins (ILs), e.g., IL-1 to IL-10; superoxide dismutase; T-cell receptors; surface membrane proteins; decay accelerating factor; viral antigen such as, for example, a portion of the AIDS envelope; transport proteins; homing receptors; addressins; regulatory proteins; antibodies; chimeric proteins, such as immunoadhesins, and fragments of any of the above-listed polypeptides. Nucleic acid and protein sequences for these proteins are available in public databases such as GenBank.

In some embodiments, the host cells express more than one exogenous protein. For example, the host cells may be transfected vectors encoding different proteins of interest (e.g., cotransfection or infection at a multiplicity of infection of 1000 with one vector encoding a first protein of interest and a second vector encoding a second protein of interest or serial transfection or infection) so that the host cell contains at least one integrated copy of a first vector encoding a first protein of interest and at least one integrated copy of second integrating vector encoding a second protein of interest. In other embodiments, more than one protein is expressed by arranging the nucleic acids encoding the different proteins of interest in a polycistronic sequence (e.g., bicistronic or tricistronic sequences). This arrangement is especially useful when expression of the different proteins of interest in about a 1:1 molar ratio is desired (e.g., expressing the light and heavy chains of an antibody molecule).

In some preferred embodiments, the vectors are engineered to express an immunoglobulin (e.g., IgG, IgA, IgM, IgD, lbE and sIg). Examples of such vectors are provided in Figures 7-16 (SEQ IN NOs: 4-13). When expression of immunoglobulins with a J chain (e.g., IgM) is desired, different approaches may be utilized. In some embodiments, a single retroviral
vector is used. In some embodiments, the J chain is placed under the control of the LTR promoter. In some embodiments, the resulting vector (see Figure 21, SEQ ID NO:37) comprises the following elements in operable association: 5'LTR, MoMLV extended packaging region J chain gene, internal promoter, signal peptide, heavy chain gene, IRES, light chain gene, RNA export element, MoMLV 3' LTR. In other embodiments, two separate retrovectors are used, one for expressing a J chain chain and the other for expressing the heavy and light chains. Representative vectors are provided in Figures 22 (SEQ ID NO:38) and 23 (SEQ ID NO:39). In some embodiments, the heavy/light chain vector is used to make a cell line comprising multiple copies of the vector (e.g., through high multiplicity of infection transduction or serial transduction or a combination of the two). A clonal cell line is then selected and transduced with the J chain chain vector. In some embodiments, the vector encoding the J chain contains a selectable marker (e.g., blast) that is different from the selectable marker in the heavy/light chain vector (e.g., neo). Individual clonal lines expressing functional IgM are then selected. It will be recognized that the order of transduction can be altered (i.e., the cells can be transduced with J chain vector first, and heavy/light chain vector second).

In still further embodiments, ribozymes are expressed in the host cells. It is contemplated that the ribozyme can be utilized for down-regulating expression of a particular gene or used in conjunction with gene switches such as TET, ecdysone, glucocorticoid enhancer, etc. to provide host cells with various phenotypes.

The transfected host cells are cultured according to methods known in the art. Suitable culture conditions for mammalian cells are well known in the art (See e.g., J. Immunol. Methods (1983) 56:221-234 [1983], Animal Cell Culture: A Practical Approach 2nd Ed., Rickwood, D. and Hames, B. D., eds. Oxford University Press, New York [1992]).

The host cell cultures of the present invention are prepared in a media suitable for the particular cell being cultured. Commercially available media such as Ham's F10 (Sigma, St. Louis, MO), Minimal Essential Medium (MEM, Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium (DMEM, Sigma) are exemplary nutrient solutions. Suitable media are also described in U.S. Pat. Nos. 4,767,704; 4,657,866; 4,927,762; 5,122,469; 4,560,655; and WO 90/03430 and WO 87/00195; the disclosures of which are herein incorporated by reference. Any of these media may be supplemented as necessary with serum, hormones and/or other growth factors (such as insulin, transferrin, or epidermal growth factor), salts (such as sodium chloride,
calcium, magnesium, and phosphate), buffers (such as HEPES), nucleosides (such as adenosine and thymidine), antibiotics (such as gentamycin (gentamicin), trace elements (defined as inorganic compounds usually present at final concentrations in the micromolar range) lipids (such as linoleic or other fatty acids) and their suitable carriers, and glucose or an equivalent energy source. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. For mammalian cell culture, the osmolality of the culture medium is generally about 290-330 mOsm.

The present invention also contemplates the use of a variety of culture systems (e.g., petri dishes, 96 well plates, roller bottles, and bioreactors) for the transfected host cells. For example, the transfected host cells can be cultured in a perfusion system. Perfusion culture refers to providing a continuous flow of culture medium through a culture maintained at high cell density. The cells are suspended and do not require a solid support to grow on. Generally, fresh nutrients must be supplied continuously with concomitant removal of toxic metabolites and, ideally, selective removal of dead cells. Filtering, entrapment and micro-capsulation methods are all suitable for refreshing the culture environment at sufficient rates.

As another example, in some embodiments a fed batch culture procedure can be employed. In the preferred fed batch culture the mammalian host, cells and culture medium are supplied to a culturing vessel initially and additional culture nutrients are fed, continuously or in discrete increments, to the culture during culturing, with or without periodic cell and/or product harvest before termination of culture. The fed batch culture can include, for example, a semi-continuous fed batch culture, wherein periodically whole culture (including cells and medium) is removed and replaced by fresh medium. Fed batch culture is distinguished from simple batch culture in which all components for cell culturing (including the cells and all culture nutrients) are supplied to the culturing vessel at the start of the culturing process. Fed batch culture can be further distinguished from perfusion culturing insofar as the supernate is not removed from the culturing vessel during the process (in perfusion culturing, the cells are restrained in the culture by, e.g., filtration, encapsulation, anchoring to microcarriers etc. and the culture medium is continuously or intermittently introduced and removed from the culturing vessel). In some particularly preferred embodiments, the batch cultures are performed in roller bottles.

Further, the cells of the culture may be propagated according to any scheme or routine that may be suitable for the particular host cell and the particular production plan contemplated.
Therefore, the present invention contemplates a single step or multiple step culture procedure. In a single step culture the host cells are inoculated into a culture environment and the processes of the instant invention are employed during a single production phase of the cell culture. Alternatively, a multi-stage culture is envisioned. In the multi-stage culture cells may be cultivated in a number of steps or phases. For instance, cells may be grown in a first step or growth phase culture wherein cells, possibly removed from storage, are inoculated into a medium suitable for promoting growth and high viability. The cells may be maintained in the growth phase for a suitable period of time by the addition of fresh medium to the host cell culture.

Fed batch or continuous cell culture conditions are devised to enhance growth of the mammalian cells in the growth phase of the cell culture. In the growth phase cells are grown under conditions and for a period of time that is maximized for growth. Culture conditions, such as temperature, pH, dissolved oxygen (dO₂) and the like, are those used with the particular host and will be apparent to the ordinarily skilled artisan. Generally, the pH is adjusted to a level between about 6.5 and 7.5 using either an acid (e.g., CO₂) or a base (e.g., Na₂CO₃ or NaOH). A suitable temperature range for culturing mammalian cells such as CHO cells is between about 30° to 38° C and a suitable dO₂ is between 5-90% of air saturation.

Following the polypeptide production phase, the polypeptide of interest is recovered from the culture medium using techniques that are well established in the art. The protein of interest preferably is recovered from the culture medium as a secreted polypeptide (e.g., the secretion of the protein of interest is directed by a signal peptide sequence), although it also may be recovered from host cell lysates. As a first step, the culture medium or lysate is centrifuged to remove particulate cell debris. The polypeptide thereafter is purified from contaminant soluble proteins and polypeptides, with the following procedures being exemplary of suitable purification procedures: by fractionation on immunoaffinity or ion-exchange columns; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; and protein A Sepharose columns to remove contaminants such as IgG. A protease inhibitor such as phenyl methyl sulfonyl fluoride (PMSF) also may be useful to inhibit proteolytic degradation during purification. Additionally, the protein of interest can be fused in frame to a marker sequence that allows for purification of the protein of interest. Non-
limiting examples of marker sequences include a hexahistidine tag, which may be supplied by a vector, preferably a pQE-9 vector, and a hemagglutinin (HA) tag. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (See e.g., Wilson et al., Cell, 37:767 [1984]). One skilled in the art will appreciate that purification methods suitable for the polypeptide of interest may require modification to account for changes in the character of the polypeptide upon expression in recombinant cell culture.

The host cells of the present invention are also useful for expressing G-protein coupled receptors (GPCRs) and other transmembrane proteins. It is contemplated that when these proteins are expressed, they are correctly inserted into the membrane in their native conformation. Thus, GPCRs and other transmembrane proteins may be purified as part of a membrane fraction or purified from the membranes by methods known in the art.

Furthermore, the vectors of the present invention are useful for co-expressing a protein of interest for which there is no assay or for which assays are difficult. In this system, a protein of interest and a signal protein are arranged in a polycistronic sequence. Preferably, an IRES sequence separates the signal protein and protein of interest (e.g., a GPCR) and the genes encoding the signal protein and protein of interest are expressed as a single transcriptional unit. The present invention is not limited to any particular signal protein. Indeed, the use of a variety of signal proteins for which easy assays exist is contemplated. These signal proteins include, but are not limited to, green fluorescent protein, luciferase, beta-galactosidase, and antibody heavy or light chains. It is contemplated that when the signal protein and protein of interest are co-expressed from a polycistronic sequence, the presence of the signal protein is indicative of the presence of the protein of interest. Accordingly, in some embodiments, the present invention provides methods for indirectly detecting the expression of a protein of interest comprising providing a host cell transfected with a vector encoding a polycistronic sequence, wherein the polycistronic sequence comprises a signal protein and a protein of interest operably linked by an IRES, and culturing the host cells under conditions such that the signal protein and protein of interest are produced, wherein the presence of the signal protein indicates the presence of the protein of interest.

B. Screening Compounds for Activity
The present invention contemplates the use of the high copy number cell lines for screening compounds for activity, and in particular to high throughput screening of compounds from combinatorial libraries (e.g., libraries containing greater than $10^4$ compounds). The high copy number cell lines of the present invention can be used in a variety of screening methods. In some embodiments, the cells can be used in second messenger assays that monitor signal transduction following activation of cell-surface receptors. In other embodiments, the cells can be used in reporter gene assays that monitor cellular responses at the transcription/translation level. In still further embodiments, the cells can be used in cell proliferation assays to monitor the overall growth/no growth response of cells to external stimuli.

In second messenger assays, the host cells are preferably transfected as described above with vectors encoding cell surface receptors, ion channels, cytoplasmic receptors, or other proteins involved in signal transduction (e.g., G proteins, protein kinases, or protein phosphatases) (See, e.g., U.S Pat. Nos. 5,670,113; 5,807,689; 5,876,946; and 6,027,875; all of which are incorporated herein by reference). The host cells are then treated with a compound or plurality of compounds (e.g., from a combinatorial library) and assayed for the presence or absence of a response. It is contemplated that at least some of the compounds in the combinatorial library can serve as agonists, antagonists, activators, or inhibitors of the protein or proteins encoded by the vectors. It is also contemplated that at least some of the compounds in the combinatorial library can serve as agonists, antagonists, activators, or inhibitors of protein acting upstream or downstream of the protein encoded by the vector in a signal transduction pathway.

By way of non-limiting example, it is known that agonist engaged transmembrane receptors are functionally linked to the modulation of several well characterized promoter/enhancer elements (e.g., AP1, cAMP response element (CRE), serum response element (SRE), and nuclear factor of activated T-cells (NF-AT)). Upon activation of a $Gq$ coupling receptor, adenylyl cyclase is stimulated, producing increased concentrations of intracellular cAMP, stimulation of protein kinase A, phosphorylation of the CRE binding protein (CREB) and induction of promoters with CRE elements. $Gq$ coupling receptors dampen CRE activity by inhibition of the same signal transduction components. $Gq$ and some $\beta y$ pairs stimulate phospholipase C (PLC), and the generation of inositol triphosphate (IP3) and diacylglycerol (DAG). A transient flux in intracellular calcium promotes induction of calcineurin and NA-FT,
as well as calmodulin (CaM)-dependent kinase and CREB. Increased DAG concentrations stimulate protein kinase C (PKC) and endosomal/lysosomal acidic sphingomyelinase (aSMase); while the aSMase pathway is dominant, both induce degradation of the NFκB inhibitor IκB as well as NFκB activation. In an alternative pathway, a receptor such as growth factor receptor is activated and recruits Sos to the plasma membrane, resulting in the stimulation of Ras, which in turn recruits the serine/threonine kinase Raf to the plasma membrane. Once activated, Raf phosphorylates MEK kinase, which phosphorylates and activates MAPK and the transcription factor ELK. ELK drives transcription from promoters with SRE elements, leading the synthesis of the transcription factors Fos and Jun, thus forming a transcription factor complex capable of activating AP1 sites. It is contemplated that the proteins forming the described pathways, as well as other receptors, kinases, phosphatases, and nucleic binding proteins, are targets for compounds in the combinatorial library, as well as candidates for expression in the host cells of the present invention.

In some embodiments, the second messenger assays measure fluorescent signals from reporter molecules that respond to intracellular changes (e.g., Ca\(^{2+}\) concentration, membrane potential, pH, IP\(_3\), cAMP, arachidonic acid release) due to stimulation of membrane receptors and ion channels (e.g., ligand gated ion channels; see Denyer et al., Drug Discov. Today 3:323-32 [1998]; and Gonzales et al., Drug. Discov. Today 4:431-39 [1999]). Examples of reporter molecules include, but are not limited to, FRET (fluorescence resonance energy transfer) systems (e.g., Cuo-lipids and oxonols, EDAN/DABCYL), calcium sensitive indicators (e.g., Fluo-3, FURA 2, INDO 1, and FLUO3/AM, BAPTA AM), chloride-sensitive indicators (e.g., SPQ, SPA), potassium-sensitive indicators (e.g., PBFI), sodium-sensitive indicators (e.g., SBFI), and pH sensitive indicators (e.g., BCECF).

In general, the host cells are loaded with the indicator prior to exposure to the compound. Responses of the host cells to treatment with the compounds can be detected by methods known in the art, including, but not limited to, fluorescence microscopy, confocal microscopy (e.g., FCS systems), flow cytometry, microfluidic devices, FLIPR systems (See, e.g., Schroedel and Neagle, J. Biomol. Screening 1:75-80 [1996]), and plate-reading systems. In some preferred embodiments, the response (e.g., increase in fluorescent intensity) caused by compound of unknown activity is compared to the response generated by a known agonist and expressed as a percentage of the maximal response of the known agonist. The maximum response caused by a
known agonist is defined as a 100% response. Likewise, the maximal response recorded after addition of an agonist to a sample containing a known or test antagonist is detectably lower than the 100% response.

The cells are also useful in reporter gene assays. Reporter gene assays involve the use of host cells transfected with vectors encoding a nucleic acid comprising transcriptional control elements of a target gene (i.e., a gene that controls the biological expression and function of a disease target) spliced to a coding sequence for a reporter gene. Therefore, activation of the target gene results in activation of the reporter gene product. Examples of reporter genes finding use in the present invention include, but are not limited to, chloramphenicol transferase, alkaline phosphatase, firefly and bacterial luciferases, β-galactosidase, α-lactamase, and green fluorescent protein. The production of these proteins, with the exception of green fluorescent protein, is detected through the use of chemiluminescent, colorimetric, or bioluminescent products of specific substrates (e.g., X-gal and luciferin). Comparisons between compounds of known and unknown activities may be conducted as described above.

C. Comparison of Variant Protein Activity

The present invention also contemplates the use of the high copy number host cells to produce variants of proteins so that the activity of the variants can be compared. In some embodiments, the variants differ by a single nucleotide polymorphism (SNP) causing a single amino acid difference. In other embodiments, the variants contain multiple amino acid substitutions. In some embodiments, the activity of the variant proteins are assayed in vivo or in cell extracts. In other embodiments, the proteins are purified and assayed in vitro. It is also contemplated that in some embodiments the variant proteins are fused to a sequence that allows easy purification (e.g., a his-tag sequence) or to a reporter gene (e.g., green fluorescent protein). Activity of the proteins may be assayed by appropriate methods known in the art (e.g., conversion of a substrate to a product). In some preferred embodiments, the activity of a wild-type protein is determined, and the activity of variant versions of the wild-type proteins are expressed as a percentage of the activity of the wild-type protein. Furthermore, the intracellular activity of variant proteins may be compared by constructing a plurality of host cells lines, each of which expresses a different variant of the wild-type protein. The activity of the variant
proteins (e.g., variants of proteins involved in signal transduction pathways) may then be compared using the reporter systems for second messenger assays described above. Therefore, in some embodiments, the direct or indirect response (e.g., through downstream or upstream activation of signal transduction pathway) of variant proteins to stimulation or binding by agonists or antagonists is compared. In some preferred embodiments, the response of a wild-type protein is determined, and the responses of variant versions of the wild-type proteins are expressed as a percentage of the response of the wild-type protein.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: M (molar); mM (millimolar); µM (micromolar); nM (nanomolar); mol (moles); mmol (millimoles); µmol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); µg (micrograms); pg (picograms); l (liters); ml (milliliters); µl (microliters); cm (centimeters); mm (millimeters); µm (micrometers); nm (nanometers); °C (degrees Centigrade); AMP (adenosine 5'-monophosphate); BSA (bovine serum albumin); cDNA (copy or complimentary DNA); CS (calf serum); DNA (deoxyribonucleic acid); ssDNA (single stranded DNA); dsDNA (double stranded DNA); dNTP (deoxyribonucleotide triphosphate); LH (luteinizing hormone); NIH (National Institutes of Health, Bethesda, MD); RNA (ribonucleic acid); PBS (phosphate buffered saline); g (gravity); OD (optical density); HEPES (N-[2-Hydroxyethyl]piperazine-N-[2-ethanesulfonic acid]); HBS (HEPES buffered saline); PBS (phosphate buffered saline); SDS (sodium dodecylsulfate); Tris-HCl (tris[Hydroxymethyl]aminomethane-hydrochloride); Klenow (DNA polymerase I large (Klenow) fragment); rpm (revolutions per minute); EGTA (ethylene glycol-bis(β-aminoethyl ether) N, N', N'-tetraacetic acid); EDTA (ethylenediaminetetraacetic acid); bla (β-lactamase or ampicillin-resistance gene); ORI (plasmid origin of replication); lacI (lac repressor); X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside); ATCC (American Type Culture Collection, Rockville, MD); GIBCO/BRL (GIBCO/BRL, Grand Island, NY); Perkin-Elmer (Perkin-Elmer, Norwalk, CT); and Sigma (Sigma Chemical Company, St. Louis, MO).

Example 1
Vector Construction

The following Example describes the construction of vectors used in the experiments below.

A. CMV MN14

The CMV MN14 vector (SEQ ID NO:4; MN14 antibody is described in U.S. Pat. No. 5,874,540, incorporated herein by reference) comprises the following elements, arranged in 5' to 3' order: CMV promoter; MN14 heavy chain signal peptide, MN14 antibody heavy chain; IRES from encephalomyocarditis virus; bovine a-lactalbumin signal peptide; MN 14 antibody light chain; and 3' MoMuLV LTR. In addition to sequences described in SEQ ID NO: 4, the CMV MN14 vector further comprises a 5' MoMuLV LTR, a MoMuLV extended viral packaging signal, and a neomycin phosphotransferase gene (these additional elements are provided in SEQ ID NO:7; the 5' LTR is derived from Moloney Murine Sarcoma Virus in each of the constructs described herein, but is converted to the MoMuLV 5' LTR when integrated).

This construct uses the 5' MoMuLV LTR to control production of the neomycin phosphotransferase gene. The expression of MN14 antibody is controlled by the CMV promoter. The MN14 heavy chain gene and light chain gene are attached together by an IRES sequence. The CMV promoter drives production of a mRNA containing the heavy chain gene and the light chain gene attached by the IRES. Ribosomes attach to the mRNA at the CAP site and at the IRES sequence. This allows both heavy and light chain protein to be produced from a single mRNA. The mRNA expression from the LTR as well as from the CMV promoter is terminated and poly adenylated in the 3' LTR. The construct was cloned by similar methods as described in section B below.

The IRES sequence (SEQ ID NO:3) comprises a fusion of the IRES from the plasmid pLXIN (Clontech) and the bovine a -lactalbumin signal peptide. The initial ATG of the signal peptide was attached to the IRES to allow the most efficient translation initiation from the IRES. The 3' end of the signal peptide provides a multiple cloning site allowing easy attachment of any protein of interest to create a fusion protein with the signal peptide. The IRES sequence can serve as a translational enhancer as well as creating a second translation initiation site that allows two proteins to be produced from a single mRNA.
The IRES-bovine α-lactalbumin signal peptide was constructed as follows. The portion of the plasmid pLXIN (Clontech, Palo Alto, CA) containing the BCMV IRES was PCR amplified using the following primers.

Primer 1 (SEQ ID NO: 35):
5′ GATCCACTAGTAACGGCCGCCAGAATTCGC 3′

Primer 2 (SEQ ID NO: 36):
5′ CAGAGAGACAAAGGAGGCCATATTATCATCGTGTGTTTTCAAG 3′

Primer 2 attaches a tail corresponding to the start of the bovine α-lactalbumin signal peptide coding region to the IRES sequence. In addition, the second triplet codon of the α-lactalbumin signal peptide was mutated from ATG to GCC to allow efficient translation from the IRES sequence. This mutation results in a methionine to alanine change in the protein sequence. This mutation was performed because the IRES prefers an alanine as the second amino acid in the protein chain. The resulting IRES PCR product contains an EcoRI site on the 5′ end of the fragment (just downstream of Primer 1 above).

Next, the α-lactalbumin signal peptide containing sequence was PCR amplified from the α-LA Signal Peptide vector construct using the following primers.

Primer 3 (SEQ ID NO: 14):
5′ CTTGGAAAAAACACGATGATAATATGGCCTCCCTGTCTCTCTG 3′

Primer 4 (SEQ ID NO: 15):
5′ TTCGCGAGCTCGAGATCTAGATATCCCATG 3′

Primer 3 attaches a tail corresponding to the 3′ end of the IRES sequence to the α-lactalbumin signal peptide coding region. As stated above, the second triplet codon of the bovine α-lactalbumin signal peptide was mutated to allow efficient translation from the IRES sequence. The resulting signal peptide PCR fragment contains NaeI, NcoI, EcoRV, XbaI, BglII and XhoI sites on the 3′ end.
After the IRES and signal peptide were amplified individually using the primers shown above, the two reaction products were mixed and PCR was performed using primer 1 and primer 4. The resultant product of this reaction is a spliced fragment that contains the IRES attached to the full length α-lactalbumin signal peptide. The ATG encoding the start of the signal peptide is placed at the same location as the ATG encoding the start of the neomycin phosphotransferase gene found in the vector pLXIN. The fragment also contains the EcoRI site on the 5' end and NaeI, NcoI, EcoRV, XbaI, BglII and XhoI sites on the 3' end.

The spliced IRES/α-lactalbumin signal peptide PCR fragment was digested with EcoRI and XhoI. The α-LA Signal Peptide vector construct was also digested with EcoRI and XhoI. These two fragments were ligated together to give the pIRES construct.

The IRES/α-lactalbumin signal peptide portion of the pIRES vector was sequenced and found to contain mutations in the 5' end of the IRES. These mutations occur in a long stretch of Cs and were found in all clones that were isolated.

To repair this problem, pLXIN DNA was digested with EcoRI and BsmFI. The 500bp band corresponding to a portion of the IRES sequence was isolated. The mutated IRES/α-lactalbumin signal peptide construct was also digested with EcoRI and BsmFI and the mutated IRES fragment was removed. The IRES fragment from pLXIN was then substituted for the IRES fragment of the mutated IRES/α-lactalbumin signal peptide construct. The IRES/α-LA signal peptide portion of resulting plasmid was then verified by DNA sequencing.

The resulting construct was found to have a number of sequence differences when compared to the expected pLXIN sequence obtained from Clontech. The IRES portion of pLXIN purchased from Clontech was sequenced to verify its sequence. The differences from the expected sequence also appear to be present in the pLXIN plasmid obtained from Clontech.

Four sequence differences were identified:

- bp 347 T - was G in pLXIN sequence
- bp 786-788 ACG - was GC in LXIN sequence.

B. CMV LL2

The CMV LL2 (SEQ ID NO:5; LL2 antibody is described in U.S. Pat. No. 6,187,287, incorporated herein by reference) construct comprises the following elements, arranged in 5' to 3' order: 5' CMV promoter (Clonetech), LL2 heavy chain signal peptide, LL2 antibody heavy
chain; IRES from encephalomyocarditis virus; bovine \( \alpha \)-LA signal peptide; LL2 antibody light chain; and 3' MoMuLV LTR. In addition to sequences described in SEQ ID NO:5, the CMV LL2 vector further comprises a 5' MoMuLV LTR, a MoMuLV extended viral packaging signal, and a neomycin phosphotransferase gene (these additional elements are provided in SEQ ID NO:7).

This construct uses the 5' MoMuLV LTR to control production of the neomycin phosphotransferase gene. The expression of LL2 antibody is controlled by the CMV promoter (Clontech). The LL2 heavy chain gene and light chain gene are attached together by an IRES sequence. The CMV promoter drives production of an mRNA containing the heavy chain gene and the light chain gene attached by the IRES. Ribosomes attach to the mRNA at the CAP site and at the IRES sequence. This allows both heavy and light chain protein to be produced from a single mRNA. The mRNA expression from the LTR as well as from the CMV promoter is terminated and poly adenylated in the 3' LTR.

The IRES sequence (SEQ ID NO:3) comprises a fusion of the IRES from the plasmid pLXIN (Clontech) and the bovine alpha-lactalbumin signal peptide. The initial ATG of the signal peptide was attached to the IRES to allow the most efficient translation initiation from the IRES. The 3' end of the signal peptide provides a multiple cloning site allowing easy attachment of any protein of interest to create a fusion protein with the signal peptide. The IRES sequence can serve as a translational enhancer as well as creating a second translation initiation site that allows two proteins to be produced from a single mRNA.

The LL2 light chain gene was attached to the IRES \( \alpha \)-lactalbumin signal peptide as follows. The LL2 light chain was PCR amplified from the vector pCRLL2 using the following primers.

Primer 1 (SEQ ID NO: 16):

5' CTACAGGTGTCCACGTCGACATCCAGCTGACCCAG 3'

Primer 2 (SEQ ID NO: 17):

5' CTGCAGAATAGATCTCTAACACTCTCCCCTGTTG 3'

These primers add a HincII site right at the start of the coding region for mature LL2 light chain. Digestion of the PCR product with HincII gives a blunt end fragment starting with the
initial GAC encoding mature LL2 on the 5’ end. Primer 2 adds a BglII site to the 3’ end of the
gene right after the stop codon. The resulting PCR product was digested with HincII and BglII
and cloned directly into the IRES-Signal Peptide plasmid that was digested with NaeI and BglII.

The Kozak sequence of the LL2 heavy chain gene was then modified. The vector
pCRMN14HC was digested with XhoI and AvrII to remove about a 400 bp fragment. PCR was
then used to amplify the same portion of the LL2 heavy chain construct that was removed by the
XhoI-AvrII digestion. This amplification also mutated the 5’ end of the gene to add a better
Kozak sequence to the clone. The Kozak sequence was modified to resemble the typical IgG
Kozak sequence. The PCR primers are shown below.

Primer 1 (SEQ ID NO: 18):
5’CAGTGTGATCTCGAGAATTCAGGACCTCACCATGGGATGGAGCTGTATCAT 3’
Primer 2 (SEQ ID NO: 19):
5’AGGCTGTATTGGTGGATTCGTCT 3’

The PCR product was digested with XhoI and AvrII and inserted back into the previously
digested plasmid backbone.

The "good" Kozak sequence was then added to the light chain gene. The “good” Kozak
LL2 heavy chain gene construct was digested with EcoRI and the heavy chain gene containing
fragment was isolated. The IRES α-Lactalbumin Signal Peptide LL2 light chain gene construct
was also digested with EcoRI. The heavy chain gene was then cloned into the EcoRI site of
IRES light chain construct. This resulted in the heavy chain gene being placed at the 5’ end of
the IRES sequence.

Next, a multiple cloning site was added into the LNCX retroviral backbone plasmid. The
LNCX plasmid was digested with HindIII and Clal. Two oligonucleotide primers were produced
and annealed together to create an double stranded DNA multiple cloning site. The following
primers were annealed together.

Primer 1 (SEQ ID NO: 20):
5’AGCTTCTCGAGTTAACAGATCTAGGCCTCCTAGGTCGACAT 3’
Primer 2 (SEQ ID NO: 21): 5'
CGATGTCGACCTAGGAGGCCTAGATCTGTTAACTCGAGA 3'

After annealing, the multiple cloning site was ligated into LNCX to create LNC-MCS.

Next, the double chain gene fragment was ligated into the retroviral backbone gene construct. The double chain gene construct created above was digested with SalI and BglII and the double chain-containing fragment was isolated. The retroviral expression plasmid LNC-MCS was digested with XhoI and BglII. The double chain fragment was then cloned into the LNC-MCS retroviral expression backbone.

Next, an RNA splicing problem in the construct was corrected. The construct was digested with NsiI. The resulting fragment was then partially digested with EcoRI. The fragments resulting from the partial digest that were approximately 9300 base pairs in size were gel purified. A linker was created to mutate the splice donor site at the 3’ end of the LL2 heavy chain gene. The linker was again created by annealing two oligonucleotide primers together to form the double stranded DNA linker. The two primers used to create the linker are shown below.

Primer 1 (SEQ ID NO: 22):
5’CGAGGCTCTGCAACAACCACTACACGAGAAGAGCCTCTCCCTGTCTCCCGG
GAAATGAAAGCCG 3’

Primer 2 (SEQ ID NO: 23):
5’AATTCGGCTTTCATTTCCCGGGAGACAGGGAGAGGCTCTTCTGCGTGTAGTGGTTG
TGCAGAGCCTCCTGGA 3’

After annealing the linker was substituted for the original NsiI/EcoRI fragment that was removed during the partial digestion.

C. MMTV MN14

The MMTV MN14 (SEQ ID NO:6) construct comprises the following elements, arranged in 5’ to 3’ order: 5’ MMTV promoter; double mutated PPE sequence; MN 14 antibody heavy
chain; IRES from encephalomyocarditis virus; bovine α LA signal peptide MN 14 antibody light chain; WPRE sequence; and 3' MoMuLV LTR. In addition to the sequences described in SEQ ID NO:6, the MMTV MN14 vector further comprises a MoMuLV LTR, MoMuLV extended viral packaging signal; neomycin phosphotransferase gene located 5' of the MMTV promoter (these additional elements are provided in SEQ ID NO: 7).

This construct uses the 5' MoMuLV LTR to control production of the neomycin phosphotransferase gene. The expression of MN14 antibody is controlled by the MMTV promoter (Pharmacia). The MN14 heavy chain gene and light chain gene are attached together by an IRES/ bovine α-LA signal peptide sequence (SEQ ID NO: 3). The MMTV promoter drives production of a mRNA containing the heavy chain gene and the light chain gene attached by the IRES/bovine α-LA signal peptide sequence. Ribosomes attach to the mRNA at the CAP site and at the IRES/ bovine α-LA signal peptide sequence. This allows both heavy and light chain protein to be produced from a single mRNA. In addition, there are two genetic elements contained within the mRNA to aid in export of the mRNA from the nucleus to the cytoplasm and aid in poly-adenylation of the mRNA. The PPE sequence is contained between the RNA CAP site and the start of the MN14 protein coding region, the WPRE is contained between the end of MN14 protein coding and the poly-adenylation site. The mRNA expression from the LTR as well as from the MMTV promoter is terminated and poly-adenylated in the 3' LTR.

ATG sequences within the PPE element (SEQ ID NO:2) were mutated to prevent potential unwanted translation initiation. Two copies of this mutated sequence were used in a head to tail array. This sequence is placed just downstream of the promoter and upstream of the Kozak sequence and signal peptide-coding region. The WPRE is isolated from woodchuck hepatitis virus and also aids in the export of mRNA from the nucleus and creating stability in the mRNA. If this sequence is included in the 3' untranslated region of the RNA, level of protein expression from this RNA increases up to 10-fold.

D. α-LA MN14

The α-LA MN14 (SEQ ID NO:7) construct comprises the following elements, arranged in 5' to 3' order: 5' MoMuLV LTR, MoMuLV extended viral packaging signal, neomycin phosphotransferase gene, bovine/human alpha-lactalbumin hybrid promoter, double mutated PPE element, MN14 heavy chain signal peptide, MN14 antibody heavy chain, IRES from
encephalomyocarditis virus/bovine α LA signal peptide, MN14 antibody light chain, WPRE sequence; and 3' MoMuLV LTR.

This construct uses the 5' MoMuLV LTR to control production of the neomycin phosphotransferase gene. The expression of MN14 antibody is controlled by the hybrid α-LA promoter (SEQ ID NO:1). The MN14 heavy chain gene and light chain gene are attached together by an IRES sequence/bovine α-LA signal peptide (SEQ ID NO:3). The α-LA promoter drives production of a mRNA containing the heavy chain gene and the light chain gene attached by the IRES. Ribosomes attach to the mRNA at the CAP site and at the IRES sequence. This allows both heavy and light chain protein to be produced from a single mRNA.

In addition, there are two genetic elements contained within the mRNA to aid in export of the mRNA from the nucleus to the cytoplasm and aid in poly-adenylation of the mRNA. The mutated PPE sequence (SEQ ID NO:2) is contained between the RNA CAP site and the start of the MN14 protein coding region. ATG sequences within the PPE element (SEQ ID NO:2) were mutated to prevent potential unwanted translation initiation. Two copies of this mutated sequence were used in a head to tail array. This sequence is placed just downstream of the promoter and upstream of the Kozak sequence and signal peptide-coding region. The WPRE was isolated from woodchuck hepatitis virus and also aids in the export of mRNA from the nucleus and creating stability in the mRNA. If this sequence is included in the 3' untranslated region of the RNA, level of protein expression from this RNA increases up to 10-fold. The WPRE is contained between the end of MN14 protein coding and the poly-adenylation site. The mRNA expression from the LTR as well as from the bovine/human alpha-lactalbumin hybrid promoter is terminated and poly adenylated in the 3' LTR.

The bovine/human alpha-lactalbumin hybrid promoter (SEQ ID NO:1) is a modular promoter/enhancer element derived from human and bovine alpha-lactalbumin promoter sequences. The human portion of the promoter is from +15 relative to transcription start point (tsp) to -600 relative to the tsp. The bovine portion is then attached to the end of the human portion and corresponds to -550 to -2000 relative to the tsp. The hybrid was developed to remove poly-adenylation signals that were present in the bovine promoter and hinder retroviral RNA production. It was also developed to contain genetic control elements that are present in the human gene, but not the bovine.
For construction of the bovine/human α-lactalbumin promoter, human genomic DNA was isolated and purified. A portion of the human α-lactalbumin promoter was PCR amplified using the following two primers:

**Primer 1 (SEQ ID NO: 24):**

5′AAAGCATATGTTCTGGGCCTTGTTACATGGCTGGATTGGTT 3′

**Primer 2 (SEQ ID NO: 25):**

5′TGAATTCGGCGCCCCCAAGAACCTGAAATGGAAGCATCACTCAGTTTCATATAT 3′

These two primers created a NdeI site on the 5′ end of the PCR fragment and a EcoRI site on the 3′ end of the PCR fragment.

The human PCR fragment created using the above primers was double digested with the restriction enzymes NdeI and EcoRI. The plasmid pKBaP-1 was also double digested with NdeI and EcoRI. The plasmid pKBaP-1 contains the bovine α-lactalbumin 5′ flanking region attached to a multiple cloning site. This plasmid allows attachment of various genes to the bovine α-lactalbumin promoter.

Subsequently, the human fragment was ligated/substituted for the bovine fragment of the promoter that was removed from the pKBaP-1 plasmid during the double digestion. The resulting plasmid was confirmed by DNA sequencing to be a hybrid of the Bovine and Human α-lactalbumin promoter/regulatory regions.

Attachment of the MN14 light chain gene to the IRES α-lactalbumin signal peptide was accomplished as follows. The MN14 light chain was PCR amplified from the vector pCRMNI4LC using the following primers.

**Primer 1 (SEQ ID NO: 26):** 5′ CTACAGGTGTCCACGTCGACATCCAGCTGACCCAG 3′

**Primer 2 (SEQ ID NO: 27):** 5′ CTGCAGAATAGATCTCTAACACTCTCCCCTGTTG 3′

These primers add a HincII site right at the start of the coding region for mature MN14 light chain. Digestion of the PCR product with HincII gives a blunt end fragment starting with
the initial GAC encoding mature MN14 on the 5’ end. Primer 2 adds a BglII site to the 3’ end of
the gene right after the stop codon. The resulting PCR product was digested with HincII and
BglIII and cloned directly into the IRES-Signal Peptide plasmid that was digested with NaeI and
BglIII.

Next, the vector pCRMN14HC was digested with XhoI and NruI to remove about a 500
bp fragment. PCR was then used to amplify the same portion of the MN14 heavy chain
construct that was removed by the XhoI-NruI digestion. This amplification also mutated the 5’
end of the gene to add a better Kozak sequence to the clone. The Kozak sequence was modified
to resemble the typical IgG Kozak sequence. The PCR primers are shown below.

Primer 1 (SEQ ID NO: 28):
5’CAGTGTGATCTCGAGAATTCAGGACCTCACCATGGGATGGAGCTGTATCAT 3’

Primer 2 (SEQ ID NO: 29):
5’GTGTCTTCGGGTCTCAGGCTGT 3’

The PCR product was digested with XhoI and NruI and inserted back into the previously
digested plasmid backbone.

Next, the “good” Kozak MN14 heavy chain gene construct was digested with EcoRI and
the heavy chain gene containing fragment was isolated. The IRES α-Lactalbumin Signal Peptide
MN14 light chain gene construct was also digested with EcoRI. The heavy chain gene was then
cloned into the EcoRI site of IRES light chain construct. This resulted in the heavy chain gene
being placed at the 5’ end of the IRES sequence.

A multiple cloning site was then added to the LNCX retroviral backbone plasmid. The
LNCX plasmid was digested with HindIII and ClaI. Two oligonucleotide primers were produced
and annealed together to create an double stranded DNA multiple cloning site. The following
primers were annealed together.

Primer 1 (SEQ ID NO: 30):
5’ AGCTTCTCGAGTTAACAGATCTAGGCCTCCTAGGTCGACAT 3’

Primer 2 (SEQ ID NO: 31):
5' CGATGTCGACCTAGGAGGCCTAGATCTGTTAACTCGAGA 3'

After annealing the multiple cloning site was ligated into LNCX to create LNC-MCS.

The double chain gene fragment was then inserted into a retroviral backbone gene construct. The double chain gene construct created in step 3 was digested with SalI and BglII and the double chain containing fragment was isolated. The retroviral expression plasmid LNC-MCS was digested with XhoI and BglII. The double chain fragment was then cloned into the LNC-MCS retroviral expression backbone.

Next, a RNA splicing problem in the construct was repaired. The construct was digested with NsiI. The resulting fragment was then partially digested with EcoRI. The fragments resulting from the partial digest that were approximately 9300 base pairs in size, were gel purified. A linker was created to mutate the splice donor site at the 3' end of the MN14 heavy chain gene. The linker was again created by annealing two oligonucleotide primers together to form the double stranded DNA linker. The two primers used to create the linker are shown below.

Primer 1 (SEQ ID NO: 32):
5'CGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCCGGGAAA TGAAAGCCG 3'

Primer 2 (SEQ ID NO: 33):
5'AATTCGGCTTTCATTTCCCGGGAGACAGGGAGAGGCTCTTCTGCGTGTAGTGGTTG TGCAGAGCCTCGTGCA 3'

After annealing the linker was substituted for the original NsiI/EcoRI fragment that was removed during the partial digestion.

Next, the mutated double chain fragment was inserted into the α-Lactalbumin expression retroviral backbone LN α-LA-Mertz-MCS. The gene construct produced above was digested with BamHI and BglII and the mutated double chain gene containing fragment was isolated. The LN α-LA-Mertz-MCS retroviral backbone plasmid was digested with BglII. The BamHI/BglII fragment was then inserted into the retroviral backbone plasmid.
A WPRE element was then inserted into the gene construct. The plasmid BluescriptII SK+ WPRE-B11 was digested with BamHI and HincII to remove the WPRE element and the element was isolated. The vector created above was digested with BglII and HpaI. The WPRE fragment was ligated into the BglII and HpaI sites to create the final gene construct.

E. α-LA Bot

The α-LA Bot (SEQ ID NO:8, botulinum toxin antibody) construct comprises the following elements, arranged in 5' to 3' order: bovine/human alpha-lactalbumin hybrid promoter, mutated PPE element, cc49 signal peptide, botulinum toxin antibody light chain, IRES from encephalomyocarditis virus/ bovine α-LA signal peptide, botulinum toxin antibody heavy chain, WPRE sequence, and 3' MoMuLV LTR. In addition, the α-LA botulinum toxin antibody vector further comprises a 5' MoMuLV LTR, a MoMuLV extended viral packaging signal, and a neomycin phosphotransferase gene (these additional elements are provided in SEQ ID NO: 7).

This construct uses the 5' MoMuLV LTR to control production of the neomycin phosphotransferase gene. The expression of botulinum toxin antibody is controlled by the hybrid α-LA promoter. The botulinum toxin antibody light chain gene and heavy chain gene are attached together by an IRES/ bovine α-LA signal peptide sequence. The bovine/human alpha-lactalbumin hybrid promoter drives production of a mRNA containing the light chain gene and the heavy chain gene attached by the IRES. Ribosomes attach to the mRNA at the CAP site and at the IRES sequence. This allows both light and heavy chain protein to be produced from a single mRNA.

In addition, there are two genetic elements contained within the mRNA to aid in export of the mRNA from the nucleus to the cytoplasm and aid in poly-adenylation of the mRNA. The mutated PPE sequence (SEQ ID NO:2) is contained between the RNA CAP site and the start of the MN14 protein coding region. ATG sequences within the PPE element (SEQ ID NO:2) were mutated to prevent potential unwanted translation initiation. Two copies of this mutated sequence were used in a head to tail array. This sequence was placed just downstream of the promoter and upstream of the Kozak sequence and signal peptide-coding region. The WPRE was isolated from woodchuck hepatitis virus and also aids in the export of mRNA from the nucleus and creating stability in the mRNA. If this sequence is included in the 3' untranslated region of the RNA, level of protein expression from this RNA increases up to 10-fold. The
WPRE is contained between the end of MN14 protein coding and the poly-adenylation site. The mRNA expression from the LTR as well as from the bovine/human alpha-lactalbumin hybrid promoter is terminated and poly adenylated in the 3' LTR.

The bovine/human α-lactalbumin hybrid promoter (SEQ ID NO:1) is a modular promoter/enhancer element derived from human and bovine α-lactalbumin promoter sequences. The human portion of the promoter is from +15 relative to transcription start point to -600 relative to the tsp. The bovine portion is then attached to the end of the human portion and corresponds to -550 to -2000 relative to the tsp. The hybrid was developed to remove poly-adenylation signals that were present in the bovine promoter and hinder retroviral RNA production. It was also developed to contain genetic control elements that are present in the human gene, but not the bovine. Likewise, the construct contains control elements present in the bovine but not in the human.

F. **LSRNL**

The LSRNL (SEQ ID NO:9) construct comprises the following elements, arranged in 5' to 3' order: 5' MoMuLV LTR, MoMuLV viral packaging signal; hepatitis B surface antigen; RSV promoter; neomycin phosphotransferase gene; and 3' MoMuLV LTR.

This construct uses the 5' MoMuLV LTR to control production of the Hepatitis B surface antigen gene. The expression of the neomycin phosphotransferase gene is controlled by the RSV promoter. The mRNA expression from the LTR as well as from the RSV promoter is terminated and poly adenylated in the 3' LTR.

G. **α-LA cc49IL2**

The α-LA cc49IL2 (SEQ ID NO:10; the cc49 antibody is described in U.S. Pat. Nos. 5,512,443; 5,993,813; and 5,892,019; each of which is herein incorporated by reference) construct comprises the following elements, arranged in 5' to 3' order: 5' bovine/human α-lactalbumin hybrid promoter; cc49-IL2 coding region; and 3' MoMuLV LTR. This gene construct expresses a fusion protein of the single chain antibody cc49 attached to Interleukin-2.
Expression of the fusion protein is controlled by the bovine/human α-lactalbumin hybrid promoter.

The bovine/human α-lactalbumin hybrid promoter (SEQ ID NO:1) is a modular promoter/enhancer element derived from human and bovine alpha-lactalbumin promoter sequences. The human portion of the promoter is from +15 relative to transcription start point to -600 relative to the tsp. The bovine portion is then attached to the end of the human portion and corresponds to -550 to -2000 relative to the tsp. The hybrid was developed to remove polyadenylation signals that were present in the bovine promoter and hinder retroviral RNA production. It was also developed to contain genetic control elements that are present in the human gene, but not the bovine. Likewise, the construct contains control elements present in the bovine but not in the human. The 3' viral LTR provide the poly-adenylation sequence for the mRNA.

H. α-LA YP

The α-LA YP (SEQ ID NO: 11) construct comprises the following elements, arranged in 5' to 3' order: 5' bovine/human alpha-lactalbumin hybrid promoter; double mutated PPE sequence; bovine α-LA signal peptide; Yersenia pestis antibody heavy chain Fab coding region; EMCV IRES/ bovine α-LA signal peptide; Yersenia pestis antibody light chain Fab coding region; WPRE sequence; 3' MoMuLV LTR.

This gene construct will cause the expression of Yersenia pestis mouse Fab antibody. The expression of the gene construct is controlled by the bovine/human α-lactalbumin hybrid promoter. The PPE sequence and the WPRE sequence aid in moving the mRNA from the nucleus to the cytoplasm. The IRES sequence allows both the heavy and the light chain genes to be translated from the same mRNA. The 3' viral LTR provides the poly-adenylation sequence for the mRNA.

In addition, there are two genetic elements contained within the mRNA to aid in export of the mRNA from the nucleus to the cytoplasm and aid in poly-adenylation of the mRNA. The mutated PPE sequence (SEQ ID NO:2) is contained between the RNA CAP site and the start of the MN14 protein coding region. ATG sequences within the PPE element (SEQ ID NO:2) were mutated (bases 4, 112, 131, and 238 of SEQ ID NO: 2 were changed from a G to a T) to prevent potential unwanted translation initiation. Two copies of this mutated sequence were used in a
head to tail array. This sequence was placed just downstream of the promoter and upstream of the Kozak sequence and signal peptide-coding region. The WPRE was isolated from woodchuck hepatitis virus and also aids in the export of mRNA from the nucleus and creating stability in the mRNA. If this sequence is included in the 3' untranslated region of the RNA, level of protein expression from this RNA increases up to 10-fold. The WPRE is contained between the end of MN14 protein coding and the poly-adenylation site. The mRNA expression from the LTR as well as from the bovine/human alpha-lactalbumin hybrid promoter is terminated and polyadenylated in the 3' LTR.

The bovine/human alpha-lactalbumin hybrid promoter (SEQ ID NO:1) is a modular promoter/enhancer element derived from human and bovine alpha-lactalbumin promoter sequences. The human portion of the promoter is from +15 relative to transcription start point to -600 relative to the tsp. The bovine portion is then attached to the end of the human portion and corresponds to -550 to -2000 relative to the tsp. The hybrid was developed to remove polyadenylation signals that were present in the bovine promoter and hinder retroviral RNA production. It was also developed to contain genetic control elements that are present in the human gene, but not the bovine. Likewise, the construct contains control elements present in the bovine but not in the human.

Example 2

Generation of Cell Lines Stably Expressing the MoMLV gag and pol Proteins

Examples 2-5 describe the production of pseudotyped retroviral vectors. These methods are generally applicable to the production of the vectors described above. The expression of the fusogenic VSV G protein on the surface of cells results in syncytium formation and cell death. Therefore, in order to produce retroviral particles containing the VSV G protein as the membrane-associated protein a two-step approach was taken. First, stable cell lines expressing the gag and pol proteins from MoMLV at high levels were generated (e.g., 293GPSD cells). The stable cell line which expresses the gag and pol proteins produces noninfectious viral particles lacking a membrane-associated protein (e.g., an envelope protein). The stable cell line was then co-transfected, using the calcium phosphate precipitation, with VSV-G and gene of interest plasmid DNAs. The pseudotyped vector generated was used to infect 293GPSD cells to produce
stably transformed cell lines. Stable cell lines can be transiently transfected with a plasmid capable of directing the high level expression of the VSV G protein (see below). The transiently transfected cells produce VSV G-pseudotyped retroviral vectors, which can be collected from the cells over a period of 3 to 4 days before the producing cells die as a result of syncytium formation.

The first step in the production of VSV G-pseudotyped retroviral vectors, the generation of stable cell lines expressing the MoMLV gag and pol proteins is described below. The human adenovirus Ad-5-transformed embryonal kidney cell line 293 (ATCC CRL 1573) was cotransfected with the pCMVgag-pol and the gene encoding for phleomycin. pCMV gag-pol contains the MoMLV gag and pol genes under the control of the CMV promoter (pCMV gag-pol is available from the ATCC).

The plasmid DNA was introduced into the 293 cells using calcium phosphate co-precipitation (Graham and Van der Eb, Virol. 52:456 [1973]). Approximately 5 x 10^5 293 cells were plated into a 100 mm tissue culture plate the day before the DNA co-precipitate was added. Stable transformants were selected by growth in DMEM-high glucose medium containing 10% FCS and 10 µg/ml phleomycin (selective medium). Colonies which grew in the selective medium were screened for extracellular reverse transcriptase activity (Goff et al., J. Virol. 38:239 [1981]) and intracellular p30gag expression. The presence of p30gag expression was determined by Western blotting using a goat-anti p30 antibody (NCI antiserum 77S000087). A clone which exhibited stable expression of the retroviral genes was selected. This clone was named 293Gp^SD (293 gag-pol-San Diego). The 293Gp^SD cell line, a derivative of the human Ad-5-transformed embryonal kidney cell line 293, was grown in DMEM-high glucose medium containing 10% FCS.

**Example 3**

Preparation of Pseudotyped Retroviral Vectors Bearing the G Glycoprotein of VSV

In order to produce VSV G protein pseudotyped retrovirus the following steps were taken. The 293Gp^SD cell line was co-transfected with VSV-G plasmid and DNA plasmid of interest. This co-transfection generates the infectious particles used to infect 293Gp^SD cells to generate the packaging cell lines. This Example describes the production of pseudotyped
LNBOTDC virus. This general method may be used to produce any of the vectors described in Example 1.

a) Cell Lines and Plasmids

The packaging cell line, 293GP<sup>3D</sup> was grown in alpha-MEM-high glucose medium containing 10% FCS. The titer of the pseudo-typed virus may be determined using either 208F cells (Quade, Virol. 98:461 [1979]) or NIH/3T3 cells (ATCC CRL 1658); 208F and NIH/3T3 cells are grown in DMEM-high glucose medium containing 10% CS.

The plasmid LNBOTDC contains the gene encoding neomycin phosphotransferase (Neo) under the transcriptional control of the LTR promoter followed by the gene encoding BOTD under the transcriptional control of cytomegalovirus intermediate-early promoter. The plasmid pHCMV-G contains the VSV G gene under the transcriptional control of the human cytomegalovirus intermediate-early promoter (Yee et al., Meth. Cell Biol. 43:99 [1994]).

b) Production of stable packaging cell lines, pseudotyped vector and Titering of Pseudotyped LNBOTDC Vector

LNBOTDC DNA (SEQ ID NO: 13) was co-transfected with pHCMV-G DNA into the packaging line 293GP<sup>3D</sup> to produce LNBOTDC virus. The resulting LNBOTDC virus was then used to infect 293GP<sup>3D</sup> cells to transduce the cells. The procedure for producing pseudotyped LNBOTDC virus was carried out as described (Yee et al., Meth. Cell Biol. 43:99 [1994].

This is a retroviral gene construct that upon creation of infectious replication defective retroviral vector will cause the insertion of the sequence described above into the cells of interest. Upon insertion the CMV regulatory sequences control the expression of the botulinum toxin antibody heavy and light chain genes. The IRES sequence allows both the heavy and the light chain genes to be translated from the same mRNA. The 3' viral LTR provides the polyadenylation sequence for the mRNA.

Both heavy and light chain protein for botulinum toxin antibody are produced from this signal mRNA. The two proteins associated to form active botulinum toxin antibody. The heavy and light chain proteins also appear to be formed in an equal molar ratio to each other.
Briefly, on day 1, approximately 5 x 10^4 293GP\textsuperscript{SD} cells were placed in a 75 cm\textsuperscript{2} tissue culture flask. On the following day (day 2), the 293GP\textsuperscript{SD} cells were transfected with 25 \(\mu\)g of pLNBOTDC plasmid DNA and 25 \(\mu\)g of VSV-G plasmid DNA using the standard calcium phosphate co-precipitation procedure (Graham and Van der Eb, Virol. 52:456 [1973]). A range of 10 to 40 \(\mu\)g of plasmid DNA may be used. Because 293GP\textsuperscript{SD} cells may take more than 24 hours to attach firmly to tissue culture plates, the 293GP\textsuperscript{SD} cells may be placed in 75 cm\textsuperscript{2} flasks 48 hours prior to transfection. The transfected 293GP\textsuperscript{SD} cells provide pseudotyped LNBOTDC virus.

On day 3, approximately 1 x 10^5 293GP\textsuperscript{SD} cells were placed in a 75 cm\textsuperscript{2} tissue culture flask 24 hours prior to the harvest of the pseudotyped virus from the transfected 293GP\textsuperscript{SD} cells. On day 4, culture medium was harvested from the transfected 293GP\textsuperscript{SD} cells 48 hours after the application of the pLNBOTDC and VSV-G DNA. The culture medium was filtered through a 0.45 \(\mu\)m filter and polybrene was added to a final concentration of 8 \(\mu\)g/ml. The culture medium containing LNBOTDC virus was used to infect the 293GP\textsuperscript{SD} cells as follows. The culture medium was removed from the 293GP\textsuperscript{SD} cells and was replaced with the LNBOTDC virus containing culture medium. Polybrene was added to the medium following addition to cells. The virus containing medium was allowed to remain on the 293GP\textsuperscript{SD} cells for 24 hours. Following the 16 hour infection period (on day 5), the medium was removed from the 293GP\textsuperscript{SD} cells and was replaced with fresh medium containing 400 \(\mu\)g/ml G418 (GIBCO/BRL). The medium was changed approximately every 3 days until G418-resistant colonies appeared approximately two weeks later.

The G418-resistant 293 colonies were plated as single cells in 96 wells. Sixty to one hundred G418-resistant colonies were screened for the expression of the BOTDC antibody in order to identify high producing clones. The top 10 clones in 96-well plates were transferred to 24-well plates and allowed to grow to confluency.

The top 10 clones were then expanded to screen for high titer production. Based on protein expression and titer production, 5 clonal cell lines were selected. One line was designated the master cell bank and the other 4 as backup cell lines. Pseudotyped vector was generated as follows. Approximately 1 x 10^6 293GP\textsuperscript{SD}/LNBOTDC cells were placed into a 75 cm\textsuperscript{2} tissue culture flask. Twenty-four hours later, the cells were transfected with 25 \(\mu\)g of pHCMV-G plasmid DNA using calcium phosphate co-precipitation. Six to eight hours after the
calcium-DNA precipitate was applied to the cells, the DNA solution was replaced with fresh culture medium (lacking G418). Longer transfection times (overnight) were found to result in the detachment of the majority of the 293GPSD/LNBOTDC cells from the plate and are therefore avoided. The transfected 293GPSD/LNBOTDC cells produce pseudotyped LNBOTDC virus.

The pseudotyped LNBOTDC virus generated from the transfected 293GPSD/LNBOTDC cells can be collected at least once a day between 24 and 96 hr after transfection. The highest virus titer was generated approximately 48 to 72 hr after initial pHCMV-G transfection. While syncytium formation became visible about 48 hr after transfection in the majority of the transfected cells, the cells continued to generate pseudotyped virus for at least an additional 48 hr as long as the cells remained attached to the tissue culture plate. The collected culture medium containing the VSV G-pseudotyped LNBOTDC virus was pooled, filtered through a 0.45 μm filter and stored at -80°C or concentrated immediately and then stored at -80°C.

The titer of the VSV G-pseudotyped LNBOTDC virus was then determined as follows. Approximately 5 x 10⁴ rat 208F fibroblasts cells were plated into 6 well plates. Twenty-four hours after plating, the cells were infected with serial dilutions of the LNBOTDC virus-containing culture medium in the presence of 8 μg/ml polybrene. Twenty four hours after infection with virus, the medium was replaced with fresh medium containing 400 μg/ml G418 and selection was continued for 14 days until G418-resistant colonies became visible. Viral titers were typically about 0.5 to 5.0 x 10⁶ colony forming units (cfu)/ml. The titer of the virus stock could be concentrated to a titer of greater than 10⁹ cfu/ml as described below.

**Example 4**

Concentration of Pseudotyped Retroviral Vectors

The VSV G-pseudotyped LNBOTDC viruses were concentrated to a high titer by one cycle of ultracentrifugation. However, two cycles can be performed for further concentration. The frozen culture medium collected as described in Example 2 which contained pseudotyped LNBOTDC virus was thawed in a 37°C water bath and was then transferred to Oakridge centrifuge tubes (50 ml Oakridge tubes with sealing caps, Nalge Nunc International) previously sterilized by autoclaving. The virus was sedimented in a JA20 rotor (Beckman) at 48,000 x g (20,000 rpm) at 4°C for 120 min. The culture medium was then removed from the tubes in a
biosafety hood and the media remaining in the tubes was aspirated to remove the supernatant. The virus pellet was resuspended to 0.5 to 1% of the original volume of culture medium DMEM. The resuspended virus pellet was incubated overnight at 4°C without swirling. The virus pellet could be dispersed with gentle pipetting after the overnight incubation without significant loss of infectious virus. The titer of the virus stock was routinely increased 100- to 300-fold after one round of ultracentrifugation. The efficiency of recovery of infectious virus varied between 30 and 100%.

The virus stock was then subjected to low speed centrifugation in a microfuge for 5 min at 4°C to remove any visible cell debris or aggregated virions that were not resuspended under the above conditions. It was noted that if the virus stock is not to be used for injection into oocytes or embryos, this centrifugation step may be omitted.

The virus stock can be subjected to another round of ultracentrifugation to further concentrate the virus stock. The resuspended virus from the first round of centrifugation is pooled and pelleted by a second round of ultracentrifugation which is performed as described above. Viral titers are increased approximately 2000-fold after the second round of ultracentrifugation (titers of the pseudotyped LNBOTDC virus are typically greater than or equal to $1 \times 10^9$ cfu/ml after the second round of ultracentrifugation).

The titers of the pre- and post-centrifugation fluids were determined by infection of 208F cells (NIH 3T3 or bovine mammary epithelial cells can also be employed) followed by selection of G418-resistant colonies as described above in Example 2.

**Example 5**

**Preparation of Pseudotyped Retrovirus For Infection of Host Cells**

The concentrated pseudotyped retroviruses were resuspended in 0.1X HBS (2.5 mM HEPES, pH 7.12, 14 mM NaCl, 75 _M Na2HPO4-H2O) and 18 µl aliquots were placed in 0.5 ml vials (Eppendorf) and stored at -80°C until used. The titer of the concentrated vector was determined by diluting 1 µl of the concentrated virus $10^{-7}$- or $10^8$-fold with 0.1X HBS. The diluted virus solution was then used to infect 208F and bovine mammary epithelial cells and viral titers were determined as described in Example 2.
Example 6
Expression of MN14 by Host Cells

This Example describes the production of antibody MN14 from cells transfected with a high number of integrating vectors. Pseudotyped vector were made from the packaging cell lines for the following vectors: CMV MN14, α-LA MN14, and MMTV MN14. Rat fibroblasts (208F cells), MDBK cells (bovine kidney cells), and bovine mammary epithelial cells were transfected at a multiplicity of infection of 1000. One thousand cells were plated in a T25 flask and 10^6 colony forming units (CFU’s) of vector in 3 ml media was incubated with the cells. The duration of the infection was 24 hr, followed by a media change. Following transfection, the cells were allowed to grow and become confluent.

The cell lines were grown to confluency in T25 flasks and 5ml of media was changed daily. The media was assayed daily for the presence of MN14. All of the MN14 produced is active (an ELISA to detect human IgG gave the exact same values as the CEA binding ELISA) and Western blotting has shown that the heavy and light chains are produced at a ratio that appears to be a 1:1 ratio. In addition, a non-denaturing Western blot indicated that what appeared to be 100% of the antibody complexes were correctly formed (See Figure 1: Lane 1, 85 ng control Mn14; Lane 2, bovine mammary cell line, α-LA promoter; Lane 3, bovine mammary cell line, CMV promoter; Lane 4, bovine kidney cell line, α-LA promoter; Lane 5, bovine kidney cell line, CMV promoter; Lane 6, 208 cell line, α-LA promoter; Lane 7, 208 cell line, CMV promoter). Figure 2 is a graph showing the production of MN14 over time for four cell lines. The Y axis shows MN14 production in ng/ml of media. The X-axis shows the day of media collection for the experiment. Four sets of data are shown on the graph. The comparisons are between the CMV and α-LA promoter and between the 208 cells and the bovine mammary cells. The bovine mammary cell line exhibited the highest expression, followed by the 208F cells and MDBK cells. With respect to the constructs, the CMV driven construct demonstrated the highest level of expression, followed by the α-LA driven gene construct and the MMTV construct. At 2 weeks, the level of daily production of the CMV construct was 4.5 µg/ml of media (22.5 mg/day in a T25 flask). The level of expression subsequently increased slowly to 40 µg/day as the cells became very densely confluent over the subsequent week. 2.7 L of media from an α-lac-MN14
packaging cell line was processed by affinity chromatography to produce a purified stock of MN14.

Figure 3 is a western blot of a 15% SDS-PAGE gel run under denaturing conditions in order to separate the heavy and light chains of the MN14 antibody. Lane 1 shows MN14 from bovine mammary cell line, hybrid α-LA promoter; lane 2 shows MN14 from bovine mammary cell line, CMV promoter; lane 3 shows MN14 from bovine kidney cell line, hybrid α-LA promoter; lane 4 shows MN14 from bovine kidney cell line, CMV promoter; lane 5 shows MN14 from rat fibroblast cell line, hybrid α-LA promoter; lane 6 shows MN14 from rat fibroblast, CMV promoter. In agreement with Figure 1 above, the results show that the heavy and light chains are produced in a ratio of approximately 1:1.

Example 7
Quantitation of Protein Produced Per Cell

This Example describes the quantitation of the amount of protein produced per cell in cell cultures produced according to the invention. Various cells (208F cells, MDBK cells, and bovine mammary cells) were plated in 25 cm² culture dishes at 1000 cells/dish. Three different vectors were used to infect the three cell types (CMV-MN14, MMTV-MN14, and α-LA-MN14) at an MOI of 1000 (titers: 2.8 × 10⁶, 4.9 × 10⁶, and 4.3 × 10⁶, respectively). Media was collected approximately every 24 hours from all cells. Following one month of media collection, the 208F and MDBK cells were discarded due to poor health and low MN14 expression. The cells were passaged to T25 flasks and collection of media from the bovine mammary cells was continued for approximately 2 months with continued expression of MN14. After two months in T25 flasks, the cells with CMV promoters were producing 22.5 pg/cell/day and the cells with α-LA promoters were producing 2.5 pg MN14/cell/day.

After 2 months in T25 flasks, roller bottles (850 cm²) were seeded to scale-up production and to determine if MN14 expression was stable following multiple passages. Two roller bottles
were seeded with bovine mammary cells expressing MN14 from a CMV promoter and two roller bottles were seeded with bovine mammary cells expressing MN14 from the α-LA promoter. The cultures reached confluency after approximately two weeks and continue to express MN14.

Roller bottle expression is shown in Table 1 below.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Promoter</th>
<th>MN14 Production/ Week (μg/ml)</th>
<th>MN14 Production/ Week - Total (μg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bovine mammary</td>
<td>CMV</td>
<td>2.6</td>
<td>1 - 520</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>CMV</td>
<td>10.6</td>
<td>2 - 2120</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>CMV</td>
<td>8.7</td>
<td>3 - 1740</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>CMV</td>
<td>7.8</td>
<td>4 - 1560</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>α-LA</td>
<td>0.272</td>
<td>1 - 54.4</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>α-LA</td>
<td>2.8</td>
<td>2 - 560</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>α-LA</td>
<td>2.2</td>
<td>3 - 440</td>
</tr>
<tr>
<td>Bovine mammary</td>
<td>α-LA</td>
<td>2.3</td>
<td>4 - 460</td>
</tr>
</tbody>
</table>

**Example 8**

Transfection at Varied Multiplicities of Infection
This Example describes the effect of transfection at varied multiplicities of infection on protein expression. 208F rat fibroblast and bovine mammary epithelial cells (BMEC) were plated in a 25 cm² plates at varied cell numbers/25 cm². Cells were infected with either the CMV MN14 vector or the α LA MN14 vector at a MOI of 1,10, 1000, and 10,000 by keeping the number of CFUs kept constant and varying the number of cells infected.

Following infection, medium was changed daily and collected approximately every 24 hours from all cells for approximately 2 months. The results of both of the vectors in bovine mammary epithelial cells are shown in Table 2 below. Cells without data indicate cultures that became infected prior to the completion of the experiment. The "# cells" column represents the number of cells at the conclusion of the experiment. The results indicate that a higher MOI results in increased MN14 production, both in terms of the amount of protein produced per day, and the total accumulation.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Promoter</th>
<th>MOI</th>
<th>% cell Confluency</th>
<th>MN14 (ng/ml)</th>
<th># Cells</th>
<th>MN14 Production /day (pg/cell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>10000</td>
<td>100%</td>
<td>4228</td>
<td>4.5E5</td>
<td>47</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>1000</td>
<td>100%</td>
<td>2832</td>
<td>2.0E6</td>
<td>7.1</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>10</td>
<td>100%</td>
<td>1873</td>
<td>2.5E6</td>
<td>3.75</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMEC</td>
<td>_LA</td>
<td>10000</td>
<td>100%</td>
<td>1024</td>
<td>1.5E6</td>
<td>3.4</td>
</tr>
<tr>
<td>BMEC</td>
<td>_LA</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMEC</td>
<td>_LA</td>
<td>100</td>
<td>100%</td>
<td>722</td>
<td>1.8E6</td>
<td>1.9</td>
</tr>
<tr>
<td>BMEC</td>
<td>_LA</td>
<td>10</td>
<td>100%</td>
<td>421234</td>
<td>2.3E6</td>
<td>.925</td>
</tr>
</tbody>
</table>
Example 9

Transfection at Varied Multiplicities of Infection

This experiment describes protein production from the CMV MN14 vector at a variety of MOI values. Bovine mammary cells, CHO cells, and human embryo kidney cells (293 cells) were plated in 24 well plates (2 cm²) at 100 cells/2 cm² well. Cells were infected at various dilutions with CMV MN14 to obtain MOI values of 1, 10, 100, 1000, and 10000. The CHO cells reached confluency at all MOI within 11 days of infection. However, the cells infected at a MOI of 10,000 grew more slowly. The bovine mammary and 293 cells grew slower, especially at the highest MOI of 10,000. The cells were then passaged into T25 flasks to disperse cells. Following dispersion, cells reached confluence within 1 week. The medium was collected after one week and analyzed for MN14 production. The CHO and human 293 cells did not exhibit good growth in extended culture. Thus, data were not collected from these cells. Data for bovine mammary epithelial cells are shown in Table 3 below. The results indicate that production of MN14 increased with higher MOI.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Promoter</th>
<th>MOI</th>
<th>% Confluency</th>
<th>MN14 Production (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>10000</td>
<td>100%</td>
<td>1312</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>1000</td>
<td>100%</td>
<td>100</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>100</td>
<td>100%</td>
<td>7.23</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>10</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>BMEC</td>
<td>CMV</td>
<td>1</td>
<td>100%</td>
<td>0</td>
</tr>
</tbody>
</table>

Example 10
Expression of LL2 Antibody by Bovine Mammary Cells

This Example describes the expression of antibody LL2 by bovine mammary cells. Bovine mammary cells were infected with vector CMV LL2 \( (7.85 \times 10^7 \text{ CFU/ml}) \) at MOI's of 1000 and 10,000 and plated in 25cm\(^2\) culture dishes. None of the cells survived transfection at the MOI of 10,000. At 20% confluency, 250 ng/ml of LL2 was present in the media.

Example 11
Expression of Botulinum Toxin Antibody by Bovine Mammary Cells

This Example describes the expression of Botulinum toxin antibody in bovine mammary cells. Bovine mammary cells were infected with vector \( \alpha\)-LA Bot \( (2.2 \times 10^2 \text{ CFU/ml}) \) and plated in 25cm\(^2\) culture dishes. At 100% confluency, 6 ng/ml of Botulinum toxin antibody was present in the media.

Example 12
Expression of Hepatitis B Surface Antigen by Bovine Mammary Cells

This Example describes the expression of hepatitis B surface antigen (HBSAg) in bovine mammary cells. Bovine mammary cells were infected with vector LSRNL \( (350 \text{ CFU/ml}) \) and plated in 25cm\(^2\) culture dishes. At 100% confluency, 20 ng/ml of HBSAg was present in the media.

Example 13
Expression of cc49IL2 Antigen Binding Protein by Bovine Mammary Cells

This Example describes the expression of cc49IL2 in bovine mammary cells. Bovine mammary cells were infected with vector cc49IL2 \( (3.1 \times 10^5 \text{ CFU/ml}) \) at a MOI of 1000 and plated in 25cm\(^2\) culture dishes. At 100% confluency, 10 μg/ml of cc49IL2 was present in the media.
Example 14
Expression of Multiple Proteins by Bovine Mammary Cells

This Example describes the expression of multiple proteins in bovine mammary cells. Mammary cells producing MN14 (infected with CMV-MN14 vector) were infected with cc49IL2 vector (3.1 x 10^5 CFU/ml) at an MOI of 1000, and 1000 cells were plated in 25cm^2 culture plates. At 100% confluency, the cells expressed MN14 at 2.5 µg/ml and cc49IL2 at 5 µg/ml.

Example 15
Expression of Multiple Proteins by Bovine Mammary Cells

This Example describes the expression of multiple proteins in bovine mammary cells. Mammary cells producing MN14 (infected with CMV-MN14 vector) were infected with LSNRL vector (100 CFU/ml) at an MOI of 1000, and 1000 cells were plated in 25cm^2 culture plates. At 100% confluency, the cells expressed MN14 at 2.5 µg/ml and hepatitis surface antigen at 150 ng/ml.

Example 16
Expression of Multiple Proteins by Bovine Mammary Cells

This Example describes the expression of multiple proteins in bovine mammary cells. Mammary cells producing hepatitis B surface antigen (infected with LSRNL vector) were infected with cc49IL2 vector at an MOI of 1000, and 1000 cells were plated in 25cm^2 culture plates. At 100% confluency, the cells expressed MN14 at 2.4 µg/ml and hepatitis B surface antigen at 13 ng/ml. It will be understood that multiple proteins may be expressed in the other cell lines described above.

Example 17
Expression of Hepatitis B Surface Antigen and Botulinum Toxin Antibody in Bovine Mammary Cells

This Example describes the culture of transfected cells in roller bottle cultures. 208F cells and bovine mammary cells were plated in 25cm$^2$ culture dishes at 1000 cells/25cm$^2$. LSRNL or α-LA Bot vectors were used to infect each cell line at a MOI of 1000. Following one month of culture and media collection, the 208F cells were discarded due to poor growth and plating. Likewise, the bovine mammary cells infected with α-LA Bot were discarded due to low protein expression. The bovine mammary cells infected with LSRNL were passaged to seed roller bottles (850 cm$^2$). Approximately 20 ng/ml hepatitis type B surface antigen was produced in the roller bottle cultures.

Example 18
Expression in Clonally Selected Cell Lines

This experiment describes expression of MN14 from clonally selected cell lines. Cell lines were grown to confluency in T25 flasks and 5ml of media were collected daily. The media was assayed daily for the presence of MN14. All the MN14 produced was active and Western blotting indicated that the heavy and light chains were produced at a ratio that appears to be almost exactly 1:1. In addition, a non-denaturing western blot indicated that approximately 100% of the antibody complexes were correctly formed. After being in culture for about two months, the cells were expanded into roller bottles or plated as single cell clones in 96 well plates.

The production of MN14 in the roller bottles was analyzed for a 24 hour period to determine if additional medium changing would increase production over what was obtained with weekly medium changes. Three 24 hour periods were examined. The CMV promoter cells in 850 cm$^2$ roller bottles produced 909 ng/ml the first day, 1160 ng/ml the second day and 1112 ng/ml the third day. The α-LA promoter cells produced 401 ng/ml the first day, 477 ng/ml the second day and 463 ng/ml the third day. These values correspond well to the 8-10 mg/ml/week that were obtained for the CMV cells and the 2-3 mg/ml that were obtained for the α-LA cells.
It does not appear that more frequent media changing would increase MN14 production in roller bottles.

Single cell lines were established in 96 well plates and then passaged into the same wells to allow the cells to grow to confluency. Once the cells reached confluency, they were assayed for MN14 production over a 24 hour period. The clonal production of MN14 from CMV cell lines ranged from 19 ng/ml/day to 5500 ng/ml/day. The average production of all cell clones was 1984 ng/ml/day. The α-LA cell clones yielded similar results. The clonal production of MN14 from α-LA cell lines ranged from 1 ng/ml/day to 2800 ng/ml/day. The average production of these cell clones was 622 ng/ml/day. The results are provided in Table 4 below.

<table>
<thead>
<tr>
<th>CMV Clonal Cell Line Number</th>
<th>MN14 Production (ng/ml)</th>
<th>Alpha-lactalbumin Clonal Cell Line Number</th>
<th>MN14 Production (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>19</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>88</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>134</td>
<td>12</td>
<td>0.7</td>
</tr>
<tr>
<td>34</td>
<td>151</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>32</td>
<td>221</td>
<td>28</td>
<td>55</td>
</tr>
<tr>
<td>23</td>
<td>343</td>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td>27</td>
<td>423</td>
<td>8</td>
<td>81</td>
</tr>
<tr>
<td>4</td>
<td>536</td>
<td>13</td>
<td>154</td>
</tr>
<tr>
<td>41</td>
<td>682</td>
<td>48</td>
<td>159</td>
</tr>
<tr>
<td>45</td>
<td>685</td>
<td>7</td>
<td>186</td>
</tr>
<tr>
<td>40</td>
<td>696</td>
<td>36</td>
<td>228</td>
</tr>
<tr>
<td>11</td>
<td>1042</td>
<td>39</td>
<td>239</td>
</tr>
<tr>
<td>8</td>
<td>1044</td>
<td>51</td>
<td>275</td>
</tr>
<tr>
<td>5</td>
<td>1066</td>
<td>31</td>
<td>283</td>
</tr>
<tr>
<td>19</td>
<td>1104</td>
<td>54</td>
<td>311</td>
</tr>
</tbody>
</table>
Example 19
Estimation of Insert Copy Number

This example describes the relationship of multiplicity of infection, gene copy number, and protein expression. Three DNA assays were developed using the INVADER Assay system (Third Wave Technologies, Madison, WI). One of the assays detects a portion of the bovine α-lactalbumin 5' flanking region. This assay is specific for bovine and does not detect the porcine or human α-lactalbumin gene. This assay will detect two copies of the α-lactalbumin gene in all control bovine DNA samples and also in bovine mammary epithelial cells. The second assay detects a portion of the extended packaging region from the MLV virus. This assay is specific for this region and does not detect a signal in the 293 human cell line, bovine mammary epithelial cell line or bovine DNA samples. Theoretically, all cell lines or other samples not infected with MLV should not produce a signal. However, since the 293GP cell line was produced with the extended packaging region of DNA, this cell line gives a signal when the assay is run. From the initial analysis, it appears that the 293GP cell line contains two copies of the extended packing region sequence that are detected by the assay. The final assay is the control assay. This assay detects a portion of the insulin-like growth factor I gene that is identical in bovine, porcine, humans and a number of other species. It is used as a control on every sample that is run in order to determine the amount of signal that is generated from this sample for a two copy gene. All samples that are tested should contain two copies of the control gene.
DNA samples can be isolated using a number of methods. Two assays are then performed on each sample. The control assay is performed along with either the bovine $\alpha$-lactalbumin assay or the extended packaging region assay. The sample and the type of information needed will determine which assay is run. Both the control and the transgene detection assay are run on the same DNA sample, using the exact same quantity of DNA.

The data resulting from the assay are as follows (Counts indicate arbitrary fluorescence units):

- Extended Packaging Region or $\alpha$-Lactalbumin Background counts
- Extended Packaging Region or $\alpha$-Lactalbumin counts
- Internal Control background counts
- Internal Control counts

To determine net counts for the assay the background counts are subtracted from the actual counts. This occurs for both the control and transgene detection assay. Once the net counts are obtained, a ratio of the net counts for the transgene detection assay to the net counts of the control assay can be produced. This value is an indication of the number of copies of transgene compared to the number of copies of the internal control gene (in this case IGF-I). Because the transgene detection assay and the control assay are two totally different assays, they do not behave exactly the same. This means that one does not get an exact 1:1 ratio if there are two copies of the transgene and two copies of the control gene in a specific sample. However, the values are generally close to the 1:1 ratio. Also, different insertion sites for the transgene may cause the transgene assay to behave differently depending on where the insertions are located.

Therefore, although the ratio is not an exact measure of copy number, it is a good indication of relative copy number between samples. The greater the value of the ratio the greater the copy number of the transgene. Thus, a ranking of samples from lowest to highest will give a very accurate comparison of the samples to one another with regard to copy number.

Table 5 provides actual data for the EPR assay:

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Control Counts</th>
<th>Control Background</th>
<th>Net Control</th>
<th>Transgene Counts</th>
<th>Transgene Background</th>
<th>Net Transgene</th>
<th>Net Ratio</th>
</tr>
</thead>
</table>

Table 5
Counts | Counts | Counts | Counts
---|---|---|---
293 | 116 | 44 | 72 | 46.3 | 46 | 0.3 | 0
293GP | 112 | 44 | 68 | 104 | 46 | 58 | .84
1 | 74 | 40 | 34 | 88 | 41 | 47 | 1.38
2 | 64 | 40 | 24 | 83 | 41 | 43 | 1.75
3 | 62 | 44 | 18 | 144 | 46 | 98 | 5.57

From this data, it can be determined that the 293 cell line has no copies of the extended packaging region/transgene. However the 293 GP cells appear to have two copies of the extended packaging region. The other three cell lines appear to have three or more copies of the extended packaging region (one or more additional copies compared to 293GP cells).

**Invader Assay Gene Ratio and Cell Line Protein Production**

Bovine mammary epithelial cells were infected with either the CMV driven MN14 construct or the α-lactalbumin driven MN14 construct. The cells were infected at a 1000 to 1 vector to cell ratio. The infected cells were expanded. Clonal cell lines were established for both the α-LA and CMV containing cells from this initial pooled population of cells. Approximately 50 cell lines were produced for each gene construct. Individual cells were placed in 96 well plates and then passaged into the same well to allow the cells to grow to confluency. Once the cells lines reached confluency, they were assayed for MN14 production over a 24 hour period. The clonal production of MN14 from CMV cell lines ranged from 0 ng/ml/day to 5500 ng/ml/day. The average production of all cell clones was 1984 ng/ml/day. The α-LA cell clones showed similar trends. The clonal production of MN14 from α-LA cell lines ranged from 0 ng/ml/day to 2800 ng/ml/day. The average production of these cell clones was 622 ng/ml/day.

For further analysis of these clonal lines, fifteen CMV clones and fifteen α-LA clones were selected. Five highest expressing, five low expressing and five mid-level expressing lines were chosen. These thirty cell lines were expanded and banked. DNA was isolated from most all of the thirty cell lines. The cell lines were passed into 6 well plates and grown to confluency. Once at confluency, the media was changed every 24 hours and two separate collections from each cell line were assayed for MN14 production. The results of these two assays were averaged.
and these numbers were used to create Tables 6 and 7 below. DNA from the cell lines was run using the Invader extended packaging region assay and the results are shown below. The Tables show the cell line number, corresponding gene ratio and antibody production.

**Table 6**

<table>
<thead>
<tr>
<th>CMV Clonal Cell Line Number</th>
<th>Invader Gene Ratio</th>
<th>MN14 Production (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.19</td>
<td>104</td>
</tr>
<tr>
<td>7</td>
<td>1.62</td>
<td>2874</td>
</tr>
<tr>
<td>10</td>
<td>2.57</td>
<td>11202</td>
</tr>
<tr>
<td>18</td>
<td>3.12</td>
<td>7757</td>
</tr>
<tr>
<td>19</td>
<td>1.62</td>
<td>2483</td>
</tr>
<tr>
<td>21</td>
<td>1.53</td>
<td>3922</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>0.23</td>
<td>443</td>
</tr>
<tr>
<td>31</td>
<td>3.45</td>
<td>5697</td>
</tr>
<tr>
<td>32</td>
<td>0.27</td>
<td>346</td>
</tr>
<tr>
<td>34</td>
<td>0.37</td>
<td>305</td>
</tr>
<tr>
<td>38</td>
<td>1.47</td>
<td>2708</td>
</tr>
<tr>
<td>41</td>
<td>1.54</td>
<td>5434</td>
</tr>
<tr>
<td>49</td>
<td>2.6</td>
<td>7892</td>
</tr>
<tr>
<td>50</td>
<td>1.56</td>
<td>5022</td>
</tr>
<tr>
<td>Average of All Clones</td>
<td>1.48</td>
<td>3746</td>
</tr>
</tbody>
</table>

**Table 7**

<table>
<thead>
<tr>
<th>α-LA Clonal Cell Line Number</th>
<th>Invader Gene Ratio</th>
<th>MN14 Production (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4.28</td>
<td>3600</td>
</tr>
</tbody>
</table>
The graphs (Figs. 17 and 18) show the comparison between protein expression and invader assay gene ratio. The results indicate that there is a direct correlation between invader assay gene ratio and protein production. It also appears that the protein production has not reached a maximum and if cells containing a higher invader assay gene ratio were produced, higher protein production would occur.

Invader Assay Gene Ratio and Multiple Cell Line Infections

Two packaging cell lines (293GP) produced using previously described methods were used to produce replication defective retroviral vector. One of the cell lines contains a retroviral gene construct that expresses the botulinum toxin antibody gene from the CMV promoter (LTR-Extended Viral Packaging Region-Neo Gene-CMV Promoter-Bot Light Chain Gene-IRES-Bot Heavy Chain Gene-LTR), the other cell line contains a retroviral gene construct that expresses the YP antibody gene from the CMV promoter (LTR-Extended Viral Packaging Region-Neo Gene-CMV Promoter-YP Heavy Chain Gene-IRES-YP Light Chain Gene-WPRE-LTR). In addition to being able to produce replication defective retroviral vector, each of these cell lines also produce either botulinum toxin antibody or YP antibody.

The vector produced from these cell lines was then used to re-infect the parent cell line. This procedure was performed in order to increase the number of gene insertions and to improve
antibody production from these cell lines. The botulinum toxin parent cell line was infected with a new aliquot of vector on three successive days. The titer of the vector used to perform the infection was 1 \times 10^8 \text{ cfu/ml}. Upon completion of the final 24 hour infection, clonal selection was performed on the cells and the highest protein producing line was established for botulinum toxin antibody production. A similar procedure was performed on the YP parent cell line. This cell line was also infected with a new aliquot of vector on three successive days. The titer of the YP vector aliquots was 1 \times 10^4. Upon completion of the final 24 hour infection, clonal selection was performed on the cells and the highest protein producing line was established for YP production.

Each of the parent cell lines and the daughter production cell lines were examined for Invader gene ratio using the extended packaging region assay and for protein production. The Bot production cell line, which was generated using the highest titer vector had the highest gene ratio. It also had the highest protein production, again suggesting that gene copy number is proportional to protein production. The YP production cell line also had a higher gene ratio and produced more protein than its parent cell line, also suggesting that increasing gene copy is directly related to increases in protein production. The data is presented in Table 8.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Invader Gene Ratio</th>
<th>Antibody Production (Bot/YP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bot Parent Cell Line</td>
<td>1.12</td>
<td>4.8 \mu g/ml</td>
</tr>
<tr>
<td>Bot Production Cell Line</td>
<td>3.03</td>
<td>55 \mu g/ml</td>
</tr>
<tr>
<td>YP Parent Cell Line</td>
<td>1.32</td>
<td>4 \mu g/ml</td>
</tr>
<tr>
<td>YP Production Cell Line</td>
<td>2.04</td>
<td>25 \mu g/ml</td>
</tr>
</tbody>
</table>

Example 20

Transfection with Lentivirus Vectors

This example describes methods for the production of lentivirus vectors and their use to infect host cells at a high multiplicity of infection. Replication-defective viral particles are produced by the transient cotransfection of the plasmids described in U.S. Pat. No. 6,013,516 in 293T human kidney cells. All plasmids are transformed and grown in E. coli HB101 bacteria.
following standard molecular biology procedures. For transfection of eukaryotic cells, plasmid DNA is purified twice by equilibrium centrifugation in CsCl-ethidium bromide gradients. A total of 40 μg DNA is used for the transfection of a culture in a 10 cm dish, in the following proportions: 10 μg pCMV△R8, 20 μg pHR', and 10 μg env plasmids, either MLV/Ampho, MLV/Eco or VSV-G. 293T cells are grown in DMEM supplemented with 10% fetal calf serum and antibiotics in a 10% CO₂ incubator. Cells are plated at a density of 1.3x10⁶/10 cm dish the day before transfection. Culture medium is changed 4 to 6 hrs before transfection. Calcium phosphate-DNA complexes are prepared according to the method of Chen and Okayama (Mol. Cell. Biol., 7:2745, 1987), and incubated overnight with the cells in an atmosphere of 5% CO₂.

The following morning, the medium is replaced, and the cultures returned to 10% CO₂. Conditioned medium is harvested 48 to 60 hrs after transfection, cleared of cellular debris by low speed centrifugation (300g 10 min), and filtered through 0.45 μm low protein binding filters.

To concentrate vector particles, pooled conditioned medium harvested as described above is layered on top of a cushion of 20% sucrose solution in PBS and centrifuged in a Beckman SW28 rotor at 50,000g for 90 min. The pellet is resuspended by incubation and gentle pipetting in 1-4 ml PBS for 30-60 min, then centrifuged again at 50,000xg for 90 min in a Beckmann SW55 rotor. The pellet is resuspended in a minimal volume (20-50 μl) of PBS and either used directly for infection or stored in frozen aliquots at -80°C.

The concentrated lentivirus vectors are titered and used to transfect an appropriate cell line (e.g., 293 cells, Hela cells, rat 208F fibroblasts)) at a multiplicity of infection of 1,000. Analysis of clonally selected cell lines expressing the exogenous protein will reveal that a portion of the selected cell lines contain more than two integrated copies of the vector. These cell lines will produce more of the exogenous protein than cell lines containing only one copy of the integrated vector.

### Example 21

**Expression and Assay of G-protein Coupled Receptors**

This example describes the expression of a G-Protein Coupled Receptor protein (GPCR) from a retroviral vector. This example also describes the expression of a signal protein from an IRES as a marker for expression of a difficult to assay protein or a protein that has no assay such
as a GPCR. The gene construct (SEQ ID NO: 34; Figure 19) comprises a G-protein-coupled receptor followed by the IRES-signal peptide-antibody light chain cloned into the MCS of pLBCX retroviral backbone. Briefly, a PvuII/PvuII fragment (3057 bp) containing the GPCR-IRES-antibody light chain was cloned into the SruI site of pLBCX. pLBCX contains the EM7 (T7) promoter, Blasticidin gene and SV40 polyA in place of the Neomycin resistance gene from pLNCX.

The gene construct was used to produce a replication defective retroviral packaging cell line and this cell line was used to produce replication defective retroviral vector. The vector produced from this cell line was then used to infect 293GP cells (human embryonic kidney cells). After infection, the cells were placed under Blasticidin selection and single cell Blasticidin resistant clones were isolated. The clones were screened for expression of antibody light chain. The top 12 light chain expressing clones were selected. These 12 light chain expressing clones were then screened for expression of the GPCR using a ligand binding assay. All twelve of the samples also expressed the receptor protein. The clonal cell lines and their expression are shown in Table 9.

<table>
<thead>
<tr>
<th>Cell Clone Number</th>
<th>Antibody Light Chain Expression</th>
<th>GPCR Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>19</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>22</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>24</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>27</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>30</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>45</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>46</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>50</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Example 22

Multiple infection of 293 cells with replication defective retroviral vector

This example describes the multiple serial transfection of cells with retroviral vectors. The following gene construct was used to produce a replication defective retroviral packaging cell line.

5' LTR = Moloney murine sarcoma virus 5' long terminal repeat.
EPR = Moloney murine leukemia virus extended packaging region.
Blast = Blasticidin resistance gene.
CMV = Human cytomegalovirus immediate early promoter.
Gene = Gene encoding test protein
WPRE = RNA transport element
3' LTR = Moloney murine leukemia virus 3' LTR.

This packaging cell line was then used to produce a replication defective retroviral vector arranged as follows. The vector was produced from cells grown in T150 flasks and frozen. The frozen vector was thawed at each infection. For infection # 3 a concentrated solution of vector was used to perform the infection. All other infections were performed using non-concentrated vector. The infections were performed over a period of approximately five months by placing 5 ml of vector/media solution on a T25 flask containing 30% confluent 293 cells. Eight mg/ml of polybrene was also placed in the vector solution during infection. The vector solution was left on the cells for 24 hours and then removed. Media (DMEM with 10% fetal calf serum) was then added to the cells. Cells were grown to full confluency and passaged into a new T25 flask. The cells were then grown to 30% confluency and the infection procedure was repeated. This process was repeated 12 times and is outlined Table 10 below. After infections 1, 3, 6, 9 and 12, cells left over after passaging were used to obtain a DNA sample. The DNA was analyzed using the INVADER assay to determine an estimate of the number of vector inserts in the cells after various times in the infection procedure. The results indicate that the number of vector insertions goes up over time with the highest level being after the 12th infection. Since a value of 0.5 is approximately an average of one vector insert copy per cell, after twelve infections the average
vector insert copy has yet to reach two. These data indicates that the average vector copy per cell is a little less than 1.5 copies per cell. Also, there was no real change in gene copy number from infection #6 to infection #9. Furthermore, these data indicate that transfection conducted at a standard low multiplicity of infection fail to introduce more than one copy of the retroviral vector into the cells.

<table>
<thead>
<tr>
<th>Table 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Line or Infection Number</td>
</tr>
<tr>
<td>293</td>
</tr>
<tr>
<td>Infection #1</td>
</tr>
<tr>
<td>Infection #2</td>
</tr>
<tr>
<td>Infection #3</td>
</tr>
<tr>
<td>Infection #4</td>
</tr>
<tr>
<td>Infection #5</td>
</tr>
<tr>
<td>Infection #6</td>
</tr>
<tr>
<td>Infection #7</td>
</tr>
<tr>
<td>Infection #8</td>
</tr>
<tr>
<td>Infection #9</td>
</tr>
<tr>
<td>Infection #10</td>
</tr>
<tr>
<td>Infection #11</td>
</tr>
<tr>
<td>Infection #12</td>
</tr>
</tbody>
</table>

Example 23

Production of YP antibody

This Example demonstrates the production of Yersinia pestis antibody by bovine mammary epithelial cells and human kidney fibroblast cells (293 cells). Cells lines were infected with the α-LA YP vector. Both of the cell lines produced YP antibody. All of the antibody is active and the heavy and light chains are produced in a ratio approximating 1:1.
Example 24
Transduction of Plant Protoplasts

This Example describes a method for transducing plant protoplasts. Tobacco protoplasts of Nicotiana tabacum c.v. Petit Havanna are produced according to conventional processes from a tobacco suspension culture (Potrykus and Shillito, Methods in Enzymology, vol. 118, Plant Molecular Biology, eds. A. and H. Weissbach, Academic Press, Orlando, 1986). Completely unfolded leaves are removed under sterile conditions from 6-week-old shoot cultures and thoroughly wetted with an enzyme solution of the following composition: Enzyme solution:

\[
\begin{align*}
\text{H}_2\text{O}, & \quad 70 \text{ ml;} \\
\text{sucrose}, & \quad 13 \text{ g;} \\
\text{macerozyme R 10}, & \quad 1 \text{ g;} \\
\text{cellulase}, & \quad 2 \text{ g;} \\
"\text{Onozuka" R 10} (\text{Yakult Co. Ltd., Japan}) & \quad \text{Drisellase (Chemische Fabrik Schweizerhalle, Switzerland),} 0.13 \text{ g;} \\
& \quad \text{and } 2(\text{N-morpholine})-\text{ethanesulfonic acid (MES),} 0.5 \text{ ml pH 6.0}
\end{align*}
\]

Leaves are then cut into squares from 1 to 2 cm in size and the squares are floated on the above-mentioned enzyme solution. They are incubated overnight at a temperature of 26°C in the dark. This mixture is then gently shaken and incubated for a further 30 minutes until digestion is complete.

The suspension is then filtered through a steel sieve having a mesh width of 100 µm, rinsed thoroughly with 0.6M sucrose (MES, pH 5.6) and subsequently centrifuged for 10 minutes at from 4000 to 5000 rpm. The protoplasts collect on the surface of the medium which is then removed from under the protoplasts, for example using a sterilized injection syringe.

The protoplasts are resuspended in a K₃ medium [sucrose (102.96 g/l; xylose (0.25 g/l); 2,4-dichlorophenoxyacetic acid (0.10 mg/l); 1-naphthylacetic acid (1.00 mg/l); 6-benzylaminopurine (0.20 mg/l); pH 5.8](Potrykus and Shillito, supra) that contains 0.4M sucrose.

To carry out the transformation experiments, the protoplasts are first of all washed, counted and then resuspended, at a cell density of from 1 to 2.5x10⁶ cells per ml, in a W₅ medium [154 mM NaCl, 125 mM CaCl₂ x 2H₂O, 5 mM KCl, 5 mM glucose, pH 5.6], which ensures a high survival rate of the isolated protoplasts. After incubation for 30 minutes at from 6 to 8°C, the protoplasts are then used for the transduction experiments.

The protoplasts are exposed to a pseudotyped retroviral vector (e.g., a lentiviral vector) encoding a protein of interest driven by a plant specific promoter. The vector is prepared as
described above and is used at an MOI of 1,000. The protoplasts are then resuspended in fresh K₃ medium (0.3 ml protoplast solution in 10 ml of fresh K₃ medium). Further incubation is carried out in 10 ml portions in 10 cm diameter petri dishes at 24°C in the dark, the population density being from 4 to 8x10⁴ protoplasts per ml. After 3 days, the culture medium is diluted with 0.3 parts by volume of K₃ medium per dish and incubation is continued for a further 4 days at 24°C and 3000 lux of artificial light. After a total of 7 days, the clones that have developed from the protoplasts are embedded in nutrient medium that contains 50 mg/l of kanamycin and has been solidified with 1% agarose, and are cultured at 24°C in the dark in accordance with the "bead-type" culturing method (Shillito, et al., Plant Cell Reports, 2, 244-247 (1983)). The nutrient medium is replaced every 5 days by a fresh amount of the same nutrient solution. Analysis of the clones indicates that express the gene of interest.

Example 25
Stability of Vector Insertions in Cell Lines Over Time

Two cell lines that contain gene inserts of the LN-CMV-Bot vector were analyzed for their ability to maintain the vector inserts over a number of passages with and without neomycin selection. The first cell line is a bovine mammary epithelial cell line that contains a low number of insert copies. The second cell line is a 293GP line that contains multiple copies of the vector insert. At the start of the experiment, cell cultures were split. This was at passage 10 for the bovine mammary epithelial cells and passage 8 for the 293GP cells. One sample was continually passaged in media containing the neomycin analog G418, the other culture was continually passaged in media without any antibiotic. Every 3-6 passages, cells were collected and DNA was isolated for determination of gene ratio using the INVADER assay. Cell were continually grown and passaged in T25 flasks. The results of the assays are shown below:

<table>
<thead>
<tr>
<th>Cell Line and Treatment</th>
<th>Passage Number</th>
<th>INVADER Gene Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMEC.Bot #66 + G418</td>
<td>10</td>
<td>0.67</td>
</tr>
<tr>
<td>Cell Line and Treatment</td>
<td>Passage Number</td>
<td>INVADER Gene Ratio</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>BMEC/Bot #66 - G418</td>
<td>10</td>
<td>0.89</td>
</tr>
<tr>
<td>BMEC/Bot #66 + G418</td>
<td>16</td>
<td>0.67</td>
</tr>
<tr>
<td>BMEC/Bot #66 - G418</td>
<td>16</td>
<td>0.64</td>
</tr>
<tr>
<td>BMEC/Bot #66 + G418</td>
<td>21</td>
<td>0.62</td>
</tr>
<tr>
<td>BMEC/Bot #66 - G418</td>
<td>21</td>
<td>0.58</td>
</tr>
<tr>
<td>BMEC/Bot #66 + G418</td>
<td>27</td>
<td>0.98</td>
</tr>
<tr>
<td>BMEC/Bot #66 - G418</td>
<td>27</td>
<td>0.56</td>
</tr>
<tr>
<td>BMEC/Bot #66 + G418</td>
<td>33</td>
<td>0.80</td>
</tr>
<tr>
<td>BMEC/Bot #66 - G418</td>
<td>33</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 12

<table>
<thead>
<tr>
<th>Cell Line and Treatment</th>
<th>Passage Number</th>
<th>INVADER Gene Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>293GP/Bot #23 + G418</td>
<td>8</td>
<td>3.46</td>
</tr>
<tr>
<td>293GP/Bot #23 - G418</td>
<td>8</td>
<td>3.73</td>
</tr>
<tr>
<td>293GP/Bot #23 + G418</td>
<td>14</td>
<td>3.28</td>
</tr>
<tr>
<td>293GP/Bot #23 - G418</td>
<td>14</td>
<td>3.13</td>
</tr>
<tr>
<td>293GP/Bot #23 + G418</td>
<td>17</td>
<td>3.12</td>
</tr>
<tr>
<td>293GP/Bot #23 - G418</td>
<td>17</td>
<td>2.91</td>
</tr>
<tr>
<td>293GP/Bot #23 + G418</td>
<td>22</td>
<td>3.6</td>
</tr>
<tr>
<td>293GP/Bot #23 - G418</td>
<td>22</td>
<td>2.58</td>
</tr>
<tr>
<td>293GP/Bot #23 + G418</td>
<td>28</td>
<td>2.78</td>
</tr>
<tr>
<td>293GP/Bot #23 - G418</td>
<td>28</td>
<td>3.44</td>
</tr>
<tr>
<td>293GP/Bot #23 + G418</td>
<td>36</td>
<td>2.6</td>
</tr>
<tr>
<td>293GP/Bot #23 - G418</td>
<td>36</td>
<td>2.98</td>
</tr>
</tbody>
</table>

These data show that there are no consistent differences in gene ratio between cells treated with G418 and those not treated with antibiotic. This suggests that G418 selection is not necessary to maintain the stability of the vector gene insertions. Also, these vector inserts appear to be very stable over time.
Example 26
Transduction in the Absence of Selectable Marker

This example describes the transduction of host cells with a retroviral construct comprising a gene of interest and lacking a selectable marker. The retroviral vector utilized expresses the gene of interest from the CMV promoter (LTR-Extended Viral Packaging Region-Neo Gene-CMV Promoter-Gene of Interest-WPRE-LTR). A Neo (-) version was constructed by removing the Neo gene with a BsaBI/NruI restriction digest, followed by re-ligation.

A. VIP Co-Transfection

The Vector Initial Production (VIP) method was utilized to generate host cells expressing the gene of interest. This method utilized initial co-transfection of the plasmid encoding the gene of interest and pHCMV-G DNA into 293GPSD cells to produce pseudotyped virus. The procedure for producing pseudotyped virus was carried out as described (Yee et al., Meth. Cell Biol. 43:99 [1994]).

Approximately 16 T150 Flasks were seeded with 293GPSD cells such that the cells were 70-90% confluent on the day of VIP co-transfection. The media in the 293GPSD flasks were changed with harvest medium 2 hours prior to transfection. 293GPSD cells were then co-transfected with 864 pg of plasmid DNA and 864 µg of VSV-G plasmid DNA using the standard calcium phosphate co-precipitation procedure (Graham and Van der Eb, Virol. 52:456 [1973]). Briefly, pHCMV-G DNA, construct DNA, 1:10 TE, and 2M CaCl₂ were combined and mixed. 2X HBS (37°C) was placed into a separate tube. While bubbling air through the 2X HBS, the DNA/1:10 TE/2M CaCl₂ mixture was added drop wise. The transfection mixture was allowed to incubate at room temperature for 20 minutes. Following the incubation period, the correct amount of transfection mixture was added to each culture vessel. The plates or flasks were returned to 37°C, 5% CO₂ incubator for approximately six hours. Following the incubation period, the transfections were checked for the presence of crystals/precipitate by viewing under an inverted scope. The transfection media was then removed from culture vessels by aspiration with a sterile Pasteur pipet and vacuum pump and fresh harvest medium was added to each
culture vessel. The culture vessels were incubated at 37°C, 5% CO₂ for 36 hr. Vector was then concentrated as described in Example 27.

B. Generation of Host Cells Expressing the Gene of Interest

The culture medium containing virus encoding the gene of interest was used to infect the 293 cells as follows. Cells were grown in the absence of Neo selection during all stages of the infection, growth, and clonal selection. 200 µl containing 1000-5000 cells of a diluted (dilutions were made in media containing polybrene at a final concentration of 8 µg/ml) 293 cell suspension were plated in 2-6 wells of a 96 well plate. Cells were incubated at 37°C & 5% CO₂ for 1-4 hours until cells have plated. The media was removed and 50-100 µl of concentrated vector was added to the desired number of wells. Cells were incubated at 37°C & 5% CO₂ for 1 hour. Media containing polybrene was added back to a final volume of 200 µl. Cells were incubated at 37°C & 5% CO₂ overnight. At 30-40% confluency, wells were pooled and passaged to 6 well plate and subsequently T25.

Cells were then diluted into 96 well plates to a concentration of one cell per well in order to perform clonal selection. Cells from the T25 flasks were counted and then diluted to 5 cells/ml. 200 µl of the diluted solution was added to each well of a 96 well plate. Plates were incubated at 37°C & 5% CO₂ until cell are confluent and are then screened for protein production using ELISA.

C. Results

Copy number was determined using the method described in Example 19 above. The top 24 clones were chosen based on ELISA assay from cultures in 96 well plates. The clones were expanded to 6 well and then T25 flasks. The productivity per day was determined by ELISA assay and the top 10 clones were expanded to T150 and frozen.

Figure 20 and Table 13 show the results of this experiment. Cell lines derived from colony number 13, which lacked a selectable marker, shows an expression level of 3 pg/cell/day. The other cell lines containing a copy number of 1 (colonies 14A, 37, and 40) showed a lower level of expression. This example demonstrates that cell lines derived from integrated vectors lacking a selectable marker and grown under non-selective conditions a) express protein from an
exogenous gene, and B) express protein at a higher level than in the presence of a selectable marker.

<table>
<thead>
<tr>
<th>Colony #</th>
<th>pg/cell/day copy number</th>
</tr>
</thead>
<tbody>
<tr>
<td>14A</td>
<td>1.14</td>
</tr>
<tr>
<td>61(2 copy)</td>
<td>1.9</td>
</tr>
<tr>
<td>13(Neo-)</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0.7</td>
</tr>
<tr>
<td>11</td>
<td>1.5</td>
</tr>
<tr>
<td>15</td>
<td>1.3</td>
</tr>
<tr>
<td>17</td>
<td>4.6</td>
</tr>
<tr>
<td>28</td>
<td>0.9</td>
</tr>
<tr>
<td>29</td>
<td>0.92</td>
</tr>
<tr>
<td>32</td>
<td>1.9</td>
</tr>
<tr>
<td>37</td>
<td>0.52</td>
</tr>
<tr>
<td>40</td>
<td>2.61</td>
</tr>
<tr>
<td>43</td>
<td>4.3</td>
</tr>
<tr>
<td>45</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Example 27

Concentration of Pseudotyped Retroviral Vectors

The VSV G-pseudotyped viruses were concentrated to a high titer by one cycle of ultracentrifugation. However, in certain embodiments, two cycles are performed for further concentration. The culture medium collected and filtered as described in Example 26 which contained pseudotyped virus was transferred to Oakridge centrifuge tubes (50 ml Oakridge tubes with sealing caps, Nalge Nunc International) previously sterilized by autoclaving. The virus was sedimented in a JA20 rotor (Beckman) at 48,000 x g (20,000 rpm) at 4°C for 120 min. The culture medium was then removed from the tubes in a biosafety hood and the media remaining in the tubes was aspirated to remove the supernatant. The virus pellet was resuspended to 0.5 to 1% of the original volume in 0.1X HBSS. The resuspended virus pellet was incubated overnight at 4°C without swirling. The virus pellet could be dispersed with gentle pipetting after the
overnight incubation without significant loss of infectious virus. The titer of the virus stock was routinely increased 100- to 300-fold after one round of ultracentrifugation. The efficiency of recovery of infectious virus varied between 30 and 100%.

The virus stock was then subjected to low speed centrifugation in a microfuge for 5 min at 4°C to remove any visible cell debris or aggregated virions that were not resuspended under the above conditions. It was noted that if the virus stock is not to be used for injection into oocytes or embryos, this centrifugation step may be omitted.

In some embodiments, the virus stock is subjected to another round of ultracentrifugation to further concentrate the virus stock. The resuspended virus from the first round of centrifugation is pooled and pelleted by a second round of ultracentrifugation that is performed as described above. Viral titers are increased approximately 2000-fold after the second round of ultracentrifugation.

Amplification of retroviral sequences in co-cultures may result in the generation of replication competent retroviruses, thus affecting the safety of the packaging cell line and vector production. Therefore, the cell lines were screened for production of replication competent vector. The 208F cells were expanded to approximately 30% confluency in a T25 flask (~ 10^5 cells). The cells were then infected with 5 ml of infectious vector at 10^5 CFU/ml + 8 ug/ml polybrene and grown to confluence (~24 h), followed by the addition of media supplemented with G418. The cells were then expanded to confluence and the media collected. The media from the infected cells was used to infect new 208F cells. The cells were plated in 6-well plates at 30% confluency (~ 10^6 cells) using the following dilutions: undiluted, 1:2, 1:4, 1:6, 1:8, 1:10. Cells were expanded to confluence, followed by the addition of G418. The cells were then maintained under selection for 14 days to determine the growth of any neo resistant colonies, which indicate the presence of replication competent virus.

### Example 28

**Cell Line Stability Analysis of GPEx Created CHO Cell Lines**

This example describes a comparison of cell line stability in the presence and absence of selection.
A. Methods

Two T75 flasks per cell line were set up for the stability test: one in the presence of selection (G418) and one without selection. The seed for each set of T75s was a T150 of each cell line in log phase. One ml from each T150 was used to inoculate into 9 mls of PFCHO media (HyClone, Ogden, UT) (non-selected) and into PFCHO+G418 (400 µg/ml). Every 2-3 days 1 ml of media was collected for protein determination and cell counts. Media samples were kept at -20°C for the duration of the experiment. Cells were then passaged 1:10 into new flasks. The assay was terminated after completion of 40 generations. All the media samples collected over the 40 generations for each cell line were then assayed on the same ELISA plate for protein expression. Protein production was measured in picograms/cell/day. The analysis was performed on five cell lines (#1, 42, 137, 195 and 233). Protein assays were performed using an ELISA assay. Cell counting was performed using an Innovatis Cedex Model AS20 using manufacturers recommended procedures. The data is shown below.

B. Results

<table>
<thead>
<tr>
<th>Sample Collection Date</th>
<th>Productivity (pg/cell/day) Cells Grown in G418</th>
<th>Productivity (pg/cell/day) Cells Grown without G418</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/30</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>12/3</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>12/6</td>
<td>0.18</td>
<td>0.25</td>
</tr>
<tr>
<td>12/10</td>
<td>0.28</td>
<td>0.10</td>
</tr>
<tr>
<td>12/13</td>
<td>0.40</td>
<td>0.08</td>
</tr>
<tr>
<td>12/16</td>
<td>0.86</td>
<td>0.10</td>
</tr>
<tr>
<td>12/19</td>
<td>0.64</td>
<td>0.05</td>
</tr>
<tr>
<td>12/23</td>
<td>1.05</td>
<td>0.10</td>
</tr>
<tr>
<td>12/26</td>
<td>0.98</td>
<td>0.13</td>
</tr>
<tr>
<td>12/30</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>1/3</td>
<td>0.77</td>
<td>0.25</td>
</tr>
<tr>
<td>Sample Collection Date</td>
<td>Cell Line #42:</td>
<td>Productivity (pg/cell/day)</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cells Grown in G418</td>
</tr>
<tr>
<td>11/30</td>
<td>0.25</td>
<td>0.39</td>
</tr>
<tr>
<td>12/3</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>12/6</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>12/9</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>12/12</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>12/16</td>
<td>0.23</td>
<td>0.43</td>
</tr>
<tr>
<td>12/19</td>
<td>0.44</td>
<td>0.37</td>
</tr>
<tr>
<td>12/23</td>
<td>0.29</td>
<td>0.20</td>
</tr>
<tr>
<td>12/26</td>
<td>0.36</td>
<td>0.47</td>
</tr>
<tr>
<td>12/30</td>
<td>0.35</td>
<td>0.27</td>
</tr>
<tr>
<td>1/3</td>
<td>0.33</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Cell Line #42:

<table>
<thead>
<tr>
<th>Sample Collection Date</th>
<th>Cell Line #137:</th>
<th>Productivity (pg/cell/day)</th>
<th>Productivity (pg/cell/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cells Grown in G418</td>
<td>Cells Grown without G418</td>
</tr>
<tr>
<td>11/30</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>12/3</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>12/6</td>
<td>0.09</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>12/10</td>
<td>0.13</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>12/13</td>
<td>0.10</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>12/16</td>
<td>0.14</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>12/19</td>
<td>0.16</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Productivity (pg/cell/day)</td>
<td>Productivity (pg/cell/day)</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cells Grown in G418</td>
<td>Cells Grown without G418</td>
<td></td>
</tr>
<tr>
<td>12/23</td>
<td>0.19</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>12/26</td>
<td>0.17</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>12/30</td>
<td>0.17</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>1/3</td>
<td>0.36</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>1/6</td>
<td>0.24</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>1/9</td>
<td>0.11</td>
<td>0.09</td>
<td></td>
</tr>
</tbody>
</table>

Cell Line #195:

<table>
<thead>
<tr>
<th>Sample Collection Date</th>
<th>Productivity (pg/cell/day)</th>
<th>Productivity (pg/cell/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cells Grown in G418</td>
<td>Cells Grown without G418</td>
</tr>
<tr>
<td>11/30</td>
<td>1.03</td>
<td>0.30</td>
</tr>
<tr>
<td>12/3</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>12/6</td>
<td>0.18</td>
<td>0.22</td>
</tr>
<tr>
<td>12/10</td>
<td>0.23</td>
<td>0.39</td>
</tr>
<tr>
<td>12/13</td>
<td>0.24</td>
<td>0.77</td>
</tr>
<tr>
<td>12/16</td>
<td>0.18</td>
<td>0.76</td>
</tr>
<tr>
<td>12/19</td>
<td>0.37</td>
<td>0.85</td>
</tr>
<tr>
<td>12/23</td>
<td>0.73</td>
<td>0.30</td>
</tr>
<tr>
<td>12/26</td>
<td>1.03</td>
<td>0.51</td>
</tr>
<tr>
<td>12/30</td>
<td>0.97</td>
<td>0.27</td>
</tr>
<tr>
<td>1/3</td>
<td>0.54</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Cell Line #233:

<table>
<thead>
<tr>
<th>Sample Collection Date</th>
<th>Productivity (pg/cell/day)</th>
<th>Productivity (pg/cell/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cells Grown in G418</td>
<td>Cells Grown without G418</td>
</tr>
<tr>
<td>11/30</td>
<td>0.38</td>
<td>0.17</td>
</tr>
<tr>
<td>12/3</td>
<td>0.12</td>
<td>0.04</td>
</tr>
<tr>
<td>12/6</td>
<td>0.15</td>
<td>0.13</td>
</tr>
</tbody>
</table>
To determine whether neo selection had an effect on protein expression over 40 generations, analysis of variance was performed on the data. The model included the following variables: antibiotic selection, line, generation and interations between each variable. The data indicate that there was no effect of including G418 in the media (p>0.10) on cell productivity over the 40 generations. The p-values for each cell line are shown in the table below. There was also no significant decrease in cell productivity over time in any of the cell lines grown with or without G418.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>42</td>
<td>0.29</td>
</tr>
<tr>
<td>137</td>
<td>0.15</td>
</tr>
<tr>
<td>195</td>
<td>0.53</td>
</tr>
<tr>
<td>233</td>
<td>0.27</td>
</tr>
</tbody>
</table>

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology,
protein fermentation, biochemistry, or related fields are intended to be within the scope of the following claims.
The claims defining the invention are as follows:

1. A method for transducing host cells comprising:
   a) providing:
      i) at least one host cell comprising a genome, and
      ii) a plurality of retroviral vectors encoding a gene of interest operably linked to a promoter internal to the retroviral 5' and 3' long terminal repeats and a segment encoding a secretion signal sequence;
   b) contacting said at least one host cell with said plurality of retroviral-vectors under conditions such that said host cells are transduced to produce transduced host cells, wherein said conditions comprise contacting said host cell at a multiplicity of infection of from about 10 to 1000;
   c) repeating steps a) and b) a plurality of times to provide host cells comprising from about 20 to about 100 integrated retroviral vectors, wherein said host cell does not produce infectious retroviral vectors.

2. The method of Claim 1, wherein steps a) and b) are repeated at least 3 times.

3. The method of Claim 1, wherein steps a) and b) are repeated at least 4 times.

4. The method of Claim 1, wherein steps a) and b) are repeated at least 5 times.

5. The method of Claim 1, wherein steps a) and b) are repeated at least 6 times.

6. The method of Claim 1, wherein steps a) and b) are repeated at least 7 times.

7. The method of Claim 1, wherein steps a) and b) are repeated at least 8 times.

8. The method of Claim 1, wherein steps a) and b) are repeated at least 10 times.

9. The method of Claim 1, wherein steps a) and b) are repeated at least 20 times.

10. The method of Claim 1, wherein steps a) and b) are repeated between about 3 and 20 times.
11. The method of any one of claims 1 to 10 wherein said retroviral vectors utilized in steps a) and b) are produced from packaging cells transfected with an envelope plasmid and a vector plasmid.

12. The method of any one of Claims 1 to 11, further comprising step:

d) transducing said host cells comprising multiple integrated retroviral vectors produced by steps a) and b) with vectors produced from packaging cells produced by transducing said packaging cells with a retroviral vector encoding said gene of interest and transfecting said packaging cell with a plasmid expressing an envelope protein.

13. The method of Claim 11, wherein said packaging cells express retroviral gag and pol proteins.

14. The method of Claim 13, wherein said packaging cells are 293-GP cells.


16. The method of Claim 15, wherein said G protein is VSV-G protein.

17. The method of any one of Claims 1 to 16, wherein said retroviral vector comprises MoMLV elements.

18. The method of any one of Claims 1 to 17, wherein gene of interest is operably linked to a segment encoding a signal sequence.

19. The method of any one of Claims 1 to 18, wherein said retroviral vector encodes at least two genes of interest.

20. The method of Claim 19, wherein said at least two genes of interest are arranged in a polycistronic sequence.

21. The method of Claim 20, wherein said at least two genes of interest comprise immunoglobulin heavy and light chains.
22. The method of any one of Claims 1 to 21, wherein said retroviral vector is a lentiviral vector.

23. The method of any one of Claims 1 to 22, wherein said host cell is selected from Chinese hamster ovary cells, baby hamster kidney cells, human 293 cells, and bovine mammary epithelial cells.

24. The method of any one of Claims 1 to 23, further comprising clonally selecting said transduced host cells.

25. The method of Claim 24, further comprising culturing said clonally selected host cells under culture conditions such that a protein of interest encoded by said gene of interest is produced.

26. The method of Claim 25, wherein said culture conditions are selected from the group consisting of roller bottle cultures, perfusion cultures, batch fed cultures, and petri dish cultures.

27. The method of any one of claims 24 to 26, wherein said clonally selected host cells synthesize greater than about 1 picograms per cell per day of said protein of interest.

28. The method of any one of claims 24 to 26, wherein said clonally selected host cells synthesize greater than about 10 picograms per cell per day of said protein of interest.

29. The method of any one of claims 24 to 26, wherein said clonally selected host cells synthesize greater than about 50 picograms per cell per day of said protein of interest.

30. The method of any one of Claims 1 to 29, wherein said retroviral vector further encodes an amplifiable marker.

31. The method of Claim 30, wherein said amplifiable marker is selected from the group consisting of DHFR and glutamine synthetase.
32. The method of Claim 30 or Claim 31, further comprising the step of culturing said transduced host cells under conditions that allow for amplification of the integrated retroviral vectors.

33. The method of Claim 32, wherein said conditions comprise culturing said transduced host cells in the presence of a selection agent selected from the group consisting of methotrexate, phosphinothricin and methionine sulfoximine.

34. The method of Claim 21, wherein said immunoglobulins are selected from the group consisting of IgG, IgA, IgM, IgD, IgE and sIg.

35. The method of any one of claims 1 to 34, wherein said host cell is transduced with at least two different vectors encoding different genes of interest.

36. A transduced host cell produced by the method of Claim 1, wherein said host cell comprises from about 20 to about 100 integrated retroviral vectors.

37. A method for transducing host cells comprising:
   a) providing:
      i) at least one host cell comprising a genome, and
      ii) a plurality of retroviral vectors encoding a gene of interest; and
   b) contacting said at least one host cell with said plurality of integrating vectors under conditions such that said host cells are transduced to produce transduced host cells;
   c) repeating steps a) and b) a plurality of times to provide host cells comprising greater than about 20 integrated retroviral vectors;
   d) clonally selecting a host cell expressing said gene of interest; and
   e) purifying a protein of interest encoded by said gene of interest.

38. A transduced host cell produced in accordance with the method of Claim 37, wherein said host cell comprises greater than about 20 integrated retroviral vectors.

Dated 11 May, 2009
Catalent Pharma Solutions, Inc.
Patent Attorneys for the Applicant/Nominated Person
SPRUSON & FERGUSON
Figure 4
SEQ ID NO:1
Hybrid Human-Bovine Alpha-Lactalbumin Promoter

1 GATCAAGCTTGGGGTCTCTGTACATGGAAGACTACGTGGCAGGCTGAGACCTTCC
51 AATACCTTTGCGCCACCTGATGCACGAGGAGAAGGCTGACTTCATAGTTGGTTAA
101 TATCTGGGAAAGAGTTGAGCAGGAGGAGTAGGAGGAGGAAGGCTGACACTGAT
151 GGTGCATATGAGATGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
201 ATGAGGGAGGATGGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
251 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
301 TCATCTTTTGGCGGATTGGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
351 GAGAAGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
401 ATTCCTTTTTTGGCGGATTGGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
451 TGAAGGTTCCTTTTTGGCGGATTGGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
501 TAAAGGTTCCTTTTTGGCGGATTGGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
551 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
601 GTGGACGCTCTTGATGGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
651 TTGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
701 GGACTCTTTTTTTTGCGGATTGGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
751 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
801 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
851 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
901 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
951 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1001 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1051 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1101 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1151 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1201 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1251 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1301 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1351 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1401 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1451 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1501 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1551 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1601 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1651 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1701 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1751 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1801 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1851 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1901 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
1951 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
2001 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
2051 TGGCGAGATGGGAGGAGGAGGAGGAGGAGGAGGAAGGCTGACACTGATCTA
2101 C

1 - 1525   Bovine alpha lactalbumin 5' flanking region (-2000 to -550 from the bovine alpha-lactalbumin transcription start point)
1526 - 2056  Human alpha-lactalbumin 5' flanking region (-600 to +15 from the human alpha-lactalbumin transcription start point)
2057 - 2101   Multiple cloning site
Figure 5
SEQ ID NO:2
Mutated PPE Sequence

1  GATTACTTACTGCCAGGTGCTGGGGCTTCCGAGACATCGGAAACATCT
51  ACACCACAACACCGCCTCCGGACGGGTGAGATATCCGCAGGAGCCG
101  GCGGCTGTTATTACTAGCGGAGGATCCGATTACTTACTGGCAGGCTGGGG
151  GCGCTTCGAGACTCGGGGGAATCTACCGACACACACAGCCCTCTGCCC
201  AGGGTGGATATCTGCGGGAGCGGCGGCGGCTGTAATTACACGG

1 - 119  Mutated PPE
120 - 126  Linker
127 - 245  Mutated PPE
Figure 6
SEQ ID NO:3
IRES-Signal Peptide Sequence

1  GGAATTCTGACCTCTCCCCCCCCCCACATGGTGCTGGGAGGAA
51  CTTGGAAATAGGAGGCGGTGCCCTCCTTATATGTGCTCCTCCATTT
101  TGCAGGCTTTTGGCAATGGGAGGCGGAGAAGCTGGCCCCCCTATGTT
151  AGGACGTCTTCTAGGGGCTTTCCTGTGACCAAGGAGCTGACCTG
201  GGTGGAATTGCCTGAAGGAGAGCTTCTTCTGGAGCCTTGGAGA
251  CACAGGTCTCTGACCCCCTTTGGAGCAAGGAGGACCCCCTACCTGGGGC
301  AGGTGCTCTGGCCGCCAAGGAGCAGCGTATATAGGATACAGCTTGGCAAGGGG
351  GGCACACCCCGAGCTGCAGGTGACGTTGAGGTAGTTGTTGGAAGGACTCA
401  AATGCTCTCTCTCAAGGGTTATGCACTACAGGCTGCTGGAGCTGGCA
451  GATCCCGATTTATATGTGCTAGTCTCTGGGCTCCCTGCAGTGGCTCTTC
501  ATCTATTTATAGGAAGGGTTTTAAAACAGGTGTTAAACCTGCTAGGGCCCGGAAACCCAGGG
551  AGGTGCTCTTCTCTTGTTGGAAAACAGGTGTTAAACCTGCTAGGGCCCGGAAACCCAGGG
601  TCTGCTCTGTGAGACCTTTATCCATATGCCAACCCGGCGGGCGCCATTG
651  GNTATCTCGAGCTCGAGGAGCTT

1 - 583  IRES
584 - 640  Modified bovine alpha-lactalbumin signal peptide coding region
641 - 680  Multiple cloning site
CMV promoter/enhancer

MN14 antibody heavy chain gene signal peptide start codon

MN14 antibody heavy chain gene start codon

EMCV IRES

Bovine alpha-lactalbumin signal peptide start codon

First codon mature MN14 antibody light chain gene

MN14 antibody light chain gene stop codon

MoMuLV 3' LTR
Figure 6b

2851 CTCTCTGTCCTCCTGTAGGGCACTTCTATTTCCATGTGCACACGGCGGACATCC
2901 AGCTGAGCGAGATGCTCTAACGACTTCGGAGTTGGAGATAGGGTC
2951 ACTATAGCGCTGTAAGTCCAGTTCAAGGTGTTTTTATACGTCGAAATCAAC
3001 GAACATGCTTGGCTTGCAATCCCGAAGAAACCCGGAATAGCGACTGAACCTG
3051 GATGACAATGGGACATGCACATAGCGAGTTGCTCTCCTGATGGATTTCTCT
3101 GGCACAGCGAGTCTGGGACAGATTTTTACTTTTACACTACAGCTCCTCTCAACC
3151 AGAGACACATGCAACATATTATTGTCACCAATACCTCCTCTGCTCTGGACGGT
3201 TCCTGCGAGGGGACACRAAGATGTCGCMYCAAAAGAACACTGGTGCTGTCGCCACTCTT
3251 GCTCTGAATCCTTGGGGACTCTGATGCTGATGGGCTGATGGGGTCTCCGCTCCGCTCC
3301 TGTGGTGTGCTCTGCAGTTATACCTCTATTATCCGAGAGAAGCGAAATAGACGT
3351 GGAGGTGGATACGCGCCCTCCTTCATTGCTGGGAACTTCCCGAAGAGATGCTCAAC
3401 GGCAGGACAGCAAGCACACACCTACAGCCCTACGAGACACCCCTGAGCGCTCC
3451 GGCAGGACAGCAAGCACACACCTACAGCCCTACGAGACACCCCTGAGCGCTCC
3501 ATCAGGCTTTAGGACCTCGGCAGCTCAAAAGCGAAGCTAGGGGAGAGGGT
3551 TAGAGGTCTATGGGCTCTCTAGTCGACATGTGATATAAAAGAAGTATTATTT
3601 TAGGCTTGGTTGAAAGGCGGATAGGGGAAATGGAAGGGCCGCCCTTGAGCTGGTGGC
3651 AGCTATGTCTAAGTACCGGTTTTTGCAAGAGCAGGGAAAATGACGTAACGTC
3701 AGAAGACAGAGATGTCGCTACGATACGAGCATGACGGACAGTGAACGCCGCTGAGATA
3751 TGCGTCAACACAGGAAATCTCTGCTGAMCGACTGCGCCGGCTCAAGGC
3801 CAGGAGAGAATGACGACACGCCACCTACTAGGGCCGAAACGGGATACTGCTGGTA
3851 AGCAGGTCCTCGGCCGCCGCTCAAGGACAGAAAGTGGCCCCACGAGCG
3901 GCCACGGCTCCAGACGGTTCTAGAGAACATCAGATTTCCAAGGCTGGCC
3951 CCAGGACCGTAAATGAGCCTGGCTTTTGAGCTACTACCAACTACAGTCT
4001 GCCCTCTGGCTCTTGCTGGGCGCCCTCCTGCAGTTGACTGACGTCGCC
4051 CCGAAGCCACCCTCTCAGGGGGCGCCAGTCTCGGCACTGATACAGCGG
4101 GCTGCCGCTTGGGCTCTCTGGGAGGCTCTCCTTGGATTTGGATTCTACCTCCGCTCAAG
4151 TCCGGTGCTTGGGCTCTCTGGGAGGCTCTCCTTGGATTTGGATTCTACCTCCGCTCAAG
4201 GTCTTTGAGTT

1 - 812  CMV promoter/enhancer
852 - 854  LL2 antibody heavy chain signal peptide start codon
8247 - 2249  LL2 antibody heavy chain stop codon
2261 - 2836  EMCV IRES
2837 - 2839  Bovine alpha-lactalbumin signal peptide start codon
2894-2896  First codon of mature LL2 antibody light chain gene
3551 - 3553  LL2 antibody light chain gene stop codon
3622 - 4210  MoMuLV 3' LTR
Figure 9b

2851 CACAGGTTAACCACTCCGGATCAGGAGAACGTTACCAAAGAAGCAG
2901 GTCAAGCTGGACCTGGCGGAAACGAGGGATGCGAAGGACCCAG
2951 GAGAGGAGCCAACTGGCCCGGAGAAGGGGAGGACGAGCGATAG
3001 CAGGCGGTCCAACCACTGCAGGCAAAGGGCGGCGGACAACCTG
3051 GAGACAGGAGGCGGCAGGACAGGAGGAGGAGGAGGAGGACCTG
3101 CGAGGCTTCTTGTCCACTTTGTGCTTTTACTTTTTTCTGCTG
3151 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3201 TGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3251 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3301 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3351 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3401 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3451 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3501 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3551 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3601 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3651 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3701 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3751 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3801 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3851 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3901 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3951 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4001 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4051 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4101 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4151 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4201 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4251 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4301 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4351 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4401 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4451 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4501 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4551 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4601 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4651 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4701 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4751 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4801 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4851 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4901 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
4951 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5001 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5051 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5101 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5151 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5201 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5251 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5301 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5351 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5401 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5451 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5501 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5551 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5601 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5651 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5701 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG

1 - 1457  Mouse mammary tumor virus LTR
1475 - 1726  Double mutated PPE sequence
Figure 9c

1752 - 1754   MN14 heavy chain signal peptide start codon
3156 - 3158   MN14 heavy chain stop codon
3170 - 3745   EMCV IRES
3746 - 3748   Bovine alpha-lactalbumin signal peptide start codon
3803 - 3805   First codon of mature MN14 light chain gene
4442 - 4444   MN14 antibody light chain gene stop codon
4487 - 5078   WPRE sequence
5133 - 5372   MoMuLV 3' LTR
Figure 10d

9051  CGATTGACTGAGGCCGGGTCACCGTGATCCCAATATAACCTCTCTGCA
9101  GTTGNACTCCGACTGAGGCCGGGTCACCGTGATCCCAATATAACCTCTCTGCA
9151  TGATTGACTCCGACTGAGGCCGGGTCACCGTGATCCCAATATAACCTCTCTGCA

1 - 658  MoMuSV 5' LTR
659 - 1468  Extended packaging region
1512 - 2306  Neomycin resistance gene
2661 - 4396  Bovine/human alpha-lactalbumin 5' flanking region
5084 - 5327  Double mutated PPE sequence
6207 - 6209  MN14 antibody heavy chain gene signal peptide start codon
6611 - 6613  MN14 antibody heavy chain stop codon
6625 - 7200  EMCV IRES
7201 - 7203  Bovine alpha-lactalbumin signal peptide start codon
7258 - 7260  First codon of mature MN14 antibody light chain gene
7897 - 7899  MN14 antibody light chain gene stop codon
7938 - 8529  WPRE sequence
8600 - 9138  Moloney murine leukemia virus 3' LTR
Figure 11c

5651 GCTGTTCCCTTGAGGGTCTCCTGAGTGGATTGGACGTCGAGCCGGG
5701 GGTCTTTTCATT

1 - 2053  Bovine/human alpha-lactalbumin 5' flanking region
2093 - 2336  Double mixed PPE sequence
2387 - 2443  cc49 signal peptide coding region
2444 - 3088  Bot antibody light chain Fab coding region
3112 - 3686  EMCV IRES
3687 - 3745  Bovine alpha-lactalbumin signal peptide coding region
3746 - 4443  Bot antibody heavy chain Fab coding region
4481 - 5072  WPRE sequence
5118 - 5711  Moloney murine leukemia virus 3' LTR
Figure 12b

2851 GAAGCCCTCGCAAGTTAACCTGATGTTCTTCTGCGCCGCAAGGAAGCTGAT
2901 CCACAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
2951 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3001 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3051 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3101 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3151 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3201 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3251 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3301 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3351 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3401 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3451 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3501 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3551 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3601 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3651 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3701 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3751 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3801 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3851 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3901 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
3951 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4001 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4051 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4101 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4151 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4201 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4251 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4301 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4351 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4401 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4451 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4501 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4551 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4601 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4651 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4701 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4751 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4801 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4851 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4901 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
4951 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
5001 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
5051 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC
5101 CAGCAGGAGCTGATACGGAAGGAGCAGATAGATGAGATGGTTCGAC

L - 589
659 - 897 Retroviral packaging region
1034 - 1714 Hepatitis B surface antigen
2279 - 2595 RSV promoter
2951 - 3745 Neomycin phosphotransferase gene
4537 - 5130 MoMuLV 3' LTR
Figure 13b

2851 ACGTCTGGTCCTGGGACCAAGCTGGCTGGTGAAGGCGGCCGAGCCCAA
2901 ATCTCAGACAAAATCAGAGATGCCACGTCGCTGGCAAGCTGAGACCTG
2951 TGAGATCGGGACCGGGAGTAACTTCGAGCTGCTGAGAGTGGGAGAACTC
3001 ATGTCTCCGGAGGCCGAGGCGAACTCAGGCTGGCGCTGAGCTGGAGAG
3051 CGAAGACCCCTGAGGTTCAACTCGACGTGCTGGTGGAGCGCCCTGAGG
3101 ATATACGCAAAAGCAAGTGGGGCGAGCAATATCAGGGGCAGCGCTGCTG
3151 GCTGCCGCGTTCGAATCAGGATCGGCTGAGCAAGCTGAGACCTG
3201 CGGCGCGGCGGCGGGAGGCTGAGGATAAACGCGGCGGCGGCGGCGGCG
3251 CGGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
3301 CCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
3351 GTATATCTGGTGGGCAAGGCTGGCTGGGAGGCGAGAACCTGAGAGAC
3401 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
3451 CTGGGCGGCCGAGAAGCGGCTGAGGATCGGCTGAGCAAGCTGAGACCTG
3501 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3551 ACTATACGCAAAAGCAAGTGGGGCGAGCAATATCAGGGGCAGCGCTGCTG
3601 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3651 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3701 AAATATCTGGTGGGCAAGGCTGGCTGGGAGGCGAGAACCTGAGAGAC
3751 AAGAGCCGAGGCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3801 ACTGCGGGGCGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
3851 CCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
3901 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
3951 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4001 TGGGCGGCCGAGAAGCGGCTGAGGATCGGCTGAGCAAGCTGAGACCTG
4051 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4101 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4151 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4201 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4251 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4301 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4351 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4401 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4451 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4501 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4551 CGGCGGAGGAGGCCGAGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4601 GCGGCGGCGGCGGGCGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGG
4651 GTATATCTGGTGGGCAAGGCTGGCTGGGAGGCGAGAACCTGAGAGAC

1 - 2055 Bovine/human alpha-lactalbumin 5' flanking region
2098 - 4011 cca49-IL2 coding region
4068 - 4661 MoMuLV 3' LTR
Figure 14a
SEQ ID NO: 11
Alpha-Lactalbumin YP Vector
### Figure 14c

<table>
<thead>
<tr>
<th>Figure 14c</th>
<th>5651</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 2053</td>
<td>Bovine/Human Alpha-lactalbumin 5' flanking region</td>
</tr>
<tr>
<td>2093 - 2336</td>
<td>Double mutated PPE sequence</td>
</tr>
<tr>
<td>2403 - 2459</td>
<td>Bovine alpha-lactalbumin signal peptide coding region</td>
</tr>
<tr>
<td>2460 - 3137</td>
<td>Yersinia pestis heavy chain Fab gene coding region</td>
</tr>
<tr>
<td>3167 - 3742</td>
<td>EMCV IRES</td>
</tr>
<tr>
<td>3743 - 3799</td>
<td>Bovine alpha-lactalbumin signal peptide coding region</td>
</tr>
<tr>
<td>3800 - 4441</td>
<td>Yersinia pestis light chain Fab gene coding region</td>
</tr>
<tr>
<td>4461 - 5052</td>
<td>WPRE sequence</td>
</tr>
<tr>
<td>5098 - 5691</td>
<td>Moloney murine leukemia virus 3' LTR</td>
</tr>
</tbody>
</table>
**Figure 15**

SEQ ID NO:12

IRES-Casein Signal Peptide Sequence

| 1  | GGAATTTCGCCCCCTCTCCCTCCCCCCCCCTGTAATGCGGGAAGCCG |
| 51 | CTTGGAATGAGCGGCGGCTGCGGTTTCCTACTAGTATTGTCACCTCAGATA |
| 101| TGGCGTTCACTTTGCAATGAGGCGGCGGAACTGGCCCTGTCTTCG |
| 151| AGCAAGATTTTCTAGGGGTCTTCTGGGGCACCAGAAGCAAGGCTCT |
| 201| GTGGATTGCTGTAAGGGAAGCGCACTTCTCTGGGAAGCTTCTGAGAAAG |
| 251| CAACGCTGCTAGGACCCCTTTGCGGACCGGAAACCCCCACCTGGGAC |
| 301| AGCTGCGCTCGGGCCAAAAGCCACGTATGAGTACACCTGCAAGGCC |
| 351| GGGACAGACCCCAAGTGGCACTTTGAGGTGGTATGGTGAGGGAGAGC |
| 401| AATGCGCTCCTCAAGCGTATTTCAACAAGGCGGCTGGAAGAAGAAAGCCAG |
| 451| CTACCCATTTTCGGATAGGTATCTGCTGCTGCTGCTGCTGCTGCTGCTG |
| 501| AGAGTTTTGTGGTGAATAGTTAAAGAAAGCAGTCTAGGGCCCTGGACCCAGG |
| 551| AGTGCTGTTTCTCTGGAAAACACGNTGATATAATGGCCCTGCTCATCCT |
| 601| TACCTGCTCTGGGGCTTTGCTCTTTGCGCGGGCGGAGGHTAATCTAGGTC |
| 651| TCGAGCTCGGAAGGCTT |

1 - 583  IRES

584 - 628  Modified bovine alpha-S1 casein signal peptide coding region

629 - 668  Multiple cloning site
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TTGGAAGACCCACCCCTAGAGGCTAGCTACTAAGAAAAGCCCTTT</td>
</tr>
<tr>
<td>51</td>
<td>TACGAGGATAGGAGAAACAGCCCTGTCGAGAATGAGAAAGCTT</td>
</tr>
<tr>
<td>101</td>
<td>CCAGGCTGGCTAAAAGAGACCGCGCAACCCGCGCCGCGCG</td>
</tr>
<tr>
<td>151</td>
<td>AGCTTCTGGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>201</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>251</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>301</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>351</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>401</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>451</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>501</td>
<td>GACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
</tbody>
</table>
Figure 16c

Moloney Murine Sarcoma Virus 5' LTR

Moloney Murine Leukemia Virus Extended Packaging Region

Neomycin Resistance Gene

CMV Promoter

ce49 Signal Peptide Coding Region

Bot Fab 5 Light Chain

EMCV IRES (Clontech)

Modified Bovine α-LA Signal Peptide Coding Region

Bot Fab 5 Heavy Chain

Moloney Murine Leukemia Virus 3' LTR

1 - 589

659 - 1468

1512 - 2306

2656 - 3473

3516 - 3572

3573 - 4217

4235 - 4816

4817 - 4873

4874 - 5572

5662 - 6255
Figure 17. CMV construct containing cell lines.
Figure 18: α-Lactalbumin construct containing cell lines

α-Lactalbumin Promoter Cell Lines

\[ y = 893.21x - 164.21 \]

\[ R^2 = 0.9438 \]
Figure 19d

9101 TCATCATTGGAAAAAGTCTTCCGCGGGAAMCTCTCRAAGGATCTTACCG
9151 CTGGTGAGATCCAGTTCCGATGTACCCACACTTGCGACCCACAATCTGATCTTC
9201 AGCATCTTTATTTTCACCAGGTTCCTTGGGGCGCGAAAAAACAGGGAGGCG
9251 AAAAAACGCCCCAARAAAGGCTGAAATGGGCGCGAACGNNAAATGTTGCGAATCTC
9301 ATACTCTTCCTTTTCCTCAATATTTGAAACGATTTTACCTACGGGCTATTTGCCT
9351 CATAAGCCCCGATACATATTTGAAATGATTCTAGTAAATACAAACAAATAGGGG
9401 TTCGGGCCTCTTTTCTTTGGGGAAAAAGCGGACTCTCGTCGTAGACAAAGACATT
9451 ATATATCATGCTATTAAACCTTTAAAAATAGGCTATCTCGGGCCCTTTCCG

Features:
149-737 Moloney murine sarcoma virus 5' LTR
807-1616 Extended Packaging Region
1680-1735 EM7 promoter (bacteriophage T7 promoter)
1754-2151 Blasticidin resistance gene coding sequence
2310-2440 SV40 poly A signal and site
2603-3420 CMV IE promoter
3675-4988 G-protein-coupled receptor (GPCR)
5071-5646 IRES
5647-5703 Bovine a-lactalbumin signal peptide
5704-6372 'humanized' antibody light chain
6553-7146 MoMuLV 3' LTR
7683Origin of replication
9302-8442 b-Lactamase coding sequence
Figure 20

Protein Comparison

Clone #

pg/cell/day

0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

14A
61
13(N-)
5
11
15
17
28
29
32
37
40
43
45
47
Figure 21(A). Gene Sequence of a Single Retrovector IgM Construct (SEQ ID NO:37)

LTR = Human CMV/Moloney Murine Sarcoma Virus Long Terminal Repeat Fusion
Ψ EPR = Moloney Murine Leukemia Virus Psi Region and Extended Viral Packaging Region
JC = J Chain Gene
CMV = Simian Cytomegalovirus Promoter Region
SP = Heavy Chain Signal Peptide
HC = Mu Heavy Chain Gene
IRES = Internal Ribosome Entry Site/α-Lactalbumin Signal Peptide
LC= Kappa or Lambda Light Chain Gene
RESE = RNA Export and Stability Element
LTR = Moloney Murine Leukemia Virus Long Terminal Repeat

37/53
FIGURE 21(D)

GAACAGATGGACAGCTGAATATGGGCAACACAGGATATCTTGATATGTAAGCAGTCTCTGC
CCCGGCTCAGGGCCAAGAAGACAGATGGGCTCCCAAGATGCGCTGCCAGCCCTCAGAGTTTCT
AGAGAACCAGATCGTGGTCCCCAGGACCCAGCTGAAATGACCCCGTGGGCTTTAT
TTGAAGCTAAGCAATCAGGTCCCTGCTTGGTGTGGGCGGCGTTGGGTCTGCCGGCTGCTGGCTCTCCGGAGCT
CAATAAAGAGCCCAAGACCCACCTCCTCGGGGCGCCAGTCCCTCGGATTGACTGAGTCGC
CCGGGTACCCTGTTATCCAAATACCCCCCTTGTGCAAGTCTGACCCGACTGTGGTGTCCTGCT
GTTCCCTGGAGGGTCTCTCCTGAGTGGACTACCCGTACGCGGGGCTTTTCTATT
Figure 22(A). Gene Sequence of a Double Retrovector IgM Heavy and Light Chain Construct (SEQ ID NO:38).

LTR = Human CMV/Moloney Murine Sarcoma Virus Long Terminal Repeat Fusion
Ψ EPR = Moloney Murine Leukemia Virus Psi Region and Extended Viral Packaging Region
Neo = Neomycin Resistance Gene
CMV = Simian Cytomegalovirus Promoter Region
SP = Heavy Chain Signal Peptide
HC = Mu Heavy Chain Gene
IRES = Internal Ribosome Entry Site/α-Lactalbumin Signal Peptide
LC = Kappa or Lambda Light Chain Gene
RESE = RNA Export and Stability Element
LRT = Moloney Murine Leukemia Virus Long Terminal Repeat

GTTGACATTGATTATGTACTAGTTATTAAATAGTATACGATACCTAGGGTACATTAGTTCAT
AGCCCCATATATGGAGTTCGCCTCAATCACTCTCTCTGTTAGCCTACATCACTGAGG
CGCCCAAGACCCCCGCCCCCTAGTGCAATATGAAGTATTTACCCATGACGCTGACCTGAC
CTAGCTGTATATGATATACCTGTGTACCTTTGCGGTTGCTAGTATCATCACTGAGG
CGTGGATAGCGGTTCGCTACACGGGATTTCCAAGTCTCACCACCCATGACGCTATTGAC
GAGTTTGTGTTGCGACAAAAATCTACACGGAGATTTCCTCCGAAATGTGTAACACTCCGGCC
CATTTAGCCAAATGGGCGGATAGCACTGAGGCTGTGGAGGTCTATATATAGCACGAGCTCAAA
TAAAGAAGCCCCAAACCCTCCCTCTCCTCGCGCCAGCTTCCGATACTGCGTCGCGCGCC
GGTACCCCTTTCCCTTTTTAAAGCCTCTCTGCTTGGGATCGAGCTCGCTCGCTGCTGTGCT
TCTTTGGAGAGGTCTCTTGATGATGTAACATCCCAACAGGGGGGTCTTTTCTTCTGCGG
GCTGTCGTCGGGGATTGGAGCCCTCGGCGACCACCCCGCCGCAACCTGGAGAGCTCGCTCGCTG
CGTCCCTTCTGCGGAACCTCGGCGGCAACCTGGAGAGCTCGCTGCTGCTGCTG
AGTTGCCCACCAATTCATTCTGTGGCTGATTGGCTAGCTCGTCGCTGCTGCTG
GCCCCTGGGCTCTGCTTACGTGCTATGCATCTCCTGCTATCTGGCAACCGGCTTTGGGA
CTGCAAGCTCTGCTGCAACCGCCGGGCAACCTGGAGAGCTCGCTGCTGCTGCTGCTGCTGCTG
CTGTTTGGCGCGCCAGCTGAGGCGGAGTGCTGTGAATCGGCAACCCGCTTTGGGA
CTGTTTGGCGCGCCAGCTGAGGCGGAGTGCTGTGAATCGGCAACCCGCTTTGGGA
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
TTTGGGTTGAGGAAAGCTGAGACGACCTAAAGACGTCCCTCGGCTTCTCGTGATTTTGGC
FIGURE 22 (B)

GATCAAG GTTCTTTTACACCTGGGCGCATGGACACCAGACCCAGGTCCCTTACATCTGTA
CTGGAGAAACCTGTTGACTTTTTACCCCCCTCCGCCCTGAGTCACAGGCTCTTTCACCTAAG
CCTCCGCCCTCCTCTTTATCCATCCGGGTCCCTCCCCGTCAAGCACCTCCTCTGCTGAC
CCCGGCCTATCATCTCCTTTATACCGCCACTCTCTCTCTTCTTAGGCGCGGAAATTCGCA
TCTGATCAAGGAGACAGGATGAGATGCCTTGCAGATGGAAAGAGATGAGATGGCAGCC
AGGTTCCTCGGCCGTCTTCGAGAAGCTATTCGCTGACTAGCTGGAGCTGGCAACAGCAAA
TCGGCTGCTGCTAGATCCCGCGGCTGCGCTCGGACTGCGGACCGCGCCGGCTTTTTTTT
GTCGAGACCGACATCGGCGCTCCCTGCAAGTGGACTGACAGGAGAGGACGACGCGGCCTATC
GTGCTGCCGACAGGACGGGCTTTCTGGGCACGCCTCTGCGCTGAGCTCTGCAAGGGG
GAAGGGAAGAGCTGCTGCTATTTGCGCAAGTGGCCGGAGGAACGCTGGGGGAGGCAGCTGCTG
GCTCCCTGGCGAGAAGATACCATCTACCTGTCAGTGCAATGCGCCTGCAGATGGCCATG
TCGCTACCCGCTCATTCCGCTCCACGAAAGCAGGAGACATCGCTCGGAGGAGGACGTACTC
GAATGGGAAGCGGTCTGTGCTGATCAAGTATGGGCTCTGGCGCCAGCATCAGAGTGGTCGAC
CGTAAAATGTCGAGAACGCTGGGGCGGAATGGGGGCTCGCTCCTGCGTGTTTACGG
TATCGCGGCTCCATTGCACATGCGCTCCTCCTACTACCTGCCTCTGAGATTTCTCCT
GAGAACCTTGATTGTGCAAGATCATTACATTATGTATTGCCCTATGCAATTAG
TATTCCCAATATGGGTTCATTCTGGCATGATCTAGACCTGCTCCTCCCTGCAAATGGGCCCTAC
ATACATATATAGGCGCTCTACTATACCGCCCATAGCCACTCCCCATGTACGTCAATGG
TCCTCTATATAATGCGCTCTCTATGATCGCTATATGGGCGGCTCTATATGGGCTTATATGGC
GGCCCTCCCCATGTACGGCTTCTACTAGAACTGGCCGGCTGGCTGCTGTATGCGCTGCA
TCATTACGCCACCCGACCTAGTGCGGCCATGCTGTGCTCCATCTACTGCCTCGT
TCAGGCAGGCTGCTCAATGTGAGCTGGTATGCTGCTGCTGCTGCAATGCTCCACTGAGG
CTATTAGACATTACCTGTCGAGGAATGGGCGGCTGCGGGAGGACGGCTGCGGAACGAG
CATCTCTCTCTCTGTCGCGCCCTCTCAATGGAGTTGCTGCTGTGCAACCACCGTCATGG
GCTCTGATTTGCAACCACATATGGTGATGTGCTGGTAGACGGCAGGCAAAAGGGCTTT
GAGTTAGCTGAAAGCTATACATGATGGAATAGCAGCTAATTACTACGCTGCTCTTAAA
GAGATAAGTTACATATCAGGAAAGACCGGAGTTCTTTGTCGACATTCTGGCGGATTAAAAG
CGCCCTGAGACCCGAAGAGACCGGATCTTTTGTGCTAAGCCCGTTTATTCTGGCCCTCCTC
TGTTTTGCTATTAGGGCGACAGGCGCCCGGCCACATGCAGCTATTCGACGCGGCGGGCC
CCCAACCTCCCTCTCTCTGCTCTCTTAGAATCCCGGTGATAGGCAGCGGGG
CGGTGGGTCTGCTGCTGCTAGGACAGCTTCTCTCCGCGGACTCTACACTTCTGTGGAGAATAC
AAAGAAACCCTGCACTAGCAGCTACCCCGGGCTTCCTCACATGCTGGTACAGAGGGCGCA
GTACGACGACACTCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CCACGTGGTTGCGAGCTGCGTCAATGCTGGCCATTGCGCCGCAAGAGGAG
AAACAGTGGTTGCGAATATGCGAGACCCACGGCACAAAGAAAGAAGACGGCTGCTCTT
CCCTGGCGAGCTGCGTCAATGCTGGCCATTGCGCCGCAAGAGGAG
CTCGAGCGACACTCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
FIGURE 22 (D)

CGCCGCTGCTTTCGCCGCTGGACAGGGCTCCTGGCTTGGCACAATTCG
TGGTGTTTGTGGGAAAATCATCTGCTTCTTCTCTTGCTGGCTGCTGGGCCACCTGG
ATTCTGGGGAGCTCTTCTGCTACGTCCCTTTGCGCCTCCCTCAATCCAGGGGACCTTC
TTCCCGCCGTCTGCCTGCGCTCTGGCCCTCTTTCCGCTCTTGCTGGCGCCCTCAGA
CGAGTCGATCTCTCTTTGGGCGGTCCCTCCGCTACGATAAAAAATAGATTTTATTTA
GGTCTCCAGAAAAAGGGGGAAAGAACCACCTGTTAGGTGGCCAGCTAGCTTTAA
GTAAGCCCATTTTGGCAGGCATGAAAAATACATAACTGAGATAGAGAATTCGATC
AAGTCAAGAACAGATGAAAGCCTGGAATATTGCCCCAACAGGAGATATCTGTGTAAGCA
GTTCTGCCCCTGGCTGAGGGCCAAAGAAGAGATGGGAAAGCTGAAATTTGCCCCAACAGG
AATATCTGTTAGACGTTCCTGCGCCTGGGGCTAGGCGCCAAAAGAGATAGGTCGCTCCAGAT
GGGTCCAGGCCTAGCAAGTTCTAGAGAACCATGACATGTTCCAGGCGGCGCCCAAGG
ACCCGAAATGACCTGTCTTTTATGAACTAAACGCTACTGCTGCTTCCTCTCTGT
TCGCGCTCTCTGCTCCTGGGCAATAAAGAGGCCACACAACCTCAACTGCGGGCGC
CGATCTGCTCCATGTGAGTGGTCCGCGGTACCGTGATGCTCCAATTAACCCCTCTCGAG
TTGCATCCGACTTGTTGCTCCTGGCTTTGGAGGGTCTCCTCTGAGTGATTGACTA
CCGCTACGCGGGGCTTTCTATT
Figure 23(A) Gene Sequence of a Double Retrotector IgM J Chain Construct (SEQ ID NO:39).

LTR = Human CMV/Moloney Murine Sarcoma Virus Long Terminal Repeat Fusion
Psi EPR = Moloney Murine Leukemia Virus Psi Region and Extended Viral Packaging Region
Blast = Blasticidin Resistance Gene
CMV = Simian Cytomegalovirus Promoter Region
SP = J Chain Signal Peptide
JC = J Chain Gene
RESE = RNA Export and Stability Element
LTR = Moloney Murine Leukemia Virus Long Terminal Repeat

GTGGACATGTGATTATTTGACTAGTTATATTAACTGTAATCTCAATTACGGTCTATTTGTTCT
AGCCCAATATATGAGTTCTCGGTACATAGCTAATAGCTAATGCTAAATATGCTACCGTCCAGCC
GCCAAGTTGTCCATCGGCATTAATGAGCAGAATTCTGTCATCAATGCTCCTTACCTGACCTA
GTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
GCCTCTTGAGGAGGCACCTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
GGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCCCGATGCAATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA
AGCCCGATGCAATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA
AGCCCGATGCAATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA
AGCCCGATGCAATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA
FIGURE 23(C)

AAGCAGTTCCTGCCCCGGGCTCAGGCAAGAAGAATGCTCCCCAGATGCCTCCAGCC
CTCAGCAGTTTCTAGAAGACATCAATAGTGGTCCCAGGGGTGCCCCAGGACCTGAAATGA
CCCTGTGGCTTAATTTAAGTTAACCAATCACTGGTGGGCTCTGTTCCTGGCGCCGCTTC
TGCTCCCCGAGGTCAATAAAAGAGCCCAACTTCCACCTGAGGGGCGCCGAGTCCTCCGA
TTGACACTAGTGCCGGCTCTCTTCAGACTCCAAACTATGGCTAAGAGCTCCACCGGTG
GGGTCTTTCATT
FIGURE 24 (A)

Mouse dihydrofolate reductase retrovector gene construct (SEQ ID NO:40).

<table>
<thead>
<tr>
<th>LTR</th>
<th>Human CMV/Moloney Murine Sarcoma Virus Long Terminal Repeat Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ψ EPR</td>
<td>Moloney Murine Leukemia Virus Psi Region and Extended Viral Packaging Region</td>
</tr>
<tr>
<td>DHFR</td>
<td>Mouse Dihydrofolate Reductase Gene</td>
</tr>
<tr>
<td>CMV</td>
<td>Simian Cytomegalovirus Promoter Region</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple Cloning Site (Gene Insertion Site)</td>
</tr>
<tr>
<td>RESE</td>
<td>RNA Export and Stability Element</td>
</tr>
<tr>
<td>LTR</td>
<td>Moloney Murine Leukemia Virus Long Terminal Repeat</td>
</tr>
</tbody>
</table>

1 GTTGACAATTGATTAATTGACTAGTTTATATAGTAATCAATTACGGGTCTATTAGTC
ATAGCCCATATATAGGATTGCGTTACATAACTTACGTAATAATGCGCCCCCTGCTGTA
CCGCGCCAGGACCCCGCCATGGCAATGAATGCGTATGTTGCTGCTGCTGCTGCTGCTG
AATAGGGAATTCCTATATCGACCTAGGATTGTACACTCCTGATGCCTGATGCCTGATGC
GAGTACATCAAGTGATCTAATGCGCAAGTAGCTCCGGCGCGCCCTGATTGAGGATGTG
TGCCCGGGCTGATTATGGCCAGAATGACATGACCTTGGGCTTCTCTGCTTCTTCTGGAC
CAGTGACATGATCGTCTGCTATCTGATGACGCGCAGCGCAGCAGCAGCAGCAGCAGCAGC
CCATTGACCGAAATGCGGCGGAGATCGTACGGGTGAGGCTATATATACGAGAGCTC
ATAATAGGAGGACCCATACCGCCCTGGCGAGTCGTTCTCCGGAATGCGTGTCGGC
CGGGTCCAGTACAACTCCGAGCTGTTCGTTTGGCTACGAGTTCGTTCTGCTGCTGCTG
GTTCTCTGGGAGGATGCTTCTGGTGTGATTTACCATACCCGAGGCGGTGCTTCTGGT
GGGGCGGCTGGGGATGGAGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
TATGCTGACAGCTCTGATCGTGTGCTCAGCGTACGCTGACGCGGACGCGGACGCGGACG
ATGCCGGTGGGGCTCTGCAGGGGATGCGATGGGATGCGATGGGATGCGATGGGATGCG
AACTGAGACGTCTGAGACCCAACCCCGCCGCAACCTGGGAGACGTCGAGACGTCGAGACG
GCCGGTTTTTGTGGGCGGACCTGAGGAAGGAGTGCGATGGGAATCCGACCGGGCTTGGT
TAGTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
GCTTCTGGTTGGGAAACACGACGCGCGCTTGTCGTCGCTGTCGCTGTCGCTGTCGCTG
TGTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
ACCACTCTTAAAGTGGTACATCGTCAGCTGGAAAGATGTCGAGCGGATGCGTCGTCG
CCAGTGGTGAGATCTGCTACGAGAAGAGACGTGGGTATCCCTGCTGCTGCTGCTGCTG
CCCTATACGTCAGACGCGACGCGACCTTATAACGAGACCTACATACCCCGAGGTT
AAGATCAAGGCTTCTCTCCACGCTGGCCGCATGGCAACCGGAGCCACGAGTCTCCCTACG
GACCTGGGAAGCCTTTGGCCTTTTGGACCCCGCTCCCGCTGCTGCAAGCCCTTTTTGTAACCCCTA
FIGURE 24 (C)

TCAGATGTTTCAGGGTGCCCCAAGGACCTGAATGACCCCTGTGCCCTTAATTGAACTAA
CCATACAGTTCGCTTCTGCTGTTCTGTTGCAGCTTTTGCAATAAAG
AGCCGCAAACCAACCCCTCACTCAGGCGCCACCAGTCTCCGATGACTGACTCGCCCGGGGTAACC
CGTGTATCCAATAAACCTCTTGGCAATGACTGACTCTGCTGCTGATTGACTACCCCGTCACGGGGGTCTTTTATT 4354

1 - 763  HCMV-MOMUSV CHIMERIC 5'LTR
833 - 1642  EXTENDED PACKAGING REGION
1712 - 2272  MOUSE DHFR GENE CODING REGION
2355 - 3038  SIMIAN CMV PROMOTER
3029 - 3114  MULTIPLE CLONING SITE (SITE OF GENE INSERTION)
3121 - 3721  RESE
3761 - 4354  3'LTR
Figure 25(A)

Glutamine synthase Retrovector gene construct (SEQ ID NO:41)

LTR = Human CMV/Moloney Murine Sarcoma Virus Long Terminal Repeat Fusion
Ψ EPR = Moloney Murine Leukemia Virus Psi Region and Extended Viral Packaging Region
DHFR = Mouse Glutamine Synthase Gene
CMV = Simian Cytomegalovirus Promoter Region
MCS = Multiple Cloning Site (Gene Insertion Site)
RESE = RNA Export and Stability Element
LTR = Moloney Murine Leukemia Virus Long Terminal Repeat

1. GTTGCACAATTGATTATTGACTAGTTATATAGTAATCAATTACGGGTACTATTGTTCC
   ATAGCCCAATATATGAGGTCCGGTACATTAACATTACGTTAAATGCCCCTGGCTGCTGA
   CCGGCCACAGCCGCCGCCCAATTGAGCTACAATATGACGTATGTTCCCCATATGTAACGCC
   AATAGGGACTTCTCATATAGCTATGACTGAATGTGAGTAGTTACGTAACCTGCCCCTTGG
   CAGTACATCAAAGTGATCATATGCTGCAATGCTGCCCATTGGTACGTAAGCTGCTGAAA
   TAGGCTGGCCTGAGTATTGAGTGTCAATGCTGAGCTAATGAGTGGACTTTTCCTACTTTGGA
   CATCTACGTATATTAGTCTACTGCTATTACGATGTGCTGAGTTTGGCAGTATACATCAATG
   GGGTGCTAGGGCTAGTTGACTCCAGGGAGTTTCCCAAGTCTCCACATCCACTATCGTAAT
   GGAGTTGTTTTGCTGCAATATACAGGACCTTTGCAAAATATGTGACAGAATCACGC
   CCCATAGAGCGAAAATGGCCGGTAGAAGTACGAGTGGAGCTATATATAAACAAGCTC
   AAATAAAAAGCGCAACACACCGCTATCCGCGCTGCGCACGTTCTCCTGGATAGGCTGGGCG
   CCGGTACCCGATTCCCATAAAGACTCTTCCTCCCTGCTTTGCTATCCGAAATCTGGCTGCTGCT
   GTTCCCTCTGGAGGAGGTTCCTCCTCTGAGGTAGTTGACTACCACGACGGGCTGTTTCTATTG
   GGCGCTGCTCCACGAGTTGGAGGCCTCCAGGCCAGACGCACACCCGACCCCGAGGAG
   TAAGCTGCAGGCTCAACTTAACGTGTCGCTGGTGCTGGTGCTTGTTGGATGATTGTT
   ATGGCGCTGCTGCTGTACTAGTTAGCTAATAGCTCCTGTATGTGCTGGGACCGCGTGGG
   AACCTGAGGATCTCGAACAACCGCGCCACGACACCTGAGGAGGCTGCCACCTGGGAG
   CCCGCTTTTGGCTGCCGACCTGAGGAGGAGTGATGCTAGAAATCCGACCGCGCTCAGG
   TAGTGGTGGTCTGGATGAGCAGAGAAGCTAAACGCTCCGCTCCCTCGAGAATTCTTTT
   GCCTTTCTGGAACCGGACGCGCCTGCTTGTGCTGTCAGCGTGACATGCTTCT
   GTGTGTGTGGTCTGACGTACTTGTCTGTTTTGTGCTGAATAATTGCGGACACTGTGTT
   ACCACACCCCTGATAGTGGCATTGGTACCTCTGCTGCTGGCAATAGCGGAG
   CCGGCTTAGATGTCAGAAGAAGACGCTGTGTTACCTTCTGCTGCTGCGAATGCGCA
   CTTTAGCTGCTGATGCGCGACAGACGCAGGACCTTTAACCAGAACCCTATACCCAGG
   AAGATCAAGGCTCTTTCRCCATGCGCCGGCAGTGGCACCAGAGCCAGGGCTCCACTCGT
FIGURE 25(C)

TCTGCACGCAAGCTTCTTCTAGTCCCTCTGGGGCTCAATTCCAGGACCTTCTT
CCGGGCGTGCTGGCCGCGCCTTTCCGCTTTCCGCTTCCCCCTCTACGAG
AGCTGAGATCTCCCTTTGTCGGCGCCTTCCCCCATCGATAAAATAAAGATT
TTTATTTATCCAGAATAGTTTCTTGTTTTCCAGTCAGTTAAGTT
AACGGCCATTGAGCAGCATGGAATAGCAGCTGAGATAATCCTCAGCTGAGAT
CCCTGTCCCCCGCTCCGGGCAAGACAGATGGAACTGAGATATGTTGGGCAAAAC
AGGATATCCTGGTTAAAGCTTCTGCTCCCAGATGC
ATCTCTGGTAACGAGTTCCCTCGCCCCGGCTCCAGGGCCAAGACAGATGCTGCCAGATGC
GGTCCAGCCGCTCCAGGTCTTCTCTAGAGAACATCAGATGTTTCCAGGGTGCCCAAGGAC
CTGAAATGACCTGGCTTATTAGAATCAAATCAGATGGCTCTGCCTCGCTTCTGTTCC
GCCGCTTTCTGCTCCCAGGCACACTAATAAAGAGGCCCAACACCCCTACTCGGGCGCCA
GTCCCTCCAGATTCTAGTGGTCGCCCCGGGTACCCGTTGATCCAATAAACCCTCTTGAGTT
GCATCCGACCTGCTGGTCTCGCTCCCTCTGGGAGGTCTCCCTCTGAGTATTGACTACC
CGTCAGCGGGGTCTTTTCATT

1 - 763     HCMV-MOMUSV CHIMERIC 5' LTR
833 - 1642    EXTENDED PACKAGING REGION
1677 - 2795  MOUSE GLUTAMINE SYNTHASE GENE
             CODING REGION
2805 - 3472  SIMIAN CMV PROMOTER
3473 - 3558  MULTIPLE CLONING SITE (SITE OF GENE
             INSERTION)
3565 - 4165  RESE
4205 - 4797  3' LTR
Bleck, Gregory T.
Bremel, Robert D.
Miller, Linda U.

Production of Host Cells Containing Multiple Integrating Vectors by Serial Transduction

GALA-08484
10/759,315
2004-01-16
41
PatentIn version 3.2
1
2101
DNA
Artificial Sequence

Synthetic

1

gatcagtcct ggggtgtcat tgaaggact gatgctgaag ttgaagctcc aatactttnn 60
cacccgatg cgaagaacct gactcatgta taagaccttc atactgggaa agattgaagg 120
cagagggaga aggggttgag caggtgggta cagtggttc ggaatcctaa accaaccct 180
agtggtttgta gcaagcttccc aagagttggg atagggcaggg gagaagctgg ctggtgact 240
cactgctgtt gcaagagtgt ggcactactc gactgatgta agctgctgatg tagtgtaatc 300
catggtcacag aatatatgat aaaaaaggg aagagttggt cacctactttgga caggtgttga 360
ggatataaaa gtttagaata ctcttctttt ggaagcttta aattatatca ttgatgtagt 420
tacctgtgc aataataaaggt atgcagctttt aagactgcat gtagagagaa tgcagctttg 480
cattcctggag gtctagactg ctggtaatttta cctggtataa ccaattatat acctacatgg 540
tggtttata gcagctgctg tggtagcctag aatgtaaggg aaccaattatg aactagttatct 600
tctaaactctg catgcttcgg cagcagctgg ttaatgctatcttte ggtctgagat cttgctatt 660
aggtctttgg tcttatattgg tggggtttccc tgggttgcttta gatgtataag ggcttgct 720
catgtttttg gatctgttcc ggatccttgg tttcccttgag aagagatgg gacccaccctc 780
caacctccact tagatctctt acctgtggaaa ttctcttttct ggcgagctgct cagctgct 840
agtccatggag gttgcagagtc aacgctcagac aacagtactctttt tcagctgtttctttttttctt 900
atgggctctt ttcggtgttg gattggaаг ттccaatgctc ggtctgtgcc aaggctggaa 960
gttttttcacc attcctgac accaggggaag cccaagaata ctggagttgg tagccatttc 1020
cacccaggg gatcttttccg aatcacaagag tttgatcctgactc tttcagggga 1080
attttttac cagcagacta cctgatgct cttctttttttt ttaatatttttctt ggtctgact 1140
agtctttttt cctttttttttt aagagttggg tccacatttt ttaaccctttt ggtctgact 1200
ttgaggtgt cttaaagttt atatattttt gacagcattaattgac gatgttggta 1260
776451Seq

gatttacaat gtgtatctg gctatattagt ggtatttggt gttgagggatg gggaggctga 1320
tagcacttaca gaggccagct agatactgtc atacacactt ttcaagttctt ccaatttttgt 1380
gaaatagaaa gtctctggat ctaagttata tgtgattctc agtctcttgat gcatattctt 1440
atttcatctcc tgacacactca aacaagaaacc aaagatatca gggacacttg ggcccccttc 1500
tgctctggggt gaggccgcca tgcataatgt ctctggcctt gttcatatggc tggattggtt 1560
ggacagtgac cagctctgtat cctggcagct tggcagtgta tgcacatccg cccctcttca 1620
cattctgcat gtctcttaggg gggaggggag aagctcggtta tagaacccttt attgtatattt 1680
tgatgctcct cactcttttat atgccctccca tggcccttttt tggctctccaa gtaaccagag 1740
acagtcttcc ccagaaacca cccctacaag aacaaagggga taaaacaaaggc caaatgggaa 1800
gcaggcatc gtgttgaaact ctttctggcc agagaaacaat acctgctatg gactagatac 1860
tgggaaggg aagagaauag taggggtgaat tatggaagga agctggccagg ctcagcgttt 1920
tgctctgggc atgcaccgtc cttctctcatt cctctctcat atgtagggct tggtaaccaga 1980
gccctctgagg cttcctgcat gaataataat atatgaaact gagtgaattct tccatttacag 2040
gttcttgggg gcgcccgaatt cagctcgggt aacccggggat ctcgaggggg gcggccggtac 2100

c 2101

<210> 2
<211> 245
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 2
gattcttac tggcaggtgc ttggggcttc cgagacaatc gccaacatct acaccacaca 60
acacccggcc gcacagggttg agatactggc cggggacggcg ggggtgttca ttacaagcga 120
gatccgatt acttactggc aggtgctggg gggctccgag acaatcgcga acactctacac 180
cacacaacac cgccctcgacc aggttggatct atcgccgcgg gccgcggccg tggtaattac 240
aagcg 245

<210> 3
<211> 680
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 3
ggagattgccc cctctccctc ccccccccc aacgttactg gcggaagccg cttggaataa 60
ggccggttg ggtttgtaata tatttttta ttccacattat tgccgctttt tgggaatttg 120
aggccccgaa acctcggccc gtttctcttg acagacattc ctagggtgct ttcccccttcc 180
gccaaaggaat tgcgaattgct gttgaagagc cagttcttct ggaagcttct 240
776451seq
tgaagacaaa caacgtctgt agcgaccttc tgcaggcagc ggaaccccc acctggcgac
aggtgcctct gcggcacaara gccacgtgta taagatacac ctgcaaaagg gcgcacaacc
cagtcgcacag ttgtgatttg gatagttgtg gaaagagctca aatggcttcc ctaagcgta
ttcaacaaggg ggctgaagaga tgcccagaaag gtaccacatt gtatggtgatc tgatctggg
ctcgcgtgca ctacgttttac atgtgttttag tcaggtttaa aaaaacgtcct aggcctcccg
aaccacgggg acgtgttttt cctttgaaaa acacgatagt aatatggcct cctttgttcc
tctgctcctg gtaggcatcc tatcccatgc cacccaggg ggccgcatgg gatattctga
tctcagctc gcgtgagcctt

<210> 4
<211> 4207
<212> DNA
<213> Artificial Sequence

<220> Synthetic

<400> 4
cggatccggc cattagccat attatctcatt ggttatatatg cattacatca taatatgtcc ccagttgttac 60

<400> 4
tggccattgc atacgtggtga ttcctatcct catatatgta tttatatttg cttattgcttac 120

<400> 4
acattaccgc cattgtgagcta tgtatatttg actatcattt aataagtaatc aattacggtt 180

<400> 4
tcattatgct atgcctggata ttgaggttgc gcgcgttttc gattttcctc 240

<400> 4
cctggtgac gcggccacag cccggcccca ttcgcgctta taattcgcata tggccacata 300

<400> 4
gtaacgcaca taggagccct cccatgcgct gtatgctgcc ggacggagtc 360

<400> 4
cctattccatc cctgtcatgt gccagacgtg gccctattgt tgttta 420

<400> 4
gcctgctggca taattccacag tattgacatc tctctattc 480

<400> 4
catagctcatc actgtattgt cttatgctact ctcgtgtacct gctgtgtccttg gcgtatactc 540

<400> 4
aatgggcatg tagagccctg ttacgtcagc ggatttccca gttccacacc cattgagcgc 600

<400> 4
aatgggctgt tggattccca gaaaaatcag gcggacttcc caaagttcag ctcacgactc 660

<400> 4
cccccatagc gcaagacggt cggtgctgtt gggtctatat cagcagagct 720

<400> 4
cgtttgtagt aaccctcagct gcctggagaa gcgcatccac gcgttgctttg gcgctactgc 780

<400> 4
agacaccggg acgcagcctt cctgccgggg ccacacgttc tggacgcttc cccgggaatt 840

<400> 4
caggacctca ccattggatt gagctgctat ccctccctct gttgacacat acgtcagactg 900

<400> 4
gtccactggag ggtcagatc gtgcaggttgg tgtgccaccc cggccctttc 960

<400> 4
tgcgctctgt cctgccgctg atctgcgctc gatttccacc actattggtgt gattgtgggtt 1020

<400> 4
agacagggac ccgaaaggg cttgggtctgt atggagctaa ttcatccaga tagcagttacg 1080

<400> 4
attaactag cgcctctctt aagaagtaga ttttaaatat ccgcagacca cccgaagaac 1140

<400> 4
agctttatc tcggtccttc ctggtgggtc tataggggcc aaggaccccc ggtccacagtc 1200

Page 3
776451Seq
gcctgctgaa taacttctat cccagaggg ccacaagtaa gttgaaggtt gataacgcc 3360
tccacgcgg taaccccaag gagagtgtca cagacgagga cagcaaggac agcacctca 3420
gcctcagcag caccttgagc ctgagcaaaag cagactacga gaaacacaaa gtctagcct 3480
gcgaagtcac ccactcagggc ctgagctgcg ccgctcacaac gagcttcacac agggagagt 3540
gtttagagatc taggcctcctt agttcgacat cgataaaata aaagatttta tttagctccc 3600
agaaaaaggg gggaatggaa gagcccccct ttagttttgg caagctagct taagtaacgc 3660
cattttgccac ggcattgaaa aatacataac tgaagaataga gaagttcaga tcaaggtcag 3720
gaacagatgg aacagctgaa tatggggcaca acaggatatac tggtagtaagc agttcctgcc 3780
cggcctcagg gcacaagaca gatggaacag ctgaatatgggc gcaaaacag ggtaaccaag 3840
taagcagttc ctgccccgggc tcagggcaca gcacagatgg tcccccagatg cggccaggcc 3900
cctcgagatt tctagagaaac catcagatgt tttcagggggc ccccaagggc ctgaaatgac 3960
ccttgctcct attagacta acaaacagtt tcgctttcgc ttctggtttgc cgcgcttttg 4020
tccccgggggc tcaaaaaag ggccccacaac cccctactcg gggcgccagt ccttccagttg 4080
actgagttgca cccggttacc atgtatccca taacccctct tcgactgtgca tccgacctcg 4140
ggctctgctg ttctttggga gggtctcctc tgtagtgttg actacccgtc agcgggggtc 4200
tttcatt 4207
<210> 5
<211> 4210
<212> DNA
<213> Artificial Sequence
<220> Synthetic
<400> 5
ggatccggccc attagccata ttattcattg gttatatagc ataaatcaat attggctatt 60
ggccattgca tagcgtgtat ccctacatac atatgtacat ttatatgggc tcatgtccca 120
cattacgccc atgtgtgatg tggattattga ctcagttta tatagtaatc attacggggt 180
cattagttca taggcctcatat atggagttcc gcgttacata acttacggtg aatggccgccc 240
cgtgctgacc gccccacgc cccgccccct tagcgtcatat aatgaacgtaa gttcccacatag 300
taacgcctat aggagcttcc ctttgcagtc atatggtgga gtatttacgg taaactgccc 360
actgggcagt atacacagtg tcatatagtc caagttcgc cccctatttg cttcaatgacg 420
gttaatggcc gcgcctggcat tatgccccagt atcagacctt atgggacctt cctacttggc 480
agtcatctca cgttaggtac atcgcatta ccatatggtat gcgggtttgg cagatcatca 540
atgggggtgg atagcggtttt gacgtcaggg gatttccaaag tttccacccct atttgcgcta 600
atgaggttt gtttgggccac aaaaactcaac gggactttttc aaatgtcgcc aacaactccg 660
ccccattgac gcctagggcc ggtaggcatg tagcggttggg ggtcttatata agcagagcctc 720
gtttagtgaa cccctcagttc gcctggagac gcctacccag ctgtttgtagc ctctcatagaa 780
776451Seq

gacacgggga cccgatccag cttccgccggc ccacagcttct ccgacggatcc ccggaattcc 840
aggacccctac catggagtgg agctgtatca tctcttcttt ggtagcaaca gctacaggttg 900
tccacctcca gttcagcctg ttcaaatcag ggggtgatagt caagaaacct gggtcatcag 960
tgaaggttcc tgtcaagggct tctggtcaca ccttctactg ctactggtcgt caactgggtc 1020
ggcaggcac ggtgcaaggtt ctggaatgga ttggatacat taatcctagg aatgattata 1080
tgtgatcaaa tcaagaacttc aaggcaacgg ccacacaaacac tgcgacagaa tccaccaata 1140
cagctctacat ggagctgagc agccctgaggt ctgagggacac ggccttttttt tttttgtgca 1200
gaagggatat tactacgtttc tactggggccc aaggcaccac ggtcaccgttc tctctcagct 1260
ccacacaggg cccatcggtc ttcccccttg cacccctccct caagagcacc tctgggggca 1320
cagcggcctt gttgtcccttg tcaagagct ttcctccccg gaccgcagtc gggtgctgga 1380
actcaggggc ctcgacccag ggctgtccaca cctttgccgc tgttccatacg tctctcagag 1440
tctacctccc cagcagctgtgc tctgggggac cctcaccacag ctggggaact caacacac 1500
ttgtgcacca aaacacacca tgcacccgct gcccacaccc tcaacctctgg gggggagcgt 1560
cagtctttcct cttctccccca aaacccaagg acacccctcat gatctccggg acccctgagg 1620
tcactctgggt gttgtccccg acagcaccgc aagcctccag ggtcagttcc aactgggtacg 1680
tggacgctggt gaggagccat acatcccaaa ccaagccggc ggaggagcag tacaacagca 1740
cgt cacacgctgt ctcacggttc ttgacccgga ttgggtggtat ggcaaggtggt 1800
acaagtgcgaa ggtctgcacc aacacccccc cagcccccctc cgagaaaaac atctccaaag 1860
ccaagggcga gccccgagaac cccacaggttg acacccctgc ccctcccggg gagaggagta 1920
ccaagagcgt ggcagcgcttg acctgctcttg tcaaaaggtt ctatccacgc gacatcgccg 1980
tggagtgggg gagcaatggc cagccggaga aacaactcata gaccacggct cccgtgctgg 2040
acctcgacgg ctcctccttc cttttatatgca agctcaccgt ggacacagac aggtggcagc 2100
agggaaagctg tctctccatgc tccgtgtatgc acgggtcttc gccaacacac tacacgcgaa 2160
agaccccttc cctgctcttc ggggaatgaa agccgaatcc gcccctcttc ccctcccccc 2220
cccaacgctta tctgccgagc cccgttgagca taaggccggt gtgcgttattg tcatatatga 2280
ccccccacca taattgcgttc tttggaaata gttgaggccc ggaaactctg ggcttgcttc 2340
ttgacagcca ctctcatcagt ctctggtgcct cccgatcagac gaatgcaaggt tctgttgaat 2400
gtgcgtgagg aaggagcccc tcagtgctctctgtgcaagct tcttggaaac taacacagtgc gtagcgacc 2460
ctttgcaggg ccggacccccc ccacctcgcc gcagagttgcct cttgccggc acagccagct 2520
gtataagata cacctgcgaa ggccggcaca cccatgcgtc agcgttggag tttgatatgtt 2580
gtgaaagagct ctaaatggtct ctcctcagct gtagtacaac aaggggtcaag gattgcggcc 2640
aaggtacccc attgtatggg atttgtgatct gggtcctggt gcacatgtctt tacatgtttt 2700
tagtcaggttt aaaaaaacag ttcaggccc ccaaaccagc gggacgttgtt tttcttttga 2760

tagtcaggttt aaaaaaacag ttcaggccc ccaaaccagc gggacgttgtt tttcttttga 2820
776451Seq

```
aaaaacagat gataatatgg ccctcctttgt ctctctgtctc ctgtgaggca cctcttatccta 2880
tgcaccccaag ggcacacacag agctgacccca gttctctcatca ttcttgagcgc catctgtggg 2940
agatagggctc actatagcgtc gtaagttcaagct ctaaaagtgttc ttatacagttg caaatcaca 3000
gaaactcttg gcctggttacc agcgagaacc agggaaagca ccataatgcgc tgaatctactg 3060
ggcatccacat aggaatatcgg tgcctccttc gcgcttcctt ggccagcgggt cttgggacaga 3120
ttttactttc ac cacagcct ctcctcaacc cagaaacatt gcaacatatt attgtcaaca 3180
atacctctcc tctgtgacgt gtcgggtggagagc cagatcacaac gaaacctgagc 3240
tgcaccatct gttctctactc tccccgcaacct gggtgacaag ctggaaatctcg gaactggtcc 3300
tgttgtgtgc ctgctgtaata actttcactcc cagagagggcc aaagttacagt ggaaaggtgga 3360
taacgcctcccc caatccggtgt actccgagga gattgtacaagag cggcagggaca 3420
cacctacgc ctctaccagca cccctgacgct gcaggaaagac gactacaggac aacaacaagt 3480
cctccgctgc ggtcaacgcc gaagctccgg cagcctcccc gtgcaacagca gcttcacag 3540
gggagaggtgg tagagattctc gcctcctctctg tcgacacatcg ataaatctaaag atttttttctta 3600
tagtttccag aaaaaagggg gaataagagag cccacccccg tgggagggg cagctaggtta 3660
agtaacgcga ttttccaagg cattcggaaaa taacataactg aggataagtaga agttcagatc 3720
aagggtcagg acaagatggaa cacgctgataa tggggcaaac aggataacctg ttgtaagcag 3780
ttcctgcggc cgcggctaggc ccaacagtccg tgaagcttcctga ataatggggc caacggctc 3840
atctggtgta agccgacctt gccccggcctc agggggaaag acagatggtc cccagagtcg 3900
gttccagcctc cagcagttttc tagagacacc tcaagatgttt ccaagggtcc ccaagagacc 3960
gaaatgcacc tggtgctttcat ttgaactaac caatcaggtc gctttcagcct tctgttcgcc 4020
cgcocttcggt cccccgagcctc aataaaagag cccacaaccc ctcactccgg ggccagttcc 4080
tcccattgcag tggaggccgg gggtacccgt cttcataactc aaccctttgg cagttgcactc 4140
cgactttgtggt ttcctggttt ccttggaggt cttctctctc agtgattgcac taccggtcag 4200
gtctttcatt 4210
```

<210> 6
<211> 5732
<212> DNA
<213> Artificial Sequence

<220> Synthetic

<400> 6
cgacgcttggc agaataggttt gaactcccgagagttgcctca cacctaggg agaaagcagcc 60
aaggggttttgt ttcctcaccaa gacagccagc cctctgcaaca aacggatgag cccatcagac 120
aaagacaatgctcatttcttgt gcgaataaactt ggcggatagc tgcagttctg ggggctattg 180
ggggagtttgcg ggtctggtgc tcggagggct ttcaccccttg actcttttctaa taataactct 240
tctgtgcaag attacaattc caaacaatttc gagaactcga cctctctctc gaggcaagga 300

Page 7
776451Seq

ccacagccca acccctctta caagccgcat cgattttgtc atccgagaaat agaaataaga 360
atgcttgctaa aaattatat ttttacctaa aagaccaatc caataaggtag attattagtt 420
actatgtaaa gaaatgaatct attatttttt agtactatttt tttactcaaat tcagaagttta 480
gaaatgggaa tagaaatag aaagagacgc tcaacctcga ttgaagaagaac ggtgcaagga 540
tatttgacca caggccctagaa agtaaaaagaag gggaaagaag ggtttttttgt ccaataagaga 600
gacaggtggt ggcaacagg gacctatagg ggacccctaca tctcacagcc aacagatgccc 660
cccttaccat atacaggaag atatatctta atgtgggata ggtggtttac agtcaatggc 720
taaaaagtgt tatataagct cccccccccgg gttggaagac tggccgagac tagacccctct 780
tggtttgatg tgttctcaagaa aaaaagagac gacatgaaac aacaggcac tcgattatat 840	tatcttggaa cagagatgca atttttgggga aagatgtttcc ataccaagga ggtagacagtg 900
gctggactaa tagaagctta ttcctgcaaaaa actatggcga taagtatatta tgagatgacct 960
tgatggtccc aaccctgccgg ttcagtaagt ttgtggttac aacatgtttct 1020
taaacaaggg atgtgagaca agttggtttcc tgaactttgttg tgttactaaag gttcttcatc 1080
tgacgcttaga gttgctttatt ttcctatggttt gttttggaaat tttatcaaat ctattatgaa 1140
tgentatgtaa aacacagata taaaagagttg ctgtagttttttt gagtaaatct gcaacagttc 1200
tacaaacctc cccccccctt ttttggtttg tttctccttcg gttgcacccgc cgatctatct 1260
cccttaccat cccagagggct ccccccggcca cccgggggcc cccctaggtgc gggcgactgcg 1320
gcagcgtggc ccccaacagg gacccctcgga taagtgaccc ttgtcttattt ttctctatct 1380
ttgttctgctt cttgtttttatt ctcctattctg ttgtggctatc ttcacaagag cggaaacgagc 1440
tcaccctcagc gaacacacgt ggcggggggt gcagcagatcc gattaatcag cccgaatttc 1500
tgggggtcttc ccagcaacatc gcgaacatat cccacacaca acaccctcctc gcaccaaggtg 1560
agatccgac ggagcggcgg cggatggttaa tttaacgcag gatccgatata cttactcgcag 1620
ggtgctgggg gttctcggag ccataacgga caatcgcgaa catctacacc acacaacacc ccctcggac 1680
gggtgagata tgggccccgg gcgcggcgggt ggtagattaca agccgagatcc cccggaatttc 1740
aggacaacctc catggagattg agctgtatatca ttcctctctt ggagacacaac gttacaggctg 1800	tccacctcgg gttccaaactt gttggagagg tggagaggtgt tgtgcaactt gggcgctcgc 1860
tgcgcctgtac tgcgtcggcca ttcggccttgg atttcaccac atatggagg agttgggtgta 1920
gacaggcacc tgtggagagt ttcggagaaaat ctatcaggtg agctcagatc 1980
ttaaactagcg ccgtttcttta aagagatgatat ttcacacatt cccgagaaca ccaccagacaat 2040
catttgctct ccaaaagac ccaccatgac ccggttttat tttttcggaa 2100
gctttactcttg gcttttcccc cttggtttttttt atttggggcgg aagggaccgg gttcgcgtct 2160
tccacgcctc caccaagggc ccctcgggtctcc cccctctggg gccacctccttc aagacaccc 2220
tggggggcagc agcggcctctgg ggtgtctctgg tcaagagacta tttttccggaa cccgttagcgg 2280
tgtctgctggaa tccagggcgcg tggacagacgc gcgtgcagac cttccggggtct gtcctcagttg 2340
cgaagtcacc catcagggcc tgagctgccc cgtcacaaga agcttcaaca ggggagagtg 4440
ttagagatcc ccgggtgtgc aggaattcga tatcaagctt atcgataatc aaacctctgga 4500
ttacaaatt tgtgaaagat tgacctgtat ttctaaaatg gtggctcctt ttacgctatg 4560
tggatacgct gctttatagt cccttgatca tgtctattgt gccgtctattg ctctctattt 4620
cctctctctg tataaatctt gttgtcgttc ttcttatag gagttgttgcc cctgttgctcag 4680
gcaacggtggc gtgtgtgcct atgtgttggc tgacgcaacc cccactgttt ggggaccttg 4740
caccacctgt cagctccttt ccgggaacct cgttttcctccccccctt ctctctatgg ccaacggcga 4800
actccatgccc gcctggtccttg ccggctgtgg gacagggggct cgctggtcttg gcactgacaa 4860	tccgcttgggt tttgcgggga aatcatcgtc cttttcttgg cttggctctgct gtgttgccac 4920
ttgattcccg ccgcggagct cctctctgta cgtcctcttc gcctctcaatc caggggaacct 4980
tccctccgc cgcctgtgct gcggctctcg gcctctctcc gcgtctctcgc tgtgctgcttc ctgccttca 5040
gagcagtcgg cgtcctctttt ccggcctgtggt cgataacccc acatcgcata 5100
aatataaaga ttttatttag cttccgacaaa aaggggggaa tgtgaaagacc caccctctagg 5160	tttgcaagtc tagcttaagtt aacgcctattt tcagaagccat tggaaaaatatataactgaa 5220
atatagagaagtt ctagttcagat tgcgctggga ccggacgagc ccggctgtgct ccagggggaac 5280
atatctgtggta taagcagcttc ccggctggtc ggccggaacc cggggaacag cccagtggaa 5340
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5400
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5460
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5520
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5580
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5640
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5700
atatgtggcca cagcagcattc ccggctgtgct cggggctctgc cccgctggtc ggtgaacgct 5732
<210> 7
<211> 9183
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 7
aaagacccca cccctaggtg gcaagctagc tttaagtaacg ccacttggca agggcatgga 60
aatattataa ctagataatag aaaaaactgag atcaagcttg ctacggagaag cagacagttga 120
tacacagcag atatactgtg gcaagctggtt ctcggtctgct cgggacgcaac aacagtgctga 180
agacagctga gctgtgctggc aaacagagta atcgtggtatg gctgtgtgtct cccgtggtcgg 240
ggctggttc cagatggtt ggctctgcgg ctcaggttccgcc agcctgtttc actgttacct 300
cagatcttac caggtggcttc caagacgctt aaatactgacc ttgtaacttat tggaaactaac 360
caatcagttc gctttctgct tctgttccgga cggctttccgct ctccgagctct aataaaagag 420
cccacaaccc cctactcgcc gcgcacagtttc ctcgatagcc tcggctgcgcc ggtaccgcgt 480
atttccataa aagcctcttg catgttggcat ccgaaatcgtg gtctcggctgt tccttgaggag 540
ggttctcctt gatgcattgga ctaccacagca gggggttctt ctatgtggcc gctgctccgg 600
gatggtgaga ccctctgccca gggacaccagc accaccaccc gggggttaga cttgcacagca 660
acttatctct gcgttgcagga ttgtctagtg tctctatggtt atgtttaacg cttgctgcttg 720
tactatctag ctaactagtct cgtgatctgga cggacccgagt gtggagaactga caggtcttga 780
acacccggcc gcaacccctgag gacagctccc agggaccttg ggggctgcctt ttggtgcccg 840
acctgagga gcaggtcgag atggcacaatgg ccacccgtcag gatataaggtc gctgtgtagga 900
gacgcagacc taaaagagtt cccgctccgg tcggataatct ttggttctggt ttggacacga 960
agccgcgagc tgtgcttgctg cagcgcctgct gcagcctgctg tcggatgctct ctctcgatct 1020
agtgttccatc attgtgtgca cagactgtga ccacactcctt aaggggagccc 1080
ttagttcact ggaagagagt gcaagccggtc gttcacaaccc agtcggtagga gtgtaaagag 1140
agagtgtaag tgtcattcct cttctgacga aagcctcggcata ttgaccgtcag atgacgacgaa 1200
gacgcagacc ctacaagtt gacgacagtcag cctgcacaccc actacggtcag cttgcgctgcag 1260
cggcatgcc accacccgac ccgatccctct atcgtgcacg cggcagaccc tcgggctgcag 1320
cccctccctct gggtcaacgc ctcaagctcact tccctctctc ctggccacctg ccgagcagcag 1380
gccccgtcttc tccccctgga acctctctgtc tcggacccgct ctcgatccttc ccctctatc 1440
gccctcactc tccccctgag cgccgcacagt ccgatctgcga cagagacagcag gatgagagatc 1500
gtttgcata tattgacaca atggattgca gcagagtcct cggcagcgtct ggtgtgagag 1560
gctattgcgtctgagatgg cacaacagact accgggtgctgc ttctgtgcgc ccttggtctcc 1620
gctgtcagcg cagggccgag gcgttttcttg tggcacaagcc gacggttgccgc ctggcgttaa 1680
tgaactgcagt gcagagccag gcgcgtatctc cttgtcggggc acgacgcgggg ctttcttgagc 1740
agctgtgtcct gcaggtgtctc tcggaaagggc agggacttgg ctgctgttgg gcagagtgcc 1800
ccccgcgagcgt ctcctctgact cttacctctgc tctctgaggcgc gggcgattcag gatataactag 1860
tgcatagcgg cgctgtgctgata ccgctgttgac cgcctcagc ccagacccgaa 1920
acatcgactgc gcgcagcgac gcgcgctgact ctttgctgtac cctaccagtata 1980
ggcgaaagatc ccacccgctgact cgcgctgact cggcagcagc ccagactggtcc cgggcgatcg 2040
gccgacggc gggagatgct gcgtgtcacc gttggttgcccg tatctgcagcc 2100
gggaaattcgc gccttctactg cttattcagc gcctttgctgctg cttgctccgatc cttctctggtg 2160
tcgccttgact ccttctctgct cggctttgatc tgggttggtg cggcagcactgc 2220
ctctggcgcct gccttttggcgt atccggtgcac tggggcgctgc ccctctctct 2280
cgtttgtagct ggtggttgact cttgggctgcg cttttgattc gctgggctgac 2340
cccacccctgc ccaccaagact tttcctgtcc accggcgcgct tcctgataaag gttgggcttcc 2400
776451Seq

ggaatggttt tcggggagcg ccgctggatg atccctccagc gcggggatct catgctggag 2460
tttctgcgcc accccgggct cgatccctct gcggaggttg tcaagctgtcg cctggggcgtg 2520
gagcagcctcg ccgagttctta ccggcagttgc aataccgctg gcattccaga aaccagcagc 2580
ggctatccgct gcataccagc ccggcaagtgc caggagtggt gcaggccagct gcggcgttttg 2640
gttcgagggcg atctcaggaac taagcgaaat gcgaagacca agacgaaaaa atgcccacaca 2700
tgaggaataac cgatctcttc attaacaatc tcaagccagt tattctggtc taaaaacgaga 2760
agttcacaacc agataacagat catatacagt gtctctccca gaggttcatg actgaacact 2820
cgttccgaga aatcaggttggt atcaagctctgg ggtggtacat gaaagacagt atgctgaagt 2880
tgagctccca atacctttggc cactctgatgc gaagaactga cttacgtagt aagacccgta 2940	
tactgaaaaa gattggaggc aggagaggaa ggatgtagac aggatggaag agtgggtgaggg 3000
aatccacacac tcgtagagca tgagtttggc caagctctca gcagttgtga atgggcaggg 3060
agaacctggtgc tggctgcagct ctgggtttgg ccaaagagtgg gacactacag atgtgcacta 3120
tgtaactgta atgtttatac atgtagtaga atatctggag aaaaaagagg agatgtgcttgcc 3180
tgtttggtag aagatgtttgct gataataaag ctttagataac ctttagtttg gaagtcctta 3240
attatatctct taggtatgggt accacactca atataagaaaa tcaggtctta gaagactag 3300
tagagagagta gacacctggtgc atacccgaaag tcaggtctca gattgtatag ctgtttataag 3360
caatatataa ccaaatatat ggatattatac atgtagagata gatgccagca atattgagga 3420
acatattaga actaatctcc taataactctc atgtttccagc acaactctgtg taagacctcag 3480
gtggcagaac ttggctttata ggcgtcaggt gtatattttg gggcttccct ggtggtctcag 3540
atgttgtata gctctgcctgc aatggtggggt atctggtttgc gatgccctggt ttgggacagat 3600
ccccctggaga aggaatgggc aacccactctc agatctctta cctgggaata cccatggaca 3660

gaggagccttc gtagactaca gtcctatggga tttgcaagaggt ttagacacaa tctgggcacat 3720
aagcacagca cagtacatga tatacctgtg aggtgaaagt aaggtgaagggt tcaatgcagag 3780
gttctctgctc tgtcagagac atctttttac ctttagctgg gcaaggggac ccaagaatatc 3840
tgggttgggt gcctatattc ttcttcaggg gatcctcccc tcccaaggat tgaactggag 3900
tctcctgctgt ctcacttgga ttctccacca cgtcaactac caggtgtaga ctactcttcata 3960
attaaagtgc tggcttctcc ctttctccac ctttcctaca aaggttgggt cactcttctta 4020
taacctttcct tgtggcttacct tcagagcttg tcaagcttta tattttttag aacactattta 4080

tgtagttgtg tatgcttttag attttcactgt gttgtacgtag cttattttact ttggttttgatg 4140
tttgggatgg gcagctgcag actatctcga aggggcagct gatagctgca tacacactct 4200
tcaagttttgc ctttctttctg aaatagaaag cttttctggt gtaagtttacag ctggtgattca 4260
gttctcctgtt gttcttatac tttactctct cgtggtgttc cagaactccag caggtgacag 4320
gcagacatctg gttgtttttcat gcctggttctgt aagttgctctgc gacattttgtc ctggggtcctg 4380

ttcagattgct ttcttggttg gacactggcct gacgctctgatc ctgggacagct tgcagtgatg 4440
gacatacacc ccccctccac attttctgatg tttcttaggg ggaaggggga agctcgggtat 4500
agaacccctta ttgtatatttc tgattgcctc acctctttata ttgcccccccct gcccttctttt 4560
gttccctcaag taaccagaga cagtgctttcc cagaaccaac cctacaaga acaaaagggct 4620
aaacaaggcc aaatgggaaag cagcatctat gtgtggactc tttctgcgca gagaacattt 4680
cctgctatgg actgatatact gggagagggga aaggaagaat aggttggttatt atgggaagga 4740
gcttgagagcc tcaaggctttc tgctttgagca tgaccgcttct cctctctacat tctctctaga 4800
tgtagggctt ggtaccagag cccctgaggg tttctgcagat aatataaaata atagtaactg 4860
agtgatgcct ccacattcagg tctctgggag gcggagatct cagctgggta cccgggagtc 4920
tcgcagggatc cgattctctca ctggcagcgtg tcgggggcttc cggagacatgc gccgaacatc 4980	
tacacaccac aacaccgctct gcaccagggtt gagatctgag ccggggagcc gcgctggtata 5040
attacaacgag agatccgatt acttactgag gcggctgctgg ggtctccagg acaactgcga 5100
acatctcaca cacacacaccc gcctcgagcc aggttgagtat atcgccgggg gacgcgggag 5160
tgtagaattta acgagagaca cccgggagatt caggagctca ccatggtgatg gacgctgtatc 5220
atctctcttt cggtagacac agctcaggtt gtccactccgg aggtccaaact ggtgggagac 5280
ggtggaggctt ttgtgcaacc tgccgggctc ctggcgctgt ctggcttcgcc atctggtgctc 5340
gatttcaaca catattggtat gaggcttggtg agacagagcc ctggaaaggg tctgggtagtct 5400
attggagaata ttcattcaca tagcagtaag attaactcatg cgccgctctct cagggagata 5460
tttaaatcct cgcggagaca cggccagaac acattttttcc tgcgaatgga cagcctgaga 5520
cccgagacac ccgggggctta tttttgtgca agcctttttct tcggcttcctc ctgggtttgct 5580	
tatiggggcc aagggcaccct gcggcagcttc tctcagctct ccacacaagg cccatcgggtc 5640
ttccccctgg ccacccctccat ccacagcacc cttgggggca cagccggcctt ggtgtgccttg 5700
gtcaagggct atctccccgga acgggtgagc ggtgtgtgag gcctaggagc ctggacccagc 5760
ggctgtaaca cttccccggc tggccccctgct tccctaggac tctactctcct cagcagctgt 5820
gtacgccggt cttccacagag cttggggacc cagacctaca tctgtcaagct gatctacaag 5880
cccgacacac ccaaggtgga cagagagttt gagcccaaat ctttgtgaca aactcacaac 5940
tgcgccaccgt ggccaccaac tgaactcttg ggggagaggt cagttcctct cttcccccccc 6000
aaccacaagg aacacccctat gatctccccg accccttgagg tcactctggt ggtgtgtgggct 6060
gtgagccagc aacacccttg cggctagttc aactgtggttg tggagccgctt ggaaggtgcat 6120
aatgcacaca gaaacccgcccc gcggagagcag tacaacacgag cgtacccgtgt ggtgcaggtgc 6180
tcctgcgttc tcggcagggat ggcagagagtt acaagtgcac gttctccccac 6240
aaagcccccc cagccccccc ctgagaacaact tctctccaaag ccagagcggc gccccggaga 6300
ccacaggtgt acacccgctcc cctccccggcc gagagagtt gagaagacacca cggcagccttg 6360
acgctccctg tcgaaggttt ctatccccgac gacacgcggc tgggtggtgga gaggaaatg 6420
cagcgggaga acactacaac gaccagctct cccgtgctgg aacctcgaggg ctctctctctc 6480
776451seq

ccttatagca agctcaagct ggacaagagc aggtggcagc aggggaaggt cttctctatgc  6540
tccgttagtc aagcaggtcttc gcacaaccac tacacgcaga agagcctcttc ccctgcttccc  6600
ggggaatgaa agccgaatcc gcccctctcccc cttccccccc cctaaagttta ctggccgaag  6660
cgccttgaa taaggcgggtgt gtcggtttgt tctatgcttt ttttacccaac cccttcgtcgc  6720

tttggcataa tgaggcccagc gcgaaaccggt cccctgttctct tggagcgacga ttctttaggg  6780
ttcctccccct ttcgccaaag gaatgcaggg tctgtgtgaatt gtcggtgaggg aagcagttcct  6840
tctaggagct tcttgaagac aaaaacgcttc tgtagcgacc ctttgcaggg acgcagaacc  6900
ccacccgtgg cacaggggtcc tcctgcggccca aagggccaggt gtataaggata cccctgcataa  6960
ggcgagcaaa cccacctgccc acgttggtagg ttggatatgt ggagaaagag tcataatggct  7020
cttccccagc gattcaacca cccctgcttgg ggcggctcag cggcttttcttgaaa  7080
atctctgttg ggccctccggtg gcacatgcttt tacatgtttg tggctgcagggg ttctaatggct  7140
cttagcccc gcagacggcag ctttttcttgg aaaaaaacggt gataatctgg  7200
cctcccccttgc tctctgtgctcc tggtagggca ttccttttaca tcgacccagc gcgacatcc  7260
agctgaccca gagcccaagcg agcctgagcg cccctgctgga tggacgagtg aggcatcact  7320
gttaagccac tcagacttgtgg ttacacttctgg atcagcttga tccagcagagg aggcttaagg  7380
cctccaaaggt gctgtactcag tggacacccca cccctgcttcggct gtcggacaga cagcagttct  7440
gcggtagcgg ttagcgtacc gacttcacctt tccacatcag cagacctccag cagacagcaca  7500
tgccccacta ctcactgcccc aatatagccc ttctatggcttc gttcggccca gggacccaagg  7560
tggaatcaca aagacacttgg tggctgcacat cttgctctcttt cttccccggca cttggatagcg  7620
agttgaaatc tgggactccc ccctctgtgg aatactctcgaa cctctctccag gagaagttgca  7680
ccaaagttca gttggaagtggt gataacccccc tcaatctgggg taacttcctcag gagaagtctca  7740
cagagcagga cagcaaggaag acacactcaca gcctcagcag cccctgctgcgg tctagcaag  7800
ccacagcaga gaaacaaaaa gtcacagcgcg cgcaggtacac ccctcaggggc tggctgcgcc  7860
cgctcacaa cagctcttccag aggggagagt gtttagacagc cccctgctgct gcaggaattcg  7920
atatcagcct cttcgataat ccaacccctgg attacaaatat tttgatgagaaa ttggacttgtaa  7980
ttcttaacta tgttggctctct tttactgtatg gttgtacagc tccttaattg ccctttgtatc  8040
atgctatattg tcctcttctgtg cttctctctc ttctctcttt gtataatcct tggttttctgt  8100
ctctctatga ggattggttgg ccccttgcctga gcagaagcttg ccgttggttgct aactgtttttgg  8160
tgagccagcaac cccacactgtgt ggggagattg ccacaccactg tcagctctcttt tccgggacct  8220
tcctttctcttt ttcctcactatg gggcagctgg gacactcagc acctctctctt ggcctgtcgt  8280
ccagaggggc tggctgttcttt gcctcagcata atccctgtgtt gttgtccgggg aatactcatgt  8340
tcttctcttg gctgtctcgcgt gttgtgtgctt cttgagattt ccgctggtgagcccagcctgc  8400
agcttcctctgc ggcctctcactt ccagcgagcc cttccccctcc cgggcctgtgg ccggcctctgc  8460
ccacccgccttt gcgctctctgctg cttgccctcccc agaggagctgc gatctctcttt tggggccgctc  8520
776451Seq

caaagagttc aacaggaatt aagtgttgaat gcactgttatt cccccctgtcc c3120
cccctccccc cccctacggtt actggccgga acgcgttggga ataaaaggccg t3180
tctatatgtt aatctgccacc atatctggctt ctttggcaaa tggaggggcc c3240
gcctgtcttt ctgtcagagcc atctctctgg gttctctcccc ttcgcccaa g3300
gcttctgtgg acttcggtaa gaaagcagttt ctttggagcc ttctttagaag c3360
tctgtagcgc ctttggcagg cacggaaccc cccaccttgg cgacaggtgc ctcctgcggc c3420
aaaagccagc tgtataagat aacactgcgaa aggcggcaca aaccagcagt gc3480
gtggtgatgt tgtggaaaga gtcaaatgcg ttcctctcaag gtatattaac a3540
agagtggcca gagggtacc cattcgtatgg gatcttgatc gggtgccccgg tc3600
ttacatgtgt ttagtcagg ttttataaac acttctggcccc ccggacccaac gccggcggtg c3660
tttctcttt agaacaacga tgaataatag gcctctcttg cttctctgtct ctcctgtagc c3720
atctctattc agttcaccctc gcgccaggtt gagctttacgc agttctgggcc aga3780
aacccagggg cctccagttca gttgtcctgc acagctcttg gctttcaacat ta3840
ttatatgact gggttaagca aagggcctgaag ccggcctggg aagtgatttg a3900
cctgcgctatg ggaatactga atatgacccg aagttccagg gcaagggcacc tataacagca c3960
zacacatcct caacacacgt caacacgag ctcagcggcc tgtacatctg gaagactgcc c4020
gcttattact gtgctagtgtg agggggactg ggggtttcttt acttggggcca agggacttg c4080
gtcactgtct ctgcacccaa aacgcaccc ccatctctgt acctacgtcgc ccccggtacct c4140
gctgtcccaaa ctaactctcat ggtgacccctg ggtgctcttg tcaagggctaca ttcctctgag c4200
ccagtgcacg tcgactctgaa ccctctctac ctggtcagca cttcccagct c4260
gtccctgcaag tcgacctcccaacct ctacatgtga ctgctccccca caccagctcc c4320
ccagcggaga cctgcacctg aaacagttgg ccccgccgca gcagcaggaa gggtgacagc c4380
aaaattgtgc ccagggtatt tactagtgga ggtggaagta gccacacatca ccatcaccat c4440
taatctagag ttaagcccgg gcctggagatct cggacatctg aattcaaccttc tggattaca a4500
aatttggtga aagattgagt gtattctcctt ctattgtgtgct ctttttaccgc tattggtgata c4560
cgctgtctttatatcgtcat atcgtcttcct gttgcctctgct atgcctcatga ttttctctcc c4620
ctttgtataa ccccttgtac ggtagacgg tggtagtggt tggccggttg tccaggcaag c4680
tggtggttgg ccactttct ctggctggac gccccggttc cttggcacc c4740
cgtcagccct cctgtctgt.cgtttcctctt tcctctcttct ccctccccctt aattgcacagc c4800
cgcgcgctg cctggcgtct gcggcsgaggg cggctcggttg tggggcactgc aacttccgt c4860
gttggtttgc gggaatcatgc tgcctctctc tggctgctgc gcctggtttg cccactggat c4920
tgctgcgggg acggatctcctgtca gctgcggcctt aatccacgagg acctcccctt c4980
cgcggccttg ctcggcggcttg ctcggtccctc ttgcccgcct ttcagcagcag 5040
tctggatctcc cttttggccc ggccctccgc ttagttgataa aataaaaaat gttttttag c5100
cctcaggaaaa aggggggaaat gaaagacccc acctgttaggt ttggcaagct agcctaaagt 5160
acgcctatttt gcaagggcatg gaaaaataca taacttgagaa tagagaagtt cagatcaagg 5220
tccaggaaacg atgtaaatggg cccaaacagga tatctgtggt aagcagtttcc 5280
tgcctgctggac aacaggttga aacagctgaat atggggccaa caggatatct 5340
gttgtaaagca gtttcctgccc ggcctgccag ccaagaaacag atgttccccc gatgcggcct 5400
agccccctgc gatgtttctaga gaaacatcag atgtttttcag ggtgcggccaa ggacctgga 5460
tgacccctgtg ccttatatttg aactaacaat cagttccgcttt ccgctttctg ttcgcgctgct 5520
tctgctccc gacgtctcaata aagagccccca ccccccctca cttcggggccg cagttccctcg 5580
atggactggag tggcgggtggg acccgtgtat ccaatcacaacc cccccactccgtgctcagcagc 5640
ttctgtgcttc gctgttcctcttg gggagggctct ccctcgtagtgg attgactacc cgtcaggcggg 5700
gttctttcat t 5711

<210> 9
<211> DNASynthetic
<213> Artificial Sequence

<220> 400
<223> Synthetic

<240> 9
tttgaagac cccaccctgtat gctggcagct gatgcttaagt aacgcctactt tgcaagggcat 60
ggaaaaatag ataacttgaga atagaaaaagt tcaagatcag gtcaggacca aagaaacagcg 120
tgaataccta aacaggatact tgggttaaggc gtttctctgcc cccggctcagg gccaagaaca 180
gatgagagacct gaggtgagag ggcacaaaaac gataatctgtg gtgaagactgt gttgccccgg 240
cctcgggaccag acagacagtgt cttccctgcat ggtctgccagt cctcagctgt tttttagtgaa 300
tcacagagtc gtttccaggtt gcccccaagga cctagaaatag acctgtctcc ttattttgaaa 360
taacattagactg tggcggctttct ggtttctggtt ccggccggcttc gctgttcctaga gttcaataa 420
agaccccaaa acccctcactt ggggctcggga gtttctccagt aagactgcgtcc gccccgggtac 480
cctgttcccc ctttttacgcct gatcagcttg ttcggttctt tcttcttcttg 540
ggagggtctcc cttccaggttag tggactacc accgacgggcttttctttttg gggggctctg 600
ccggggacctt gagaaccccttg cccaggaccacc ccagcgcgggtaaagcgctggcc 660
agcaacatttt cggtgtcttctgg caggtttctgt aagctttctgt ggatgtttg agtgctgctgc 720
tctgtcacttt gttgctaatc agctttctgt tggcggagacc cggccggttgwtctcagt 860
ctgaacctcc gcgccccaaa cctgggagact tccccccgctttttttctgttgg 840
ccggacccga gggaggagagt cgtggtgaggaa tcggacccccg tcaggtatag tgggtcttgg 900
agggagagag aacccacccctt agctttcccccttcgccggttctgctga 960
ccggagccggc ggtcttttctgcgtctcagc cagctttggggtaa ggcagctggggg 1020
ccctgcaccc gacatggaga cacaacatcc aggtttctctg tggcggcttc cgtggttatac 1080
ggcggggttt ttcttggtga caagaatcct cacaatacca cagagtctag actcgtggtg 1140
gactcttccc aattttctag gggagcacc cagcgttctt ggccaaattt ccgacttccc 1200
aacctcaaat cactcacaaa cctcttgctcc tccaattttg ctctggctatc gctggatggtg 1260
tctcgccgct ttttcatat tccctctcat cctgctgcta tcgctctatc tctggttggt 1320
tcttcctggac taccaggtta tgggtagccgt tgctttctctta ttcccagggaa catcaactac 1380
cagcagggga ccatgcaaga cctgacgat tcctgtcctaa ggaaactctta tgctttccc 1440
ttgttgctgt acaccaacctt cggacggaata ctgcaccttgt attcccacccc catcactcctg 1500
ggccttcga agatctctat ggagttgggc ttgaagggct ctcagtttttct 1560
agtgccacct ttcctaggtt tcgtaggcct ttcccccaact gttggctcctt cagttatagt 1620
gatgatgtag tattgggggc caagctcgtta caacatctttt actctctcttt taccctctatt 1680
accaattttc ttttgtttttt tgtatacatt ttaaaacccca ataaacaacaa aagtgggggc 1740
tactcctctta acctcattat cgtgatgggg gatcattttc gacaagacat 1800
attgtcataa aaataaaaga ctagtttttga aacactgcctg taataagacacttattgatgg 1860
aaagttgccc agagacacgtt ggtgtcttttg ggctttgtctg cccccttttac acaattgtggc 1920
tatcctgctt taatgcctttt atagcatgt atacaaatcta agcaggcttt cacttttctcg 1980
ccaaactaca aggcctttttc tgtgaacaaac tatttgaaacc tttaaccggt gcccgcgaa 2040
cgggtcaagtt tcctgcaagtt tggcgtgac gcaaaacccca ctggatgggg cttggctatc 2100
ggccataagcc gcatgcgagg accttctggg ccctcttttc gatccatact gccaagctc 2160
tagcagctttg tttctgctgc aggctggcttg aggcggacact tatctggcacc gacaacctctg 2220
ttgctctctc tcgaaatata acctccttttc catgggtgctg aggtgtgtcat gccaacgtgga 2280	ccccctcagg atataaagtgt ttgctcttttg catagggagg gggaatagta gctttatgca 2340
atacaccttgt agtctttcga catggttaacg atgagttaac aacatgcctt aacagggagg 2400
aaaaagcacc ggtcagctggg atttggtgaa gtaaagttgt atcgatcgtgc cttattagga 2460
aggcaacaga cagctttcag atggattgga cgaaccactg aatctggcat tgtcagagata 2520
atgttatatta agtgctctagc tcgatagcgc aaacgctcatt tttggcataatt caccacatt 2580
gtgtgacact tttaaagctt caacgctcgg caagcactca gggcgcaagg gcgtgtaaag 2640
gaaggcaac acgtagtaaaac cgcctcgcga gaaacggtgc tgaaccggga tgaattgtcag 2700
tactggggtct atctggaacaa gggaaacgcg aagcgcaaaa gaagacaggc tagctttgcag 2760
tggtgctacaa tggcagatgc tagactgggct ggattttatgg acagcaagc caacaggaat 2820
gccagtgggg gcgcctttcttg tgaaggtggg aagacctgtg ccagtaaactt gatgctgttt 2880
cttgccgcca agaggttatg ggcacagggg atcaagaatct gtaaagagaga caggatgagg 2940
atctttccgc atggatgaac aagatgattg gcacgcaggt ccctccggccg cttggctgga 3000
gaggttatttc gcgtatgact ggccacaaca gacaatccgg gcgtctgtatg ccgcctggtt 3060
cggctgtca gcgcaggggc gcccgggcttt tttgtctcaag accgacctgt ccggtgcctc 3120
gaatgaacctc caggacgagg cacgcgcttc atctggtgctg gcaccagcgg gcgttcccttg 3180
cgcagcttgt ctgcagcttgg tcactgaagc gggaaggagc tggctgtcatat tggggcaagt 3240
gccgggcagg gatctctctgt catctcacttc tgtctctgcc gagaaagtat ccatcatggc 3300
tgatgcaagt gcgggcgttgc atacgctttga tccgctactcc tgcgcaattc accaccaagc 3360
gaaaactcgc atcgacgagc cacgtactcgg gatggagaccc ggtctttgctg atcacagatga 3420
tctggacgaa gagcatcagg ggtctcgagcc agcggcagctg ttcggcaggcc tcaagccgagc 3480
catgccgcac gcggagagtc tcgtctgtgac ccattggcgat gcttgcttgc gcaaatcatc 3540
ggtgggaaat gcgcggctttctt cttgattcag gacgtgtggcc gcggctgggtg tggcggagcgg 3600
catatcagac atacgtttgg ctacccgtgga tattgctgaa gagctttggcg gcgaatgggcc 3660
tgaccgcttc ctccgtttctt acggtatcgc cgcggcctgag tgcagcgcgag tgcgctttta 3720
tgcgctttctt gagcagttcttc tctgaggcggg acctctgggt tcgaaatgcg cagcaacagc 3780
acgcccaacc ccagttcagct cttccaggcc ccctcttattag aagttggcgcct 3840
ttcggaaatcg ttttcgggga gccgaggcttg atgatcctacc agcgcggaggg gttctcaagtg 3900
gatctccttg ccaccccaaa cccctgcccc attattggttt ggaactaacca tgggggggaat 3960
tgcgctgtgga ataggaacag gcagactgtgc ttcataggc ccacctcagat ttcagagcgt 4020
ccagcccgga gtacagagtt atctcaggag ggtggaaaaa ttcaattcctc acctgaaaca 4080
gctctctcttc tctctctggtt caggttctctc acagaatacgag agggcctctag acctgttatt 4140	tcgaaagaa ggaggctgtt gttgctttctt aaagaagaagaatgttctctctt atgcgggacca 4200
cacaggtactc gtgagagaca gcattggccaa atgtgagagag accgtaatctc agagacagaa 4260
actgtaagg gcaactcctag tgcagtcgggg gggcagttgtt aacagatcccc cttggtttac 4320
cacccgtgata tcctactatt cccgaggctc cctgtaccct ctaatgatttt tcgtctcctgg 4380
accctgcatc caaaggggtggt ggtgctgctt ctcagcattt tggtaaagagc agggagcattg 4440
tctagttttcg actcacaatt ataccaacatgt ctcagcattt gcagcctattc gtagataaa 4500
ataaaggttt ttattattgc ccagaaaaaa gggggaagttt aacagatcccc cttggtttac 4560
tggcaagctgtc tttgaaattag ccagcttggg cagggcgattttt aaaaatctcctaatgagtat 4620
agagaggtctt agatcaaggt cagggctcgc gaaaaagctgt caatatttggc cacaacaggt 4680
atctgtggta agcgcgttct gcgcgggttgc aggggcaaga acagatggag cagctggtta 4740
tggcccaac agagatcctg ttgtaacaggt ttcctctcgc gcctcggcgc caagacagaca 4800
tgtcctccacag atggcgtcctag ttgcttcttcag agcagtttgg tctttcggagt 4860
gtctctctcttc gactgctgtggttgc cctccgctggct cggggttcggct gttctcttcag 4920
tctgcttcttctgtc tgcgcgcgctt tcgcttccggc agctcctaata aagagcgcac aaccccttcac 4980
tggggtgcgg gttggtccgga ttcgctgttct tcgatattttt ccctcggggta cccccgtttttc 5040
tctttcaggt ttggcctctgcttcttg ggaggttgcttc tcctgaagttga 5100
tgtcagcatc tggcttccttc tcgggcttcggg gttctctcattt 5130
776451Seq

<210> 10
gatcagtcct ggtggtcat tgaaggact gatgctgaag ttgaagctcc aatactttgg
ccacctgagt cgaagaactg acctactgtGA taagacccctg atacctggaa agattgaagg
ccggaggaga agggagcaca gaggagggaa gatgttggag gaatcaccaag cttcagatggac
atgagtttga gcagcgttcctc agaggtttgt gatggccagg gaaagctggc gtgcggactgt
ccatggggtt gcacagagtt ggcacactct actgctgctga aacctgcgtg tagttgtaatc
catgccagt aatataagtt aaaaagaagg aagagtttgc cctgtatcttg aagagttgta
ggatataaa gtttagataa ccttttagttt ggaagttctta aatatattac ttgaagttgg
 ttccacacgc aataaagaa actagcgttct gtagagagaa tgagccctgg
catccagaga gcttacacgt attggttata gccagttaata ccaaatatat accaatatat
 tggattatata gcacactgct tagtggccagc aatgtaagag aaccatttag aactagttac
cttaaactcta catgttcagc gagacactgtg ttattagttct gtagcacagattttatat
aggctctaggg tgttatttgtt ggggctttccc tgttggctcaa gatggttaag tagttgctgt
caatgttggtg gatcttggttt gatcctctgg cttggagaga tccccctggag aagggaaattg
cacccaccact tagtactcttt acctgtgaaat tccctagcag acagacgcgtc tgaagctcgt
agttcagaggg attgcaaacag tggcagcacact caagcagcact acagttcacgt
atatcactgtg gatttggttg gaagcagtaat ttcattcagc tggcttccagt ttcattcagc
 gattcttttttc cattttgtgag ccacggagag ccaccaagatt gtagcattgtt ctagcagattt
catttttc attctcttttc acatcctttgc acattttttgc ctataatttc ctagagatgt
 gattttgc attcttctgg gaattttgctg agattttaag ggttggttgtc ggctgcacgg
taatccttgc attttttttt gtttagccaat tgaacagtattagtttttta 1260
gattttatc attttttttt attctttttttt ttctttttttt gaatattttt tggctttttttt
 gattttttt gaatttttttttttatttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
acagtcttc ccagaacca a cctacaaga aacaaagggc taacaaagac caaatggga

1800
gccaggatcat ggtttaaatc ttttcgaccgc agagaaaca atacctgctag gactgataac
gaacagaag ggagagga aagagaaag tagggagagatt tatggaagga agctgacgaggg
tctctctgac agtaccagtc tctctctcatt cctctctctag atgtagggct tggatacaga
gcctgctacc ccctctgcat gaatataaat atataaaccag taaatgatgtc cccatcctcag
gccttgccgg ggcgccgattt cagagctcgg cccgggggat ctcgagaagc ttcataactag
gaatggagct gggctttctt cttctctctag tcaagtacac caaggtgtcctt ccctccagtt
cagtggacgc gactgacggtcg agagctgtggg aacccgctggg ccctactgag gttctctgag

gaaagctcttg gctacactctt cactgacgat gcaatctact gttgtaaaca caagcccagca
cagggctacc gaaagcttggg atatattatttt cccggaaatgt atgatatttata atacaatgag

gaggtctcagg caaagggcact acctggctgca aacaaaccctt ccacgacagtcc tctacctag

tcctactgag gcctgactcacc ccctctctcag gagccggaggg cacggagggc
ggaggtccgg ggagggagg gctgagttcat gtagtctgac agatctctcatt ctctctcact
gtacacagt gcggagaggt tcaagtgtgc ttgaaaggtc gtaagctcctt tttatatatg

ggtaatcaca aagaactatt ggcctcggtac cagcagaaac gaggaggtcgt cctctaaagt

tctgattatac ggcagtcgctct taggggtactt tcctgcctctt aagctaatactt caggcaagtga

tcttcggcag atcctctct ctcctctctccag atctgcttcagt cagcagtttcg
gaggtcagcc gttttgaggt cttgagaggtg tcctggccca cttaaaactgg

cctgagacgg accgcaagcc atcagctctcg cagcagctacat gcggcaggt cttctacaac
tgggagcgtta atggagagg cattacttca ctaaacgcctt ccacgacaccg

tctgtaacag cggagaggtt ggtcagctcc cttccccaaa caaaaccac cggacacccct

gtccagttcc gacagccctg gttcagcatgc gtggggtgtgg acgtgagcct cgaagacctt

gaggtcaagg tctacagcag cttgagacgct tctgggactgct ataaggctaa cacaagccgg
cagaggtcctt ctcgctgctggt gcatagactg ctcgacctag cgcagccagc

gggtgcgtga atgggcaggt gattctctcc acaagacccc cccagcccccc

cacctgaaag caggctccaa agccgaagg cagccccgag aaccacaggt gttacacctgt

cccatcctcc cggagatgact gcaacagaac cagttcagcc tgaacctgcct gttcaggagc
tttaatccac ggcacagtcc gctggtgagtgg gagagagcaag gcacgcaggag cacaacactc

eaagacacgcct cccccggcgtc gcagccgacgc ggctctccct tttctctacag caagccacctc

gtgagagagtt gcaggggaaat gttctctctct ggctcgtctag gcatagagct

cgcagcaca ctaacctgcg caaagagcctc ccctctctctctt ccggtgattac ggatatttaatt actgatcgagatcc caaaccctc
tgtctagaag aagaactcaa acctctggag gaagtgcctaa atttagctca aagcaaaaaa c 3840
ctttcactta gaacccagga ctttaaagc gatatacaag taatagttct ggaactaaag 3900
ggatctgaaa caacattcat gtgtgaatat gctgtgtaga cagcaaccat tgtgataattt 3960
tgaaacagat ggattaccct ttgtcaaaac atcatctcaaa caactaatgg aagcttga 4020
acatcgataa aataaaaagat ttttatttgt cttccagaaaa aaggggggaat gaaagacccc 4080
accttgaagtt ttgccaagct agcttaagta acgccatattt gcaaggtatg gaaaaataca 4140
taactgagaa tagaagaagtt cagatcaagg tcaggaacag atgggaacgc tgtaatgtgg 4200
ccaaacagga tatctgggtg aagcagtctcc tcgccccgcct cagggcacaag aacagtggg 4260
acagctgaat atgggccaaa cagagatctct tgtgtaagca gttcttgccc cggctcaggg 4320
ccaaacagca atgtgcctcusa gatgcgggtc agccccctcag agttttctgag gaacactcag 4380
atgttccccag ggtgcctcusa ggaacctgaaa tgtacctgctgt ccctttttga actaaccaat 4440
cattcgcttt ctgtcctcttg ttcgccccgtct tcgcttcctcc gagctcaataa aagagcccc 4500
ciaaaccaatc ctcgggcgcgc cagctctcctg atggactgag tcgccccgggt accggtgtat 4560
ccataaaacc ccttctcgagt gcctgacggg ggtggtcttc gctgcttctctt gggagggctt 4620
cctctgagtg attgacattcc gcgcagcggg ggttctttct t 4661

<210>  11
<211> 5691
<212> DNA
<213> Artificial Sequence

<220> Synthetic

<400>  11
gatcagttcct gggtggtcat tgaaaaggact gatgtgatga tggagctcctt aatacctttg 60
ccacccgtgt cgaaagaactc actctgtgta taagcctctgt atactggaag agattgaagg 120
caggaggaga agggtgacca gaggatggaa gatgtggatg gaatcaccct tccgtatggc 180
atgagtggaa gcagttcgctg agagttctgt aatggggcag gaaagcttgg gccttgcagt 240
ccatggggtt gcaagagagt gcaccatcct gatgcagctca gtaacactt ctagtgaactc 300
cattgctgag aataatagtt aaaaaagagg aaggttgctgc cctgtatttt aaggttgatta 360
ggatataaaa gtttgagaata cctttgatgtt ggaagtctta aattttatc ttaggtgagg 420
taccctcctc aataataaac atcagccttgg agagactgtc gttagagagaa tgagctctgg 480
cataccagaa gctaatcagct atgggttata gcgtttataa ccaaatatat accaatatat 540
tggatatata gcatgagcct tgtccgccagc aatgtgaggg aaccatttag aactagtctc 600
taaactcta catctccag cacactgatca ttaagcctca ggtcctgaat cttgctttat 660
aggctctagg tgttatattg gggcttcctcc tgggtgctca gattgtaaag tgtctgcccc 720
catactgggtt gatctctggct cgtccctgctg cttggaagaa ccaccttggag aaggaattaag 780
cacccaccc tagtaccttt actctgggaaa cttccatgacc agaggagcct tgtagcttc 840
agtccatggg attgcaaaaa gttgaacaca actgagcacac taagcacagc acagtacagt 900
atataccctgt gagggtgaagt gaagtggaag ttcaatgcag ggtcttctgc agttgagaaaa 960
gattctttac catcttgagcc accagggaag cccaagaata ctggagttgg tagcctattc 1020
cttcctcagg ggatctttcc atccccagaag ttaagctgga ggtcttctgc aaaaaattggg 1080
atttcctaac acgtaggata ccagaggtgt actacactca aataaagtt acgattaaggg 1140
agttttttcct ccctttttccaa aaagttgggg tcaacttttt tttaacctctt ggtgcccctact 1200
cctagagggct ctttttctac atataataat cattagttttt gttgttctttttt tatagtttt 1260
gagtttctata ggtgtatctg gatatttttg ggtttttttt gtttgggtagt gggagggctga 1320
tagcatctgcaaggccagct atagatcttc ctaacaatat cattttttttt gttgttttttt 1380
agaataagaca gttccttggt atgaattata tttgatctg gttatcattctt 1440
attcctactcc tggagacactca acaaggaacc aagatataca gggagacactt ttttttgg 1500
tgcactctggg gatgcctgctc tctggggcctt tggatatgctg cggattttttttt 1560
ggagatgtgc cagcctgtgat cccggagctg tggaggtttag cggagcattagc cccctcccc 1620
cattctcact gttcctcaggg ggagaagggag aagctggtta tagaactttt tatttttttt 1680
cctgagtgcct cactcttctat atgtgccccca tggctcttttt ttttttttctaa aagcagtt 1740
acagtgctcc cccagacacca cccctacaaga aacaaagggc taaacacaaca aataggagaa 1800
gcagtagatc gttggttaact ttttttttccaa aagagaacct aacctgtatag gctagatcac 1860
tggtagaggg aggagaaaga tagggtgaat tatggaagaga gctggcaggct ccagtctttttt 1920
cctgtctcggc atgcacactc tctctctattt cttcctctattt agttgattggt ctggagatgc 1980
gccccctcagg gtttctctat gaaatataat atatgaaact gatgtatctttt ccccccccc 2040
ggtttcctggc ggcgacagtcc aacccggggct cttcagggact ggttttttctt 2100
acctgcaggt gttggggttct tccgagacaa tgcgcgaactc ttaacccaca aacaccggcc 2160
tcgaccaggc ttgagatactg cccgggggacg cggcgggtct gtttgatatg ggtgtttttt 2220
actttactgg caggtggcctg gggctttgaga gacaatacgt gcacatccaca ccacaacaaa 2280
cggcctcgac cagggtggata tacgagccgg ggagcggggcg gtgtgtaatta caagcagagat 2340
ctcgagttta cagattctgg cttctctagg gcagggatcc cgggaatcgg gcgcgtggcaca 2400
ccatctcttc cccttttcttt tagagctctat atactcattcc aacccaggccc 2460
agcttcacat gcagctggctgc gggcttcgacg tggtagaacc tggagtcttc gcagtagtatc 2520
cctgcaagct tttttctgtact acctttctaca gcactactttt aacgtggtgtt aagcgagaggc 2580
cctgcaaggg attcttctgg attgcatggag ttttccctctg aataattttt acactgtaca 2640
atgagataatc aagggcacag gccaactctga cttcagacca attctctcagc acagcctaca 2700
tgcacactca acagctggcc ctggtactgc tttttttttt gcagggactc agggtggtacc 2760
atgacattgct catctctgggcc caagggcaac cttcttacact cttcctcagcc aaaaaagcaca 2820
ccccactctgt ctactccactcg ggcctctggttc tcgctctgccca aacacttccct actggtgacc 2880
776451seq
tgggatgcct ggtcaagggc tatccccctg agccacagtc agtgacctgg aaccttggat  2940
tccctgccac cggtggtcac accttcctcg cttgcgtctca gttggtgact tcaacactga  3000
gcagctcagt gactgctccc tccagcactt gcgcacgaga aacgctcacc tgcaacgttg  3060
ccacccggcc cagcagccac aagttggaca gaaataattg gccacaggag tggctactgtg  3120
gagttggagc tagctaaagg aagttgcgag ggtatcggcgg aagattgcttc ctctgcccc  3180
ccccccccctta aagttctctgg cgcagcggcc tgggaataaag ggcgggttgcc gttgtgtctat  3240
atgttatatt ccaccatatt gcgctctttt ggcaatgtga ggccggcggaa acctggccct  3300
gtctttctgg cagcacctcc tagggctcttc tccctctcttg ccaagaggaat gcaaggcttg  3360
ttgaaattcgg tgaaggaagcg agttccctcttg gaagcttcttt ggacaacaaac aagctcagtt  3420
gcgacctttt ggcggcgagcg gaacccctaa cacggccagca ggtcttctcg cggccaaag  3480
ccacggtatat aagataaccc tgcaaaagggg gcacaaaaacc aagttgacgt alltgattgg  3540
ataagtgtgg aagttgctcc aagtgccttca tcaagcgttat ctaacagaggg gctgaaggt  3600
gcccaagagg taccacccatttg tatggtgattt gctggtgctc atggctttca  3660
tgttgattagt cgaggttaaa aaaaaagtctta gccccccctga accacggggga cggttgggcttcc  3720
ctttgaaaaa cacagtgata atatggctcct ctttgctctct ctctgcctcact taggcactct  3780
attccatgc acaccgagcgg acattgtgct gacacaaacct ccaagcaatc tggctgcctc  3840
tccaggggag aagttcaca cgcagcgcctgc gcccacccctta agttgaagtt catcaactcg  3900
gtacccacgc aagttcgagca cccctccaa aagatggatt ttagacacat ccaacagtggc  3960
tttggagttc cctgctctgtct cagtggcgag tgggctttgg acctttcact ctctcacaact  4020
cagcgagtct cgggagttggg gctgtcgcct actacaacct tattactgc cagcaggtgg gtagttactc  4080
cactggctgt ggccgaggca accgtgcgct attgtttggct cagcgctctc aacgcttactc  4140
cattttccca ccacctgcagtc gacagttac aatccgagtt cgcctccagtc tgtgctttttc  4200
gaacacttc taccacccaa acataatgtg caagggttga atgtgttgag ctgcagcaaca  4260
aaatggcgtc ttgcaagagt ggactgtgatc gcgcagcaca gagcagcaact acagctagag  4320
cagaccccttc acgttgacca aggacagga tgaacgagac atacgcttta cctggtgaggc  4380
caactacaag acaataacct caaccattgt caagagccct aacaggaatt tagtttaata  4440
gggagagcttc gcaacagctt aacatcaccct tggattacca aatatggtaa agattgactg  4500
tgttctttaa ctagtggctct cctgcttcttg ctttctgtta ctggctttta atgctttttg  4560
atcatgctat tgtcctttctg atggctttta ctttccctcc tttttctttc ttctgtttggc  4620
tgggatgttc tgggctttgt ggccggcttg tcagcaagcac tgccgttgctg tgcaacttgtg  4680
ttgtgacgc acacccacct gggtggccag ttgccaaccac ctgtcagctc cttcggcgga  4740
cttttgcttt ccctccctctc acttggccaggg caacagccatc cgcgcatgtc ctgcccgcgt  4800
ggtggcagcc gctggtgcttg tgggcttcct acaattgccc ggtggtgttg gggaatcata  4860
cgtctttccc ttgagcttggc gctgtgcttg gcaaacgtgt tctgccggga acgttcctct  4920
776451Seq

gctacgtccc ttgcgccttc aatccagccg acctctctcct cccgcgcctgc ctgcgcgctc 4980
tgctgctct tccgctcttt cgccttgcgc ctcagacgag tcggatcttc cttttggcgg 5040
cctccccgcc gtatcgaaaa aataaaagat tttttttagt ctccagaaaa aggggggaat 5100
gaaagacccc acctgtaggt tggcagctc agcttaagta acgccttttt cgaagggcatg 5160
gaaaataaca taactgagaa tagagagatt cagatcaagg tcagaacagc atggaaaccgc 5220
tgaatctggt ccacacacga tatcttatgg aacgcaatccc tgcgcgctc cggccggaag 5280
aacagatgga acagctgaat atgggccccaa caggatattc gttgtaagca gttctgcgccc 5340
cggctcaggg ccaagaacag atgcctccca gatgcggctc acgcctccagc agttttcaga 5400
gaaccagtg gggttttgaa gcacatcagc atggcccaac gaggttcggaa tcggcagtgga 5460
actaaccaat cagttcggct cctgcttggg tgcgcgcgggt ctctgctcggc gacgctcaata 5520
aaagacccca caacccccac ctcggtggccc aacgtcgttgc attgagtcgg tcggccggtt 5580
accggtgta ccaataaaacc cttcttacgt gtcgatgcag ctggtgctctc gctgcttcctt 5640
gggacggct tccctgacgtg attgacattc cgtcagcggg ggtttttctat t 5691

<210> 12
<211> 668
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 12

ggaattcgcc cctcttcccctcccccccccct aacggtacttg gccgaagccg cttggaataa 60

ggcccgttggt ctggtgtctca tatttttttt tccacacatc tgccgctcttt tgtgcaattgtg 120

gggccgcta aacctggtccc gtctttcttg agacgccttc ctgctggtct cttttttcttc 180

gccaaagaa tcgaaggtct gtggatgcgc tgtgaagggcg cagttctctct ggaagctttt 240

tgagagccaa caacgtctgtg agccgacccttg tcgagggcgc ggaaccccccc acctgggcagc 300

tagtgctcttt ggcggcacaac gcaagcgtgta taagatacact ctsacagcagc ggcaccaaccc 360

catggtcgcag ttgggtcttt gatactttcg gaaaggtctca atttggttcttc ctcgaagctga 420

ttcacaaaggg ggcgtgaagg tgcgcagagag atcctccatt tgtatgggtt gccagcgggg 480

cctcgggtcag cagttcttact atggtgtttag ctaggtttaa aaaaaagtct aaggccccgg 540

aacacgggg acgttgggtttt ccctggaaaa agacgtatg aatattggct tgcgtcatcct 600

tacctgcttt gttgctgttg ctcttgcggcg ggcaccggga tatcttagac tcgaagctgctc 660

gaaagctt 668
2005208507
03 Oct 2006

223> Synthetic

<400> 13

7TGGAGGAG TCGGCCGAC
GGGCCGACCG
GATGTAACG
TGCAGTCAAG
GGCCAGGCA
ATACGTGC

776451Seq

03 Oct 2006
atctggacga agagcatcag gggctgcgcg cacgccaagct gttgcccagg ccaaaggccg 2040
gcatgccccga ccggcaggat ctctgcgtga cccatggcga tggctgtcttg ccgaatatca 2100
tgggtgaaaaa tgccgctttt tctcgattca tcgactgtgg ccggctgggt gtggcggacc 2160
gcatcagaga catagctgtg gcataccgttg atattgctga agagcctttgc gcgcaaatgga 2220
tgacccgctt ctctgtgcct tacggtatcc cgctcctgcc ccgctgagcgc atctctgtct 2280
atcgcctctct tacgagagttt ttctgagcgg gactcttggtt tcgcaaatgta ccgaccacagc 2340
gacgcccaac ctgcatcaca gaggattcga ttcaccgcgc gcccttcattg aaaaaattgg 2400
ttcggaatca ctccccgagg acgccaggtg gatgtctcctc cagccgaggg atctctgtct 2460
ggcgtcttct gccacccccg gcgcgcatcct ctcggcagatg tggctgctatgc gtcggctgtg 2520
gtcggcagac tgcgtccaggt tctactggca gtcgcaatcc ctcggcagatgt cagagcaacc 2580
cagcggtcat ccgccgcttc atgccccccag atgctccaggg tggggagcc gcatacgccgc 2640
ctggtgctag gggagccttc cttgatttcat tgggtatatata gcttttttcat 2700
aatacgtgca ttggtcctag ttcatacttt ccttttttttta caatgtggtg ggtatgttcc 2760
gtcatacatc ccaatccag atctggtcat attgattatt gactagttat taataaaggg 2820
caattacggg gtcattacgt tagccatctt aatatgaagtt cccgctttggc taacctcag 2880
ttaatatggc gcgcgcctga cggcccccaag acccccccggt atggctctct atatgcagtt 2940
atggttcatc atgaacgcca atagggacct tccattgacg tcaatggttg ggtatccttc 3000
ggtaatctgc ccactttgcca gttcatacag tgcctctttt gccaagtacgg ccccttttgtt 3060
acgtcaatga cggtaaatgg ccggctttggc attatcgcctt gcgttggagt atatggaggt 3120
tttctacctg gcagtaacac ttcggtttttc ttcgtctgatc tccgattgttt gtcgctgttc 3180
ggctcgcctct gtaatgcggct ggtagccggc tgtcatttac gggactcttcg cacgttctcc 3240
ccattgactg caatggaggt ttggttttggc accaaaaatc aagggacatct ccagagtttgc 3300
gtatacaactc ccgccccattg acgcaatttg gcgcgtagggc ttgctgtggtt caggtctattg 3360
taaacagagc ttgctctgatg aacgctccaga tccgcttcggc aagcacatcct cggctttttg 3420
acccctcatg aagacaccgg gacgcattcc gcgcctccccg ccaccagcct ctgggagtcgg 3480
ccccggagggt ttgctggacc cacatggcgc gcaaccttttg caggtgagtc tgcgttcttg 3540
tccctctgtcc agtaactaca ggtctctccgt cgcatactcc gatgaccacc agcttctctgt 3600
ccctatctgc atctgtggga gaaactgctca ttaatacagc tcgagaagttt ggaatattgc 3660
acaatatattc acatggttcatc gaggagaata gcctttcccc gcctgttctata 3720
atgcaaaaaac ctttaaccgct gtgggtgtccct caggttctacgg tggctgtgagt ccaggaacc 3780
aataatcatct caaggtcaac acgcttgccag ttcgaggtttt ttgaggttat tactgctcaac 3840
atttttgagg tctccctgtgg agctttgtcttg gaggcaacaa gttcggacatc aaacgggctt 3900
atgcgtcacc aacttctatcc atctctttctcc ctccgagcttg gcagtttaaca tcggggaggg 3960
cctctgtgct gtggctctttt gaaacatact acccccaagaa catcaatgctt gaaagttcga 4020
ttgatggcag tgaacgcaca aatgacgcct taacgaacttg gagctagtcag cacagcaacag 4080
acagccacct caacagtcag accagccacct cacagcaacct gaagcacgtg ataagcgaaag 4140
acagctatata ctgctagggc actcacaaga cattcaacac acceaccttc accagcttcc aagacgccttca 4200
acaggaatga gtttggaagac catcagtttc ccctgtaatcc gcccaccttc ctccctcccc 4260
cctaatgctta ctgctgcaag ccgctttgaa taaggcgggt gtcgctttggt cttatagttta 4320
cttccacca tattgctgac ttttgcccag gtaagggcgc gagaacccgg gcttgctctcc 4380
ttgacagagca tccccaggg cccttccct ccctgcaagca cagatccgct cttgcaagcc 4440
gttgaagaggg tcaatgctgg ctttctcaagc gattcaacag gattcgagtt gttgcctgttt 4500
tgggaagagt ccattgctgg ctttctcaggg gatcacacttt gctgccctgg gccgactgctc tatactggtt 4560
agtctgaggt tataaaaccc tctagcggccc ccgaacacag gggacgctggt tttccctttga 4620
aaacacaggt gataaatggt cccctctttgt ctttctgcct cttgtagggc tccctattccca 4680
tgccaaccag gggcagtttc gtcctagca gatctgaggg cagcttggtga agccaggggc 4740
tcagctcaggg ttcgtcctgggc cggtctctgg gtcagcgcta gctgcccttt tttatgcagc 4800
atttgtaggt gtaaaaactgc tctagcggccc ccgaacacag gggacgctggt tttccctttga 4860
gaatactgtaa tagaaccagg agtgcacctc ccagccagct ataacagcag acacatcctc 4920
cacacagtc aacctgagcc tcagcagctt gacatctgag gacactgcgc gcttattactg 4980
tgtagtggga gggacgctgg gttttcctta cttggccaga gggacgcttg gcctgtgccct 5040
tgcagcctcaaa acagcaccct cattctgatt cctcactgaggg tcgcctgtaag ccgcttttgcag 5100
taacctcaggg tggcacccctgg gattgctgag tcccccagag cagctgcctga 5160
gactctggaac tccctgcattg ttgcagcggc gttgcaacac cttccagcgtg ttcctgcgatc 5220
tgactctgtctgctta taccgtgagc cttctcctcct cccagcttccg cccacactggag cttgagtttgcag 5280
agtctcaggg caccagggcc caccagccag cagcagagaa gtcgagaaag cttgagggag 5340
cagggattgt atcgatgttg gtttgaggtag ccacccatcc catcaccatt aatctgagtt 5400
taagggcagg ctcgagatcct gcctctcctag tgcacaagctc taaataaataa aatgttatagttt 5460
agatgacgagg tttttgcagaa gcatccagaa ctaataactgc gtaagatgctg cttatttctatcataactgc 5520
aagggctttgcagaa ctaataactgc gtaagatgctg cttatttctatcataactgc gtaagatgctg cttatttctatcataactgc 5580
aatcttgggag cagagactcct gcctctcctag tgcacaagctc taaataaataa aatgttatagttt 5640
agatgacgagg tttttgcagaa gcatccagaa ctaataactgc gtaagatgctg cttatttctatcataactgc gtaagatgctg cttatttctatcataactgc 5700
agatgacgagg tttttgcagaa gcatccagaa ctaataactgc gtaagatgctg cttatttctatcataactgc gtaagatgctg cttatttctatcataactgc 5760
aaggtcaggag cagattggag ggaattgttg ccacccatcc catcaccatt aatctgagtt 5820
attctgaggt gtcaggtcct ccctccaggg cacagctttcc gcctctcctag tgcacaagctc taaataaataa aatgttatagttt 5880
aatcttgggag cagagactcct gcctctcctag tgcacaagctc taaataaataa aatgttatagttt 5940
attcttgggag cagagactcct gcctctcctag tgcacaagctc taaataaataa aatgttatagttt 6000
aatcttgggag cagagactcct gcctctcctag tgcacaagctc taaataaataa aatgttatagttt 6060
776451Seq
cgtttctgct cccccagctc aataaaagag cccacaacc ctcactcggg gcgcaggctcc 6120
tccgattgac tgaagtcgccc gggtacccgt gtatccaaata aaccccctttg cagttgcactc 6180
cgaaccttgag tctcgctgttt ccttggaggg gtctcctctcg agtgattgac tacccgctcag 6240
cgagggcttc tcat t 6255

<210> 14
<211> 43
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 14
cctttgaaaaa cacgatgata atatggcccc tttgtctctct ctg 43

<210> 15
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 15
ttcgcgagct cgagatctag atatccccatg 30

<210> 16
<211> 35
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 16
cctaacggtgct ccagcgcagc atccagctgaa cccag 35

<210> 17
<211> 34
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 17
caggagaata gatctctcaac actctccccct gtgg 34

<210> 18
<211> 51
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 18
cagtgtgatc tcggagatcc aggacctcag caggggatgg agctgtatca t 51
Artificial Sequence

Synthetic

DNA

19

aggctgtatt ggtggattcg tct

20

gttaacagat ctaggcctcc taggtcgaca t

21

cgatgtcgcac ctaggaggcc tagatctgttt aactcgaga

22

cgaggctctg cacaaccact acacgcagaa gagcctctcc c tgtctcccc g gaaatgaaa
gccg

23

24

776451Seq
DNA
Artificial Sequence
Synthetic

aaagcatag ttctgggcct tgtacatgg ctggattggt t
41

DNA
Artificial Sequence
Synthetic

1gtaatccgg gccccccaaga acctgaatg gaagcatcag tcagttctcat atat
54

DNA
Artificial Sequence
Synthetic

ctcaggtgt ctacgctgac atccagctga cccag
35

DNA
Artificial Sequence
Synthetic

ctgcagaata gatctctcaac acttccctct gtgg
34

DNA
Artificial Sequence
Synthetic

cagtgtgatc tgcagaatcc aggacctcag catgggatgg agctgtatca t
51

DNA
Artificial Sequence
Synthetic

1gtgtcttccgg gtctcaggct gt
22
<210>  30
<211>  41
<212>  DNA
<213>  Artificial Sequence

<220>  Synthetic

<400>  30
agcttctcga gttaacagat ctaggcctcc tagg tcgaca t  41

<210>  31
<211>  39
<212>  DNA
<213>  Artificial Sequence

<220>  Synthetic

<400>  31
cgatg tc gagcc gac ctaggc gcca tagat cgt gg aactcgaga  39

gccg

<210>  32
<211>  64
<212>  DNA
<213>  Artificial Sequence

<220>  Synthetic

<400>  32
cgaggctctcg caca acaccact acac gcgaa gac gccctctcc ctg tct cccg gaaat gaaa  60
gccg  64

<210>  33
<211>  72
<212>  DNA
<213>  Artificial Sequence

<220>  Synthetic

<400>  33
aattcggtctt tcattt cccg ggacac ggg agag gcctctt ctgcgt tag tgg tgt gc  60
gagcctcgtg ca  72

<210>  34
<211>  9511
<212>  DNA
<213>  Artificial Sequence

<220>  Synthetic

<400>  34
gaatta at ccagt c a accga aact gtcct ccaaa tgtgt cccc tcacact cccc  60
gaattc gcgg gctct gtgcct ctagac ccac tcac tcat tccccac ac caggcc gagc  120
aagcgc gcgg cccct ccgtt ttcttgcttt tga aag acc cc cac cgt agg tgc aagcta  180
Page 33
gcttaagtaa cgccaactttg caagggcatgg aaaaatacat aactgaaat agaaaaattc 240
agatcaagggt caggaacaaa gaaacagctg aataaccaac aggatatctg tggtaagcgg 300
ttcctgcccc ggctcagggc caagaacaga tgaagacagt gagtggatggg ccaacagga 360
tatctgttgt aagcagttcc tgcccccggct cggggcccaag aacgatggtt cccccagatgc 420
gttcaccagcc tccagctttt ctatgtaaat atcagatgtg tccaggttgc ccacaggacc 480
tgaatagcc cctgtaccttt atttgaacta accatactgt tgcctctctg cttctgttctg 540
cgctccctcg cttcctccagc tcaataaaag agccccaaac cccccacact cgcgcgcagtt 600
cctcctatag actgctgctgc cccggtaacc gtatctccca aaaaaccttt tgcgttttgcc 660
atccgagatcg tggctctgct gttctctggg agggtctctc ctgagtgtatt gactacccac 720
gacgaggggtc tttcattttgg gggctctgccc gggatttggg gaccccctgcc cagggaccac 780
cgcaccaccc ccgggagcta agctggcagc caacctatct gtgtctgtcc gattgtctag 840
tgtctattgt ttgatgtttag gcctctgccg tcgtactagt agctaaactag ctctgtatct 900
ggccgaccccg tgggtggaacct gcagagtctt gcacaccccg ccggacacccg gggacacgc 960
ccagggaccc ttctcttgccc cgcacctggg aagggagctg atgtggatatc 1020
cgacccgctc aggatatctg tggctctgtag gagacgagaa ccctaaacag tccccgcctc 1080
cgctcaatttt tgcgtctccct gtttggacac gagaagcgcgc gtttctgctg ctgacagcgc 1140
gcagcatgct tctgctgtgt tctctgctctga ctgctgctgtt gcattttgct gaaaattagg 1200
gccagactgt tactactccc ttaagttttg ccattgcttc ctggaaagat gtcagcgcca 1260
tcgctcaaa ccagtcggtca gatgtcaaga agagacgatt ggattacccc tgctcctgcag 1320
aatggccac cccttaagctg ggtgagccgc gcagacgcac ctttaaccga gccctcatca 1380
ccagggtaaa gagcaaggttt tttctactct gcggggatgg ccctggctccc 1440
acacagtgac ctggggacgc ttggtcttttg accccccctcc ctgggtcaag cccttttgtac 1500
accctaaagcc ttcgcctcctt ctctctcctaat ccgcagcgc ctctctctctt gaaacctccc 1560
gttcaccac ccgctccgct tcccccaatct cgcctccccc tggcggcagaa 1620
tttcgcgggtcg atcaagagac agagatgagg agctttgtata tcctcttttgg gatcgcctca 1680
gcagctgtgac atctgcagag ccgggcaagcc ccgggttcag cccccctatcc 1740
aggaactaaa ccctgggcaa gctcttttcctt caagaagatccc acacccctcat tgaagagaca 1800
acggctcaaa ttaacagaccc ccacacccctg caagacatc gctggcgcagcg cgacgtcttc 1860	tctagccag ccgccactttt cactctgttc aatgtatatc attttcttgttg gggacctctg 1920
gcgcggagtgc tgcgtctggg gacgctctgg gcgtcgccacgc cggtggaacct gacctgatc 1980
gctcggcgtgg ggataagagac cagggctatc ttgagccccc gcggcgaggt ggcacaggtg 2040
ccttcctgat gctcagctgtt gatccaaagcg atagtaaggg acaagtctgttg acaggcgcagc 2100
gcagttgtgg aacctccagtaag gttatagttg gggagggctg acacacgcttgc 2160
ggccgggag caggaactgc agctgtctag agatctgcag ctccacgcgg ccctctatatga 2220
aaggttgggc ttgggatcgc ttttccggga cgcgcggctgg atgatcccct ccgcgggtgg 2280
ttcctatgtg gatctttcttc cccacccca ca tttttttttt gcatgctata atggctata 2340
ataaagcatc atccatcaca ttttccaaaa taagaagcttt tttttctgtg attctgttag 2400
ttgcttttccc aactatcata catgatccttc tctatcacttc acgtggtctg gcagcgtctg 2460
ccttgaggtctg gacgcacctcg ggagtctctta ccgcggctgc aactcgc accgcgttg 2520
aaccacgac gcctatcgcgc ccatctcgct gccgacgttg caggaggtgg gaggcaacag 2580
ggccccgtttg gtcgaggcgcc atccggcact tagccacatct attctgtatg tattatagct 2640
aatatatg tgcgatgttc ccacctcata ggtggtatcc atatacatatt atgtatcactt 2700
ataatgcttc atgctccaa actccgctac atggacatg atagttatatt atgtatcactt 2760
agaatcaat ctaggggctca ttattttctct gccctatatt gcgattctcg gtcatatacat 2820
atctagcttaa tggccgcccct gcggctggg ctcgctagc ccagcgcctgg aatccgcttg 2880
tcgagctagt tccctatgag gcgcggcttc aaccgtccttg cggcggtttg 2940
atttcgctta aactgcaccc gtcgctacat atcaatgtgtc tcatatggtta gcctgtgccc 3000
catttgacgt caatgacgtt gtagggcccg cctggtcatca tgctgctgtcc atgtacatct 3060
gggaacttccc tattttgggag tctattgct catattggtac cggctattc atgtgatcgc 3120
cttttgcgct ctcctacaag gcagccgtgtt acgggttggc ttcatacgag ctcaggcttg 3180
tcccccccag tgcgctgctg ggcgggtttag ttgctgacaca aacatcagcg gaccttcccc 3240
aatgtcgatg aacttccgac ccacctgacgc aatgggctgg cagggatgtg cgtggtgag 3300
tctataaa aacttttttgt cctattggtc gcctctcgcc cttattttgg caccgctcag 3360
cttcgtctgc cttcgtctgc tattttgatt cctatttatt tatttttttttt 3420
agtataagca tctaggtccg cgcagcagcgt tccagcttgcag gaaacgcggg aacgctgcag 3480
acgcaatatt aatgtgctgtg cctcactttc aggcacccca gcccttttac gccctttttc 3540
cggctctgtat gtttggtgatag ctcaggtgctct gataacagca acagctattc 3600
accatgat cgcctgactg ctcgctgtag tggcgctacag cgcctttttc 3660
tgctggtatg cccatcgagg caccggccgg gctcgttgctg gcaccacatt 3720
ugaaccagc gcgcggcacc gcggagctggt gcgttcaggg aaggaagggg ctggtggcag 3780
tgggcatctg cttgctcttt ctcgctgttg ccacggccgtg tgggaattgt ctcggtcatca 3840
cagccagctag cgtcgtctgct gtcgctggcc caaccacctc tataatcttc 3900
cctgtgctag ccggcctcag cggcgtcttt ctcctgtggt gcaccgcggg ccatatttccc 3960
tggagatatt gcaccttttgc aactcttggt gcaacgttctgc gcaccctatt gatgtgtcgtt 4020
gcgcctgggc atctgttcagg acctgtgtcc gtcgctgtct ccagccgcat gcggcatctt 4080
ctgcgtcttt cagtaccagt ctcggtcagg ggaaggggg ctgctaccaga 4140
tggtgtggaat gttgctcagg ttcctcctct gtctcctgcc ctataatgtg ctcgctttcc 4200
ccacccacca ggaaggcctca aactgtgata cccatgtgag ctcgctgtgac ttccttcaca 4260
accaagccta tgccattgcc tcttcctctc cgttccccctg gtgatcatgg
4320
tcttctgctta cttccaggttc tttcaggagg ccacaaagggc gctccaggaag a tgtgacaatt
4380
tcttgagccgg cttccatcatg tcccagctta gcacagggtga gcagagatgagg ccggagcgggg
4440
atggacctcg cagatctcttc aaggtgcttt gcagagagca caaagccctc caagacgtag
4500
gcatcactcat ggccacttttc acctctctgg gcgtcgccttc cttctcatgctt caactgtgct
4560
atgtgactcca ggataaccttc atcccgtaagg aaagtttcacat ccttcctaatat tggatagcct
4620
atgtcaatttc tgtgttcatat cccccctattc actggcggag cccagatcctc aggattgct
4680
tcccaggagtct tgtgtgccctg gcacagtttct ctttgaaggc ctttgcaatt tcgtactcca
4740
gcaacggeaa cacagggagg cagaggtgatt atcaacggtga acagagaaaa gaaataaaac
4800
tgctcttgtga agacactcaca gcggacggaag acctctgtgg gcctacaagtgt cactgtccta
4860
gcgataacacta gttattcaacaagg ggaggaattactgattacaa aaactcactgcct ctctcgagaa
4920
tcgagggcgg gaccacctat cactacacag tcgcagccccgg ggaactcaagag atgacagatg
4980
acaaataag gccttatctac acctgcggccg ccgcctgcagct atgcactctga ccgccgcctcg
5040
agggccgcaaa gcggcgactcc cccgggaatttc gcctctcccc ccccccctcct cctaaccgta
5100
tctggccgaag ccgctttgaa taaggccggtgt gcgcgttggtct ctatagttat ttttctccaa
5160
tatggccgttc ttttcgcaata gtgaagggccc ggaaacctgg cccctgtcttc ttcgcagcaca
5220
ttctctaggcc ttttccccctct ccgcccaaaag gatgtcaagg tctgttgaaatgcgtggaagg
5280
aagcagtttcct tgtgtaagact ttttcgaagac aacaacagtc tgtaagcaggct cttttccaggg
5340
agcggaaccc cccacccctgc gcacagtgcctg tctgcggccaa aaaaaccagct gtataagaata
5400
ccacctgaaa ggcggcaccaca cccccagctgc acgttgtagt gttgatgattt gttgaaagag
5460
tcaaatggctct ctcttcacagc gtatctcaaca ggggctggaag ggtgcaccag aaggttacccc
5520
atttgatggct atctgcttccgg gcacagcttct cttctctttgg ttttagttgtta ttcgggtgtggaat
5580
taaaaaacctgctcctccc cccgaaccacg gggacgtggtt ccctctccatg aaacaacagtat
5640
gataataggg ctccttttctct ctctctctgtgt gcggtaggca tctctattca ggcaccacgccg
5700
gccgagctcca cccagtctcc agactccctgg gctgtgctctct gcggcgaggag gcctggacccac
5760
aactgcaagt aggccgagagg tttttgttac agcctcaaca ataagactta tttagtttgc
5820
tatcagcaga aaccagcagc gctctctaaag ctctctgattt actggccactc taccgggaga
5880	tccggggtcct cttgaggccgatt cgagttcgagc gcggctttggag cagatcttcc ttcgaccaac
5940
agcgacgtgc aggtggaagat ttctactgtca gcaaatatta tagtactcag
6000
acgttggcggc aagggaccgaag ggtggagatct aacgccaatct tgggctgcaacct atctgtctcc
6060
acctctcgcgt catctgtgcta gcagtttaaat ccctctgattt gcctgctgcttc gttgggcagagtc
6120
aatcctcttc atcccgacagc ggcacagttg cattggagagc tggtgataatct ccctccaatgag
6180
ggtaactccc aggagagttgt cacagacagcg gacagccagag aacgcactct cagcctgcacgc
6240
agcaccctgaa gcgcgtgaacag aacagacactg gagaaacaca aacatctacgc ctggccagagttg
6300
acccatcagg gccctagacgc gcccgtcaca aagagcttca acaaggggag agtgttagtt 6360
cctagaatt aattagggg aggtaatcggc tgtcggacaa gcagttcgctt 6420
tataaccgctg tggactggaat aacccctaggc ttacccaact taatgccttt gcagcacatc 6480
cctctccccgc cagcctcttc tttgctgacatg gataaaatat gaaatctttta ttgaccttca 6540
gaaaaaggg ggaatgaagaa acccaccctg tagggttggc aagtctagtt taaatgaccc 6600
atgttgcgaag gcataagaac taatataact gagataagag aagtttcatga caaggtacgg 6660
aacagctgga acagtctaat gttggccaaa caggatatct gtgtgtaagca gttctctgccc 6720
cggtctcaggg ccaagaacag atggaacagc tgaatatgagg ccacacagga tatcttctgt 6780
aacagcgttcc tgcggcagct cagggccaga acagatgtgt ccccaagtgc gttccagccc 6840
tctacgcttt ctgtagaacc atctagattt ttcaggtctgc cccagaggcc tgtaatgacc 6900
cctgtgcatct gttgttaaatt ccacatcgtt gcttggtgctgc ctctttcgccgc ggcgttctgc 6960
tccccggact ctaaaagaga gccaacacac tgtcctccgag gcgcacggct ctcctcattga 7020
cctgtgaggct ccggatctccct tgatatatcc aaaccccaact gcgttctgttg cccagtcttg 7080
gtctgtcggc tcttggggag ggtctcttcct gatgtaaattg ctaacgccggt gcggggtacct 7140
ttcatttgggg gcggcctgctg gcagttgggag acccctgccc agggacaccac gacccaccag 7200
cgagctgaata gttggtgctgg gcgcgggttt gcggtgtgaac ggcgttaaacc tttgacacat 7260
gcaagcctcc gcagacgtcga cagcgctttgt gtaaaggcag gcgggacgca gacaagcccg 7320
tcagggcgcgc tcagcggtttgt tggcggggttg tgcggcagcga gcctagaccc acgtagctag 7380
cagatgcgcga gttcataactt gccttaactt gcgggactcag agcagattgt actgagagtg 7440
caccatagcc ggggtgaaat acggcagcagg ttcgtaagga gaaaatccag ctcagggcgc 7500
tctctggcatt cctgctcctac cgtactcggc gcgcggtctg cttcggctgc ggcagcggta 7560
tcagctcact ctaaggccgggt aataagggtaa ttcacagagaata caggggataa gcagaggaag 7620
aacatgtgag ccaaaggcga gcaaaaaaggc aggaacgcga taaaaaaccgc cttgctgggc 7680
ccccccccata ggctcggccc cctgcgcggcgc atcagaggtc atcagcaattgt 7740
tggcagaccc gcagagagact ataaagatac cagggcttttc cccctgggaag ctctctccttg 7800
cgctctctgt tttgctcaca ggctgccttcc ggtgcttggcttt cccctgggga 7860
agcttgccgc ctttcatact ccagctcgtgt aggtatctta gttcgggttg aattcgtcttcg 7920
tcaagcttgg gcgtgtgctga cagagatccgc aagctgttcag ctttcagctgt 7980
aacatctgctgt agttctcaccac cccgtggtgg cagcagcttcat gcgcctggtgac gacagccact 8040
ggttaacagga ttcgagatcc gcagttgtga ggcggttgctaa caggttcttctt gaaatggggt 8100
ccctacacag gtcctcagaga aagcagatga ttggtatattc gcggctttttgc gaagccagtt 8160
acetttggag aagaggtagg tagcttcttc ttcggtggcc aaccaccggg ttgtgacggtt 8220
gggtttttttt ttcgagatcc gcagatttcc gcgcacaaga aagagcttca agaatgactt 8280
ttgatctttct cctggggttgc tgtaagctctt tggacacaga cctcaggttta agggatgg 8340
gtcatgagat tatcaaaaag gatcctcacc tagatccttt taaatattaa atgaagtttt 8400
aatcaatct aaagtataata tgaagtaaact tgggctgaca gttaccaatg ctaaatcagt 8460
gagggccact tagtcctcagct gtgtcttatttt ccttcatcctca tagttgcttg acctccccgtc 8520
gtttagataaa cttgcatcagc ggaggctta ccacatgtgcc ccagttcgtgc aattgcatacg 8580
cgagcaccac gttcaccggc tccagatttta tcgcaaatata accgagccag cggagaagggcc 8640
gacgcagaa gttggtccgct gaaaccttatcc gcctcctaccc agtctattaa ttgtgtgccgg 8700
gaagcagtag agaagagtcttc gccagtttaa agtttgccga acggttggtgc cattgctgca 8760
ggcaagcttg ggtcagcctgct gtgggtgtgt atgtgttgtcgaca tccacaggtocag tcccaacga 8820
tccagggcag ttacagtgtc ccccatgttg tgcaaaaaaag cggtgtagctc ctctggctct 8880
cctgcagctg ctcagaagtaa gttggccgca gtgggtatcacc tctgatttat ggcagcactg 8940
cataatttccc ttacgtgtact gcaatccgta agatgtctttt cttgtgacttg tgaagtatac 9000
acccagctca tctgagaata tggatggtgc ggacaggtct gctctgtgcc ggccgtcaaca 9060
cgggataata cccgcgacca tagcagaact ttaaaagtgct tcattcattgg aaacagtctt 9120
tcggggcagaa aacctcctag gatctttacgg ctgggtttagat ccaagccccactt 9180
cgtgacacca actgtctcttc agcacttttt acttttcacca gcgtttttcgg gtagacaaaa 9240
acaggaaggc aaaaatgcccc aaaaaagga ataagggcga cacggaatgg ttgaaatctc 9300
atactctctcc ttttttcaata ttatttgaag ttttatcagg gttattgtct catgacgcga 9360
tacatatgctt aatgtatatgaaaaaataa caaatagggg ttccgccaccc atttccccgga 9420
aaagtgccac ctgacgctcta agaaaccatt attatcatga cattacacta taaaaatagg 9480
cgtatccaga ggcctctttcg tcctcaagaa t 9511

<210> 35
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 35
gatccactag taacgggcccc cagaaatccgc 30

<210> 36
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 36
cagagagaca aagaggccca tatattcctg ctgttttttca aag 43

<210> 37
<211> 7197

Page 38
2005208507     03 Oct 2006

<212> DNA
<213> Artificial Sequence
<220> Synthetic
<400> 37
gttgacattg attattgct agttattaat agtaatcaat taccgggtca ttgttctata  60
gcccatatat ggagtttgcgc gttcataata ttacggtaaa tgtgcccgcct ggtcgaccgc  120
ccaacggccc ccggccattg acgctaataa tgcactgtat tcccatagta acgcacatag  180
ggacctttca ttgacttcga attttaggtta aactgcccac ttggtcagtc  240
atcaagtgtta tcatatgcca agtacgcccc ctattgacgt caatgacggtaatggcccgc  300
cctggcatta tgccagttc acagctttta tctttgacag tacattcag  360
tattagcat cgctattacc atggtgatgc ggtttttggca gtcatcataat gggcgtggat  420
agcgggtttga ctccacggga ttccccagtc ttcccccaat tcagttcataa gggagtttgc  480
ttttgaccca aatacgaagc gacctttcaaat aatgtgcttaa cacaacttcgg ccctgacacg  540
aaatgacgagc tagcagctgta cgggtggagg ccaccctataa cacaacttcg cagaagtccaa taaaaagccc  600
cacaaccccct acctgagcgc gcccagctttc cgataagctt agctgccccg ggtacgcttat  660
tcccaatataa gcccttctttc ttgctttgatc gaatcatgttc ctggcttgctc tttggagagc  720
tctcctcttgta gttgttgcact accacgcgc ggcctggctccttt attgggtggc tctctggcgc  780
tttggacacc ccggccagcg gagccacccg cccacaccgc gaggtcagagc ggccacgcaac  840
ttatctgtgt ctttcgagtt cttgctgttc tatcttggct gaattgcgcc tcgcgtcgtta  900
cattgtaattc aactagtttc gtatcctgctt gacccgcttt ggaacttacgtt aagttctgacac  960
accgggccc gcacccggttg gacgtccggc agcggtctttt gggcgggagtg tgggccccac 1020
tttacgaagtg ggctgtaattt gccgcgctgc acggcggagtg tgggagtgca ttttgagggag 1080
cgggaacataa cacaacccct gccctcgcttc tgcatacttt cggaggtgatat gggaggtgag 1140
cggcgcgctct gtctgtgctcg cgccgctgttt gatctcgcttt tcggctgtttcg tggctgtcgt 1200
gttttctgtat gttgtgcttaa aataggccca gactttctat cttgctttct tgggagtttgag 1260
aggtctttgg aagatgtcag aggctggatgc tcacacccag tctggtgagtc tcaagaagcag 1320
agcttggggt aacctctctct gttgcagatgt gcacactccct aacctctttt ggcgcgtgaga 1380
cgggcaccttc aacccgcaggg tcaccccaag gtttaagactc aaggtctttt tcacccgtcgg 1440
gcatggaccc caagaggttg cttccctcag ctggacgtgg gagaacctgg ctgggtgaccc 1500
cggcctggtctg ccacgtccgc ccggcccctt ctccttggtg ctccacccgc ccgcgtcctctc 1560
cctctgcttc ccccttgcaacc ttctctcttc gcctgccggtcg gatcccttctc tttatacttc 1620
cctcactctct tctctagggcg cccgctaaccc gacgcgtccttc tttttcaagcagcagtctgta 1680
agatggaaga caacagtctct ctctgtggtgt ttgtgctgtcctg gggcttcagc 1740
tgaaagccca agagagatgaa gaagttgcttg tcgtgacacca caaatgtaatg tgtgcgccgga 1800
ttacccag gatcactccgct tctccgcaag atccctaagtg gcagattgtg gagagagacaag 1860
tccgaattat tgttcctctg aacaacagg agaatatcct tcaccccacc tcaccattga 1920
gaaccagatt ttgtgacctat tgtctgacc cctgtcaaaa atgtgtacctt acagaagcgg 1980
agcttgataaa tcagatatgt actgctacc cagagcaatct ctgtgatgaa gacagtgcta 2040
cagagacctg ctacaccttac agcagaaaca agctgttacct acgcttggttc ccacctctat 2100
atggttggta gacccaaatgg tgtgaaacac ccttaacccc agatgccctgc tatctctgac 2160
aagatcccta tggctattgg ccagttctaa tactatgtat tggccccatt ccatatacga 2220
ttccatatat gggtttttctct attgacgttag atagccccctc ccaatgggctg gtttctatata 2280
ccatatatggt ggcttttctaa taccgccccat agcgcacctcc cacattgacgt caatzgcttc 2340	
tatatggttt cttccctata gacgctatat gggccggtctc attgacgtat agtgcccttc 2400
ccccattgac gcctaatcag gtaatggtgc ccgtctgctc ctattgcccatt gcacccctata 2460
ggccacccca gcctagacgt caatgggatag gctctattgc cattctatac gttttctacg 2520
ccccctattag aagtctcatag ccggttgaag ctccccctgc agttggtcag tgtgctctcc 2580
tagtaaattcg gcaagtctat tacatttgga agtacgcacag ggtctattgg cagtaacccc 2640
attgacgtcatc atggcggtaaa atggcccgccg atggtgctcaag agtcctaccc tattgacgtc 2700
aattggggagg ggcaatgacg caaatgggcc tccctattgc gtaaatggtgc ggtagggcttg 2760
ccaatggttaa ggctctattat cagcagtccttt gtttgggagg cacccctcttt gctctggggac 2820
gtccggaggg ctcggaggtct tagggccctca ccatgggattg gagctggttact aatcctcttc 2880
tggtgctcacc agctcagaggt ggccctctccg aagtccacatg ggtggagaggc ggtggaggttg 2940
tttgctgacctag tgtgctcccttg ggtgctttccg aagttgagac ggtgctgggttc cctgggagtgc 3000
catattggtat ggttggggtg agcagagccac tctgggaaagc tcttgagttg attgagaaaa 3060
tctccggagcac tagcagtcacg attactagtg ggcgctctctt aaaggtatag tttttatatg 3120
cggcagacca gcggacaaac ccctgtggcc tcaaatggtg cagccttagag cccgaaagc 3180
ccggggctta ttttttgtca agcccttttact tgggttctccg ctgtttttgtcattttgggccc 3240
aagggccgcc ggctcagcctct tttccccaggta gttgctctggc cccacccccc tttccctccc 3300
tctctttttgtgaa atttccccctgc tggccagcta gcagctggtgg cgtggttctgc cttgcagagg 3360
actttcctttc cactttccctac attttgtgtct ggaataacca gaaacactct gacatcagca 3420
gtacccgggg cttccctacat gttcctgagag gggccagacta ccagctcacc ctacaggtgc 3480
tgtgctccct taaggagcttc atgcagggca cagagcaaca cgttggttctg aaagttccagc 3540
acggcaacgg caaacaaga aagaaagctcg cttccctcagt gattgcctag ctggctctccaa 3600
aagttgacgtgcttcgctcca cccgcgcggag cttctctcttg ccaacccccg cagtcccaagc 3660	
taccttgccca gcggacagggg ttcagctcccc gcagattacta ggtggtcctgtgctggccgag 3720
gggaaacagtg gggtttctgctg gttccacacgg accaggtgca ggctgaggcc aaagatgtcg 3780
ggccccacagctacaaggttag acacagcacg taccataaca agagagcgac tgtgtctgccc 3840
agagcatggtt cacctgcccgc gttgattcaca ggggcctgTac cttccagccag atgtcgtctc 3900
ccatgtgtgt ccccgatcaa gacacagcca tccgggtcttt cgcccccccc ccactctttg 3960
ccagcatcctt cttccaccaag tccacaaagt tgacctgctct ggtccagcaac ctgacacacct 4020
atgacagcgt gaccatctcc tgggccgccc agaatggcga agctgtgaaa accacacaca 4080
acatctccga gaggccaccc aatgcacatc tcagctggcct ggtgagaggcc agcatctgcg 4140
aggatagctg gaatccgggg gagaggttca cgtgcacccgt gaccaccacca gacctgcctt 4200
cgccactgaa gcagacactc tccgggcccc aaggggtggc cctgccaggg cccgatgcct 4260
acttgctgcc accagccggg gagcaggtga acctgcggga gttggccacc atcagctgcc 4320
tgggtacggg cttctctccc gcggaagctct tcggtcagtg gatgcaaggg ggccagccct 4380
ttgtccccgg gaaagtattg accacgcccc caatgcctga gccccccggc ccaaggccgt 4440
acttcggcca cagatcttgg accgttgtccg aagaggaatg gaacaggggg gagacctaca 4500
ccggcctggtgc ccatgagggc ctgccaacaac ggggtcaccg gaggccctgg gacagactca 4560
ccggtaacc caccctctgac aacgtgctgct tgtgcctgac tgtcacaagc gccacacttcgct 4620
actgagactct tgcagaaaac cgggtctagg ttttgggaatt cgcctctctc cctcccccct 4680
cctacgctta cttgcsccag cggctttgaa taaggtgggt tgtgctggttg ctatattgta 4740
ttttccacca tattgcgttc tttggtcaat tgtgagggcc ggaaacttg gctctgtcttc 4800
ttgacagaca ttctctagggg ttctttccccct tctgcccmaag gatgcaaggg tgttggatt 4860
gtcgctgaag aggcaagttcc tgtggaacct tgtctgaagac aaacaagctc tgtcagcacc 4920
cctgcagcgc ccacacccag cggcctttgc tagcaggccag tctgctggcc caagccacgt 4980
gtataagata caccctgcaaag gcgggcacaa cccagttgctg cctgctgtgag tttgatatgt 5040
gtggaaagag tcaaatgtcct cttcctccaggc tatttcaaca aggggtcggga ggtgcccag 5100
aaggtacccct atgtgttgag ctctgctatg gcggctcgggt gacatgctt tacatgttgt 5160
tagtcgaggt taaaaaacgg ttgtccccat ggcaccacag gggcctgtgt tttcttttta 5220
aaaaacagat gataatagg ccctcttttg ctctctctgt cttggtcgac cctctattcca 5280
tggccacagg gcgcagacct gcgacttgg aagcccaacc gcctgtagcc ccagcctgggg 5340
tgacagagtgc accatcactct tgaagggcag tcaagctgtg ggtacctttgt tagctctgta 5400
ccacgcagaag ccaaggtgag cttccaaagct gctgtattact tgtgaatcactc cccggcacaac 5460
tggtgtgcca acacggatttc ggcggcgctg tagcagcttaa gcctcaccct tctaccatcg 5520
cacccctcag ccagccactc tggcccaatct ctaactgcccag aatagctagc tctatcggc 5580
gttcccggcag gggacagagg tggaaataca acgaaacttgt gtcgacacat cttgtctctcat 5640
cctcccccca tctgtatgag agttgaaatcc tggacactcag ttcggtttgt gctgtactgaa 5700
taaccttcat ctccagggag ccaaatctaca tgtgaagctgt gataacgcac tccaatctggg 5760	taacctccag gagaggtctca cagagcagag cagacacctca gcctcagcag 5820
cacccctgagct cgacagaaag cagactacga gaaacacaaa gtctagcctgg cgaagcgtc 5880
ccacgagggc ctgagctgcgg cctgctacaaaa gagcttcaac agggagagtt gttgatcttg 5940
776451Seq
ttaacccctag gtcgagttc gacatcgata atcaacctct ggattacaaa aattgtgaaa 6000
gattgacctgg tattcttaac tatgtgcttc cttttaagct atgtggtata ctgcgttttaa 6060
tgctttgtaa tctagctatt gcttccccga tggcttcttc tttctcctcc ttgtataaat 6120
cctgggttgtg gtttccttttt attgaagttgt ggcccggttgt cagggcaaggt ggcgttggtg 6180
gacatgtgtt tgtgcagcga aaccccccaagt gttggttgatt tgcacaccct gttgagcctcc 6240
tttccgaggac ttctgcttctt ccccccccta tgtcagctgg gcgggtcctgag gcgggtcctgcc 6300
ttgcccgcctgg ccgctgttgg gggtgtgctg tgtggcagcag ggaggtcctt ccacggtggga 6360
ggaaatcatc gctcttccttc ttggtgctgc gcttggttgc cactctgattct tcagctggga 6420
cgtctttcctg ctaagctcccct cccgcctcctc aatccagggc ccttctccct ccgcgctgc 6480
tggcgcctgct ggcggctttct cccgcctgatgcc gcggctgctcc tcagacgagt ggagtttccc 6540
tttgggccccgc ctccccccagtc gtagaaaaattt tgggtcctcc agaagaaagg cccccccagcg 6600
ggagttgaga aacccccctgcctt gtaaggggtg ccaggtctagt gaaaggttgt cttttttgaa 6660
ggcatggaga aataaatatat gagaagattag ggagttcagc ctacagctgagagag 6720
aacacgctgaa tatggtggaa caagagcatct ccgctgctctg gatgcgcctccc 6780
gccaagaca gatgggacgt ctaatgttgg gcagcaacag gaattctggtt aagagcggtt 6840
cctgcccgcct cagggccataa gcagagaggg ctgctcgagct gctcttccaggg cctctttcgtt 6900
tctggggagact ctcagttttc aatccagagtca cttgtggggt cccagctgctct 6960
attgtggact caacatagt ctgcgctttctg ccgctgcttcg cgcgctgcttc ctccccggccc 7020
tcaataaaag gcggagcaac cctctcactgc cgggccccatt cccctcattgc aacgcaagttgcg 7080
ccgggtaccc gttcataccaa taaacccctct tgcagttgga tgcctcgcctg 7140
tctcctggga ggttgctctctc tgaagttgattactccggctgct cttcattttgggctgag 7197

<210> 38
<211> 7516
<212> DNA
<213> Artificial Sequence

<220> Synthetic

<400> 38
gttgacattt attattgacct agttatagaa agtaatatc taccgggtcca tgtgtctata 60
gcccatatatt gagaatttcc gttacataac ttacggtttaa tggccggcctt ggctgaccgc 120
ccaacgcacc cccgccctatt gcgtcaataa tgcagttgtatt tctccatagta aacgcaatag 180
ggacttttcca ttgacgttaa tgggtggagt atttacggtta aactgccccac tgtgcagttac 240
atcaagtgtat ctatagcca agtagcccccct catattgagct caatgacggtt aaatgccccg 300
cctggcatta tcacccagcttc tgtaccttttt ggtactttccc tctttgaggct tacattctag 360
ttatttctcagc tctatttcttgtgtctgtgc gttttttgtt gctacataattg gggccgtggagttct 420
aggggttttga ttcagcccgg tgtccacttcg tccaccccatt tgcagttcaat tggagttgt 480

Page 42
tttggcacc aaatcaacgg gaccttcacaa aatgtcgtaa caactccgcc ccattgacgc 540
aaatggcggc taggcagtta cggggtgagag tgtatataag cagagctcaa taataagagcc 600
cacaacccct cactccggcgc ggccagtcttc cgtcagactg cgctgccccc gtacccggt at 660
tcccaataaa gctccctgctg gttgctgatcc gaatctgtgt ctgcgtgcttc cttggaggg 720
tctctcctga tgtgttgact acccagcagc ggggtctttcc attttgtgag tcgtccggga 780
tttggagacc cttgcgccagg gaccaccgc acaccacccgg gaggtaagct ggccagcaac 840
ttatctgtgt cttgcagatt gtctagtgtc tgtttttgtg tgtatgcgcc tgtctctgtgta 900
ttagttagct aactagctct gtatctggcg gacccgtggtg gaaactgacg agttctgga 960
acccgcggcc aacccctgga gacctcccagg ggcctttgag ggcagtttct gtggccggac 1020
tctgagaggg cagttcgatgt agaatccggag cccgcagagc ctttgtgtcc cttggagaga 1080
cggaaccata aacacccttc cgcctctggcc tcgactttgg cttttgcgt gaaaccggaga 1140
cggcgcgtct tgtctctgctg acgcgtctgc cactgcttcct tgtgtgcttc tgtctgtctg 1200
gtttgttat tgtctctgaa attagggcgc gaccttgctacc acctcccttaa gtttctgctt 1260
aggctcactgg aaagatgctg agccggtcgac tcacacaccg tcggtagatg tcacaagagag 1320
acgttggttt acctttctctgt ctcgagatag gcacaaccttt aacgctggag gcggcggaga 1380
cggccaccttt aacccgaggc tctcaacccca ggttaagctc aagttcttttc cactctgggc 1440
gcattgcaac ccagacccgg cccctctcat cgtgagcttgg gaaagcctgg cttttgcacc 1500
cccctctcgg gtcagccccc tttgactaccc taagctgctcc cctctctccct ctcacatccgc 1560
ccggtctcttt cccccttggac ccctcttgctgt gacccggggc gatcctcccc tttatccagc 1620
tccactccct ttcttaggc gggaattcgc gatctgtatca agacacagga tgaggatctg 1680
tcgcatgtg tgaacaagat ggtattcagcg caggtcttcc gcggcgttgg gttggagaggg 1740
tattcggcta tgactgggca caacagacaag cccggtgcctc tgggtgctgct gcggggcgcgc 1800
tgctcagcgc gggcgccccc gtcttcttttg ctaagacccga cctgtcgcgg gcctgtaatg 1860
aatcgcagga cgaggccacgc ggcctatgcgt ggctggccagc gacggcgggt ctttgccgac 1920
cgctgtctcg cggctgtctcgt gaaacgggaa gggagtgggt gctattttgg gaaagtgggcc 1980
ggcaggactct cctgcgtatct cacctgtgctg ctgacgacgg gccgagatcg atcggtgtag 2040
caatcgcggcg gctgctagccgt ccctctccgg ctcacatcgc attgcagcac caagccgagg 2100
atcgacactg cgacgacgcg acctgcgtgctt aacgcgcttgcttg gattgcctgag 2160
acgaagagca tccaggggtct gcgcgccaggg aactgttgcgc cagctgcagcg gcggcgatgc 2220
ccgacgggga gatcttctgc tgtctccatag gctgactctgg cttgcgcgaat atcatctgtgg 2280
aaaatggcgg cttctcttgg gcatactgct ggtggtgccttc ggtgctgggc gacgcagctc 2340
agagcatacg gttggtctacc cggtgatattt ctcgaaagtt gtcggccggag tcgggtgctgg 2400
gttctctcgt gttttaaggct atgcgcgctc ccgattgcgc ggcctacgcc ttctatgccc 2460
tttgtgcagca tttactccag gcgtactcag caggtgccat cagctatgctt actatgtatt 2520
776451Seq

ggccctatgc catatagtat tccatatatg ggttcttctta ttgacgtgta gtaggcccctcc 2580
caatggcgcc ttccatatac catatggtgg gccctctaat accgccccata gccaactccc 2640
catggtggtc gatgtttcttc atatatggttc tttcccatttg acgtcatactg ggccgtccata 2700
ttgacgttag tggccgctctcc ccatgtggac ctacattcag ttaattgccc ggcttgctctaa 2760
atgcccatgg aagtcaatag accaccaccac cattgactgc aatgggatgg ctcaattgccc 2820
attcatatcc gtccctcaagc cccctatttg gctcaatgc ggttaatggc ccacttgcca 2880
gtacatcaat atctatataa agtaacctgg caagtacatt actattggaa gtacgccagg 2940
gtacatgccc agtacctccca ttgacgtcctaa tggccgcttaa tggccgctgca tgggtgctaa 3000
gtacatccctt attgacgtca atggggaggg gcaattgcagc caatggtgcct tccattgacg 3060
taaatggccg gtaggccgtgc ctaatggtag gtccatatataa gcaatgctcg tttaggggaac 3120
cgccattctct cttgagccag tcggagccag tcggagccagc gcagacccct agatgggtgg 3180
agctgtctca ccttccttcct gtgacgctttc gctgcaagc tctagacgttg ttcacgctcgcag 3240
agctgtctca ccttccttcct gtgacgctttc gctgcaagc tctagacgttg ttcacgctcgcag 3300
ttcggtcttcg attttcaccac atattgggag agtgggtggag gacaggccacc ttggaaaggt 3360
cttgagttcct tttgagcttt gtagagcaat tcattccgat acagctgatcg ttaactatgc gcggctcttct 3420
aaggtataag tgttcatatct gcgtgaacac gcgtgaacac gcggctgccttacttgca gctggtctctc 3480
agctgtgtga ccggggtccag cggggtctctt ttttggccttt gcacgtctctt gggggttctcc 3540
tggctgttctt attggggcct cgggctctct ccctcctcttt tcggagccag tcggagccagc 3600
ccaccctttt tccctctttc cttcctggag catttctgtgct cgggagggcg cgggagggcg 3660
gttggtctgc tgcacgagtt ctccttttcc gactctcatc caatcactgtgg gaaatccagat 3720
aaacacatctg acatacgagc tacccggttgc ttcctctcag cttcgtgaag gcggacagtc 3780
gcagccacct cacacggtctc cgcgtctttct cggagctctg cggcggcttg tgggctgtggag 3840
gttggtgtca aagttcaccag ccccaagggc aacaacagaa aagacgtgact ctctctctctt 3900
attggccgagc tcgctctttctt ctctctctctt cggctttcctt gccggtctgtg gacggagtctt 3960
aaccgctctca actgccgactct acctctcttt ctactctctctt cccggtctgtg gacggagtctt 4020
gtgggtcttcgc gtcggttcctg tgggtctgcc gtaaccagtcg ggcggcagtc gcggcttgccg 4080
gcgagggcct cagcaggtcc gcccagcctt gacgtgtggtt gaatgctcgc aggagcttccgt 4140
agagcgtact gcgcggtcttc cgcgtctctctt cccggtctgtg gacggagtctt 4200
tccacgagat atcgtccttc gcgtctctctt cccggtctgtg gacggagtctt 4260
gccatccttc cattcctcctc ctcgctcggag cccagacact aagccagcttc ggcgtctgtg 4320
gtacagcacttg accagctcgtc ccatcctcctc ctcgctcggag cccagacact aagccagcttc 4380
gctgtgaaa cccacacacac ccaccttccttc ccctctctctt cccggtctgtg gacggagtctt 4440
gtggtactgtg ggtcatcatctg aatccggtgg aggagtccctt cgcgctgctctt 4500
aacccagcgagctggttcctc gcggcgtgtg cagccactctt cccggtctgtg gacggagtctt 4560
2005208507     03 Oct 2006

cagccagccc ccgatgctca ctgggtgcct cccgccccggg agcagctgaa cctgcggggag 4620
tcgcccaac ttacgtgcct tggagcacggc tttctctcccg cggacgtctt cgtgcaagtgg 4680
atgcagaggg ggcaagccctt gttccccggag aagtagtgtga ccagcgcccc cactgcctgag 4740
ccccagcccc caggccggtta ctctgcccac agcatcctga cctgtgcgca cagaggaattg 4800
aacacgccccg aggctcacaac ctgctgtgcc catgagccgc tgcctcaacag gtctacgcag 4860
aggacagggt acaagtcacgc sggtaaaccc accctgtaca acgtgtcccc ggtcatgtcc 4920
gacagcagtgc cagcctgtctg ctgagatctcc tcgagaaacc ggtcctaggt ttgggaatcc 4980
gccccctctcc cccccccccc cttaagcttc gggcggaaag cgtctgaaat aagcgcaggtg 5040
tggtttggtct tatatttttat tttttcacatt attggcgtctt tttggcaatg tggagcccg 5100
gaaacctggcc cctctgtctct gtagcagcat ttcttaggggt cttcccccct tcgccaacag 5160
aatgcaaggt ctgtgtgaattc ctgtagaagga acagccttct ctggaagctt cttgaagaca 5220
aacacgtcct gtgcagcagcc cccggcagcccc ccacctgagcg caggtgctct 5280
tcgcgcccag aacgcacggt tagataacatg acctgcaagag ggccgacaaa cccaggctc 5340
cgttgtagtg tggatagttg tggaagaggt caaatggtct ctctctgacgc tattcaaca 5400
ggggctgaaag gatgccccga agtgacccca ttgatgagga tctgactggcg ggcctcgggtg 5460
 cacagcttcct cctagctgctt cagtcagagtt aaaaaacgt ctggccccc ggacccacgg 5520
 ggcagtcgtgct ttctcttgaa aaacagcagt atataagtcg ctcctcttcttc ttctctgctcc 5580
tggtaggctcat ccatctccat gccaacccag cggacatcga gttgacccag agcccaacga 5640
gcctgcacgc cagcctgggtg gacagagtgca cctatcactc taaagccagtt caggtgtagg 5700
 gtaactctgtg acgctggtac cagcaggaag ccagtagtgct cagcagctt cagctgctct 5760
 ggacatccag cccgccccaggg ggtgtgcgca gcagattcag cggtagcgggt acgccgctcg 5820
acttcacctt cccatcagcg agcctccagc cagagcagcat cccacccctc tattgccccg 5880
aatataagct cttatcggctgc ttcgagcagcag ggaacacgat ggaatccaa cgaactggctg 5940
ctgacacatc tgtctcttcac ttcctccgctt cttggagcac gtagtatggta ggaatctctt 6000
catttgctgt cttgctgataaatctttactt ccaaggacgc caaagtacag tggaggttgtg 6060
atatgcaccc cctccgaggg aacgccacag agaagtgctc aagcgagcag cagcagcagc 6120
gcacctacag cccagcaggc acctgcagcg tgaagcaacg agactacgag aacacaaag 6180
tctagcctgc cgaagacggt ccagaggtccgc tgcgtgcgaa gcgctacaaag gccttcacas 6240
ggggagagtgg ttgagctgtg ttacctccagg cttgcaggtcg acaatacgtgaa tcacatcctctg 6300
 gattacaaaa ttttgtgaag aatgactggtt attttaact aaggtgctcc ttttacgctaa 6360
 tttggatatc cttctttgtat cttgctttgta ctccgctgtt gctcctttctct 6420
 ttctctctctct tatttatatg cttgatgtgtg cttcttttattag ggaggtgctg gcccgggtgctc 6480
agggcaagtct gcgtgggtgctca cactgtggtct gttggctggct cttgggcccatt 6540
gccacaccc tgcagcctcct ttcgggagactttgccttcc cccctcttat gttggcagcgc 6600
776451seq

gaaacctcatcg ccgcctgcct tgcccgcgtgc tggacagggg gccgccttgtt ggacactgac 6660
aattccgtgg tgtgtgctgga aaatcatcag tcctttctcttg ggtctgtgcctg ccctgttgcc 6720
acctggatcc tcgcggcggag gcctctctgc agctgtccctt cgcgccctaa tcagcggagac 6780
cctccctcccc gcggcctgtc gcggcctgtc gcggccttcttc cgtgccttccc ccctgccttc 6840
cagcagagtgc ggtacctccttt tcggccgccct ccctcgcacat ctgaataataa aagattttatt 6900
ttagtctccta gaaaagaggg ggaattaaag acaccaactcg tagggtttggc aagctagcttt 6960
aagtaacgcct atttttgcaag gcattgaaaa atacataaact gagaatagag aagttcagat 7020
caggttcagg aacagatgga acagctgtaat atgggcacaa aacagttcatt gtggtaagca 7080
gttcctgcct ccggctcaagg ccaagacgca agttgaacagc gtaatataagg ccacacagga 7140	
tatctgtgtt aacagtccttc tgccccggcgt cgggtccaaag aacagatggtg ccccgatgct 7200
gttcagcgct tccagcgattc tctgagaccc atcagagttt tccaggtgtgc ccaagagacc 7260
tgaatatgcct cttcgcctct ccgctactcta cattccagctt ggctcctcgg ttcctgtttgc 7320
gccgctctcgg cttccgagct caataaaaga gccccaaacc cttccacttgg gcgcgcggct 7380	
cctccgattga tcggctcgcc ggggtaccgg tcgttcacat aaacccctctt gcagttgcct 7440
cgcagttggt gtctgctgtgt tccttgggag ggtctcctctt gcagtgattgg cttcaggctc 7500
ggcaaggggtc ttcatt 7516

<210>  39
<211>  DNA
<212>  RNA
<213>  Artificial Sequence

<220>  Synthetic

<400>  39
gttgacattg attattgact agttattaat agtaatcaat taccgggttca ttacttcata 60
gcccatatat ggatcttcgc gcctcaatac ttacggttaa tggccccgct gccgtgaccgc 120
ccaaagaccc ccgcctttttg agctcaaaaa ttgagcgtatg ttcccatagta acgcgaacag 180
gaccaccttc ttgagctcca tgggtggagt attaccggt aacgtccccc cttgctgacg 240
atacagttta ttacatgcac agttaccccc cttattgagtt caatgaccggt aatggccgg 300
cctggctatta tgcgcaggtact tgaatttttt ggaaccttcttt ctctacatct gcagcgagtt 360
tattttctgt ctgctattacc atggtgtagtc gtttttggca ctcgtatcaat ggacggttga 420
agcggtttttag cttccgaggct ttcttcaagtt cctcccccttt tgcgctgatcg ggagtgggt 480
 ttggccacca aatacaccgg acgtcttcca aattgtctgttta caacctccgccc cacctgcccg 540
 aatggtgggg cgggtctgatta cctctgtgttgg ctatataag cagagctcag aaaaagagccc 600
cacaacccct taccgcggcc gcagccttct tctatagactc gcgtgcgcgg gttacccgtat 660
tcccaataaa cccctttttc tgttgcctcctaatcgctgcttt cttgggaggg 720
 tttccctctga tttagttgact accacacgcgcc ggggtcttttc atttgggggct tctgccggga 780
2005208507     03 Oct 2006

776451Seq

tttggagacc cctgcccagg gaccaccgc ac caccaccgg gaggtaagct ggcacagcaac  840
ttatctgtgt ctgctcggatt gtcagctgta tattttttgt gttatgcgcc tgctctgtta  900
ttagttagct aacctacttt ctatctcggag gaccgcgtggt ggaacagtac agttctgaac  960
accggccgc aacccctggga gaggtctccag ggcttctggtt gccgcttttt gttggccgcac 1020
tgagaggaag agatcaggt ggaacctgc acctgcacag gaccacggaag tattgtgttc tggtaagaga 1080
cgagaacctca aaacagtccc gcgcctctgtc tgaatgttttg ctctgtgttt gcacggagaag 1140
ccgcgcgtct tgtctgtcgc agcgcgtgcag cattgctctct tgtgctctct gtcgactgt 1200
gttcttgat ttgtgtaaa attagggcga gactgttacc anctctcttaa gtttgacctt 1260
agttcctcgg aagatgctg acgcggctgc ttccaacacc tggtagatg tcaagaagag 1320
acgttgtggtt acctttctgc tgtgcagatg gcccaactttt aacgctggat ggcgcgcaga 1380
cgccaccttt aacccgagacc tctacaccca gggtaagatc aaggtctttt cacctgccc 1440
gcatggacac ccagacccct ctcctcctcat cgtacggctg gaaacagctg ctttttacacc 1500
ccctcctctgg gtatagccct ccttccctctc ctcctcatccct ctcctcatcgc 1560
ccgtccctcc ccccctgaaac ctctctgttc gaccgccgct cgatccctcc tttacccagc 1620
ccctccctct ccctctagcc cgccggactg ccagatccct cccctacagc gagaacacce 1680
ccctccctga aagagcaacg gcctcaacca caccgcatcc cctctctgga gactacacgc 1740
tgcccagcgc agctctctct acgcagcgcgc gcattctcctc tgggtctcaat gtatattcatt 1800
ttactgctggg acctttctgc agatcgtgat gctggtccgac tgctgtgctg gcggctcagt 1860
gcaacctgac ttgtatcgtc ggcgtcggaa atgagacag ggcacatttg agccctcctgc 1920
gacggtgccg aacaggtctct ctgcgtctgcct ctcgtctgctc tggacaaagac 1980
gtgatggaca ggcagacgcc gttttgtgctc tgtaatggct gctcctcttct tattgtgtgg 2040
agggctaaag ctctttgcct cttatggtat ggcgcaggtt caattctatg tattgcaccc 2100
atgtcaacata ctattccccat tattgggggt cattttgtcg tagatagccc cttcacaatgg 2160
ggcgtcctat ataccatata tgggcctttcc taattaccgc ctcggtacac cctccctattg 2220
cgtctcaaggt ctctctattat ggctttttctg attgatgtaa ccttttctgg 2280
tatctgccc cccccctctc agctaatggtag aacgctcggactgctc gtttgcgccct 2340
atgtccgctc attggcctgac ccagttaccg gtttctgtgtg atgtctctctg ggcacacacc 2400
atccgtttcc acgcggccctta tcagctgctact gcacgttaaag tggccacact gcagcctcct 2460
caatatatat taataagata cttgcagtaa cttcattcatt gggatgagc ggcaggtcatt 2520
tggtcgatct cccatgtgct ccaattgggc gcagattctgtg ccaattccat 2580
ccccattgcc gccatggggg agggccaatg acgcgaattgc gcgtctgtcatt gacgtaaatg 2640
ggcgcgaggc gcgcctaatct gcaggtggtg ataagaatct cttgctttgag gcagccgcttc 2700
tgtggctgctg gacgcgtgag gagctcggaa gctttgaagtc aagatagaga accattctgt 2760
tttttctggg gtccctggcc gtttttattaa ggctgcctctg gcacggac gccagcagat 2820
aaggattgtt cttgttgaca acaaatgttaa gttggtcccg gattacttcca ggtcataccg 2880
tttctccgaa gatcctaatg aggacattgt ggagagaaaac atccgaatta ttgttccctc 2940
gaacaacagg gagaatatct ctgatcacc ctgccattg agaacagatg tgtgtacca 3000
tttgtctgac cttgtaaaa aatgtgatcc tacagaagtg gagctggata atcagatagt 3060	tactgtctacc cagagcaata ttgtgatgga agacagtgcct acagagacct gctacaatct 3120
tgacagaaac aagtgtctaca cagctgtggt cccactcgtatatatggtggt agaccaaataat 3180
ggtggaaaca gccttaaccc cagatgcttg ctatcctgac taagatcgtg taaaccgtcc 3240
tagtttaaa ctcgaagttcg acatcgataa tcaacctctg gattacaaaaa ttgtgaaag 3300
attgacgtgtt attcttaacct attggtgcccc ttttacctga tgtggatacg ctgcttttaat 3360
gcttttgatat catgctataa cttccgctaat ggtttctatt ttctctctct tgtataaatc 3420
tctgtgtctg acagtctatg aggagttgtgc ggccgtgctgc aggcacagtg gcgtgggtgtg 3480
cactgtgttt gtaagcgcac ccccactcctt ttcgggccatt gcacccacacct gtcagctcct 3540	ttcgggca tgtccttcccc ccttccccat tagccagccgg gtaacatcgcc ccgcttgccct 3600
tgccccgctgc tgagacggggg ctggctgtgtt gggcaagtcc aatccggtggt gtgttgctggg 3660
gaatcatcg tcccttctct ggctgctgcctg tctgttgtgcc acctggattc tgcgcgggac 3720
gtctcttcctg cagctcctca ccgccctcga cccctctccc ggccccctgc 3780
gccggcctctg ggcgcctttcc ccttccgccc cagacaggtc gatttcctct 3840	ttgggcggcc tccccgcacag gataaaataa aagatttttaa tgtgtcttcc aaaaaaggga 3900
ggaatggaag accccacctg taggttgggc aagctagctt aagtaacgcc atttttgaag 3960
gcattgaaaa atacataact gagaatagag aagttcagat cagggcccag acagatggga 4020
acagcctaat atgggcacac caggatattc ctggtaagca gtctctgccc cggctcaggg 4080
ccaagacag atggacagcg tgaatattggg cccaacaggag tatttctgtgt aagcagttcc 4140
tgccccggct cagggccaga aacagatgtg ccccgatgct gttccagccc tcagcaagttt 4200
tagagaacc atcagatgatt tccaggtgac ccaagagacc tgaaattgacc tctgtgcctta 4260	tttaaatc tccatcgatt gctctctctgc tttgttctgc gcccttgcagct 4320
cataaaaga gcgccacacc cttcctcgg ggccgacgtc ctttgatgga ctgagtgcgccc 4380
cgggtacccg tgtatcctaat aaaccctctct gcagttgcat cccaccttttggtgt cttctgctgt 4440	tcttggggag gtctctctctc gatttgattc ctacccgctca gcgggggtctt tttcatt 4496

<210> 40
<211> 4354
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic
<400> 40
gttgacattg attatggact agttttaaat agtaatcaat tacggggtcga ttagttcata 60

Page 48
gcccatatat ggagtccgcg gttaacataac ttacggtaa tggccgcct gcgtgacgc 120
ccaacgcacc ccgcgcatg acgtcaataa tgacggtatgt tcccatagta acgcacactag 180
ggaccttcca ttgacgtaaa tgggtggaggt atttaacagta aacgcccccc ttggcagttac 240
atcaagtgtat cctatgagca agtaacgcctct tatttgacgt caatgaacggtt 300
ccctggcatttc gtcaccagtc atgcacctattt gggacctcccc tacattgacg 360
tattgtcct cgtattacc atggtgatgc gtttttgggca gtcatcatcaat gggcgtggtatt 420
agccgatttgca ctcacgagga ttctccaagtc tccaccccat tgacgctaattt ggaggttattt 480
ctttgacca aatacagcgc gactttcctaa aatgtcgttaa caacctccgac ccattcgagc 540
aaattggcggtagggattgt aaggttgagaattctataaa cagacgctcaaa taaagagccc 600
cacacacaccc cactcggcggc gccaagctttc cgtgacggtgtt gttaccgtat 660	
tcccaaaaaa gctcttttcct gccgcttttc caacaggtgttt cttggggaag 720
tctctctgtg agttaccagcc aaccagcagcg ggtgcctccttc cttaaggagc 780
cttggacacc cctgccccggt gacccacgcgg cccacacccgct aggataagct gcaccgcaac 840
ttatctgtgt cttgctgattc tatggtctgtt gttatatgctgc tgcgctctgta 900
cattgtagct aactagctct ctatcctggcg gacccggttt gtaactacgag atgtctgcaac 960
acccggcgcac aacccgggtgag gcaccttcggg ggtgggcttctggggggc 1020
tctgagcttc gacgacgtggtg gaaacggagcttgctggttc tgagtgagtaa 1080
cagacaacca aacagcttcc gcgccttcgcc tgaatattttt cttccggttt ggacccgcaacag 1140
cgggcgtctt cttcctgctgct gagctggcagc atcgcctttc tttgttatgcgtgacgtttcgtctgcttgttgagta 1200

gtttcctgtat ttggtcgaata attaggccca gactgttacc aacctccttacaat gtttgacccaaatgctcactgtg 1260
aggtctcactg gaaagttgacgtg agcacgtgct gacccaacagtgcttgcaattgt actacacgac 1320
agtctgggttt acctctctgct ttcgcaagatgt ggcaactccttta aacgcctgggt gcggctgagaag 1380
cggcagctttt aacccgagacc tcatcacaaca gtttaagatc aacgttttctt cacctggcccc 1440
gcatggaccc cagacacgacc tcccctcatact cttgaccttgg gcaggctttcc cttaaggaccc 1500
ccctccctgg tgtaaggccct ttgtaacccct ttaagccctcgg cccttctctct ctctcc matlab 1560
ccctcgctcct ccctcttggaac ctctctgttct gcacgcgacctg cgtattccctt tttatccacgctc 1620
cctctctccct ttctgactgc cggagatccct cggactgagct aagtagttgtagttagtg 1680
ctttgcaaaaa aagtctgtttc cccgctgcctt cttgcggctat gcatgtttga ccattgaact gcctcacacgctc 1740
cgtgctccaca gatattgggaa ttggcaagaa cggagacccct cctggtgcctcgctcaggaa 1800
caggttcagc tactcctaca aacgctctcttg ggcggagcga acacgaaatctctggtttagttgtgtagctac 1860
ggtatttatttt ggttaggaaaactctggtcccctc tatttcgttgag aagaattgcctttaaaggagttggagt 1920
cagaattaact atagttctca gtagagacat ccaagacacc cccagagggg ctcatttttctccttatt 1980
tgccaataagtttggtaggtt cctttagact tgttgaacaa ccggaattggt caagtttaaat 2040
agacatlgttt tggataagttgc gaggcatttc tttttacagggacacgatga atcaacgagg 2100
ccacctcaga ctctttgtga caaggtcatc gcaggaatctt gaagatgaca cgttttttccc  2160
agaaattgat ttggggaaatataaacttcct ccacgaataac ccagggctggcc tctcttgaggt  2220
ccaggagga aaagccatcagataaagtattg caagttctac gagaagaaag actaaccagga  2280
agatgtccttc aagtttcctctt cttctcctccat aaagctatgc ttttttataa gaccatggga  2340
ccttttgcggg ctctttagacta cttatgctatc ggccagcagtt caatactacttg tattttgccc  2400
atgaccata tatactcata tattttttttt cctattgagac tagatacctc gctcccaatgg  2460
gccgtcccat ataccata ttaggtcttcc taataacgcgt catagccactt cccccattga  2520
cgtcatcggtt cttcatatatt gttcttctttct atggagctcata tattgggcggt cctattgcag  2580
tatatgggcc cttccccctatt gacgtcactat acggtaaaatg gccgcctcgg gctcaattcct  2640
atgagctca atagagcacc ccaccatcag taatttggccgcc gcgtaggtgt ccaaggtacat  2700
ccccatggcc gtcacatgggg aggggcaatgt acggcataattgg gccttcctat gacgttaaatg  2760
ggctggtaggc gtgcactaatgt gaggctctatataaagctcatt ctcggtttagg gaacctcctat  2820
tctgccccgg gcacgtgcaag gagctgaaacag cttcctagac aattgacgagc taggcctgcg  2880
gccgctgcga ccaagggccc agatctgttta accggtcctta gtttttaaac caggttcgac  2940
atcgataattc aaccctttgga ttacaaatttt ttggaagatg taacttggtatatctatttcctat  3000
gttgcccttt ttacgctatg tggaatactgt gtttttatagc ctttgtatctact gtaattttgct  3060
tccgtatggg ctctcttccttt ccattttcat ttattttggtta cttttttttgag  3120
agaggctggc ccgtggtgcct ctggatctttc ctcggaggtgg gcattttgctctg cggcccaacc  3180
cccccatggtt gggctgatcga caccacatggt cagctttccttt cccgggcatttt cgctttttccc  3240
ctccctttttt ccacgcggga acactacgctc gctctctcgcgt ccggctcgctg gacaggggtctg  3300
ggctggtttg gcacgtgacaa cttccgagtttg ttggtgccgg gactcctgcgctgct ccctcctttgg  3360
tgctctgcct gggcgcctca gcaaggggttg cttctccccgctcttggctat gttttttgca  3420
gccctccaac cagggccacttt ctctctcctcgg gctctctcgcg gctctctcccg  3480
cgtctctccg ctccctccatg ccagaggtcggt atcctccctttt gggccgtcctc cccgcatacgga  3540
taaaatagat gtttcttcctttt gcaagggggtt aatggaagagc cccacagttgtaaag  3600
gttgctggcct tttgcggatcct cttctctgtccg ccggatttcgct gtagggactt cctcctccctg  3660
gctcctcaattct ccagcgtgagcc gctcgctctccg gctctctgctg gctctctcccg  3720
tacaatcagatg ttcttttcctt tttcttcgctg cttttttgca ggttggaactt cctcctccccg  3780
gaatagaggg ttcagagatc caaatagggagttctttctcagcttaaccgcaggedacccagcc  3840
ggatatcggtgt ggaactgtcgct cttggtctcttgc ctcttggacgc tggccaaacaca  3900
agaatctgtt gcaggtgtcct gctcgcgcctc gccggcgtcag tggcctctcctgagggcagaaaa  3960
aatatgggct aaacaggtatcctccttcctg agcccggctgcgcgccagggcag  4020
ccagagttgtc cctccagtccagcagccactt cagaggtgtcgt cttggtctctgagggcagaaaa  4080
cagggcgcctc caagagggcagg gaaaactactt ttgctctgagggcagaaaa  4140
776451seq
ccttcgacct ctttgccgcc cttttgcttc cccggagctca ataaagagc ccacaacccc 4200
tcactcgggg ccgcagtcct cccagttgact gatgcgccgg ggtaccctg tatccaaata 4260
acccctttgc agtgtgcattc gacattggtg ctctgctttc cttggaggag ttcctctcga 4320
gtgaattgac aaccgctcag ggggtctttt catt 4354

<210> 41
<211> 4978
<212> DNA
<213> Artificial Sequence

<220> Synthetic

<400> 41
gttgacattg attgtaattat agtatcatat tgcggtgtaaa ttcgtctaaa 60
gcccatatat ggacgctgct gtcatataac cccttgcattt ggtgtaataa 120
ccaaagaccc cggcctccgg acgctcataa tggactggttt cctccatagta agcgcattag 180
gactttgcc ttgctgctaa gttggtggag tttacgga aactgcccac ttgccagtac 240
ataaacgtta tcataagcct aagcagctcc ccctgtacgct caattgacag taagtgccgg 300
ctgtggcaga agatggacag cacgggtcgt gactgggtcc tcttgcgtcg tcttgcgtcg 360
tttgacgcc aatcgctcct attgctcagc cggcctggcc gtcattggttt 420
agggcaggg ctcgctctctt ttcgctctcc cggccctttt ggtgtcctttt 480
agtactgct gccagccg aatcgctcct attgctcagc cggcctggcc gtcattggttt 540
aatcggagc cgttcgctcgg ccgctcctct cggcctggcc gtcattggttt 600
cacacccct cctttgtgttg cggccctttt ggtgtcctttt 660
tcccataaa ggccttccttt gttgctctcc aatctgtggtt ctctgtgtctt ttgctgtggt 720
ttcctcctcg atgtattgcct accacagacg ggggtctttc attttgggag tcgctcggga 780
tttggagacc cttgcccagg gaccacgcac ccaccacagg gaggtaagct ggccagcaac 840
ttagtgctgct ccagctcttc ggtctagtttc tatgctagttt gtttatgctgct tcgctgctg 900
cagtttctct taactttctct gcattctgct gcagcttctgtt cggagcagc cgtgctgctg 960
accgagcgc caccagctgg gaagtgccag cggcgtttagg gcggggtttgt 1020
tgaggaagg gatccctgcg ggtacagc tcgcttgcgt ccgcttcttc gaggttagga 1080
cgagaactat ccagacttcg cttgccctgg tcttcctttt ggacgcagag 1140
cggccggtt tttctttgct gcggctgctg ctttcttcttc gcggctgctg 1200
gtttctgtat ttggtgaaaa ttagaggcca gactgctttc actttcctaa ttgctgtctt 1260
agggctggg aagattagtcg agcggtcagc tcaacaccc tctggttagt tcaagaagag 1320
aggtttggtt actcttctct tctgcaagtgc gccaaccttc tagttgcgggt gaccggtcga 1380
cggccaccc aaccgagacc tcataacacc ggattagacc aatgcttccac ccctgctgcc 1440
gcatgggcac ccagagcagg tccctctcat cttgagcctgg gaggctcttt tttgctgac 1500
776451Seq

ccccctctgg gtcaagcccct ttgtacaccc taagacctcgg cctcctcttc ctccatccgc 1560
cccgctctcc ccccttgaac ctctctgtttg gacccgcccc ctgatctcttc tttatccagc 1620
cctcaactct tctcttaggcg ccggaatcctc gatctgtacca agagacagga tgaaccatgg 1680
cacccctcacg aagttccccct tggacaaag cactcaacag atgttctatct ccccttcgccc 1740
aggggtcagaa agtcccaagcc atgtatattc ggggttagtgac taccggacag gacagtgcgtct 1800
gcaagacccg tacccctggac tgtggacccca agtgtagtggga agagttaacctc gattggaact 1860
ttgatgggctc tagtacctttt cagtctgaag gcctcaaacag cgacatgtac cttccatcccg 1920
ttgccatgtt ttcggagccc ttgagcaaaag gaccccaaccac gctgagctta ctgtaaggttt 1980
tcaagttaaa cccggaagct gcagaagacca acttgagggct cattgctataa cggatatagc 2040
acatggttac caacacacac tccccgttttg gaatggacga ggaatatattt cttatgagga 2100
cagacgcgcca aacccctctgct tgggcttccc aatggtctcctc tggacctccaa ggcgggtatt 2160
actgcgggtgt ggagccagac cgagcgcaac cgtaagcaac gctgagctta cctacccggc 2220
cctggtttgta tggctggagtc aagatcaccgg gagaacaatgc gaggatttgc cctgccgact 2280
gggatcatttt cattgagggag ctaagatattgc ggatagcccc 2340
gtttatacttt gccatggtgg tgcgaagact tttgaggtgtat acgaacacttt gaccccaagc 2400
ccattccccag gaaatggaat gttgctagct gcgtccacca cttgacacct caggcagttg 2460
ggaggagagag tggctggagac tgcagcagag ccgccattga cctgtgggagc ccaggggcccctt 2520
agtaccaacat ccggcctcac gatcccaagg ggggtctggga cagcccggcg cggctgactg 2580
gattccagca aacctccaat atcaacagct ttctcctgccag tgcggctcaca cgcgggtgcac 2640
gtactccgcat tcccggcagct gtcggccagct gacggtcgaag gacgaggcggc tcatcttggaa 2700
cctgtcctca tttcgacaca ttcgagtctga cggaaacagtc ctggccacag tgcgttccttca 2760
acgaaacagcc gacgcagaccc ttccataaca agaactaagg atcccatatttt catttggcaca 2820
gggttcataac tagtattttgg cccccctgcc ctagttgcct cctatctcttg cttttctattt 2880
gacgtgtaga gcctctccca attgggcggct ccctataccac tataggggctc ttcctaatatt 2940
ccgccgatgc acctctcctac tttcgacgtaa tgggtctctatt atatgggctt tctctattgac 3000
gtctataggg cgggcttcatt gacgtatatg gcgctcttccc cattgacagtct aattacggtta 3060
aatggccgcc ctcggctcaat cccctctagar gcgatccag tcctgatccaa ccacccacca cctgacgctaa 3120
tggggtgggt ctaccgcacc atccatgtggt cccttgggcc cctatctctgg ctaatgacgg 3180
ctgggtctcc atttcgagtg tatttaatag ctatgtgcctc taaaagcttgcc taactcaatact 3240
tattggagtt atagcagggc tattggccag taccctcatt gacgcttaatt gcggttaaatg 3300
gcgcggcatgt ctgtcgaagg ctaaacttccac tgaacgtcaat gggaaggccc aatgacgcaaa 3360
atgggcgttc cattgagcttc aatggccggtcg aggccgactt cattataagcc 3420
aatgctcttg ttagggaaccc ccacccctcc ctggggacgtgc ggagccagtt tgaagctcttct 3480
agacacatgga cgcgctggccg tgcggccgcgg tgcagccagg gccccagatct gttaccgcggt 3540
cctaggtta aacctgagtt cgacatcgat aatcaacctc tgattaca aatggtgaa 3600
agatggactg gtatctttaa ctatggttct ccttttacgc tatgtggata cgctgcttta 3660
atgcctttgt atcatgctat tgcctccggt atggctttca tttttcttct cttgtataaa 3720
tccctggttc tgcctcttta tgaggatttg tcgccccggt tcaggcaacgc gggcgtggtg 3780
tgcactgtgt ttgctgacgc aaccccccac ggtttgggga ttgccccacc acgctgacct 3840
cctttcggga ctttcgcttt ccccccctct atggccacgg cggaactcat cgccgtgcctg 3900
cctgcggcct gctggaacagg ggctcggctg ttgggcaagt acatatccgt ggtgtttgtcg 3960
gggaaatcat cgctcccttc ttcggtgctc gcctgttttg ccacgctgat ttcgcccggg 4020
acgtccctct gctacgttcct tgtgccccct aacctccaggg accctcccttc ccgcggccctg 4080
tgcgcggctc tcgcggccct tccgctgctt cgcctccgcc ctcagccagag tcggatctcc 4140
cctgaggcgg cctcccccgc a cgcataaat ataaggattt tttagtttct cagaaaaaggg 4200
ggggaatgaa agacccccc accgtgagttg caagctgatcg ttaagtaacgc ccttttgca 4260
agcatggaa aatacataa ctgagaattag agaagttcag atcaaggtca gaaacagagt 4320
gaacagctga atatgggcca aacaggtatat ctgtggtaag cagtttctgc ccgggctcag 4380
ggccaaagac aggatggacag tgcataat ggcccaacag gatactcttg gagtaagcgtt 4440
cctggccggg cgcttagggc ggaacagatgt catcccaagat ggtgcctcgcagt cctcgaggt 4500
ctttagagaa ccatacagtgt ttttcagggt gccccaagga cctggaatgta cctctgtgcc 4560
tatttgacta aaccaatcag ttcgcctttct gcctggtttcc gcgcgctcttc gccgcttcag 4620
tctcaataaa gagccttcatg ccctctcactc gggggtgctg tgcctggatt gactgagctg 4680
cccggttacc cgtgtatcct ataacaacttc ttgcagttgc atccgacctg tggctcctgt 4740
gttcttggg agggtctcct ctgagttgat gactacccct cagcgggggt cttttcatt 4798