Title: POLYCISTRONIC EXPRESSION OF ANTIBODIES

Abstract: Described herein is a novel expression system for producing multiple gene products of interest from a single polycistronic construct. In particular, the expression system contains a polycistronic vector capable of expressing functional antibodies in eukaryotic host cells, which vector contains at least the following elements operably linked in the 5' to 3' orientation: a promoter operable in a eukaryotic cell; a DNA sequence encoding at least the variable region of an antibody light chain; an internal ribosome entry site (IRES); and at least one DNA sequence encoding an antibody heavy chain. Also disclosed are mammalian cells containing the polycistronic expression vector, and a method of producing functional antibodies in mammalian cells transfected with the polycistronic expression vector.
POLYCISTRONIC EXPRESSION OF ANTIBODIES

RELATED APPLICATION

[0001] This application relates to and claims priority to provisional patent application 60/331,481 filed on November 16, 2001, and provisional application 60/400,687, filed on August 5, 2002, both incorporated by reference in their entirety herein. Additionally, this application claims priority to PCT/US2002/02373 and PCT/US2002/02374 both filed January 29, 2002.

FIELD OF THE INVENTION

[0002] The present invention relates to a novel polycistronic expression system and methods for producing antibodies in eukaryotic cells using this expression system. More specifically, the invention relates to a eukaryotic polycistronic expression system in which antibody heavy and light chain genes are transcribed from the same promoter, and, preferably, the antibody heavy and light chain genes are separated by one or more internal ribosome entry sites (IRES’s).

BACKGROUND OF THE INVENTION

[0003] Methods for expressing genes of choice in recombinant host cells utilizing polycistronic expression vectors are well known. Historically, polycistronic expression vectors incorporating a desired product gene sequence at the 5’ end of a transcribed region and a 3’ selectable marker gene sequence have been employed. Such vectors display inefficient translation of the 3’ selectable gene, while preferentially translating the desired gene sequence at the 5’ end of the polycistronic mRNA. Recombinant host cells expressing high levels of the desired gene product are obtained via a single step method comprising culturing initial transfectants in a selectable medium. A drawback of such vectors is the unpredictable influence an upstream reading frame may exert on translation of the selectable marker sequence. See Kaufman, Meth. Enzymol., 185:487 (1990).
[0004] For example, U.S. Patent No. 4,713,339, issued to Levinson et al. (assigned on its face to Genentech, Inc.) discloses a polycistronic expression system capable of producing a gene product of interest in eukaryotic host cells. In the system patented by Levinson et al., the gene sequence of the second cistron encodes a protein that provides for the detection of transfectants or transformants that express a desired gene of interest encoded by the first cistron. Certain growth conditions induce amplified expression of the detector gene in a host cell, which thereby enhances the expression of associated sequences encoding the desired gene contained on the polycistronic transcription unit. In particular, Levinson constructed vectors containing a polycistron in which a sequence encoding hepatitis B-surface antigen was positioned upstream of a sequence encoding the screening marker dihydrofolate reductase. Due to the polycistronic arrangement of the Levinson vectors, unequal expression of the downstream marker gene sequence of the second cistron occurs in comparison to the expression of the first cistron sequence.

[0005] Additionally, polycistronic expression systems have been used to express multichain polypeptides. For example, polycistronic expression of multichain polypeptides is reported in US Patent No. 6,060,273 to Dirks et al.; US Patent No. 6,033670 to Bublot; US Patent No. 6,096,505 to Selby et al.; US Patent No. 6,143,520 to Marasco et al.; US Patent No. 6,153,199 to Audonnet et al.; and US Patent No. 6,156,558 to Johnston et al.

[0006] Investigators have determined that levels of the second gene in the polycistron are improved by the incorporation of an internal ribosome entry site (IRES) between the genes in the polycistron. IRES elements, first identified in picornaviruses, mediate the initiation of translation by directly recruiting and binding ribosomes to a message, bypassing the 7-methyl guanosine-cap involved in typical ribosome scanning. The presence of an IRES sequence can increase the level of cap-independent translation of a desired protein. Early publications descriptively refer to IRES sequences as “translation enhancers”. For example, cardioviral RNA “translation enhancers” are described in U.S.

[0007] Some IRES containing reporter cistrons have been patented, such as the XIAP IRES (US Patent Nos. 6,171,821 and 6,159,709 to Korneluk). One IRES sequence having known use in polycistronic expression vectors is that of herpes virus; other viruses may be used as well (See US Patent 6,193,980). Korneluk discloses bicistronic vectors constructed by inserting β-galactosidase and chloramphenicol acetyltransferase reporter sequences into a plasmid having a CMV promoter, such that the two cistrons are separated by a 100 bp intercistronic linker region containing an IRES sequence. The first cistron, encoding β-galactosidase, was translated via a conventional cap-dependent mechanism. The second cistron, encoding chloramphenicol acetyltransferase, was translated only when the preceding linker region contained the IRES site. Thus, Korneluk showed that IRES sequences can mediate the translation of a second open reading frame in bicistronic mRNA constructs designed to measure cellular responses to stress. However, only marker sequences were utilized in the second cistron, as the state of the art recognized the inefficiency associated with expression of second sequences in bicistronic arrangements.

[0008] Additionally, the co-expression of amplifiable markers using polycistronic expression systems have been described. For example, WO 92/17566 discloses a method of co-transfecting host cells with an intron-modified selectable gene and a gene encoding a protein of interest. The intron-modified gene is generated via insertion of an intron into the transcribed region of a selectable gene such that the intron is correctly spliced from the mRNA with reduced efficiency. Inefficient splicing of this nature results in low amounts of selectable marker protein being produced from the intron-modified selectable gene in comparison with unmodified selectable gene sequences. While reduced amounts of selectable gene are produced concurrently with the protein of interest, this model does not employ a transcriptional linkage between the selectable gene sequences and the desired gene sequence of interest, as the two sequences are driven by separate promoters.
[0009] Antibody production using a dicistronic expression vector has been disclosed, employing vectors exhibiting transcriptional linkage between dihydrofolate reductase (DHFR) and a nucleotide sequence encoding a desired antibody product. See U.S. Patent No. 5,561,053 to Crowley. The method of Crowley utilizes DNA constructs having a single promoter to drive expression of a selectable marker sequence and a single sequence encoding a protein of interest. The heavy chain of the desired antibody was inserted downstream of a selectable gene, DHFR, which was placed within an intron at the 5' end of the DNA construct. The intron was positioned between the cytomegalovirus immediate early promoter (CMV) and the sequence encoding the antibody heavy chain. The light chain of the antibody was placed into a second vector constructed to place the light chain sequence under control of the SV40 promoter/enhancer and the selectable hygromycin B resistance gene sequence driven by the CMV promoter/enhancer and SV40 poly-A. Vectors containing the light and heavy chain sequences were linearized and co-transfected into host cells for expression and subsequent disulfide linkage between light and heavy chains. Thus, in order to obtain a complete antibody structure according to Crowley, two vectors must be constructed for separate expression of the heavy and light chain components of the antibody. This method is inefficient due to the use of multiple vectors.

[0010] The method of Crowley exemplifies the limitations on multiple subunit protein expression imposed by the positional effect associated with traditional polycistronic expression systems. The imbalance of expression between the 5' and 3' genes in a polycistron reduces the efficiency with which commercial scale production of multichain proteins or similar products of interest can be generated.

[0011] U.S. Patent No. 6,060,273 to Dirks discloses multicistronic expression units that allow equimolar expression of genes located in corresponding cistrons. According to Dirks, bicistronic expression vectors may be configured for expressing two genes of interest, such as PDGF-A and PDGF-B, with the assistance of an IRES. IRES-dependent translation in the bicistronic expression vectors is aided by a *Xenopus laevis* 5'UTR β-globin sequence that enhances
expression of the second cistron such that equimolar expression of cistrons 1 and 2 is achieved. The bicistronic vectors may be arranged to as follows: promoter-first cistron-IRES-β-globin sequence-second cistron.

[0012] While the teachings of Dirks suggest a method of overcoming the inefficiency associated with polycistronic expression of some multichain protein subunits, the teachings are not considered by those skilled in the art to be a satisfactory solution to the unpredictability of downstream cistron expression. Mizuguchi et al. (IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. Mol. Therapy, 1(4): 376-382 (April, 2000)) investigated the efficiency of IRES-dependent second gene expression in comparison with cap-dependent first gene expression in vitro in several cultured cell lines as well as in vivo in mouse liver. IRES-dependent second gene expression ranged from 6 to 100% of first gene expression, depending upon which cell types and reporter genes were used in vector constructs. In addition, the selection of which gene was positioned first in the bicistronic vectors affected the expression of gene positioned downstream.

[0013] Moreover, Borman et al. noted that the efficiency of IRESes to drive cap-independent translation of a second cistronic sequence was greatly affected by the type of cell chosen as the expression host. (Comparison of picornaviral IRES-driven internal initiation of translation in cultured cell of different origins. Nucleic Acids Res, 25(5): 925-932 (1997)). Dramatic variations in activity were noted for individual IRES elements of vectors transfected into different cell lines.

[0014] These concerns with the unpredictability of downstream translation of sequences in polycistronic expression systems are particularly pertinent to the efficient production of antibodies, as the ratio of light chain to heavy chain expression is key to ensuring proper antibody folding and secretion. In this regard, Horwitz and coworkers concluded that because some antibody light chains are naturally poorly secreted, the co-expression and proper association of light and heavy chains is important to the efficient secretion of at least some whole antibodies. (Chimeric immunoglobulin light chains are secreted at different levels: influence of framework-1 amino acids. Mol. Immun. 31(9): 683-699 (1994)).
Interestingly, Kolb et al. reported inefficient expression of the second cistronic gene in the following genomic DNA dicistronic expression vector: CMV promoter-antibody light chain gene sequence-IRES-antibody heavy chain sequence-polyadenylation signal (Expression of a recombinant monoclonal antibody from a bicistronic mRNA. Hybridoma, 16(5): 421-426 (1997)). Western blotting indicated a substantial amount of light chain was produced by this construct following transfection into murine myeloma cells, but heavy chain protein was barely detectable. The investigators specifically designed this dicistronic expression construct to inefficiently express the heavy chain due to concerns about host cell toxicity induced by unpaired heavy chains. Complete, functional antibodies were detected in the supernatant of these cells only after column purification, suggesting this IRES vector/cell combination is a poor model for producing antibodies of choice on a commercially useful scale.

To the best of the inventors' knowledge, an effective means for producing functional multichain proteins, such as antibodies, using a single polycistronic expression system suitable for commercial production schemes, has heretofore been unreported in the literature. Thus, a need exists for an efficient means of producing commercially acceptable amounts of antibodies in recombinant host cell systems, in which a single construct may be utilized for expression of two or more desired protein products, such as the light and heavy chains of a desired antibody.

SUMMARY OF THE INVENTION

The invention pertains to the expression of functional (antigen-binding) antibodies at adequate levels of expression via a polycistronic expression system in a eukaryotic host cell.

More specifically, the invention relates to the expression of functional antibody molecules in eukaryotic cells, preferably mammalian cells, fungal or yeast cells and still more preferably (Chinese Hamster Ovary) CHO cells, using a polycistronic expression system.

An aspect of this preferred embodiment of the invention is the expression of functional antibodies in mammalian cells using a polycistronic
expression system comprising a eukaryotic promoter operably linked to at least one antibody light chain coding sequence and at least one antibody heavy chain coding sequence, wherein such antibody coding sequences are separated by at least one IRES. In this expression system, the gene that is 3’--most of the promoter has at its 3’ terminus a poly A sequence, the other coding sequences in the polycistron lack a poly A sequence, and each gene is preceded by a start codon and ends with a stop codon.

[0020] Another embodiment of the invention provides a polycistronic expression unit comprising in the 5’ to 3’ direction the CMV promoter operably linked to an antibody light chain coding sequence that is flanked by a start and a stop codon, followed by one or more antibody heavy chain coding sequences. Each heavy chain coding sequence is also flanked by a start and a stop codon. Each pair of heavy chain coding sequences is separated by at least one IRES, preferably that of a cardiovirus, such as human encephalomyocarditis virus or poliovirus. The DNA sequence encoding the antibody light chain is preferably expressed at a ratio ranging between 10:1 and 1:2 relative to the expression of the DNA sequence encoding the antibody heavy chain. More preferably, the antibody light chain will be expressed at ratios relatively balanced with respect to the heavy chain, i.e., from about 5/1 to 1/1 and still more preferably about 3/1 to 1/1, and still more preferably about 1.5/1 to 1/1.

[0021] Another embodiment of the invention is a eukaryotic cell line, such as a CHO cell line, that secretes an antibody, wherein expression of said antibody is via a polycistron as described herein. In one aspect of the invention the polycistron comprises in the 5’ to 3’ orientation following the CMV promoter: a secreting signal; an antibody light chain coding sequence which comprises a start and a stop codon, but which does not comprise a poly A sequence; an IRES preferably selected from the group consisting of a cardiovirus, poliovirus and a herpes virus; at least one antibody heavy chain coding sequence each operably linked to a secreting signal sequence; and 3′ of each heavy chain coding sequence an IRES, preferably selected from the group consisting of a cardiovirus, poliovirus and a herpes virus, IRES, and a poly A sequence contained at the 3′ end of the gene located at the 3′ end of the polycistron.
Preferably, a eukaryotic cell containing the polycistron will secrete about 5-100 picograms of functional antibody per cell, per day. Preferably, at least 1-5 picograms of antibody are secreted per cell, per day over at least a continuous 3-4 day period. In preferred embodiments, the eukaryotic cell is a CHO cell or a yeast cell, e.g., Pichia.

[0022] Another embodiment of the invention is a culture of mammalian or yeast cells comprising a polycistronic expression system capable of producing functional antibodies. Vectors containing polycistronic sequences according to the invention may be introduced into the mammalian or yeast cells. During cell culturing, desired exogenous DNA sequences may be introduced to target mammalian or yeast cells, such that exogenous DNA is inserted into the genome of the mammalian or yeast cell via homologous recombination. Depending upon the sequences employed, functional antibodies may be recovered from the biomass of the cell culture or from the cell culture medium. Methods of integrating genes at specific sites in eukaryotic, e.g., mammalian cells via homologous recombination and vectors suitable for such recombinant processes are disclosed in U.S. Patent Nos. 5,998,144, 5,830,698, and 6,413,777 the entire contents of which are incorporated herein by reference.

[0023] Yet another embodiment is a method of producing a functional antibody comprising culturing eukaryotic cells, preferably mammalian or yeast cells containing a polycistronic expression system that expresses antibody light and heavy chain sequences, and recovering functional antibodies from the cell culture. The functional antibodies may be produced in batch fed cell cultures at levels suitable for therapeutic use under conditions optimized for maximal commercial output. For example, CHO cells grown in batch fed cultures in which glucose levels are continuously controlled can produce recombinant protein for at least 12 days or more. See, for example, U.S. Patent 6,180,401 for a discussion relating to the output of recombinant protein by cells grown in batch fed cultures.
BRIEF DESCRIPTION OF THE FIGURES

[0024] The invention is further illustrated in the Figures discussed herein, wherein:

[0025] Figure 1 depicts a schematic drawing of a polycistronic expression construct encoding HuCC49, a humanized anti-TAG72 antibody;

[0026] Figure 2 depicts a schematic of the NEOSPLA vector and the situs of the IRES between light and heavy chain sequences;

[0027] Figure 3 depicts a schematic of the NEOSPLA vector in which immunoglobulin light and heavy chain gene sequences are located in independent transcriptional cassettes;

[0028] Figure 4 is a Southern blot of polycistronic HuCC49 G418-resistant cell isolates 22E6 and 22B6 probed with HuCC49 heavy chain;

[0029] Figure 5 is a Southern blot of polycistronic HuCC49 G418-resistant cell isolates 25A2 and 22H10 probed with HuCC49 heavy chain and

[0030] Figure 6 shows the sequence of expression construct HuCC49 which is a humanized antibody that binds TAG 72.

[0031] Figure 7 schematically depicts construction of a polycistronic expression construct according to the invention. "L" indicates a leader sequence V_L indicates DNA encoding an antibody variable light chain and V_H indicates DNA encoding antibody variable heavy chain.

[0032] Figure 8 depicts a plasmid, PEMPTY A4, used to construct a polycistronic vector according to the invention.

[0033] Figure 9 contains the nucleic acid sequence of PEMPTY A4 plasmid.

[0034] Figure 10 depicts a plasmid, PEMPTY B, used to construct a polycistronic vector according to invention.

[0035] Figure 11 contains the nucleic acid sequence of the plasmid PEMPTY B.

[0036] Figure 12 depicts a polycistronic vector according to the invention, Polycis 2 which expresses HuCC49 Gly/Ser.

[0037] Figure 13 contains the nucleic acid sequence of HuCC49 Gly/Ser contained in Polycis 2 vector.
DETAILED DESCRIPTION OF THE EMBODIMENTS

[0038] Prior to providing a detailed description of preferred embodiments, the following definitions are provided.

[0039] "IRES" or an internal ribosome entry site, means a region of a nucleic acid molecule e.g., an mRNA molecule, that allows internal ribosome entry/binding sufficient to initiate translation in an assay for cap-independent translation, such as the bicistronic reporter assay described in U.S. Patent 6,715,821. The presence of an IRES within a mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated. IRES's were first identified in picornaviruses, and are considered the paradigm for cap-independent translation. The 5' UTRS of all picornaviruses are long and mediate translational initiation by directly recruiting and binding ribosomes, thereby circumventing the initial cap-binding step.


[0041] "Cap-dependent translation" means that 7-guano, 7-methylguanosine cap must be present at the 5' end of an mRNA molecule in order to initiate translation of the mRNA into a protein.
"Cap-independent translation" means that a 7-methylguanosine cap is not required for translation of the mRNA molecule. Cap-independent translation mechanisms include ribosome re-initiation, ribosome shunting, and internal ribosome binding.

"Cistron" means a "coding sequence" or sequence of nucleic acid that encodes a single protein or polypeptide.

"Reporter cistron" means a segment of nucleic acid that encodes a detectable gene product; which may be expressed under the translation control of an IRES.

" Reporter gene" means any gene or translatable nucleic acid sequence that encodes a product whose expression is detectable and/or quantifiable by immunological, chemical, biochemical or biological assays. A reporter gene may have e.g., one of the following attributes: fluorescence (e.g., green fluorescent protein), toxicity (e.g., ricin) enzymatic activity (e.g., lacZ/beta-galactosidase, luciferase, chloramphenicol transferase), and an ability to be bound by a second molecule (e.g., biotin or a detectable labeled antibody).

"Promoter" means a minimal sequence sufficient to direct transcription, preferably in a eukaryotic cell. Specifically, included are promoter elements that are sufficient to render promoter-dependent gene expression controllable in a cell type-specific, tissue-specific, or temporal-specific manner, or inducible by external signals or agents, such elements may be located in the 5' or 3' or intron sequence regions of a particular gene. Preferred promoters for use in the invention include viral, mammalian and yeast promoters that provide for high levels of expression, e.g., mammalian CMV promoter, yeast alcohol oxidase, phosphoglycerokinase promoter, lactose inducible promoters, galactosidase promoter, adeno-associated viral promoter, baculovirus promoter, poxvirus promoter, retroviral promoters, adenovirus promoters, SV40 promoter HMG *hydroxymethylglutarylcoenzyme A) promoter, TK (thymidine kinase) promoter, 7.5K or H5R poxvirus promoters, adenovirus type 2 MPC late promoter, alpha-antrypsin promoter, factor IX promoter, immunoglobulin promoter, CFTR surfactant promoter, albumin promoter and transferrin promoter.
“Expression vector” means a DNA construct containing at least one promoter operably linked to a downstream gene, cistron or RNA coding region. Herein, the promoter may be operably linked to one or more genes or cistrons each preceded by a start and followed by a stop codon. Transfection of the expression vector into a recipient cells, i.e., eukaryotic cell, e.g., mammalian cell, fungal cell, yeast cell, allows the cell to express RNA encoded by the expression vector. Expression vectors include e.g., genetically engineered plasmids or viruses.

“Transformation” or “transfection” refers to introduction of a polycistronic vector or construct into suitable eukaryotic cells for expression of genes therein.

“Eukaryotic cells” refers to any eukaryotic cell which produces or expresses genes of interest using the polycistronic expression system of the invention. This includes by way of example mammalian cells such as CHO, myeloma, BHK, immune cells, insect cells, avian cells, amphibian cells, e.g., frog oocytes, fungal and yeast cells. Yeast useful for expression include by way of example Saccharomyces, Schizosaccharomyces, Hansenula, Candida, Torulopsis, Yarrowia, Pichia, et al. Particularly preferred yeast for expression include methylotrophic yeast strains, e.g., Pichia pastoris, Hansenula polymorpha, Pichia guillermordii, Pichia methanolica, Pichia inositovera, et al. (See e.g., U.S. Patent 4,812,405, 4,818,700, 4,929,555, 5,736,383, 5,955,349, 5,888,768, and 6,258,559). These and other patents further describe promoters, terminators, enhancers, signals sequences, and other regulatory sequences useful for facilitating heterologous gene expression in yeast, e.g., antibody genes as in the present invention.

As is apparent from the instant specification, genetic sequences useful for producing the antibodies using the polycistronic expression system of the present invention may be obtained from a number of different sources. For example, a variety of human antibody genes are available in the form of publicly accessible deposits. Many sequences of antibodies and antibody-encoding genes have been published and suitable antibody genes can be synthesized from these sequences much as described herein. Alternatively, antibody-producing cell lines
may be selected and cultured using techniques well known to the skilled artisan. Such techniques are described in a variety of laboratory manuals and primary publications. In this respect, techniques suitable for use in the invention as described below are described in *Current Protocols in Immunology*, Coligan et al., Eds., Green Publishing Associates and Wiley-Interscience, John Wiley and Sons, New York (1991) which is herein incorporated by reference in its entirety, including supplements.

[0051] It will further be appreciated that the scope of this invention further encompasses all alleles, variants and mutations of the DNA sequences described herein.

[0052] As is well known, antibody-encoding RNA may be isolated from the original antibody-producing hybridoma cells or from other transformed cells by standard techniques, such as guanidinium isothiocyanate extraction and precipitation followed by centrifugation or chromatography. Where desirable, mRNA may be isolated from total RNA by standard techniques such as chromatography on oligo(dT) cellulose. Techniques suitable for these purposes are familiar in the art and are described in the foregoing references.

[0053] cDNAs that encode the light and the heavy chains of the antibody may be made, either simultaneously or separately, using reverse transcriptase and DNA polymerase in accordance with well known methods. It may be initiated by consensus constant region primers or by more specific primers based on the published heavy and light chain DNA and amino acid sequences. As discussed above, PCR also may be used to isolate DNA clones encoding the antibody light and heavy chains. In this case libraries may be screened by consensus primers or larger homologous probes, such as mouse constant region probes.

[0054] DNA encoding light and heavy chains may be isolated, typically in the form of plasmid DNA, from the cells as described herein, restriction mapped and sequenced in accordance with standard, well known techniques set forth in detail in the foregoing references relating to recombinant DNA techniques. Of course, the DNA may be modified according to the present invention at any point during the isolation process or subsequent analysis.
While the inventive compositions and methods are suitable for any antibody, or indeed any multichain protein, preferred antibody sequences are disclosed herein. Oligonucleotide synthesis techniques compatible with this aspect of the invention are well known to the skilled artisan and may be carried out using any of several commercially available automated synthesizers. In addition, DNA sequences encoding several types of heavy and light chains set forth herein can be obtained through the services of commercial DNA vendors. The genetic material obtained using any of the foregoing methods may then be altered or modified to provide antibodies compatible with the present invention and the desired use of such antibodies.

A variety of different types of antibodies may be expressed according to the instant invention. As used herein “antibody” and "antibodies" refers to assemblies which have significant known specific immunoreactive activity to an antigen (e.g. a tumor associated antigen), comprising light and heavy chains, with or without covalent linkage between them and thus include single chain antibodies, and the like. "Modified antibodies" according to the present invention are held to mean immunoglobulins, antibodies, or immunoreactive fragments or recombinants thereof, in which at least a fraction of one or more of the constant region domains has been deleted or otherwise altered so as to provide desired biochemical characteristics such as the ability to non-covalently dimerize, increased tumor localization or modified serum half-life when compared with a whole, unaltered antibody of approximately the same (scAb) immunogenicity. For the purposes of the instant application, immunoreactive single chain antibody constructs having altered or omitted constant region domains may be considered to be modified antibodies. As discussed above, preferred modified antibodies or domain deleted antibodies expressed using the polycistronic system of the present invention may have at least a portion of one of the constant domains deleted. More preferably, one entire domain of the constant region of the modified antibody will be deleted and even more preferably the entire Cys2 domain will be deleted.

Basic immunoglobulin structures (e.g., antibodies and the like) in vertebrate systems are relatively well understood. As will be discussed in more
detail below, the generic term "immunoglobulin" comprises five distinct classes of antibody that can be distinguished biochemically. While all five classes are clearly within the scope of the present invention, the following discussion will generally be directed to the class of IgG molecules. With regard to IgG, immunoglobulins comprise two identical light polypeptide chains of molecular weight approximately 23,000 Daltons, and two identical heavy chains of molecular weight 53,000-70,000. The four chains are joined by disulfide bonds in a "Y" configuration wherein the light chains bracket the heavy chains starting at the mouth of the "Y" and continuing through the variable region. The dimeric assemblies disclosed herein may be likened to two associated Ys (H4L4) so there will be four binding sites. Hence the term "tetravalent" antibodies.

[0058] More specifically, both the light and heavy chains are divided into regions of structural and functional homology. The terms "constant" and "variable" are used functionally. In this regard, it will be appreciated that the variable domains of both the light (V_L) and heavy (V_H) chains determine antigen recognition and specificity. Conversely, the constant domains of the light chain (C_L) and the heavy chain (C_H1, C_H2 or C_H3) confer important biological properties such as secretion, transplacental mobility, Fc receptor binding, complement binding, and the like. By convention the numbering of the constant region domains increases as they become more distal from the antigen binding site or amino-terminus of the antibody. Thus, the C_H3 and C_L domains actually comprise the carboxy-terminus of the heavy and light chains respectively.

[0059] Light chains are classified as either kappa or lambda (\(\kappa, \lambda\)). Each heavy chain class may be bound with either a kappa or lambda light chain. In general, the light and heavy chains are covalently bonded to each other, and the "tail" portions of the two heavy chains are bonded to each other by covalent disulfide linkages when the immunoglobulins are generated either by hybridomas, B cells or genetically engineered host cells. In the heavy chain, the amino acid sequences run from an N-terminus at the forked ends of the Y configuration to the C-terminus at the bottom of each chain. At the N-terminus is a variable region and at the C-terminus is a constant region. Those skilled in the art will appreciate that heavy chains are classified as gamma, mu, alpha,
delta, or epsilon, (γ, μ, α, δ, ε) with some subclasses among them. It is the
nature of this chain that determines the "class" of the antibody as IgA, IgD, IgE
IgG, or IgM. The immunoglobulin subclasses (isotypes) e.g. IgG1, IgG2, IgG3,
IgG4, IgA1, etc. are well characterized and are known to confer functional
specialization. Modified versions of each of these classes and isotypes are
readily discernable to the skilled artisan in view of the instant disclosure and,
accordingly, are within the purview of the instant invention.

[0060] As indicated above, the variable region allows the antibody to
selectively recognize and specifically bind epitopes on immunoreactive
antigens. That is, the \( V_L \) domain and \( V_H \) domain of an antibody combine to form
the variable region that defines a three dimensional antigen binding site. This
quaternary antibody structure provides for an antigen binding site present at the
end of each arm of the Y. More specifically, the antigen binding site is defined
by three complementary determining regions (CDRs) on each of the \( V_H \) and \( V_L \)
chains.

[0061] The six CDRs present on each monomeric antibody (H\(_2\)L\(_2\)) are short,
non-contiguous sequences of amino acids that are specifically positioned to
form the antigen binding site as the antibody assumes its three dimensional
configuration in an aqueous environment. The remainder of the heavy and light
variable domains show less inter-molecular variability in amino acid sequence
and are termed the framework regions. The framework regions largely adopt a
\( \beta \)-sheet conformation and the CDRs form loops connecting, and in some cases
forming part of, the \( \beta \)-sheet structure. Thus, these framework regions act to
form a scaffold that provides for positioning the six CDRs in correct orientation
by inter-chain, non-covalent interactions. In any event, the antigen binding site
formed by the positioned CDRs defines a surface complementary to the epitope
on the immunoreactive antigen. This complementary surface promotes the non-
covalent binding of the antibody to the immunoreactive antigen epitope.

[0062] For the purposes of the present invention, it should be appreciated that
antibodies expressed using the subject polycistrionic expression system may
comprise any type of variable region that provides for the association of the
antibody with the selected antigen. In this regard, the variable region may
comprise or be derived from any type of mammal that can be induced to mount a humoral response and generate immunoglobulins against the desired antigen. As such, the variable region of the modified antibodies may be, for example, of human, murine, non-human primate (e.g. cynomolgus monkeys, macaques, etc.) or lupine origin. In particularly preferred embodiments both the variable and constant regions of compatible modified antibodies are human. In other selected embodiments the variable regions of compatible antibodies (usually derived from a non-human source) may be engineered or specifically tailored to improve the binding properties or reduce the immunogenicity of the molecule. In this respect, variable regions useful in the present invention may be humanized or otherwise altered through the inclusion of imported DNA or amino acid sequences.

[0063] For the purposes of the instant application the term "humanized antibody" shall mean an antibody derived from a non-human antibody, typically a murine antibody, that retains or substantially retains the antigen-binding properties of the parent antibody, but which is less immunogenic in humans. This may be achieved by various methods, including (a) grafting the entire non-human variable domains onto human constant regions to generate chimeric antibodies; (b) grafting at least a part of one or more of the non-human complementarity determining regions (CDRs) into a human framework and constant regions with or without retention of critical framework residues; or (c) transplanting the entire non-human variable domains, but "cloaking" them with a human-like section by replacement of surface residues. Such methods are disclosed in Morrison et al., Proc. Natl. Acad. Sci. 81: 6851-5 (1984); Morrison et al., Adv. Immunol. 44: 65-92 (1988); Verhoeven et al., Science 239: 1534-1536 (1988); Padlan, Molec. Immun. 28: 489-498 (1991); Padlan, Molec. Immun. 31: 169-217 (1994), and U.S. Pat. Nos. 5,585,089, 5,693,761 and 5,693,762 all of which are hereby incorporated by reference in their entirety.

[0064] Those skilled in the art will appreciate that the technique set forth in option (a) above will produce "classic" chimeric antibodies. In the context of the present application the term "chimeric antibodies" will be held to mean any antibody wherein the immunoreactive region or site is obtained or derived from
a first species and the constant region (which may be intact, partial or modified in accordance with the instant invention) is obtained from a second species. In preferred embodiments the antigen binding region or site will be from a non-human source (e.g. mouse) and the constant region is human. While the immunogenic specificity of the variable region is not generally affected by its source, a human constant region is less likely to elicit an immune response from a human subject than would the constant region from a non-human source. Certain chimeric antibodies may be generated by first immunizing monkeys with a desired antigen, isolating antibodies raised to the antigen and substituting the constant region of the heavy and light chains with a constant region (e.g., human) having desired function (e.g., human effector function, or the like). Such "Primatized®" antibodies and methods of making same are further described in U.S. Patent No. 5,658,570, incorporated herein by this reference in its entirety.

[0065] Preferably, variable domains in both the heavy and light chains of chimeric antibodies are altered by at least partial replacement of one or more CDRs and, if necessary, by partial framework region replacement and sequence changing. Although the CDRs may be derived from an antibody of the same class or even subclass as the antibody from which the framework regions are derived, it is envisaged that the CDRs will be derived from an antibody of different class and preferably from an antibody from a different species. It must be emphasized that it may not be necessary to replace all of the CDRs with the complete CDRs from the donor variable region to transfer the antigen binding capacity of one variable domain to another. Rather, it may only be necessary to transfer those residues that are necessary to maintain the activity of the antigen binding site. Given the explanations set forth in U. S. Pat. Nos. 5,585,089, 5,693,761 and 5,693,762, it will be well within the competence of those skilled in the art, either by carrying out routine experimentation or by trial and error testing to obtain a functional antibody with reduced immunogenicity.

[0066] Alterations to the variable region notwithstanding, those skilled in the art will appreciate that modified antibodies compatible with the instant invention will comprise antibodies, or immunoreactive fragments thereof, in which at least
a fraction of one or more of the constant region domains has been deleted or otherwise altered so as to provide desired biochemical characteristics such as increased tumor localization or modified (e.g., reduced) serum half-life when compared with an antibody of approximately the same immunogenicity comprising a native or unaltered constant region. In preferred embodiments, the constant region of the antibodies expressed using the subject polycistronic vectors will comprise a human constant region. Modifications to the constant region compatible with the instant invention comprise additions, deletions or substitutions of one or more amino acids in one or more domains. That is, the modified antibodies disclosed herein may comprise alterations or modifications to one or more of the three heavy chain constant domains (C_H1, C_H2 or C_H3) and/or to the light chain constant domain (C_L). As will be discussed in more detail below and shown in the examples, certain embodiments of the invention comprise expression of antibodies having modified constant regions wherein one or more domains are partially or entirely deleted ("domain deleted antibodies"). In especially preferred embodiments compatible modified antibodies will comprise domain deleted constructs or variants wherein the entire C_H2 domain has been removed (ΔC_H2 constructs). For other preferred embodiments a short amino acid spacer may be substituted for the deleted domain to provide flexibility and freedom of movement for the variable region.

[0067] As previously indicated, the subunit structures and three dimensional configuration of the constant regions of the various immunoglobulin classes are well known. For example, the C_H2 domain of a human IgG Fc region usually extends from about residue 231 to residue 340 using conventional numbering schemes. The C_H2 domain is unique in that it is not closely paired with another domain. Rather, two N-linked branched carbohydrate chains are interposed between the two C_H2 domains of an intact native IgG molecule. It is also well documented that the C_H3 domain extends from the C_H2 domain to the C-terminal of the IgG molecule and comprises approximately 108 residues while the hinge region of an IgG molecule joins the C_H2 domain with the C_H1 domain. This hinge region encompasses on the order of 25 residues and is flexible,
thereby allowing the two N-terminal antigen binding regions to move independently.

[0068] Besides their configuration, it is known in the art that the constant region mediates several effector functions. For example, binding of the C1 component of complement to antibodies activates the complement system. Activation of complement is important in the opsonisation and lysis of cell pathogens. The activation of complement also stimulates the inflammatory response and may also be involved in autoimmune hypersensitivity. Further, antibodies bind to cells via the Fc region, with a Fc receptor site on the antibody Fc region binding to a Fc receptor (FcR) on a cell. There are a number of Fc receptors which are specific for different classes of antibody, including IgG (gamma receptors), IgE (eta receptors), IgA (alpha receptors) and IgM (mu receptors). Binding of antibody to Fc receptors on cell surfaces triggers a number of important and diverse biological responses including engulfment and destruction of antibody-coated particles, clearance of immune complexes, lysis of antibody-coated target cells by killer cells (called antibody-dependent cell-mediated cytotoxicity, or ADCC), release of inflammatory mediators, placental transfer and control of immunoglobulin production. Although various Fc receptors and receptor sites have been studied to a certain extent, there is still much which is unknown about their location, structure and functioning.

[0069] As discussed above, modification of the constant region by some of the methods described herein allows the disclosed modified antibodies to spontaneously assemble or associate into stable antibodies. Moreover, while not limiting the scope of the present invention, it is believed that antibodies comprising constant regions modified as described herein provide for altered effector functions that, in turn, affect the biological profile of the administered antibody. For example, the deletion or inactivation (through point mutations or other means) of a constant region domain may reduce Fc receptor binding of the circulating modified antibody thereby increasing tumor localization. In other cases it may be that constant region modifications consistent with the instant invention moderate compliment binding and thus reduce the serum half life and nonspecific association of a conjugated cytotoxin. Yet other modifications of the
constant region may be used to eliminate disulfide linkages or oligosaccharide moities that allow for enhanced localization due to increased antigen specificity or antibody flexibility. More generally, those skilled in the art will realize that antibodies modified as described herein may exert a number of subtle effects that may or may not be readily appreciated. However the resulting physiological profile, bioavailability and other biochemical effects of the modifications, such as tumor localization and serum half-life, may easily be measured and quantified using well known immunological techniques without undue experimentation.

[0070] Similarly, modifications to the constant region in accordance with the instant invention may easily be made using well known biochemical or molecular engineering techniques well within the purview of the skilled artisan. In this respect the examples appended hereto provide various constructs having constant regions modified in accordance with the present invention. More specifically, the exemplified constructs comprise chimeric and humanized antibodies having human constant regions that have been engineered to delete the C_{H2} domain. Those skilled in the art will appreciate that such constructs are particularly preferred due to the regulatory properties of the C_{H2} domain on the catabolic rate of the antibody.

[0071] ΔC_{H2} domain deleted antibodies set forth in PCT/US02/02373 and PCT/US02/02374, both filed on January 29, 2002 and incorporated by reference in its entirety herein, are derived from the chimeric C2B8 antibody which is immunospecific for the CD20 pan B cell antigen and a humanized CC49 antibody which is specific for the TAG 72 antigen. As discussed in more detail below, both domain deleted constructs were derived from a proprietary vector (IDEC Pharmaceuticals, San Diego) encoding an IgG1 human constant domain. Essentially, the vector was engineered to delete the C_{H2} domain and provide a modified vector expressing a domain deleted IgG1 constant region. Genes encoding the murine variable region of the C2B8 antibody or the variable region of the humanized CC49 antibody were then inserted in the modified vector and cloned. When expressed in transformed cells, these vectors provided huCC49.ΔC_{H2} or C2B8.ΔC_{H2} respectively. As illustrated below, these
constructs exhibited a number of properties that make them particularly attractive candidates for monomeric subunits.

[0072] It will be noted that the foregoing exemplary constructs were engineered to fuse the C_{H}3 domain directly to the hinge region of the respective modified antibodies. In other constructs it may be desirable to provide a peptide spacer between the hinge region and the modified C_{H}2 and/or C_{H}3 domains. For example, compatible constructs could be expressed wherein the C_{H}2 domain has been deleted and the remaining C_{H}3 domain (modified or unmodified) is joined to the hinge region with a 5 – 20 amino acid spacer. Such a spacer may be added, for instance, to ensure that the regulatory elements of the constant domain remain free and accessible or that the hinge region remains flexible. However, it should be noted that amino acid spacers may, in some cases, prove to be immunogenic or inhibit the desired dimerization of the monomeric subunits. For example, a domain deleted CC49 construct having a short amino acid spacer GGSSGGGGSG (Seq. ID No. 1) substituted for the C_{H}2 domain (CC49.ΔC_{H}2 [gly/ser]) is used as a control in the examples because it does not assemble spontaneously into a dimeric form. Accordingly, any spacer compatible with the instant invention will be relatively non-immunogenic and not inhibit the non-covalent association of the modified antibodies.

[0073] Besides the deletion of whole constant region domains, it will be appreciated that polycistronic antibody constructs of the present invention may be provided by the partial deletion or substitution of a few or even a single amino acid as long as it permits the desired non-covalent association between the monomeric subunits. For example, the mutation of a single amino acid in selected areas of the C_{H}2 domain may be enough to substantially reduce Fc binding and thereby increase tumor localization. Similarly, it may be desirable to simply delete that part of one or more constant region domains that control the effector function (e.g. complement CLQ binding) to be modulated. Such partial deletions of the constant regions may improve selected characteristics of the antibody (serum half-life) while leaving other desirable functions associated with the subject constant region domain intact. Moreover, as alluded to above, the constant regions of the disclosed antibodies may be modified through the mutation or substitution of one
or more amino acids that enhances the profile of the resulting construct. In this respect it may be possible to disrupt the activity provided by a conserved binding site (e.g. Fc binding) while substantially maintaining the configuration and immunogenic profile of the modified antibody. Yet other preferred embodiments may comprise the addition of one or more amino acids to the constant region to enhance desirable characteristics such as effector function or provide for more cytotoxic or carbohydrate attachment. In such embodiments it may be desirable to insert or replicate specific sequences derived from selected constant region domains.

[0074] Following manipulation of the isolated genetic material to provide antibodies and modified antibodies genes as set forth above, the genes are then inserted in a polycistronic expression vector according to the invention for introduction into host cells that may be used to produce the desired quantity of antibody. Constructive of such constructs is described in detail infra.

[0075] The term “vector” or “expression vector” is used herein for the purposes of the specification and claims, to mean vectors used in accordance with the present invention as a vehicle for introducing into and expressing a desired gene in a cell. As known to those skilled in the art, such vectors may easily be selected from the group consisting of plasmids, phages, viruses and retroviruses. In general, vectors compatible with the instant invention will comprise a selection marker, appropriate restriction sites to facilitate cloning of the desired gene and the ability to enter and/or replicate in eukaryotic or prokaryotic cells.

[0076] For the purposes of this invention, numerous polycistronic expression vector systems may be employed. For example, polycistronic vector may contain DNA elements which are derived from animal viruses such as bovine papillomavirus virus, polyoma virus, adenovirus, vaccinia virus, baculovirus, retroviruses (RSV, MMTV or MOMLV) or SV40 virus. Additionally, cells which have integrated the polycistronic construct DNA into their chromosomes may be selected by introducing one or more markers which allow selection of transfected host cells. The marker may provide for prototrophy to an auxotrophic host, biocide resistance (e.g., antibiotics) or resistance to heavy metals such as copper. The selectable marker gene can either be directly
linked to the DNA sequences to be expressed, or introduced into the same cell by cotransformation. Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcriptional promoters, enhancers, and termination signals.

[0077] In preferred embodiments the instant invention will express antibodies, including modified antibodies, using the subject novel polycistronic expression systems. In these novel expression systems, multiple gene products of interest such as heavy and light chains of antibodies may be produced from a single polycistronic construct. These systems advantageously use an internal ribosome entry site (IRES) to provide relatively high levels of modified antibodies in eukaryotic host cells. Compatible IRES sequences are disclosed in U.S.P.N. 6,193,980 which is also incorporated herein in its entirety. Those skilled in the art will appreciate that such expression systems may be used to effectively produce the full range of modified antibodies disclosed in the instant application.

[0078] More generally, once the vector or DNA sequence encoding the monomeric subunit (e.g., modified antibody) has been prepared, the expression vector may be introduced into an appropriate host cell. That is, the host cells may be transformed. Introduction of the plasmid into the host cell can be accomplished by various techniques well known to those of skill in the art. These include, but are not limited to, transfection (including electrophoresis and electroporation), protoplast fusion, calcium phosphate precipitation, cell fusion with enveloped DNA, microinjection, and infection with intact virus. See, Ridgway, A. A. G. "Mammalian Expression Vectors" Chapter 24.2, pp. 470-472 Vectors, Rodriguez and Denhardt, Eds. (Butterworths, Boston, Mass. 1988). Most preferably, plasmid introduction into the host is via electroporation. The transformed cells are grown under conditions appropriate to the production of the light chains and heavy chains, and assayed for heavy and/or light chain protein synthesis. Exemplary assay techniques for identifying and quantifying gene products of interest include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), or fluorescence-activated cell sorter analysis (FACS), immunohistochemistry and the like.
[0079] As used herein, the term "transformation" shall be used in a broad sense to refer to any introduction of DNA into a recipient host cell that changes the genotype and consequently results in a change in the recipient cell.

[0080] As used herein, "host cells" refers to cells that have been transformed with vectors constructed using recombinant DNA techniques and encoding at least one heterologous gene. As defined herein, antibodies or modifications thereof produced by a host cell that is, by virtue of this transformation, recombinant. In descriptions of processes for isolation of antibodies from recombinant hosts, the terms "cell" and "cell culture" are used interchangeably to denote the source of antibody unless it is clearly specified otherwise. In other words, recovery of antibody from the "cells" may mean from spun down whole cells, or from the cell culture containing both the medium and the suspended cells.

[0081] The host cell line used for protein expression is most preferably of mammalian origin; those skilled in the art can readily determine particular host cell lines which are best suited for expression of the desired gene product. Exemplary host cell lines include, but are not limited to, DG44 and DUXB11 (Chinese Hamster Ovary lines, DHFR minus), HEla (human cervical carcinoma), CVI (monkey kidney line), COS (a derivative of CVI with SV40 T antigen), R1610 (Chinese hamster fibroblast) BALBC/3T3 (mouse fibroblast), HAK (hamster kidney line), SP2/O (mouse myeloma), P3.times.63-Ag3.653 (mouse myeloma), BFA-1c1BPT (bovine endothelial cells), RAJI (human lymphocyte) and 293 (human kidney). CHO cells are particularly preferred. Host cell lines are typically available from commercial services, the American Tissue Culture Collection or from published literature.

[0082] In vitro production allows scale-up to give large amounts of the desired polypeptide produced using the polycistronic expression system, preferably an antibody. Techniques for eukaryotic, e.g., mammalian and yeast cell cultivation under tissue culture conditions are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilized or entrapped cell culture, e.g. in hollow fibers, microcapsules, on agarose microbeads or ceramic cartridges. For isolation and recovery of the
antibodies, the immunoglobulins in the culture supernatants may first be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as PEG, filtration through selective membranes, or the like. If necessary and/or desired, the concentrated solutions of multivalent antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose or (immuno-)affinity chromatography.

[0083] Disclosed herein is a novel expression system for producing multiple gene products of interest from a single polycistronic construct. Unlike previously known polycistronic constructs, the inventive expression system produces sufficient levels of desired genes in both the first and subsequent cistrons to be commercially useful. This is surprising, as heretofore, gene sequences located after the first cistron in a polycistronic expression system have been expressed at very low levels in comparison with gene sequences expressed in the first cistron. Accordingly, previous polycistronic expression systems were generally limited to expressing a marker sequence in a second cistron, not a gene sequence of interest. Alternatively, the second cistronic gene was inefficiently expressed, thereby precluding the production of detectable translation products from genes contained in each cistron. By contrast, the polycistronic construct according to the invention may contain two, three or more cistrons, each encoding a gene of interest, if so desired.

[0084] The present invention, in its preferred embodiment, expresses antibodies, which may be modified as described herein, and further includes dimeric antibodies, using a polycistronic expression system, wherein two or more antibody genes are expressed off the same eukaryotic promoter, wherein such antibody genes are separated by one or more IRES's. Particularly, as discussed previously, the invention includes the expression of domain-deleted antibodies and other modified antibodies as described in PCT/US02/02373 and PCT/US02/02374, each filed on January 29, 2002, and incorporated by reference in its entirety herein. The eukaryotic cells used for expression will preferably be mammalian or yeast cells, most preferably CHO cells and other cells that can be efficiently cultured for high level protein production. As noted
above, the obtaining or cloning of antibody heavy and light chain genes for incorporation into polycistronic expression systems according to the invention is well within the purview of ordinary skill. As noted, such antibody genes may encode mature heavy or light chain antibody genes, e.g., murine, rabbit, human, hamster, etc., or these heavy and/or light chain genes may be modified, e.g., by chimerization, humanization, domain deletion or site-specific mutagenesis.

[0085] The invention further contemplates the expression of any heavy chain and light chain sequence which when expressed using the polycistronic expression system of the invention, associate to produce a functional antibody, i.e., one that specifically binds to a target antigen, e.g., a tumor associated antigen.

[0086] As discussed, the copy number of the heavy and light chain genes in the polycistronic construct may be selected such that the preferred ratio of light/heavy chain are obtained, e.g., the light chain is expressed at levels which typically range from 10/1, more preferably 5/1, and still more preferably about 3/1 to 1/1 relative to the heavy chain.

[0087] In the preferred embodiment, the expression system produces antibodies in eukaryotic cells, preferably mammalian cells such as Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, fibroblast cell lines and myeloma cells. For example, CHO cells are employed as hosts for an expression system comprising a polycistron comprising, in the 5' to 3' orientation, at least the following sequences: a eukaryotic promoter sequence that is functional in the particular eukaryotic cell used for expression such as CMV, SV40 early or actin promoter sequences, preferably CMV; a DNA sequence encoding an antibody light chain and, preferably at its 5' end, a eukaryotic secretory leader sequence; an internal ribosome entry site (IRES), preferably that of a cardiovirus, poliovirus or herpes virus, positioned to follow the antibody light chain sequence; at least one DNA sequence encoding an antibody heavy chain, each heavy chain sequence preferably being preceded by a eukaryotic secreting leader sequence, and flanked by a 5' start and a 3' stop codon, wherein some or all DNA's encoding an antibody heavy chain are separated from subsequent heavy chain sequences by an IRES, and wherein
the ultimate antibody heavy chain coding sequence comprises a poly A sequence at its 3’ terminus. A suitable locale of the IRES site between heavy and light chain sequences is exemplified within the NEOSPLA vector of Figure 2.

[0088] As noted herein, the eukaryotic cell preferably comprises a mammalian cell and more preferably a CHO cell. In a preferred embodiment, the promoter is the CMV promoter, and the IRES is derived from a cardiovirus such as Encephalomyocarditis virus, Mongo virus, Mous-Elberfiell virus, MM virus, and Columbia SK virus, most preferably human encephalomyocarditis virus (hEMCV).

[0089] The inventive polycistron preferably comprises an antibody light chain encoding sequence and one or two antibody heavy chain coding sequences. However, polycistron constructs according to the invention may include 3 or 4 gene sequences or cistrons, for example, one light chain and two or more heavy chains, are contemplated. Additionally, the subject polycistron will preferably comprise a poly A sequence, preferably that of the bovine growth hormone (bGH) gene. The polycistron system optionally may be used in loci targeting by homologous recombination.

[0090] In this embodiment, the polycistronic construct will comprise sequences that facilitate homologous recombination at a targeted site, e.g., gene that is inactivated after recombination.

[0091] In a preferred embodiment the inventive polycistron comprises one or two copies of the heavy chain coding sequence dependent upon the stoichiometry of expression of the particular antibody heavy and light chain genes. In this regard, it is well known that in polycistronic expression systems, the second gene is typically expressed at lesser efficiency than the first gene. Accordingly, in one embodiment, the inventive polycistron, in which the first cistron encodes an antibody light chain, may encompass a second cistron encoding two or more heavy chain coding sequences, if deemed necessary, to facilitate sufficient expression of the heavy chain relative to the light chain. In general, it is preferred that the heavy chain be expressed at levels which are at least equivalent to levels observed with non-polycistronic co-expression of the
heavy and light chains when expressed in the same eukaryotic cell using a non-polycistronic expression system.

[0092] It is permissible, and in fact desirable, that more of the antibody light chain is expressed relative to the antibody heavy chain, as this is analogous to what occurs in endogenous antibody producing cells. Disparate expression levels exist because the light chain is instrumental in directing the appropriate assembly of the antibody heavy and light chains, and excessive unpaired heavy chain is thought to induce cell toxicity. The light chain is also critical in directing folding of the assembled antibody heavy and light chains to produce a functional (antigen-binding) antibody in the endoplasmic reticulum. Preferably, the antibody light chain will be expressed from about 10/1 to 1/1 relative to the antibody heavy chain.

[0093] However, levels of the heavy chain must not be de minimus, and should be present in sufficient ratios with respect to light chains to enable the generation of functional, secretable antibodies in commercially acceptable levels. Thus, it is undesirable for the heavy chain expression to be too low relative to the light chain, as underexpression results in inadequate yields of functional antibodies. For purposes of industrial utility, inadequate yields of functional antibodies render an expression system commercially non-viable, and makes the recovery of complete antibody molecules from batch cultures difficult to achieve. Generally, functional antibody is recovered from cultured cells at an amount ranging from about 5-100 picograms per cell, per day, however greater levels of expression may be achieved. For example, cultured cells may secrete at least 1-5 picograms of functional antibody per cell each day, or at least 3-10 mg/L for at least 3-4 days.

[0094] With respect to the above, it is generally unpredictable whether a given polycistronic expression system will result in adequate levels of antibody production relative to other expression systems. This unpredictability arises because, in some instances, the second desired gene in the polycistronic complex may be expressed at very low levels relative to the first gene. Therefore, preferred embodiments of polycistronic vectors should provide a ratio of antibody light chain expression to antibody heavy chain expression within the
range of about 10:1 to about 1:1. More preferably, the ratio of light chain to heavy chain gene expression is from about 3:1 to about 1.5:1.

[0095] Initial IRES constructs were created to contain an antibody light chain sequence in the first cistron, followed by two IRES-antibody CH2 domain deleted heavy chain sequence pairings, thereby ensuring sufficient heavy chain protein production to enable suitable levels of antibody to be produced and secreted from host cells.

[0096] The inventive polycistronic vectors enable the requisite levels of heavy and light chain expression to be achieved by selection of appropriate heavy chain antibody sequences, by selection of an efficient IRES, such as that of hEMCV, or by the incorporation of multiple copies of the antibody heavy chain genes. Still further, the DNA corresponding to the 5' end of the heavy chain gene may be modified by site specific mutagenesis in a manner whereby the coding structure remains unaltered around the ATG codon, typically the first 10 codons, but which modification results in altered expression of the heavy chain coding sequence relative to an unmodified heavy chain gene.

[0097] The heavy chain yield using the subject polycistronic expression system will typically be less than the light chain yield, as is the typical expression relationship in an endogenous antibody producing cell. The light chain yield to heavy chain yield ratio will be sufficient to enable protein secretion and folding. The ratio of the light chain to heavy chain expression may be varied by, for example, increasing the number of IRES-linked downstream gene sequences following the light chain sequence of the first cistron. A particular IRES and expression cell combination may be selected to optimally increase the amount of second cistron expression in a system.

[0098] An antibody that is expressed according to the subject expression system may be specific to any desired antigen. Preferably, the antibody will be a functional antibody that elicits a therapeutic effect, such as an antibody useful for treating an autoimmune, inflammatory, infectious, allergic or neoplastic disease. The antibody may be combined with other therapeutic agents for synergistic effects. For example, the antibody may be combined with a radioactive source for use as a cancer chemotherapy agent.
[0100] In general, antibodies expressed according to the present invention may be used in any one of a number of conjugated (i.e. an immunoconjugate) or unconjugated forms. In particular, the antibodies of the present invention may be conjugated to cytotoxins such as radioisotopes, therapeutic agents, cytostatic agents, biological toxins or prodrugs. Alternatively, the antibodies of the instant invention may be used in a nonconjugated or naked form to harness a subject's natural defense mechanisms to eliminate malignant cells. In particularly preferred embodiments, the antibodies produced according to the expression system of the present invention may be modified, such as by conjugation to radioisotopes. Examples of radioisotopes useful according to the invention include $^{90}$Y, $^{125}$I, $^{131}$I, $^{123}$I, $^{111}$In, $^{105}$Rh, $^{153}$Sm, $^{67}$Cu, $^{67}$Ga, $^{166}$Ho, $^{177}$Lu, $^{186}$Re and $^{188}$Re, using anyone of a number of well known chelators or direct labeling. Conjugated and unconjugated antibodies may be used together in the same therapeutic regimen, e.g., as used in the currently approved therapeutic regimen employing Zevalin for the treatment of certain non-Hodgkin's lymphomas.

[0101] In other embodiments, the antibodies of the invention may be included in compositions that comprise modified antibodies coupled to drugs, prodrugs or biological response modifiers such as methotrexate, adriamycin, and lymphokines such as interferon. Still other embodiments of the present invention comprise the use of modified antibodies conjugated to specific biotoxins such as ricin or diphereria toxin. In yet other embodiments the modified antibodies may be complexed with other immunologically active ligands (e.g. antibodies or fragments thereof) wherein the resulting molecule binds to both a neoplastic cell and an effector cell such as a T cell. The selection of which conjugated and/or unconjugated modified antibody to use will depend of the type and stage of cancer, use of adjunct treatment (e.g., chemotherapy or external radiation) and patient condition. It will be appreciated that one skilled in the art could readily make such a selection in view of the teachings herein.

[0102] As used herein, "a cytotoxin or cytotoxic agent" means any agent that is detrimental to the growth and/or proliferation of cells and which may act to reduce, inhibit or destroy a cell or malignancy when exposed thereto.
Exemplary cytotoxins include, but are not limited to, radionuclides, biotoxins, enzymatically active toxins, cytostatic or cytotoxic therapeutic agents, prodrugs, immunologically active ligands and biological response modifiers such as cytokines. As will be discussed in more detail below, radionuclide cytotoxins are particularly preferred for use in the instant invention. However, any cytotoxin that acts to retard or slow the growth of immunoreactive cells or malignant cells or to eliminate these cells and which may be associated with the polycistronic derived functional antibodies disclosed herein is within the scope of the present invention.

[0103] It will be appreciated that, in previous studies, anti-tumor antibodies labeled with the above-noted isotopes have been used successfully to destroy cells in solid tumors as well as lymphomas/leukemias in animal models, and in humans. The radionuclides act by producing ionizing radiation which causes multiple strand breaks in nuclear DNA, leading to cell death. The isotopes used to produce therapeutic conjugates typically produce high energy α- or β-particles which have a short path length. Such radionuclides kill cells to which they are in close proximity, for example neoplastic cells to which the conjugate has attached or has entered. They have little or no effect on non-localized cells. Radionuclides are essentially non-immunogenic.

[0104] With respect to the use of radiolabeled conjugates in conjunction with the present invention, the modified antibodies may be directly labeled (such as through iodination) or may be labeled indirectly through the use of a chelating agent. As used herein, the phrases "indirect labeling" and "indirect labeling approach" both mean that a chelating agent is covalently attached to an antibody and at least one radionuclide is associated with the chelating agent. Such chelating agents are typically referred to as bifunctional chelating agents as they bind both the polypeptide and the radioisotope. Particularly preferred chelating agents include 1-isothiocyanatobenzyl-3-methylthiothylene triaminepentaacetic acid ("MX-DTPA") and cyclohexyl diethylenetriamine pentaacetic acid ("CHX-DTPA") derivatives. Other chelating agents comprise P-DOTA and EDTA derivatives. Particularly suitable radionuclides for indirect labeling include ¹¹¹In and ⁹⁰Y.
As used herein, the phrases "direct labeling" and "direct labeling approach" both mean that a radionuclide is covalently attached directly to a antibody (typically via an amino acid residue). More specifically, these linking technologies include random labeling and site-directed labeling. In the latter case, the labeling is directed at specific sites on the antibody, such as the N-linked sugar residues present only on the Fc portion of the conjugates. Direct labeling may result in multi-meric antibodies linked by the labels. Various direct labeling techniques and protocols are compatible with the instant invention. For example, Technetium-99m labelled tetravalent antibodies may be prepared by ligand exchange processes, by reducing pertechnetate (TcO₄⁻) with stannous ion solution, chelating the reduced technetium onto a Sephadex column and applying the antibodies to this column, or by batch labeling techniques, e.g. by incubating pertechnetate, a reducing agent such as SnCl₂, a buffer solution such as a sodium-potassium phthalate-solution, and the antibodies. In any event, preferred radionuclides for directly labeling antibodies are well known in the art and a particularly preferred radionuclide for direct labeling is ¹³¹I which can be covalently attached via tyrosine residues. Modified antibodies according to the invention may be derived, for example, with radioactive sodium or potassium iodide and a chemical oxidizing agent, such as sodium hypochlorite, chloramine T or the like, or an enzymatic oxidizing agent, such as lactoperoxidase, glucose oxidase and glucose. However, for the purposes of the present invention, the indirect labeling approach is generally favored.

Patents relating to chelators and chelator conjugates are known in the art. For instance, U.S. Patent No. 4,831,175 of Gansow is directed to polysubstituted diethyleneetriaminepentaacetic acid chelates and protein conjugates containing the same, and methods for their preparation. U.S. Patent Nos. 5,099,069, 5,246,692, 5,286,850, 5,434,287 and 5,124,471 of Gansow also relate to polysubstituted DTPA chelates. These patents are incorporated herein in their entirety. Other examples of compatible metal chelators are ethylenediaminetetraacetic acid (EDTA), diethyleneetriaminepentaacetic acid (DPTA), 1,4,8,11-tetraazatetradecane, 1,4,8,11-tetraazatetradecane-1,4,8,11-tetraacetic acid, 1-oxa-4,7,12,15-tetraazaheptadecane-4,7,12,15-tetraacetic
acid, or the like. Cyclohexyl-DTPA or CHX-DTPA is particularly preferred and is exemplified extensively below. Still other compatible chelators, including those yet to be discovered, may easily be discerned by a skilled artisan and are clearly within the scope of the present invention.

[0107] Compatible chelators, including the specific bifunctional chelator used to facilitate chelation in co-pending application Serial Nos. 08/475,813, 08/475,815 and 08/478,967, are preferably selected to provide high affinity for trivalent metals, exhibit increased tumor-to-non-tumor ratios and decreased bone uptake as well as greater in vivo retention of radionuclide at target sites, i.e., B-cell lymphoma tumor sites. However, other bifunctional chelators that may or may not possess all of these characteristics are known in the art and may also be beneficial in tumor therapy.

[0108] It will also be appreciated that, in accordance with the teachings herein, modified antibodies may be conjugated to different radiolabels for diagnostic and therapeutic purposes. Radiolabeled therapeutic conjugates for diagnostic "imaging" of tumors before administration of therapeutic antibody may be prepared. "In2B8" conjugate comprises a murine monoclonal antibody, 2B8, (rituximab) specific to human CD20 antigen, that is attached to $^{111}$In via a bifunctional chelator, i.e., MX-DTPA, which comprises a 1:1 mixture of 1-isothiocyanato(benzyl-3-methyl-DTPA and 1-methyl-3-isothiocyanatobenzyl-DTPA. $^{111}$In is particularly preferred as a diagnostic radionuclide because between about 1 to about 10 mCi can be safely administered without detectable toxicity; and the imaging data is generally predictive of subsequent $^{90}$Y-labeled antibody distribution. Most imaging studies utilize about 5 mCi $^{111}$In-labeled antibody, because this dose is both safe and has increased imaging efficiency compared with lower doses, with optimal imaging occurring at three to six days after antibody administration. See, for example, Murray, *J. Nuc. Med.* 26: 3328 (1985) and Carraguillo *et al.*, *J. Nuc. Med.* 26: 67 (1985).

[0109] As indicated above, a variety of radionuclides are applicable to the present invention and those skilled in the art are credited with the ability to readily determine which radionuclide is most appropriate under various circumstances. For example, $^{131}$I is a well known radionuclide used for targeted immunotherapy.
However, the clinical usefulness of $^{131}$I can be limited by several factors including: eight-day physical half-life; dehalogenation of iodinated antibody both in the blood and at tumor sites; and emission characteristics (e.g., large gamma component) which can be suboptimal for localized dose deposition in a tumor. With the advent of superior chelating agents, the opportunity for attaching metal chelating groups to proteins has increased the opportunities to utilize other radionuclides such as $^{111}$In and $^{90}$Y. $^{90}$Y provides several benefits for utilization in radioimmunotherapeutic applications: the 64 hour half-life of $^{90}$Y is long enough to allow antibody accumulation by tumor and, unlike e.g., $^{131}$I, $^{90}$Y is a pure beta emitter of high energy with no accompanying gamma irradiation in its decay, with a range in tissue of 100 to 1,000 cell diameters. Furthermore, the minimal amount of penetrating radiation allows for outpatient administration of $^{90}$Y-labeled antibodies. Additionally, internalization of labeled antibody is not required for cell killing, and the local emission of ionizing radiation should be lethal for adjacent tumor cells lacking the target antigen.

[0110] Effective single treatment dosages (i.e., therapeutically effective amounts) of $^{90}$Y-labeled modified antibodies range from between about 5 and about 75 mCi, more preferably between about 10 and about 40 mCi. Effective single treatment non-marrow ablative dosages of $^{131}$I-labeled antibodies range from between about 5 and about 70 mCi, more preferably between about 5 and about 40 mCi. Effective single treatment ablative dosages (i.e., may require autologous bone marrow transplantation) of $^{131}$I-labeled antibodies range from between about 30 and about 600 mCi, more preferably between about 50 and less than about 500 mCi. In conjunction with a chimeric antibody, owing to the longer circulating half life vis-à-vis murine antibodies, an effective single treatment non-marrow ablative dosages of iodine-131 labeled chimeric antibodies range from between about 5 and about 40 mCi, more preferably less than about 30 mCi.

[0112] While a great deal of clinical experience has been gained with $^{131}$I and $^{90}$Y, other radiolabels are known in the art and have been used for similar therapeutic purposes. Still other radioisotopes are used for imaging. For example, additional radioisotopes which are compatible with the scope of the
instant invention include, but are not limited to, $^{123}$I, $^{125}$I, $^{32}$P, $^{57}$Co, $^{64}$Cu, $^{67}$Cu, $^{77}$Br, $^{81}$Rb, $^{81}$Kr, $^{87}$Sr, $^{113}$In, $^{127}$Cs, $^{129}$Cs, $^{132}$I, $^{197}$Hg, $^{203}$Pb, $^{206}$Bi, $^{177}$Lu, $^{186}$Re, $^{212}$Pb, $^{212}$Bi, $^{47}$Sc, $^{105}$Rh, $^{109}$Pd, $^{153}$Sm, $^{188}$Re, $^{199}$Au, $^{225}$Ac, $^{211}$At, and $^{213}$Bi. In this respect alpha, gamma and beta emitters are all compatible with the instant invention. Further, in view of this disclosure it is submitted that one skilled in the art could readily determine which radionuclides are compatible with a selected course of treatment without undue experimentation. To this end, additional radionuclides which have already been used in clinical diagnosis include $^{125}$I, $^{123}$I, $^{99}$Tc, $^{43}$K, $^{52}$Fe, $^{67}$Ga, $^{68}$Ga, as well as $^{111}$In. Antibodies have also been labeled with a variety of radionuclides for potential use in targeted immunotherapy. Peirersz et al. *Immunol. Cell Biol.* 65: 111-125 (1987). These radionuclides include $^{188}$Re and $^{186}$Re as well as $^{199}$Au and $^{67}$Cu to a lesser extent. U.S. Patent No. 5,460,785 provides additional data regarding such radioisotopes and is incorporated herein by reference.

[0113] In addition to radionuclides, the functional antibodies of the present invention may be conjugated to, or associated with, any one of a number of biological response modifiers, pharmaceutical agents, toxins or immunologically active ligands. Those skilled in the art will appreciate that these non-radioactive conjugates may be assembled using a variety of techniques depending on the selected cytotoxin. For example, conjugates with biotin are prepared e.g. by reacting the dimeric antibodies with an activated ester of biotin such as the biotin N-hydroxysuccinimide ester. Similarly, conjugates with a fluorescent marker may be prepared in the presence of a coupling agent, e.g. those listed above, or by reaction with an isothiocyanate, preferably fluorescein-isothiocyanate. Conjugates of the tetravalent antibodies of the invention with cytostatic/cytotoxic substances and metal chelates are prepared in an analogous manner.

[0114] Preferred agents for use in the present invention are cytotoxic drugs, particularly those which are used for cancer therapy. Such drugs include, in general, cytostatic agents, alkylating agents, antimetabolites, anti-proliferative agents, tubulin binding agents, hormones and hormone antagonists, and the like. Exemplary cytostatics that are compatible with the present invention
include alkylating substances, such as mechloretamine, triethylene phosphoramid, cyclophosphamide, ifosfamide, chlorambucil, busulfan, melphalan or triaziquone, also nitrosourea compounds, such as carmustine, lomustine, or semustine. Other preferred classes of cytotoxic agents include, for example, the anthracycline family of drugs, the vinca alkaloid family of drugs, the mitomycins, the bleomycins, the cytotoxic nucleosides, the pteridine family of drugs, diynes, and the podophyllotoxins. Particularly useful members of those classes include, for example, adriamycin, carminomycin, daunorubicin (daunomycin), doxorubicin, aminopterin, methotrexate, methotrexate, mithramycin, streptonigrin, dichloromethotrexate, mitomycin C, actinomycin-D, porfiromycin, 5-fluorouracil, flixuridine, fltoral, fluracil, mercaptopurine, cytarabine, cytosine arabinoside, podophyllotoxin, or podophyllotoxin derivatives such as etoposide or etoposide phosphate, melphalan, vinblastine, vincristine, leurosidine, vindesine, leurosine and the like. Still other cytotoxins that are compatible with the teachings herein include taxol, taxane, cytochalasin B, gramicidin D, ethidium bromide, emetine, tenoposide, colchicin, dihydroxy anthractin dione, mitoxantrone, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Also included as suitable cytotoxins are maytansinoids. Hormones and hormone antagonists, such as corticosteroids, e.g. prednisone, progestins, e.g. hydroxyprogesterone or medroprogesterone, estrogens, e.g. diethylstilbestrol, antiestrogens, e.g. tamoxifen, androgens, e.g. testosterone, and aromatase inhibitors, e.g. aminogluthethimide are also compatible with the teachings herein. As noted previously, one skilled in the art may make chemical modifications to the desired compound in order to make reactions of that compound more convenient for purposes of preparing conjugates of the invention.

[0115] One example of particularly preferred cytotoxins comprise members or derivatives of the enediyne family of anti-tumor antibiotics, including calicheamicin, esperamicins or dynemicins. These toxins are extremely potent and act by cleaving nuclear DNA, leading to cell death. Unlike protein toxins which can be cleaved in vivo to give many inactive but immunogenic polypeptide fragments, toxins such as calicheamicin, esperamicins and other
enediyynes are small molecules which are essentially non-immunogenic. These non-peptide toxins are chemically-linked to the dimers or tetramers by techniques which have been previously used to label monoclonal antibodies and other molecules. These linking technologies include site-specific linkage via the N-linked sugar residues present only on the Fc portion of the constructs. Such site-directed linking methods have the advantage of reducing the possible effects of linkage on the binding properties of the constructs.

[0116] As previously alluded to, compatible cytotoxins may comprise a prodrug. As used herein, the term "prodrug" refers to a precursor or derivative form of a pharmaceutically active substance that is less cytotoxic to tumor cells compared to the parent drug and is capable of being enzymatically activated or converted into the more active parent form. Prodrugs compatible with the invention include, but are not limited to, phosphate-containing prodrugs, thiophosphate-containing prodrugs, sulfate containing prodrugs, peptide containing prodrugs, β-lactam-containing prodrugs, optionally substituted phenoxyacetamide-containing prodrugs or optionally substituted phenylacetamide-containing prodrugs, 5-fluorocytosine and other 5-fluorouridine prodrugs that can be converted to the more active cytotoxic free drug. Further examples of cytotoxic drugs that can be derivatized into a prodrug form for use in the present invention comprise those chemotherapeutic agents described above.

[0117] Among other cytotoxins, it will be appreciated that antibodies can also be associated with a biotoxin such as ricin subunit A, abrin, diphtheria toxin, botulinum, cyanginosins, saxitoxin, shigatoxin, tetanus, tetrodotoxin, trichothecene, verrucologen or a toxic enzyme. Preferably, such constructs will be made using genetic engineering techniques that allow for direct expression of the antibody-toxin construct. Other biological response modifiers that may be associated with the modified antibodies of the present invention comprise cytokines such as lymphokines and interferons. In view of the instant disclosure it is submitted that one skilled in the art could readily form such constructs using conventional techniques.

[0118] Another class of compatible cytotoxins that may be used in conjunction with the disclosed antibodies are radiosensitizing drugs that may be effectively
directed to tumor or immunoreactive cells. Such drugs enhance the sensitivity to ionizing radiation, thereby increasing the efficacy of radiotherapy. An antibody conjugate internalized by the tumor cell would deliver the radiosensitizer nearer the nucleus where radiosensitization would be maximal. The unbound radiosensitizer linked modified antibodies would be cleared quickly from the blood, localizing the remaining radiosensitization agent in the target tumor and providing minimal uptake in normal tissues. After rapid clearance from the blood, adjunct radiotherapy would be administered in one of three ways: 1) external beam radiation directed specifically to the tumor, 2) radioactivity directly implanted in the tumor or 3) systemic radioimmunotherapy with the same targeting antibody. A potentially attractive variation of this approach would be the attachment of a therapeutic radioisotope to the radiosensitized immunoconjugate, thereby providing the convenience of administering to the patient a single drug.

[0119] Whether or not the disclosed functional antibodies are used in a conjugated or unconjugated form, it will be appreciated that a major advantage of the present invention is the ability to use these antibody constructs in myelosuppressed patients, especially those who are undergoing, or have undergone, adjunct therapies such as radiotherapy or chemotherapy. That is, the beneficial delivery profile (i.e. relatively short serum dwell time, high binding affinity and enhanced localization) of the dimeric antibodies makes them particularly useful for treating patients that have reduced red marrow reserves and are sensitive to myelotoxicity. In this regard, the unique delivery profile of the functional antibodies make them very effective for the administration of radiolabeled conjugates to myelosuppressed cancer patients. As such, the modified antibodies are useful in a conjugated or unconjugated form in patients that have previously undergone adjunct therapies such as external beam radiation or chemotherapy.

[0120] The functional antibodies produced according to the invention may bind to a tumor specific or tumor-associated antigen, an antigen expressed on specific cell types such as T cells or B cells, a viral antigen, a bacterial antigen, or to a parasite. Suitable antigens include TAG-72, CD4, CD11, CD19, CD20, CD22, CD23, CD37, CD40, CD45, CD80, CD86 and CD154. In a preferred
embodiment of the invention, the antibody will bind to an antigen expressed by a T cell or B cell, such as CD4, CD19, CD20, CD22, CD23, CD40, CD80, and CD154. In another preferred embodiment, the antibody is directed to a cancer antigen such as CEA, prostate specific antigen, HER-2 (erbB2), a tumor adhesive molecule, etc. The antibody of choice may be a human, humanized or chimeric antibody. In particular, the antibody may be a human, humanized or chimeric antibody specific to CD20 or TAG-72. CD20 is an antigen expressed by B cells that has been targeted for treatment of B cell disorders such as B cell lymphomas and leukemias. Rituxan® is a chimeric anti-CD20 antibody that has been approved by the FDA for treatment of non-Hodgkin’s lymphoma.

[0121] Tag-72 is an antigen that is known to be overexpressed by numerous human cancers including digestive cancers (gastric, colorectal, pancreatic), and reproductive organ associated cancers (prostate, ovarian, breast) as well as other cancers including e.g., head and neck cancers and lymph node metastases. (See e.g., Galietta et al., Oncol. Rep. 9(1): 135-40 (2002); Meredith et al., Cancer Biother. Radiopharm 16(4): 303-15 (2001); Karan et al., Oncol. Rep. 8(5): 1123-26 (2001); Allende et al., Int. J. Biol. Marker 15(2): 1997-99 (2000) and Altimissi et al., HNO 38*10): 364-66 (1998), all incorporated by reference in their entirety). Anti-TAG-72 antibodies have been reported to possess therapeutic efficacy in treating such cancers.

[0122] The DNA sequences encoding the antibody light chain and heavy chain(s) may comprise an intact or modified variable region and constant region. The constant region is preferably human. The variable regions may be of primate origin or of rodent origin, and may be humanized. Primate variable regions may be of human origin. Rodent variable regions may be, for example, of rat or mouse (murine) origin. As noted, domain deleted constant regions are within the scope of the present invention.

[0123] Antigens that are characteristically over-expressed by specific cancer types are well known in the art. Preferred embodiments embrace the expression of antibodies that recognize CD22 or CD20. In the most preferred embodiment, the antibody will specifically bind CD22. Even more preferably, the antibody will comprise heavy and light chains of the antibody rituximab
(RITUXAN®, IDEC Pharmaceuticals, San Diego, CA, USA), as discussed in, for example, U.S. Patent Nos. 5,736,137; 5,776,456; and 5,843,439. Examples of suitable antibodies for use in the invention are disclosed in U.S. Patent No. 6,136,310 to Hanna et al. (CD4); U.S. Patent No. 6,011,138 to Reff (CD23); U.S. Patent No. 6,113,898 to Anderson (CD80); and U.S. Patent No. 6,001,358 to Black (CD54). The entire contents of these patents are hereby incorporated herein by reference.

[0124] Although the invention has been illustrated utilizing antibodies, the invention is suitable for expressing any multichain proteins that would not compromise the viability of their mammalian cell hosts. The ratio of the gene products of the first and successive cistrons may be varied by increasing the number of subunits following the first cistron. Thus, for example, two or more cistrons, each encoding the same desired gene product, may be incorporated into a multicistronic expression vector in order to enhance the absolute amount of peptide produced by the desired gene. However, each successive cistron following the first cistron may encode distinct products, if desired.

[0125] The invention will be further illustrated by the following non-limiting Examples.

EXAMPLE 1

[0126] A polycistronic DNA expression vector was constructed according to the invention, expressing the heavy and light chains of humanized anti-TAG72 antibody, which is denoted as HuCC49.

A. Expression Vector Construction

[0127] The expression construct, depicted in Figure 1, comprises the mouse beta-globulin major promoter (beta), situated within the construct so as to drive the expression of sequences encoding the neomycin phosphotransferase gene. The neomycin phosphotransferase gene, which is of bacterial origin, is composed of two exons, N1 and N2, and contains an artificial intron. Methods of intronic insertion of selectable markers designed to enhance gene product expression in expression vectors systems are disclosed in, for example, U.S.
Patent No. 5,648,267, U.S. Patent No. 5,733,779, U.S. Patent No. 6,017,733, and U.S. Patent No. 6,159,730, all to Ref. The entire contents of each of these patents are hereby incorporated herein by reference. The antibody selected is known in the art for its suitability for treating solid tumors. A high level of tissue penetration is required in order to affect solid tumor cells, and therefore the antibody should have a relatively longer half-life within serum than is seen with Fab fragments. The domain deleted construct used in this example contains the human gamma 1 heavy chain constant domain in which most of the CH2 domain has been deleted, with the exception of the first nine amino acids of CH2 following the hinge region.

[0128] As shown in Figure 1, following the first exon of the neomycin phosphotransferase expression cassette (N1) and within the intron of the neomycin phosphotransferase gene is located a Simian virus SV40 origin of replication (SVO), which facilitates replication of the expression construct (vector) in COS cells following transient transfection. Thereafter is located an expression cassette for the gene sequences of interest, which in this case are immunoglobulin coding regions. First in the cassette is the human cytomegalovirus promoter (CMV), which drives expression of the polycistronic immunoglobulin (Ig) message. The Ig message in this construct encodes the humanized anti-TAG-72 antibody (denoted HuCC49), which is sequentially composed of (in the order presented): the light chain leader (L)/HuCC49 light chain variable (VL)/human kappa constant (Kappa)/IRES from the encephalomyocarditis virus (IRES)/synthetic heavy chain leader (L)/HuCC49 heavy chain variable (VH)/human gamma 1 CH2 domain deleted constant region (CH2 link G1DelCH2). The Ig message is followed by the bovine growth hormone polyadenylation signal (BGH). Thereafter is situated an expression cassette for a dihydrofolate reductase coding region, in which the mouse betaglobin major promoter (Beta) drives expression of the murine dihydrofolate reductase gene (DHFR). The DHFR sequence is followed by a bovine growth hormone polyadenylation signal (BGH).
[0129] Following the neomycin intron is the second exon of the neomycin expression cassette (N2), which, in turn, is followed by the SV40 early polyadenylation signal (SV).

[0130] The vector also contains sequences required for replication in bacteria, including the coIE1 origin of replication and the beta-lactamase gene to confer ampicillin resistance (Amp).

[0131] The EMCV is commercially available, and was obtained from ATCC (VR-129B Encephalomyocarditis strain: EMC (TC adapted)). The viral particles were disrupted and subjected to methods known in the art for isolating single stranded viral RNA. A specific cDNA was generated to the IRES region within the viral genome. PCR (polymerase chain reaction) amplification of that cDNA was performed in order to amplify the DNA, as well as to add 5' and 3' ends suitable for insertion between the light and heavy chain Ig coding domains. No polyadenylation signal was placed after the light chain immediately prior to the IRES. The ATG trinucleotide at position 834-836 (Genbank accession number NC-001479) was used as the start codon for the synthetic heavy chain leader sequence.

[0132] Because the literature indicates that translation of the open reading frame downstream of the IRES would be less efficient than the upstream open reading frame (translation initiated by 5' CAP), restriction sites were introduced such that the IRES and heavy chain sequences could be duplicated within the vector. Mun I and Bgl II sites were introduced downstream of the BamH I site. Following digestion with EcoR I and BamH I, the fragment containing the IRES and heavy chain gene sequence could be easily inserted into the Mun I and Bgl II sites. The resultant construct would then contain: CMV-light chain-IRES-heavy chain-IRES-heavy chain. The addition of the second heavy chain was made in order to compensate for any inefficiency associated with internal ribosome entry.

B. CHO Cell Expression

[0133] A plasmid containing the expression vector expressing one heavy chain utilizing an SV40 enhancer was transfected into CHO cells using both a
transient transfection protocol as well as a protocol designed to yield stable transfected cell lines having high expression capability. These transfections were performed in parallel with conventional expression constructs containing independent light and heavy chain cassettes (CMV-Light-BGH) (CMV-Heavy-BGH). Experimental results indicate that no apparent difference in expression levels arose, meaning no reduction in heavy chain translation as a result of inefficient IRES function was detected. Accordingly, an embodiment of the inventive polycistronic construct is expected to express other gene sequences of interest contained in first and subsequent cistrons in satisfactory amounts as well. Experiments have shown that these constructs express the downstream heavy chain gene in the bicistronic construct at levels equivalent to those obtained using conventional non-polycistronic expression systems, as discussed herein below.

**EXAMPLE 2**

A. Generation of G418 Resistant Cell Lines

[0134] The polycistronic expression construct in this example encodes the modified humanized antibody with specificity to the human TAG 72 protein. This antibody comprised of humanized murine variable domains (both light and heavy), the human kappa light chain constant domain and the human gamma1 heavy chain constant region with a specific deletion of the CH2 domain. This expression construct is referred to as “HuCC49 CH2 Linker in N5KIRESG1DelCH2”, a map of which is depicted in Figure 3, and the sequence of which is set forth in Figure 6. HuCC49 CH2 Linker in N5KIRESG1DelCH2 was transfected into the CHO-DG44 parent cell line. Transfections were performed using 0.5, 1.0, 2.0 or 4.0 µg linearized DNA per 4 x 10^6 CHO cells per 96 well plate. Dominant selection of stably transfected cells was accomplished via the expression construct encoded neomycin resistance marker and media (CHO-S-SFM II + HT) containing G418.

[0135] Viable wells were identified, expanded to small scale (120 ml) spinner culture and cellular productivity assessed by ELISA assay. A variety of G418 resistant cell isolates were obtained producing easily detectable secreted antibody. Four in particular, 22B6, 22E6, 22H10 and 25A2 were determined to
be producing 3-10 mg/L in 3-4 days, or 1-5 picogram/cell/day (pcd). Southern blot analysis, as shown in Figures 4 and 5, was performed and it was observed that each of these four cell lines possess expression construct integrated at a single site within each cell line. The lanes labeled "M" contained size markers. Chinese hamster ovary cell nucleic acid served as a negative control, and was run in the lanes labeled "CHO". Sample 14H9CD was a NEOSPLA construction.

B. ELISA Assay Ascertain ratio of Light to Heavy Chain

[0136] In parallel to generation of these cell lines, a NEOSPLA expression construct of the same antibody was constructed and transfected. HuCC49 CH2 Linker in N5KG1DelCH2, shown in Figure 3, encodes the immunoglobulin light and heavy chain genes on independent transcriptional cassettes, translation of each mediated by 5' CAP translational initiation complexes. A particular G418 resistant cell line isolated by transfection of this expression construct was termed 14H9. Analysis of culture supernatants from 14H9 and the polycistronic isolate 22B6 was performed to examine the relative ratios of secreted light chain and heavy chain. Analysis was performed by ELISA by capturing and detecting with either light or heavy chain specific secondary antibodies. Results of duplicate samples are displayed in Table 1, the data being expressed as mg/L compared to an immunoglobulin standard. Total light chain measured (κκ) is compared to total assembled antibody measured (γκ) in the ratio column.

<table>
<thead>
<tr>
<th>Construct Type</th>
<th>Cell line</th>
<th>κκ</th>
<th>γγ</th>
<th>κκ/γκ</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLYCISTRONIC</td>
<td>5F7</td>
<td>20.6</td>
<td>2.2</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>5F7</td>
<td>8.4</td>
<td>0.97</td>
<td>2.5</td>
</tr>
<tr>
<td>NEOSPLA</td>
<td>14H9</td>
<td>79.5</td>
<td>6.9</td>
<td>33.6</td>
</tr>
<tr>
<td></td>
<td>14H9</td>
<td>56.0</td>
<td>6.7</td>
<td>23.9</td>
</tr>
</tbody>
</table>

[0137] The cell line 14H9 is an unusually high producing G418 cell line as a result of a high gene copy number. Of interest here are the roughly equivalent relative ratios of free or assembled immunoglobulin light and heavy chains. This data demonstrates that, surprisingly, expression of immunoglobulin heavy chain
within the Polycistronic expression system, placed after the IRES, is not significantly reduced as compared to the NEOSPLA system. The observation that the $\kappa_k/\gamma_k$ ratios are greater than one is not unexpected and are likely a result of permissive cellular secretion of free light chain. Therefore, in this Polycistronic expression system, IRES driven heavy chain protein production appears to be very efficient.

C. Further ELISA Assays Comparing Expressed Ratio of Light to Heavy Chain

[0138] Analysis of culture supernatants from a variety of cell lines generated by either the NEOSPLA or Polycistronic expression systems was later performed to examine the relative ratios of secreted light chain and heavy chain. Analysis was performed by ELISA by capturing and detecting with either light or heavy chain specific secondary antibodies. Results of duplicate samples are displayed here and are expressed as mg/L compared to an immunoglobulin standard.

[0139] The ELISA is performed by coating with a 96 well micro-titre plate with a "capture" antibody, followed by binding of cell culture supernatant, followed by "detection" via binding of a secondary antibody conjugated to HRPO (horseradish peroxidase), followed by colorometric quantification.

$\kappa_k$ indicates capture by goat anti-human kappa : detection by goat anti-human kappa-HRPO

$\gamma\gamma$ indicates capture by goat anti-human IgG : detection by goat anti-human IgG-HRPO

$\gamma_k$ indicates capture by goat anti-human IgG : detection by goat anti-human kappa-HRPO

$\gamma\gamma$ indicates capture by goat anti-human kappa : detection by goat anti-human IgG-HRPO

![TABLE 2]

<table>
<thead>
<tr>
<th></th>
<th>$\kappa_k$</th>
<th>$\gamma\gamma$</th>
<th>$\gamma_k$</th>
<th>$\gamma\gamma$</th>
<th>$\kappa_k/\gamma\gamma$</th>
<th>$\kappa_k/\gamma_k$</th>
<th>ANTI-CD23</th>
</tr>
</thead>
<tbody>
<tr>
<td>8B9</td>
<td>4.4</td>
<td>4.3</td>
<td>3.6</td>
<td>3.6</td>
<td>1.2</td>
<td>1.0</td>
<td>NEOSPLA</td>
</tr>
<tr>
<td>8B9-5A12</td>
<td>10.7</td>
<td>5.0</td>
<td>6.0</td>
<td>7.2</td>
<td>1.8</td>
<td>2.1</td>
<td>NEOSPLA</td>
</tr>
<tr>
<td>8B9-5A12-50F1</td>
<td>55.9</td>
<td>35.9</td>
<td>43.3</td>
<td>50.4</td>
<td>1.3</td>
<td>1.6</td>
<td>NEOSPLA</td>
</tr>
<tr>
<td>8B9-5A12-50F1-500G8</td>
<td>97.4</td>
<td>69.5</td>
<td>77.8</td>
<td>93.5</td>
<td>1.3</td>
<td>1.4</td>
<td>NEOSPLA</td>
</tr>
<tr>
<td>HUCC49 NO LINKER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12E9</td>
<td>4.6</td>
<td>4.1</td>
<td>3.5</td>
<td>3.3</td>
<td>1.3</td>
<td>1.1</td>
<td>NEOSPLA</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>12E9-5E8</td>
<td>21.6</td>
<td>11.5</td>
<td>11.5</td>
<td>10.1</td>
<td>1.9</td>
<td>1.9</td>
<td>NEOSPLA</td>
</tr>
<tr>
<td>12E9-5E8-50C9</td>
<td>68.3</td>
<td>25.9</td>
<td>27.7</td>
<td>22.9</td>
<td>2.5</td>
<td>2.6</td>
<td>NEOSPLA</td>
</tr>
</tbody>
</table>

**ANTI-CD23**

| 5F8      | 4.2  | 3.3  | 3.1  | 3.0  | 1.4  | 1.3  | POLYCISTRONIC |

**HUCC49 GLY/SER LINKER**

<table>
<thead>
<tr>
<th>3C11</th>
<th>4.3</th>
<th>3.9</th>
<th>3.6</th>
<th>2.5</th>
<th>1.2</th>
<th>1.1</th>
<th>POLYCISTRONIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3C11-5B12</td>
<td>21.0</td>
<td>10.6</td>
<td>11.1</td>
<td>9.2</td>
<td>1.9</td>
<td>2.0</td>
<td>POLYCISTRONIC</td>
</tr>
<tr>
<td>3C11-5B12-50B5</td>
<td>98.3</td>
<td>49.1</td>
<td>55.7</td>
<td>46.4</td>
<td>1.8</td>
<td>2.0</td>
<td>POLYCISTRONIC</td>
</tr>
</tbody>
</table>

**HUCC49 CH2 LINKER**

<table>
<thead>
<tr>
<th>25A2-50A5</th>
<th>21.8</th>
<th>12.4</th>
<th>12.3</th>
<th>10.6</th>
<th>1.8</th>
<th>1.8</th>
<th>POLYCISTRONIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>25A2-5B3-50A5</td>
<td>96.9</td>
<td>45.0</td>
<td>46.5</td>
<td>45.5</td>
<td>2.1</td>
<td>2.2</td>
<td>POLYCISTRONIC</td>
</tr>
<tr>
<td>25A2-5B3-50A5-50F6</td>
<td>180.0</td>
<td>111.0</td>
<td>105.0</td>
<td>102.0</td>
<td>1.7</td>
<td>1.6</td>
<td>POLYCISTRONIC</td>
</tr>
</tbody>
</table>

[0140] This data further demonstrates that, surprisingly, expression of immunoglobulin heavy chain within the polycistrionic expression system, placed after the IRES, is not significantly reduced as compared to the NEOSPLA system. The observation that the k_x/k_y ratios are greater than one is likely a result of permissive cellular secretion of free light chain. Therefore, in this polycistrionic expression system IRES driven heavy chain protein production appears to be very efficient.

**EXAMPLE 3**

**A. Increases in Expression via Genomic Amplification**

[0141] Both NEOSPLA and Polycistronic expression systems contain an expression cassette encoding the murine DHFR gene. As the parent CHO-DG44 cell line used for expression is completely deficient in DHFR enzymatic activity (double deletion), amplification of the integrated target gene (murine DHFR) is possible by growth selection in media containing methotrexate (MTX). During this amplification, the directly linked immunoglobulin genes are concomitantly amplified. Thus it is possible to isolate cell lines producing elevated amounts of immunoglobulin. Our first round selection of G418 resistant cell lines is performed in 5 nM MTX. Highest level producers (as
determined by ELISA) are identified and then subjected to two subsequent rounds of amplification with increasing concentrations of MTX (50 nM then 500 nM).

[0142] Following three successive rounds of amplification in MTX, 500 nM resistant cell lines can be identified. For the Polycistronic cell lineage derived from the G418 resistant cell line 25A2 the expression levels at each step for various isolates are listed below. For comparison purposes, listed below are expression levels for an irrelevant antibody (anti-CD23) expressed using the NEOSPLA expression system. This antibody is Primatized® antibody bearing primate-derived light and heavy chain variable domains, human kappa light chain constant region and the complete human gamma1 heavy chain (including the CH2 domain) region.

[0143] A comparison of production levels of the Polycistronic System and the NEOSPLA system grown in controlled bioreactors is also included. Duplicate runs in bioreactors are represented for each system, Polycistronic and NEOSPLA.

**TABLE 3**

**Polycistronic Expression System**

<table>
<thead>
<tr>
<th>Media</th>
<th>Cell</th>
<th>mg/L</th>
<th>pcd</th>
<th>doubling</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G418</td>
<td>25A2</td>
<td>4.5</td>
<td>2.0</td>
<td>30 hrs</td>
</tr>
<tr>
<td>5 nM MTX</td>
<td>25A2-5B3</td>
<td>16.6</td>
<td>11.5</td>
<td>32</td>
</tr>
<tr>
<td>50 nM MTX</td>
<td>25A2-5B3-50A5</td>
<td>45.6</td>
<td>20.3</td>
<td>35</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500D8</td>
<td>78.0</td>
<td>26.8</td>
<td>42</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500F2</td>
<td>84.2</td>
<td>27.5</td>
<td>40</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500F6</td>
<td>89.9</td>
<td>25.1</td>
<td>35</td>
</tr>
</tbody>
</table>

In bioreactors

<table>
<thead>
<tr>
<th>Media</th>
<th>Cell</th>
<th>mg/L</th>
<th>pcd</th>
<th>doubling</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500F6</td>
<td>729.0</td>
<td>28.6</td>
<td>Growth over</td>
</tr>
<tr>
<td>15 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500F6</td>
<td>701.0</td>
<td>28.2</td>
<td>Growth over</td>
</tr>
<tr>
<td>15 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NEOSPLA Expression System**

48
**Anti-CD23 P5E8N-SHL in N5KG1**

<table>
<thead>
<tr>
<th>Media time</th>
<th>Cell</th>
<th>mg/L</th>
<th>pcd</th>
<th>doubling</th>
</tr>
</thead>
<tbody>
<tr>
<td>G418</td>
<td>8B9</td>
<td>14.9</td>
<td>4.5</td>
<td>33</td>
</tr>
<tr>
<td>5 nM MTX</td>
<td>8B9-5A12</td>
<td>11.0</td>
<td>6.5</td>
<td>40</td>
</tr>
<tr>
<td>50 nM MTX</td>
<td>8B9-5A12-50F1</td>
<td>56.1</td>
<td>22.0</td>
<td>35</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>8B9-5A12-50F1-500G8</td>
<td>53.5</td>
<td>28.6</td>
<td>44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In bioreactors</th>
<th>Cell</th>
<th>mg/L</th>
<th>pcd</th>
<th>doubling</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 nM MTX 12 days</td>
<td>8B9-5A12-50F1-500G8</td>
<td>927.0</td>
<td>29.1</td>
<td>Growth over</td>
</tr>
<tr>
<td>12 days</td>
<td>8B9-5A12-50F1-500G8</td>
<td>680.0</td>
<td>29.1</td>
<td>Growth over</td>
</tr>
</tbody>
</table>

**[0144]** As can be seen from Table 3 above, expression levels at each stage of selection (G418, 5, 50 or 500 nM MTX) are comparable between the Polycistronic and NEOSPLA expression systems. Also, production levels from controlled bioreactor growth conditions appear comparable. Therefore, in this system, placing the immunoglobulin heavy chain under translational control of the EMCV IRES downstream of the light chain leads to effective immunoglobulin production.

**EXAMPLE 4**

**Method For Generation of Expression Construct**

**[0145]** A general method for the generation of polycistronic expression construct according to the invention is described herein. In this methodology the final expression construct is generated by insertion of a DNA fragment encoding the genes of interest (i.e. immunoglobulin light and heavy chains) in a single cloning step. The DNA fragment may be generated by use of conventional recombinant DNA technologies including oligo synthesis and ligation, overlapping PCR, complete DNA synthesis, or a combination of any of these methodologies as described previously.

**[0146]** Figure 7 contains a schematic outline summarizing the steps utilized to generate a vector suitable for expression of an immunoglobulin (e.g., HuCC49) in mammalian cells. This methodology demonstrates the use of overlapping PCR to generate a DNA fragment encoding both light and heavy chain variable domains. This fragment is then cloned in a single step into a single insertion
site of a desired vector DNA to generate the final immunoglobulin expression construct. However, while this is preferred, other methods are also be used, e.g., the respective heavy and light chain genes alternatively may be inserted at different insertion sites in the vector construct. Also, the number of light and heavy chain genes in the polycistronic construct may be varied.

[0147] As disclosed, the immunoglobulin light chain variable domain is isolated or cloned from a suitable source (i.e. hybridoma, B-cell, plasmid clone) and cloned into a suitable vector such as PCEMPTY A4 containing the light chain leader (which may or may not be retained depending on situation), light chain constant domain, IRES sequence, and a heavy chain leader sequence. Also, a heavy chain variable domain is isolated and cloned into the same empty vector (not containing the light chain variable domain). Initially having the light and heavy chain variable domains contained in separate vectors is desirable as it allows subsequent nucleotide modifications to one gene variable region without potentially impacting the other variable region.

[0148] The two separate vectors containing the light and heavy chain variable domains serve as DNA templates for independent cognate PCR amplification reactions. Design of the 5' and 3' PCR primers are such that there is a nucleotide sequence in common between the two resulting PCR products. This common sequence then allows an overlapping region for a subsequent PCR amplification step. The product of this ultimate PCR amplification step then encodes the light chain leader/light chain variable/constant/IRES/synthetic heavy leader/heavy variable domains or fragment thereof, e.g., domain deleted fragments. Figure 7 illustrates one example wherein a single DNA sequence comprising light and heavy chains separated by an IRES is inserted into a single restriction site on a vector.

[0149] By the judicious design of the above 5' and 3' PCR primers, flanking restriction endonuclease cleavage sites such as Nhe I are included. These Nhe I or other restriction sites then allow cloning of the DNA fragment into a suitable vector such as PCEMPTY B which contains all other elements for efficient expression the immunoglobulin in mammalian cells. The final vector (e.g.,
HuCC49 Gly/Ser in Polycis2) is then transfected into the desired cells to produce the gene product of interest e.g., (immunoglobulin).

**EXAMPLE 5**
Increases in Expression via Genomic Amplification

[0150] HuCC49 in N5KIRESG1DelCH2 also contains an expression cassette encoding the murine DHFR gene. As the parent CHO-DG44 cell line is completely deficient in DHFR enzymatic activity (double deletion), amplification of the integrated target gene (murine DHFR) is possible by growth selection in media containing methotrexate (MTX). During this amplification, the directly linked HuCC49 immunoglobulin genes are concomitantly amplified. Thus it is possible to isolate cell lines producing elevated amounts of immunoglobulin. First round selection of G418 resistant cell lines was performed in 5 nM MTX. Highest level producers (as determined by ELISA) were identified and then subjected to two subsequent rounds of amplification with increasing concentrations of MTX (50 nM then 500 nM).

[0151] Following three successive rounds of amplification in MTX, three 500 nM resistant cell lines were identified. For the cell lineage derived from the G418 resistant cell line 25A2, the expression levels at each step are listed in Table 4. For comparison purposes, Table 5 shows expression levels for an irrelevant antibody (anti-CD23) expressed using the NEOSPLA expression system. This antibody is Primatized® antibody bearing primate-derived light and heavy chain variable domains, human kappa light chain constant region and the complete human gamma1 heavy chain (including the CH2 domain) region.
TABLE 4: Polycistronic Expression System

<table>
<thead>
<tr>
<th>Media</th>
<th>Cell</th>
<th>mg/L</th>
<th>pcd</th>
<th>Doubling Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>G418</td>
<td>25A2</td>
<td>4.5</td>
<td>2.0</td>
<td>30 hrs</td>
</tr>
<tr>
<td>5 nM MTX</td>
<td>25A2-5B3</td>
<td>16.6</td>
<td>11.5</td>
<td>32</td>
</tr>
<tr>
<td>50 nM MTX</td>
<td>25A2-5B3-50A5</td>
<td>45.6</td>
<td>20.3</td>
<td>35</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500D8</td>
<td>78.0</td>
<td>26.8</td>
<td>42</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500F2</td>
<td>84.2</td>
<td>27.5</td>
<td>40</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>25A2-5B3-50A5-500F6</td>
<td>89.9</td>
<td>25.1</td>
<td>35</td>
</tr>
</tbody>
</table>

TABLE 5: NEOSPLA Expression System

<table>
<thead>
<tr>
<th>Media</th>
<th>Cell</th>
<th>mg/L</th>
<th>pcd</th>
<th>Doubling Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>G418</td>
<td>8B9</td>
<td>14.9</td>
<td>4.5</td>
<td>33 hrs</td>
</tr>
<tr>
<td>5 nM MTX</td>
<td>8B9-5A12</td>
<td>11.0</td>
<td>6.5</td>
<td>40</td>
</tr>
<tr>
<td>50 nM MTX</td>
<td>8B9-5A12-50F1</td>
<td>56.1</td>
<td>22.0</td>
<td>35</td>
</tr>
<tr>
<td>500 nM MTX</td>
<td>8B9-5A12-50F1-500G8</td>
<td>53.5</td>
<td>28.6</td>
<td>44</td>
</tr>
</tbody>
</table>

[0152] As can be seen in Tables 4 and 5 above, expression levels at each stage of selection (G418, 5, 50 or 500 nM MTX) are comparable between the Polycistronic and NEOSPLA expression systems. Therefore, in this system, placing the immunoglobulin heavy chain under translational control of the EMCV IRES downstream of the light chain leads to effective immunoglobulin production.

[0153] While the invention has been described in connection with specific embodiments and examples thereof, it will be understood by those of average skill in the art that the invention as disclosed is capable of further obvious modifications. This disclosure is intended to embrace such obvious departures from the preferred embodiments that fall within known or customary practices in the art and may be applied to the essential features set forth hereinabove, as well as to the scope of the appended claims.
WHAT IS CLAIMED IS:

1. A polycistronic vector for expressing functional antibodies in eukaryotic host cells which vector comprises a polycistronic transcription system comprising the following elements operably linked in the 5' to 3' orientation:
   (i) a promoter operable in a eukaryotic cell;
   (ii) a first cistron comprising a first DNA sequence encoding an antibody light chain which optionally comprises at its 5' terminus a signal peptide coding sequence operable in eukaryotic cells wherein said first DNA sequence does not comprise at its 3' end a poly A sequence and wherein said first DNA sequence comprises a 5' start codon and a 3' terminal stop codon;
   (iii) an internal ribosome entry site (IRES) obtained from a member selected from the group consisting of a cardiovirus, a herpes virus and a poliovirus; and
   (iv) at least a second cistron comprising the following elements: (a) a second DNA sequence encoding an antibody heavy chain, wherein said second DNA optionally comprises at its 5' terminus a signal peptide coding sequence operable in eukaryotic cells and wherein said second DNA sequence comprises a poly A sequence at its 3' terminus only if the DNA sequence is the 3' most coding sequence in the polycistron, and further comprises a start and stop codon at the 5' and 3' termini, respectively of said second DNA sequence;

wherein the first DNA sequence is expressed at a ratio ranging between 10:1 and 1:1 with respect to the second DNA sequence in a eukaryotic host cell containing the polycistronic vector.
2. The polycistrionic vector of claim 1, wherein the first and second DNA sequences encoding respectively antibody heavy and light chain constant regions which are of primate origin.

3. The polycistrionic vector of claim 2, wherein said primate is human.

4. The polycistrionic vector of claim 1, wherein first and second DNA sequences encode respectively antibody heavy and light chain variable regions which are of primate origin.

5. The polycistrionic vector of claim 4, wherein said primate is human.

6. The polycistrionic vector of claim 5, wherein the heavy and light chain variable regions are humanized.

7. The polycistrionic vector of claim 1, wherein the first and second DNA sequences encode respectively antibody heavy and light chain constant regions which are of rodent origin.

8. The polycistrionic vector of claim 7, wherein said rodent is mouse.

9. The polycistrionic vector of claim 1, wherein the first and second DNA sequences encode respectively antibody heavy and light chain variable regions which are of rodent origin.

10. The polycistrionic vector of claim 9, wherein the DNA sequences encoding antibody heavy and light chain variable regions are of mouse origin.
11. The polycistronic vector of claim 1, wherein the eukaryotic promoter is a mammalian promoter or viral promoter.

12. The polycistronic vector of claim 11, wherein the promoter is a CMV promoter.

13. The polycistronic vector of claim 1, wherein the IRES is obtained from a cardiovirus.

14. The polycistronic vector of claim 13, wherein the cardiovirus is human encephalomyocarditis virus.

15. The polycistronic vector of claim 1, wherein the functional antibodies expressed by the polycistronic vector specifically bind to a tumor associated antigen, an antigen expressed on a B cell or an antigen expressed on a T cell.

16. The polycistronic vector of claim 15, wherein the functional antibodies expressed by the polycistronic vector specifically bind to an antigen selected from the group consisting of TAG-72, CD4, CD11, CD19, CD20, CD22, CD23, CD37, CD40, CD45, CD80, CD86 and CD154.

17. The polycistronic vector of claim 16, wherein the antigen is TAG-72.

18. The polycistronic vector of claim 16, wherein the functional antibody is a human, humanized Primatized or chimeric antibody specific to TAG-72.

19. The polycistronic vector of claim 15, wherein the antibody is rituximab or ibritumomab.
20. The polycistronic vector of claim 1, wherein DNA sequence encoding the antibody light chain is expressed at a ratio ranging between 3:1 and 1:1 with respect to the antibody heavy chain.

21. A eukaryotic cell comprising a polycistronic vector according to any one of claims 1-20, wherein the eukaryotic cell secretes about 5 to about 100 picograms of functional antibody per day.

22. The eukaryotic cell of claim 21, wherein the eukaryotic cell is a mammalian cell or yeast cell.

23. The mammalian cell of claim 22, wherein the mammalian cell is a member selected from the group consisting of baby hamster kidney cell, fibroblast cell, myeloma cell, and Chinese Hamster Ovary cells (CHO cells).

24. The mammalian cell of claim 23 which is a CHO cell.

25. The yeast cell of claim 23 wherein said yeast cell is selected from the group consisting of Saccharomyces, Schizosaccharomyces, Hansenula, Yarrowia, Pichia, and Candida.

26. The yeast cell of claim 26 wherein said Pichia strain is Pichia pastoris.

27. A method of producing functional antibodies comprising culturing a eukaryotic cells according to claim 21 in a cell culture to produce functional antibodies and recovering the functional antibodies from the eukaryotic cell culture.
28. The method of claim 27, wherein the cultured eukaryotic cells produce at least about 1-5 picograms of antibodies/cell per day.

29. The method of claim 28 wherein said eukaryotic cell is a mammalian or yeast cell.

30. The method of claim 28, wherein the functional antibodies are recovered from cell culture medium.

31. The method of claim 28, wherein the functional antibodies specifically bind TAG-72.

32. The method of claim 31, wherein said functional antibody comprises a humanized, human or chimeric anti-CH₂ domain deleted HuCC49 antibody.

33. The method of claim 28, wherein said chimeric antibody is rituximab or ibritumomab.

34. The method of claim 27 wherein the production of functional antibodies further comprises the step of homologous recombination.

35. The method of claim 28, wherein the functional antibodies are produced in batch fed cell cultures.
HuCC49 CH2 Linker in N5KIRESG1DelCH2

9233 bp

**Figure 1**
HuCC49 CH2 Linker in N5KG1DelCH2

FIGURE 3
### Polycistronic HuCC49 G418 Cell Isolates

<table>
<thead>
<tr>
<th>M</th>
<th>CHO</th>
<th>3C11</th>
<th>5A8</th>
<th>5H11</th>
<th>22B6</th>
<th>22H1</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3A2</td>
<td>5A5</td>
<td>5F7</td>
<td>22A7</td>
<td>22E6</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

*FIGURE 4*
Polycistronic HuCC49 G418 Cell Isolates

M | 23A3 | 24C2 | 25A2 | -- | 23A10 | 24E1 | 25A5 | M
--- | --- | --- | --- | --- | --- | --- | --- | ---
22H10 | 23C12 | 25A1 | 14H9CD

kb

23 -
9.4 -
6.5 -
4.3 -
2.3 -
2.0 -

FIGURE 5
661 ATCGATCTTC CATGCTTGAA CAAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTTGG
N1

721 AGAGCCGCTATT CGGCTATGAC TGGCCACAAC AGACAAATCGG CTGCTCTGAT GCAGGCGGTG
N1

781 TCCGGCCTGTTC AGCGCAGGGG CGCCCGGTTC TTTTTGCAAA GACCGACCTG TCCGGTGCCC
N1 SVO

StuI

841 TGAATGAACCT GGAGTGATG GCGGCCGCTC TAGGCCTCCA AAAAGGCTC CTCACTACTT SVO

SfiI

901 CTTGGAATAGG TCGAGGGCCG AGGCGGCCTC GCCCTCTGCA TAAATAAAA AAATTAGTCA SVO

961 GCCATGCGAG GGGCGAGAAA TGGGCCGAAC TGGGCGGAGT TAGGGCGGAG ATGGGCGGAG SVO

1021 TTAGGGCCGG GACTATGGTT GCTGACTAAT TGAGATGCAAT GCTTTGCATA CTTCTGGCTG SVO

1081 CTGGGGAGCC TGGGGACTTT CCACACCTGG TTCTGACTA ATTGAGATGC ATGCTTTGCA SVO

1141 TACTTCTGCC TGGCTGGGAG CCTGGGGACT TTCCACACCC TAACTGACAC ACATTCACACA CMV SVO

-
1201 GAATTAATTC CCCTAGTTAT TAATAGTAAT CAATTACGGG GTCTATTAGT CATAGCCCAT

                         CMV

1261 ATATGGAGTT CCGCGTTACA TAACCTACGG TAAATGGCCC GCCTGGCTGA
                         CCGCCCCAAGC

                         CMV

1321 ACCCCCGCCC ATTGACGTCA ATATGGACGT ATGTTCCCAT AGTAACGCCA
                         ATAGGGGACTT

                         CMV

                         ~~~~~~~~~~~~~

1381 TCCATTGAGC TCAATGGGTG GAGTATTTCG GTAAACTGCA CCACTTGGA
                         GATACATAG

                         CMV

                         ~~~~~~~~~~~~~

                         NdeI

1441 TGTATCATAT GCCAAGTACG CCCCTATTG ACGTCAATGA CGGAATATGG
                         CCCCGCCTGGC

                         CMV

                         ~~~~~~~~~~~~~

1501 ATATGGCCCA GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC
                         TACGTATTAG

                         CMV

                         ~~~~~~~~~~~~~

1561 TCATCGCTAT TACCATGGTG ATGCCGTTTTT GCCAGTACAT CAATGGCCGT
                         GGATAGCGGT

                         CMV

                         ~~~~~~~~~~~~~

1621 TTGACTCACC GGGATTTCCA AGTCTCCACC CCATTGACGT CAATGGGAGT
                         TTGTTTGGC

                         CMV

                         ~~~~~~~~~~~~~

1681 ACCAAAATCA ACAGGACTTT CCAAATGTTC GTAACAATCT CGCCCCATTG
                         ACGCAAATGG

                         CMV

                         ~~~~~~~~~~~~~

1741 GCGGTAAGGGC GTTACGGTGG GAGGTCTATA TAAGCAGAGC TGGGTACGTG
                         AACCCTCAGA
1801  TCGCCTGGAG AGGCCATCAC AGATCAAAGC TTTGATCTCT CACCATGGAT AGGCCAGGCCC

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

L

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

RleAI

~~~~~~~~~~

1861  AGGTGCTCAT GCTCCTGCTG CTGTGGGTGA CGGGCACATG CGGCGACATC GTGATGAGCC

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

1921  AGTCTCCAGA CTCCCTGGCC GTGTCCTGGT GCGAGAGGCT GACTCTGAAT TGCAAGTCCA

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

1981  GCCAGTCCCC GTCTATAGC GGAATCAGA AGAACTATCT CGCCTGGTAT CAGCAGAAAC

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

2041  CAGGGCAGAG CCCTAAAACGT CTGTATTACT GGGCATCCGC TAGGGAATCC GCCGTGCCTG

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

2101  ATCGCTTCAG CGGCAGCGGA TCTGGGACAG ACTTCACTCT GACAATCAGC AGCGTGCAGG

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

2161  CAGAAAGACGT GCCAGTCTAT TATTGTACGC AGTATTATAG CTRATCCCCTC ACATCGGGCC

Kappa

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

VL

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

2221  CTGGCACCAA GCTGGAACCTG AAGACGTACGG TGCTGCAACC ATCTGTCTTC ATCTPCCCGC

Kappa

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

2281  CATCTGATGA CGAAGTGGAA TCTGGAACCTG CCTCTGTGTTGT GTGCCTGCTG AATAAACTTCT

Kappa
2341 ATCCCAAGTA GGCCAAAGTA CAGTGGAAAGG TGGATAACGC CCTCCAATCG GGTAACCTCC
Kappa

2401 AGGAGAGGTGT CACAGAGCAG GACAGCAAGG ACAGCACCCTA CAGCCCTCAGC AGCACCCTGA
Kappa

2461 CGCTGAGCAGA AGCAGACTAC GAGAAACACA AAGTCTACGC CTGCCGAAGTC ACCCATACTAG
Kappa
IRES

2521 GCCTGAGCTC GCCGTGCACA AAGAGCTTCA ACAGGGGAGA GTGTTGAATT CCTCCCCCTCC
IRES

2581 CCCCCCTTTA CGTACTGGC CGAAGCCGCT TGGAATAAGG CCCGTGTGCG TTTGTCTATA
IRES

2641 TGATATTTTT CACCATATTG CCGTCTTTTG GCAATGTGAG GGGCCCGAAA CCTGGCCCTTG
IRES

AvrII

2701 TCTTCTTGAC GAGCTTTTTT CCGTCTCGG CAAAGGAATG CAAGGTCTGT
IRES

2761 TGAATGTCTGT GAAGGAAGCA GTTCTCTCTT AAGCTCTCTT AAGACAAAACA ACGTCTGTAG
IRES

2821 CGACCCCCCTT GAGCAGCGGG AACCCCCCACC CTGCCAGACG GTGCCCTGCTG GCACAAAAGC
IRES
PmaCI

2881 CACGGTGTTATA AGATACACCT GCAAAGGC GG CACAAACCCA GTGCCACGTT
GTGAGTTGGA

IRES

2941 TAGTTGTGGA AAGAGTCAGAA TGGCTCTCTCT CAAGCTATT CAACAAGGGG
CTGAAAGGAGT

IRES

KpnI

3001 CCCAGAAAGGT ACCCCATTGT ATGGGATCTG ATCTGGGCCC TCGGTCACA
TGCTTTACAT

IRES

3061 GTGTATTAGTC GAGGTTAAAA AACGTCTAGG CCCCCAGAAC CACGCGGAGC
TGTTTTTCT

L

IRES

3121 TTGAAAAACA CGATGATAAT ATGGGTGGA GCCTCATCTT GCTCTTCCTT
GTCGCTGTTT

L

VH

MluI

3181 CTAAGGCTGTC CCGTCCCCAG GTCCAGCTGG TGCAGTCCC GCGTGAAGGTG
GTGAAAACCTG

VH

3241 GGGCTTCCCCT GAAGATTCCC TGCAAGGCCAA GCGGCTACAC CTTCACGTGAT
CACGCAATCC

VH

3301 ACTGGGTTGAA ACAGAAATCCT GGACAGGCCC TGGAAGGGAT TGGAATTTCC
TCTCCGGGA

VH
3361  ACGATGATTT TAAGTACAAT GAGAGGTTCG ACGGCAAGGC CACACTGACT
       GCAGACACAT
               VH
-----------------------------

BaeI
-----------------------------

3421  CTGCCAGCAC TGCCCTAGTG GAGCTCCTCCG GCCTGAGATC CGAGGATCT
       GCAGTCTACT
               VH
-----------------------------

3481  TCTGCACAGG ATCCCTGAAAT ATGGCCTACT GGGGACAGGG AACCCCTGGTC
       ACCGTCTCCA
               CH2 Link G1DelCH2
-----------------------------

NheI
-----------------------------

3541  GCGCTAGGACG CAGGGGCCCAG CCGGACAGCC CTCCTCCAAG
       AGGCACCTCTG
               CH2 Link G1DelCH2
-----------------------------

AgeI
-----------------------------

3601  GGGGCACAGGC GCCCTGGGCG TGCCTGGTCA ACGACTACTT CCCCCAGGCG
       GTGACCGGTGT
               CH2 Link G1DelCH2
-----------------------------

3661  CGTGGAACTC AGGCGCCCTG ACCAGGCGCG TGCACACCTT CCCCCGCTGTC
       CTACAGTCTCT
               CH2 Link G1DelCH2
-----------------------------

3721  CAGGACTCTA CTCCCCTAGGC AGCGTGGTGA CGTGCCCTCT CAGCAGCTTG
       GGCACCCAGA
               CH2 Link G1DelCH2
-----------------------------

3781  CCTACATCTCG CAACGTTGAAT CACACGCCCA GCAACACCAA GGGGACAAG
       AAAAGTTGAGC
               CH2 Link G1DelCH2
3841 CCAAAATCTTG TGACAAACT CACACATGCC CACCGTGCCC CATCGAGAAA ACCATCTCCA

CH2 Link G1DelCH2

------------------------

Bsp1407I

--------

3901 AAGCCCAAAGG GCAGCCCCGA GAACCACAGG TGTACACCT GCCCCCATCC CGGGATGAGC

CH2 Link G1DelCH2

------------------------

3961 TGACCAAGAAG CCAGGTCAGC CTGACCTGCC TGGTCAAAGG CTTCTATCCC AGCGACATCG

CH2 Link G1DelCH2

------------------------

4021 CCGTGAGGAGT GGAGGAGCAAT GGGCAGGCGGG AGAACAACCTA CAAGACCACG CCTCCCGTGC

CH2 Link G1DelCH2

------------------------

4081 TGGACTCCGA CGGCTCCTTC TTCCCTCTACA GCAAGCTCAC CGTGGACAAG AGCAGGTGGCC

CH2 Link G1DelCH2

------------------------

4141 AGCAGGGGAAA CGTCTTCTCA TGCTCCGTGA TGGATGAGGC TCTGACAAC CACTACAGGC

CH2 Link G1DelCH2

------------------------

MfeI

--------

SrfI

--------

BamHI

--------

--------

4201 AGAAAGACCT CTCCCTGTCT CCGGTAATA GAGGATCCAA TTGGCCCCGGG CAGATCTGAT

HpaI

--------

HindII

EcoRV

--------

--------

4261 CCGTTAACGG TTACCAACTA CCTAGACTGG ATTCGTGACA ACATGCCGCCC
BGH

CGATGATGCA GCCGCACTG TGCCCTCTAG TTGCCGCCCA TCTGTGGTTT
GCCGCTCCCC
BGH

CGTCCTCTCC TTGACCCCTGG AAGGTGCCAC TCCCCACTGTC CTTTCTTAAT
AAAAATGAGGA
BGH

-----------------------------

AAATGGCATCG CATTGCTGTA GTAGGGTGTCA TTCTATTCTG GGGGGTGAGG
TGAGGCGAGA
BGH

-----------------------------

CAGCAAGGGG GAGGATGGGG AAGACAATAG CAGGATGCT GGGGATGCCG
TGAGCTCTAT
BGH
Beta

~~
~~
SgfI
SpeI

-----------------------------

GGCTTCTGAG GCGAAAAGAA CCAGCTGGGG CTGCGACAGCG CTGCGATCGC
CTGCGACTAGT
Beta

-----------------------------

AGCTTGTCTT CTCAATTTCT TATTTGCATA ATGAGAAAAA AAGGAAAAAT
AATTTTTAACA
Beta

-----------------------------

CCAATTCAGT AGTGGATTGA GCAATGCGT TGCCAAAAAG GATTCTTAG
AGACAGTGTCT
Beta

-----------------------------

CTCTGCAAG CATGAGCAA ACATTATCAG GAGGGAGTAC CCAGAGCTGA
GACTCTAAG
Beta

-----------------------------

CCAGTGAGTG GCACAGCATC CAGGGAGAAA TATGCTTGTC ATCACCGAAG
CCTGATCTCG
Beta
4861  TAGAGCCACA CCCTGGTAAG GGCCAATCTG CTCACACAGG ATAGAGAGGG CAGGAGCCAG
Beta

4921  GGCAGAGCCT ATAAGGTGAG GTAGGATCAG TTGCTCCTCA CATTTGCTTC TGACATAGTT
Beta

4981  GTGTTGGGAG CTGAGATAGC TTGGGCCCCG GACAGCTCAG GGCTGCGATT TCGCGCCAAA
DHFR

5041  CTTGAGGGCA ATCCTAGCGT GAAAGGCTGTT AGGATTTTAT CCCCCGCTGCC ATCATGGTTC
DHFR

5101  GACCATTTGAA CTGCATCTGC GCCGTCGTCC AAAATATGGG GATTGGCAAG AACGGAGACC
DHFR

5161  TACCTTGGGC TCCGCTCAGG AACGAGTTCA AGTACTTCCA AAGAATGACC ACAACCTCTTT
DHFR

5221  CAGTGGAAAGG TAAACAGAAAT CTGGTGATTCA TGGTGAGGAA AACCTGGTTC TCCATTCTTG
DHFR

5281  AGAAGAATCG ACCCTTAAGG GACAGAATTCA ATATAGTTCT CAGTAGAGAA CTCAAAGAAC
DHFR

5341  CACCGAGGG AGCTCATTTT CTTGCCAAAA GTTTGGATGA TGCCTTAAGA CTTATTGAAC
DHFR

5401  AACCGGAATT GGCAAGTAAA GTAGACATGG TTTGGATAGT CGGAGGCAGT TCTGTTTACC
DHFR

5461 AGGAAGCCAT GAATCAACCA GGCCACCTCA GACTCTTTGT GACAAGGATC ATGCAAGGAAT
DHFR

5521 TTGAAAGTGTA CACGTTTTTC CCAGAAATTG ATTTGGGGAA ATATAAACTT CTCCCGAAT
DHFR

5581 ACCCAGGCGT CCTCTCTGAG GTCCAGGAGG AAAAAGGCAAT CAAGTATAAG TTTGAAGTCTC
DHFR

5641 AGCGAAGAAGA AGACTAACAG GAAGATGCTT TCAAGTCTCTC TGCTCCCCCC CTAAAGCTAT
BGH

5701 GCATTTTAT AAGACCATGG GACTTTTGCT GGCTTTAGAT CAGCCTCGAC TGTGCCCTTCT
BGH

5761 AGTTGCGAGC CATCTGTTGT TTGCCCCCTCC CCGTGCCCTT CCTGACCCCT GGAAGGTGCC
BGH

5821 ACTCCCACCTG TCCTTTCTTA ATAAAAATGAG GAATTTGCAT CGCATTGTCT GAGTAGGTGT
BGH

5881 CATTCTATTC TGGGGGTGG GGTTGGGCGAG CACAGCAAGG GGGAGGATTG GGAAGACAAT
BGH

5941 AGCAGGGCATG CTGGGGATGC GGTTGGCTCTG ATGGGCTCTG AGCCCCAAAG AACCAGCTGG
Sse8387I

6001 GGCTCGAAGGCG GCCCGCTCCG GATCGAGGCG CCCTACTAACT CTCTCTCTCC TCCTTTTTCC
N2
Sse8387I  
BaiI  

6061  
TGGAGGACGA GGCAGCGCGG CTATCGTGCC TGGCCACGAC GGGGCTTCTT TGGCGGAGCTG  
N2  

6121  
TGCTCGACGT TGTCAGCTGA GCGGAAGGG ACTGCTTGCT ATGGGCGGAA GTGGCGGCGG  
N2  

6181  
AGGATCCCTG GTACATCTCA CTTGCTCTTG CGAGAAAAAG ATCCATCATT GCTGATGCAA  
N2  

6241  
TGGGCGCGCT GCATACGGTT GATCCGGCTA CCTGCCCATG CGACCACAAA GCGAAACATC  
N2  

6301  
GCATCGAGCG AGCAGGTACT CGGATGGAAAG CGGTCTCTGT CGATCAGGAT GATCTGGACG  
N2  

6361  
AAGAGCATCA GGGGCTCGCG CGACGCGAAC TGGTCGCCAG GCTCAAAGCG CGCATGCCCCG  
N2  

6421  
ACGGCGAGGA TCTCGTCTGT ACCCATGGCG ATGCCTTGCTT GGCGAATATC ATGGTGAAA  
N2  

6481  
ATGGCGCATT TTCTGGATTC ATCGACTGTC GCCGGCTGGG TTGGCGGAC CGCTATCAGG  
N2  

6541  
ACATAGCGTT GGCTACCCGT GATATTGCTG AAGAGCTTGG CGGCAATGG GCTGACCCTG  
N2  

RsrII
WO 03/048306
18/74

AsuII

TTGACGAGTT CTTCCTGAGCG GGACTCTGGG GTTCGAAATG ACCGACCAAG
CGACGCCCAA

CGAGATTTCG ATTCACCGGC CGCCTCTCTG TAAAGGGTTGG
GCTTCGGAAT

GATGATCCTT CTACGCCGGG CTACCGGGGC GGAGTCACTGC TGGAGTTCT

FseI SacII

ATCTATCTTT ATCATGTCTGG GATCGCGGCC GCGCGCCACC GCGGTGGAGC
TTTAAATTAAG

Ascl

TTTCCCTTAT GTTCGGCTTTA GTCAGGGTTA ATTTCGAGCT
TGCCGTAATC

ATGGTCATAG CTGTTTCTCTG TGTGAAATTG TTATCCGCTC ACAATTCCAC
ACAACATACC

AGCCGGAAGC ATAAAGTGA AGCCTTGGGG TGCCTAATGA GTGAGCTAAC
TCACATAATT

TGCGTTCGCG TCAGCTCCCG CTTTACGTGC CCGGAAACCTG TCAGCCGGAC
TGCAATTAATG

AATCGGCCAA CGCGGCGGGA GGAGGGTGGT GCTATTGGG CGCTCTTCCG
CTCCTCGCGC

CAGTCACTTG CTGCGCTCGG TCAGCTCCTGC GCGCGAGCC GTATCGACCT
ACCTAAGGCG

GGATATACGG TTATCCACAG AATCGCGGGA TAAGCGAGGA AAGAACATGT
GACGAAAGGG

CCAGCAAGAA GCCAGGAAACC GTAAAAAGGCG CGCGTTTCG TGCTTTTTCG
ATAGGCTCCG

SV

CGCCCGCCCA ACTGTTTTAT TGCGGCTTAT AATGTTACAA AATAAAGCAA
TAGCATTACA

SV

SV

SV

SV


CCCCCTGAC GAGATCACAA AAAATCGACG CTCAAGTCAG AGGTGGCGAA
ACCCGACAGG
ACTATAGAGA TACCAGGCGT TTCCCCCTGG AAGCTCCCTC GTGCAGCTTC
CTGTTCGGAAC
CCTGCCGCTT AGCGGATACC TGTCCGCTCT TCTCCCCCTG GGAAGCGTGG
TGCGGTCTGT
TAGCTCACGC TGAGGTATTC TCAGTTCGAG GTAGGTCTGT GCCTCAAAGC
TGGGCTGTTGT
GCACAGAACC CCCGTTACCC CCGACCCGACT CGCCTTATCC GSTAACTATC
GTCTTGAGTC
CAACCCGGTA AGACAAGACT TATGCACACT GGCAGCAGCC ACTGGTAAACA
GGATTACGAG
AGCGAGGTAT GTAGGCGGTT CTACAGAGTT CTGGAATGGT TGCCCTAAACT
ACCGCTACAC
TAGAAGAGCA GTATTTGGA TCTCGCGCTCT GCTGAAGCCA GTTACCTTCG
GAAAAAGAGT
TGGGTAGCTCT TGATCCGGCA AACAAACACG CGCTGGTAGC GGTGGTTTCT
TTGGTTGCAA
GCAGCAGATT AGCGCGCAA AAAAAAGATC TCAAGAGAT CTTTGATGT
TTTCTACGGG
GTGTCAGCGCT CAGTGGAAAG AAAACTCACG TTAAGGGATTT TGTTATCGAG
GATTATCAA
AAGGATCTTC ACCTAGATCC TTTAAATTA AAAATGAAGT TTTAAATCAA
TCTAAGATAT
ATATGAGTAA ACTTGGTCGT ACAGTTACCA ATGCTTAATC AGTGAGGCCA
CTATCTCGAC

~~~~~~~~~~~~~~~~~~~~~~~~~~ Amp ~~~~~~~~~~~~~~~~~~~~~~~~~~~
Eam11051

GATCTGTCTTA TTTCCGTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTGAAT
TAACATCGAT

~~~~~~~~~~~~~~~~~~~~~~~~~~ Amp ~~~~~~~~~~~~~~~~~~~~~~~~~~~

ACGCGGAGGCC TTACCATCTG GCCCCAGTGTC TGCAATGATA CCGCGAGACC
CACGCTCACCC

~~~~~~~~~~~~~~~~~~~~~~~~~~ Amp ~~~~~~~~~~~~~~~~~~~~~~~~~~~

GGCTCCAGAT TTATCAGCAGA TAAACCCAGCC AGCCGGAAGG GCGGAGCGCA
GAACTGGGCTC

~~~~~~~~~~~~~~~~~~~~~~~~~~ Amp ~~~~~~~~~~~~~~~~~~~~~~~~~~~

TGCAACTTTTA TCAGCTCCCA TCCAGGATTA TAATTTGTTGC CGGGAAGCTA
GAGTAAAGTAG

~~~~~~~~~~~~~~~~~~~~~~~~~~ Amp ~~~~~~~~~~~~~~~~~~~~~~~~~~~
TGCGCGTTTC AATAGTTGGC GCAACGTGTT TGCCATTGCT ACAGGCATCG TTGGTGCACG

Amp

CTCGTCGTTT GGTATGGCTT CATTACACCT CCGTTCCTAA CGATCAAGGC GAGTTACATG

Amp

ATCCCCCATG TTGTGAACAA AAGCGGTGTA GCTCCTCGGT CTCGCGATCG TTGCTCAGAG

Amp

TAAGTGGCCT GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCCATAAT CTCTTACTGT

Amp

CATGCCATCC GTAAGATGCT TTTCTGTCAG TGGTGAATAC TCAACCAAGT CATTCTGAGA

Amp

ATAGTGTTATG CGGCGAACGA GTTGCTCTTG CCCGCGGTCAT ATACGGGATA ATACCGCGGC

Amp

ACATAGCAGA ACTTTAACA AAGTGCTCCAT TGGAACAGCT TCTTCCGGGC GAAATACTTC

Amp

AAGGATCTTA CGCGCTGTTGA GATCCAGTTT GATGTAACC ACTCGTGCA GCAACTGATC

Amp

TTACGCTCTT TTATCTTCCA CCAGCTTTC GGGTGAAGCA AAAACAGGA GGCAAATATGC

Amp

SspI

CGCAAAAAAG GGAATAAGGG CGACACGGGA ATGTTGAATA CTCATTACTCT TCTTTTTTCA
SspI

ATATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT
TTAGAAAAAT AAACAAATAG GGGTCGCGC CACATTTCCC CGAAAGTGC CAC
Hybridoma or Other

PCR

Bgl II  BsiW1  Mlu I  Nhe I

L  VL  VH

EcoR I  Mlu I

Swa I  Nhe I

PCEMPTY A4

L  Kappa  IRES SHL

Bgl II  BsiW1  EcoR I  Mlu I

Swa I  Nhe I

L  VL  Kappa  IRES SHL

Light Chain Variable in PCEMTPY A4

Bgl II  EcoR I  Mlu I  Nhe I

BsiW1

L  IRES SHL  VH

Heavy Variable in PCEMTPY A4
27/74

Kappa

481 GAGGCCAAAG TACAGTGGAAC GTGGGATAAC GCCTCCTAAAT
CTCCGGTTTC ATGTCACCTT CCACCTATTT CGGGAGGTTA
Kappa

521 CGGGTAACTC CCAGGAGAGGT TCACAGAGAC AGGACAGCAA
GCCCATGAGG GTCTCTCTCA CAGTGTCTCGT TCCTGTGTGTT
Kappa

561 GGACAGCACC TACAGCCTCA GCAGCACCCT GCAGCTGAGC
CCTGTCTGGG ATGTCGGGAGT GTCTGGGGA CGCCGACTCG
Kappa

-------------------------------

Accl

-------------

601 AAAGCAGACT ACGAGAAACA CAAAGTCTAC GCCTGCGAAG
TTTCGTCTGA TGCTCTTTGT GTTTCAGATG CGGACGCTTC
Kappa

-------------------------------

SacI

-------------

641 TCAACCACGCA GCCGCTGAGC TCGCCGCTCA CAAAGAGCTT
AGTGGAAGATG CCGGACTCG AGCGGGCAGT GTTTCTCGAA
Kappa IRES

-------------------------------

EcoRI

-------------

681 CAAACAGGGGA GAGTTGTGAAAT TTCCCTCCCTC CCCCCCCCCCT
GTTGTTCCCT CTCACAACTT AAGGAGGGAG GGGGGGGGGA
IRES

-------------------------------

ApoI

-------------

721 AAGCTTACTG GCCGAAGCGG CTTGGAAATAA GCCTCCTGTTG
TTGCAATGAC CCGCTTCCGC GAACCTTATAT CGGCCCACAC
IRES

-------------------------------

761 CGCTTGGCTTA TATGTTATTG TCCACCATAT GCCTGCTTTT
GCAAACAGAT ATACAAATAA AGGTGGTATA AGCGCCAAGAAA
IRES

-------------------------------

801 TGGCAATGTTG AGGGCCGCAGA AACCTGGCCC TGTCTTCTTTG
ACGGTTACAC TCCCGGCGCT TTGGACCGGG ACAGAAGAAC
IRES

841 ACGGACATTCC CTAGTGGGTCT TTCCCCCTCTTC GCCAAAGGAA
TGCTTGTAAG GATCACCAGA AAGGGGAGAG CGGTTCCTTT
IRES

881 TGCAAGGTCT GTGGAACTGTC GTGAAGGAAG CAGTTCCTCT
ACGTTCAGA CAACTTACAG CACTCCTTCC TCTCAAGGAGA
IRES

HindIII

921 GGAAGCTTTCT TGAAGACAAA CAACGTCTGT AGCGACCCTT
CCTTCGAAGA ACTTCTGTTT GTTGCAAGACA TCGCTGGGAA
IRES

961 TGGCAGGGACG GGAACCCCCC ACCTGGCGAC AGGTGCCCTCT
ACGTCCGTCG CCTTGGGGGG TGGACCGCTG TCCACCGGAGA
IRES

PmlI EspMI

1001 GCGGCCAAAA GCCAGGCGTA TAAGATACAC CTGCAAAGGC
CGCGGTTTTT CGGTGCACAT ATTCTATGTG GACGTTCCTG
IRES

1041 GCCACAACCC CAGTGGCAGG TTGTGAGTTG GATAGGTTGTG
CGTGGTTGGG GTGACGGCTCC AACACTCAAC CTATCAACAC
IRES

1081 GAAAGAGTCG AATGGCTTCTC CTCAACCTGTA TTCAACAAAGG
CTTTCTCAGT TTACGGAGAG GAGTTGCGAT AAGTGTCTCC
IRES

Pf1MI

1121 GGCTGAAGGA TGCCCGAGAA GTACCCCCATT GTATGGGATC
CCGACTTTCTC ACGGCTTCTC CATGGTTAA CATACCCTAG
IRES

1161 TGATCTGAGGG CCTCGGTGCA CATGCTTTAC ATGTGTTTAG
ACTAGACCCC GGAACCACTG GTACGAAATG TACACAAATC
IRES
DsaI

---------

1201 TCGAGGTTAA AAAACGTCTA GGCCCCCCGA ACCACGGGGGA
AGCTCCAATT TTTTGCAGAT CCGGGGGGCT TGGTGCCCCCT
SHL

---------

IRES

-------------------------------------------
1241 CGTGGTTTTTC CTTTGAAAAAA CACGATGATA ATATGGGTTG
GCACCAAAAG GAAACTTTTTT GTGCTACTAT TATACCCCAAC
SHL

-------------------------------------------

MluI

---------

1281 GAGCCTCATC TTGCTCTTCC TTGTGCCTGT TGCTACGCGT
CCTCGGAGTAG AACGAGAAGG AACACGGACA ACGATGCGCA
NheI EssHII

---------

SwaI AscI

---------

1321 ATTTAAATGC TAGGGGCGCG CCAGCTCCAG CTTTTGTTCC
TAAATTTACG ATCGCCCGGC GGTGAGGTC GAAAACAAGG

1361 CTTTAGTGGAG GTTAAATTTTC GAGCTTGCGGCA ATATCATGGTT
GAAATCCTCTC CCAAATTAAG CCTGAAACC CGATGACCA

1401 CATAGCTGTGTTT TCCCTGTGTGA AATTGTTATCT CGCTCAACT
GTATCGACAA AGGACACACT TTAACAATAG GCGAGTGTTA

1441 TCCACACAAC ATACGAGGCG GAAGCATAAA GTGTAAGGCC
AGGTGTGTTG TATGCTCGCG CTTCGATTCT CACATTTCCG

1481 TGCGGTGCGCT AATAGGCTGAG CTAACCTCAA TTAATGCTGT
ACCCCCCGGA TTACTCCTCT GATTGAGTGT ATTAACGCA

1521 TGCGCTCAGT GCCGCGCTCTC CAGTGGGGAA ACCGTCCTGTG
ACGGAGGTGA CGGGGGAAGA GTCAGCCCTT TGGGACGAC

PvuII

---------

1561 CCAGCTGCAT TATGGAATCG GCCAACGCGC GGGGAGAGGC
GGTGAGGCTA ATACTTTAGG CGGTGAGCGG GCCCTCTCCG

1601 GGGTTTCGCTA TTGGCGCTCT CTCGCGGCTCT TCGCTCAGTG
CCAACAGCAT AACCAGGGAG AGGCGGAAGG ACGAGTGAC

---------

Col E1

1641 ACTCGCGCTCC CTCGGTCGTTT CGGCGCTCGGC GACCGGTATC
TGAGCGACGC GAGCCAGCAA GCGAGCCGC CTTCCCATAG

---------

Col E1
1681 AGCTCACTCA AAGCCGTTAA TACCGTTATC CACAGAATCA TCGAGTGAATC TCCGGCCATT ATGCCAATAG GTGTCTTCTG

Col E1

1721 GGGGATAACG CAGGAAAGAA CATGAGAGCA AAAGGCCAGC CCCCTATGGC GTCTTCTTCG GTACACCTGT TTTCCGGTCC

Col E1

1801 TTTCCATAGG CTCCGCCGCC CTGACCGAGCA TCCACAAAAAT AAAGGTATCC GAGGCGGGGG GACTGCTCGT AGTGTGTTTTA

Col E1

Drd I

1841 CGACCGCTCAA GTCAAGAGGG GCGGAAACCCG ACAGGAACAT TGGTTCTTTC CACAGCCTAC CGCTTTTGGG GTTCTCTGATA

Col E1

1881 AAAGATAACCA GGCGTTTCCC CCTGGAAGCT CCCCTCGTGCC TTTCTATGCT CGCAAAAGGG GGACCTTCCA GGGAGCAGGC

Col E1

1921 CTCTCTCTGGT TCCGCCCGGC CGCTTACCGG ATTACCTTGCC GAGGAGACAA GGGGTGGGACG GCGAATGCCC TATGGACAGG

Col E1

1961 GCCTTGTCTCC CTTCGGGAAG CTTGCGCCGGT CTCTCATAGCT CGGAAGAAGGG GAACGACCTTC GACCACGGAA AAGATATCGA

Col E1

2001 CACGCTGTAG GTATCTCAAT CCAGGTGTTAGG TGCTTTGGATC GTCCGACATC CATAGAGTCA AGCCACATCC AGCAAGCCAG

Col E1

2041 CAAGCTGGGC TGTGTCCTCG AACCACCCTG TACAGCCCCAC GTTCGACCCG ACACACGGTC TTGGGGGGCA AGTCCGGCTG

Col E1

2081 CGCTGCGCCT TATCCGGTAA CTATCGTCTTT GAGTCACACC GCGACGCGGA ATAGGCCATT GATAGCAGAA CTCAGGTTGG
ATCAATCTAA AGTATATATG AGTAAACTTG GTCTGACAGT
TAGTTAGATT TCATATATAC TCATTTGAAC CAGACTGTCA
Amp

TACCAATGCT TAATCAGTGA GCCACCTATC TCACGGATCT
ATGGTTACGA ATTATGCACCT CCGTGAGATAG AGTCGCTAGA
Amp

Eam11051

GTCTATTTTG TTCATCCATA GTTGCGTGAC TCCCCGTCGT
CAGATATAAGC AAGTAGGTAT CAACGGACTG AGGGGCAGCA
Amp

GTAGTAAACT ACGATACGGG AGGGCTTACC ATCTGGCCCC
CATCTATTGA TGCTATGCCC TCCCGAATGG TAGACCGGGG
Amp

BsaI

AGTGCTGCAA TGATACCCGC AGACCCACGC TCACCGGTCT
TCACGGACGT ACTATGGCCG TCTGGTGCGG AGTGGCCCCCC
Amp

CAGATTATAC AGCAATAAAC CAGCCAGCCG GAAGGCGCGA
GTCTAAATAG TCGTTATTTG GTCGGTCCGC TTTCCCAGCT
Amp

CGCCGAGAGT GTTCTGCCAA CTTTATCCCG CTCCATCCAG
CGCGTCTCCA CCAAGACGTT GAAATAGGCC GAGGTAGGTCC
Amp

TCTATATTATT GTGGCCGGGA AGCTAGAGTA AGTATGTCGC
AGATAATTTA CAACGGCCCT TCGATCTCAT TCATCAAGCG
Amp

Fsp1

CAGTTAATAG TTGCCGCAAC GTTGTTGCCA TTGCTACAGG
GTCAATTATC AAACCGCGTTG CAACAAACGGT AACGATGTCC
Amp

CAGCGTGCGTG TCAGCCCTCGT CGTTTGGTAG GGCTTTACCTC
TAGGCACCAC AGTGCGACGA GCAAACCATA CCGAAGTAAG
2921 AGCTCCCGTT CCAACAGATC AAGCGGAGTT ACATGATCCC
TCGAGGCCCA AGGTTGCTAG TTCCGCTCAA TGACTAGGG
Amp

PvuI

2961 CCAATGTGTGTG CAAAAAGCG GTTAGCTCCT TCGGTCCTCC
GGTACAAACAC GTTTTTTCCGC CAATCGAGGA AGCCAGGGAG
Amp

PvuI

3001 GATCGTTGTC AGAAGTAAGT TGGCCCCAGT GTTATCACC
CTAGCAACAG CTTCATTTCA ACCGGGTCA CAATAGTGAG
Amp

3041 ATGGTATTAGG CAGCACTGCA TAATTCTCTT ACTGTCATGC
TACAATACC GTGTCGACGT ATTAAGAGAA TGACAGTACG
Amp

ScaI

3081 CATCCGTAAG ATGCTTTTCT GTGACTGTTG AGTACTCAAC
GTAGGCATTG TACGAAAAAGA CACTGACCAC TCAATGAGTG
Amp

BcgI

3121 CAAGTCATTG TGAGAATAGT GTATGCGGCG ACGAGTTGC
GTTCAGTAAG ACTCTTAATCA CATACGCGGC TGGCTCAACG
Amp

AcyI

3161 TCTTGCCCCGC GTCAATTACG GGATAATACG GCGCCACATA
AGAACGGGCCC GCAGTTATGC CCTATTATGG CGCGGGTGTAT
Amp

3201 GCAGAACTTT AAAAGTGCTC ATCATTTGAA AACGTTCTTC
CGTCTTGAAGTTTCACGAG TAGTAACCTT TTGCAAGAG
Amp
GGGGCGAAAA CTCTCAAGGA TCTTACCGCT GTTGAGATCC 
CCCCGCTTTT GAGAGTTCCT AGAATGGCGA CAACTCTAGG 

Amp

AGTTTCGATGG AACCCACTCG TGCAACCAAC TGATCTTCAG 
TCAAGCTACA TTGGGTTGAGC ACGTGGGGTTG ACTAGAAGTC 

Amp

CATTCTTTAC TTTCAACCAGC GTTCTTGGGT GAGCAAAAAAC 
GTAGAAAAATG AAAGTGGTCG CAAGACCCA CTCGTTTTTG 

Amp

AGGAAGGCAA AATGCCGCAA AAAAGGGAAT AAGGCGGACA 
TCCTTCGGTT TTACGCGGT TTTTCCCTTA TTCCCCGCTGT 

Amp

SspI

GGAAATGTT GAATACTCAT ACTCTTCCCT TTTCAATAAT 
GCTTTACAA CTTATGAGTA TGAAGAGGA AAAGTTATAA 
ATTGAAGCAT TTATCAGGGT TATTTGTCTCA TGAGGGGATA 
TAACTTTCGTA AATAGTCCCA ATAACAGAGT ACTCGGCCTAT 
CATATTGAAA TGATTTTAGA AAAATAAACCA AATTAGGGGT 
GTAATAACTT ACATAAAATCT TTTTATTTGT TTATCTCCAA 
CCCGCACCAT TTCCCCGAAA AGTGCCAC 
GGCGCGTGTA AAGGGGCTTT TCACGGTG
CTGACGCGCC CTGTAGCGCC GCATTAAGGC GGGCGGGTGT
GACTGCCGCG GACATCGGCC GCTAATTCGC GCGGCCACA

GGTGGTTAAG CGCAGCGTTAG CGCTTACACT TGCCACCGCC
CCACAAATGC CGTCCGCACT GCGGATGTGA AGCCTCGCGG

CTAGCGCCCG CTCCCTTTCGC TTCTTTCCCT TCTTCTTCGC
GATCGGCCGCG GAGGAAAGCG AAAGAAAGGA AGGAAAGGAC

CCAGTTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG
GTTGCAGCAG GCCGAAGGGA GCAGTTCGAG ATTTAGGCCC

GGTGCAGGCAG TGGGCACTCAAT CCCACTACCA AGTCACCTCC

-------

CGGCGAGGTA CGGGGCCCCC CTCGATTTAA TTAATCGAGC
AGCGCGCAT GGGCCGCGGG GGAGCTAATT AATTAGCTCG

Beta

TACTAGCTTT GCCCTCTCAAT TCTTTATTTG CATAATGAGA
ATGATCGAAG CGAAGAGTTA AAAGATAAAC GTATTACTCT

Beta

AAAAAAGGAA AAATTAATTGT AACAACCATTT CAGTATGGTG
TTTTTCTCCT TTAAATTAAA TTGTCGTTAA GTCATCAACT

Beta

TTGACCCAAT GCCTTTGCAAA AAAGGATGCT TTAGAAGCAG
AACCTCGTTA CGCAACGGTT TTTCCTACGA AATCTCTGTC

Beta

TGTCCTCTGAC AGCATAGAA ACAAACATTA TTCAGAGGGA
ACAAGAGACG TGCTCATTCC TGGTGTTAAT AAGTCTCCCT

Beta

GTACCCAGAG CTGAGACCTCC TAAGCGATGT AGTGCCACAG
CATGGGTCTC GACGCTGAGG ATTCGTCAC TCACCGTGTC

Beta

CATCCAGGGA GAAATATGCT TGTCATCACC GAAAGCTGAT
GTAGGTCCTT CTTTATACGA ACAGTAGTG GCTTCGACTA

Beta

TCCGTAAGAGC CACACCCTGCG TAAGGGCCAA TCTGCTCACA
AGGCATTCGC GTGTGGGACC ATTCGGGTG AGACGAGTGT
Beta

561 CAGGATAGAG ACGGCAGGAG CCAGGGCAGA GCATATAAGG
GTCTATATCT CCCCCGTCTC GGTCCCGTCT GTATATATCC
Beta

601 TGAGGTAGGA TCAAGTCTCC ATCCAGTTTG CTTCTGACAT
ACTTCATCTCT AGTCAACGAG GAGTGAATAC GAAGACTGTA
Beta

ClaJ

641 AGTTGTGCCA GCATGGAGGA ATCGATCCCT CATGCTTTGA
TCAACACGCT CTAACCTCCT TAGCTAGGAG GTACGAACTT

N1

681 CAAGATGCGAT TGCCACGAGG TTCTCCGGCC GCTTGGGGTG
GTTCTACCTA ACGTGCCTCC AAGAGGCGGG CGAACCACC
N1

721 AGAGGTATATT CCGCTATGAC TGGGCACACA AGACAAATCGG
TCTCCGATAA GCCGATACTG ACCCGTGGTTG TCTGTTAGCC
N1

761 CTGCCCTGAT GCCGCCCTGT GCTCCGCTGTC AGGCCAGGGG
GACGAGACTA CGCCGGCACA AGGCCGACAG TCGCGTCCCC
N1

801 CGCCCCGTTC TTTTTGTCAA GACCGACCTG TCCGGTCGCC
GCCGGCCAGA AAAAACAGTT CGGGCTGGAC AGGCCACGCGG
N1

SVO

StuJ

841 TGAATGAAGT GCAGGTAAGT GCAGGGCGCTC TAGGCCTCCA
ACTTACTTGA CGTCCTTACA CGCCGGCGAG ATCCGGAGGT
SVO

SfiJ

881 AAAAGGCCCTC CTCACTACTT CTTGGAATAGC TCAGAGGCAG
TTTTCGGAG GAGTGATGAA GACCTTATCG AGTCTCCGGC
SVO
SfiI
---------
921 AGGGCGCGGC TGGCGCTGGA TAAATAAAAA AAATTAAGTCG
TCCGCCGAAG CGGGAGACGT ATTTATTTTT TTTAATTACGT
SVO
---------
961 GCCATGCGATG GGGGCGGAGAA TGGGCAGGAC TGGGCGGAGT
CGGTAGTGAC CCGCCTCTTT ACCCGCGTGT ACGCCGCTCA
SVO
---------
1001 TAGGGGCGGG ATGGGCGCGG TTAGGGGCGG GACTATGATT
ATCCCCGCCC TACCCGCGTCG AACTCCCCGACC CGTATACGGA
SVO
---------
1041 GCTGACTTAAAT TAGGATGCAT GCTTTGCATA CTTCTGCCTG
CGACTGATTAA ACTTCTACGTA CGAAAGCTAT GAAGACGGAC
SVO
---------
1081 CTGGGGGACGCT TGGGGAATTT CCACACCCTGG TTGCTGACTA
GACCCCCGTG ACCCCTGGAA GTGTCGGACC AACGACTGAT
SVO
---------
1121 ATGGAGATGC ATGCTTTTAGC TACCTCTGCC TGCTGGGGAG
TAACTCTACGT TACGGAAAACGT ATGAAGACCG ACGACCCCTC
SVO
---------
1161 CCTGGGGACT TTCCACACCC TAACTGACAC ACATGCTCA
GGACCCCTGA AAGGTGTGGG ATGGACCTGTG TGTAAGGTGT
SVO
---------
1201 GAAATTAATTC CCCTAGTTAT TAAATGTAAT CAATTACGGG
CTTAATTAAG GGGATCAATA ATATACCTTA GTAATGCCC
SVO
---------
1241 GTCAATGGTT CATAGCCCAT ATATGGAGTT CCGCTTACA
CAGTAACTCGA GTATCGGGTA TATACCTCA ACGGCAATGT
SVO
---------
1281 TAACTTACGG TAAATGGCCC GCCTGGCTGA CCGCCAAGCG
ATGGAAATGCCC ATTTACCGGG CGGACGGACT GCGGGTTGCT
SVO
---------
1321 ACCCCCGCCC ATGGACGTCAG TAAATGACGGT ATGTTCCCCAT
TGGGGCGGGA TAACTGCAGT TATATGCA GTAAAGGGTA
CMV

1361 AGTAACGCCA ATAGGACTTT TCCATTGACG TCAATGGGTTG TCATTCGGGT TATCCCTGAA AGGTAACCTG AGTTACCAC CMV

1401 GAGTATTTAC GGTAAACTGC CCACTTGGCA GTACATCAAG CCTCATAAAATG CCATTTGACG GGTAACCGGT CATGTAGTTC CMV

NdeI

1441 TGTAATCATAT GCCAAGTAGC CCCCCTATTG ACGTCAATGA ACATAGTATA CGTGTCACTG GGGGATAAAG TGAGTACTG CMV

1481 CGGTAATGGG CCCGCTTGGC ATATGCCCCA GTACATGACG GCCATTTACC GGGCGGACCG TAATACGGGT CATGTACTGG CMV

1521 TTATGGACT TTCTCTACTG GCACTACATC TACGTATTAG AATACCCTGA AAGGATGATG CGTCATGATG AATGCATATCG CMV

1561 TCATCGCTAT TACCATGGGTG ATGCGGTTTT GGCAGTACAT AGTACCGATA ATGGTACCAC TACGCCAAAA CGGTATGCTA CMV

1601 CAAATCGGCTG GGATGCGGAT TGAAGTACAG GGGATTTCCA GTTACCGCCA CCACTGCGGA AACTGAGTGC CCAAGTTGCT CMV

1641 AGTCTCCACC CCATTCAGCT CAATGGGAGT TTGTTTGGGC TTCAGAGGTTG GGTAACTGCA GTTACCCCTA AACAAAAACCTG CMV

1681 ACCAAATCA ACGGGACCTTT CCAAATGTGC GTAACAACCTC TGTTTTTAGT TGCCCTGAAA GGTATTACAG CTATTGAGG CMV

1721 CGCCCCATTTG ACAGCAATTGG GCGTACGGCG TGTAAGGTGGA GCCGGGTAAC TACGTTTTACC CGCCATCCGC AACATACC CMV
1761 GAGGTCTATA TAAGCAGAGC TGGGTACGTG AACCCTACGA
    CTCCAGATAT ATTGCTCTCG ACCCATGCAC TTGGCAGTCT
    Gamma1
PmeI

Esp3I  BglII  NheI

1801 TCGCTTGGAACGCACCATCAC AGATCTGTTTAAACGCTAGC
    AGCGGACCTCTGCGTATGG TCTAGACAAA TTTGCCATCG
    Gamma1

1841 ACCAAGGGGC CCATCGTCCTT CCCCCGGGAC CCCTGCTCCA
    TGGTTCGCGGG GTAGCCACAA GGGGGGACGT GGGAGGAGGT
    Gamma1

1881 AGAGCACTCTCCGCGGACA CCGGCGCTGG GCTGCCGCTGG
    TCTCGCTGGAG ACCCCCGGTGT CCGCGGACCC GCACCGACCA
    Gamma1

AgeI

1921 CAGGACTAC TTCCCCGAAC CGGTCGACGGT GTCGTCGGAAC
    GTTCTCGATG AAGGGGCTTT GCCACTGCCA CAGCACCTTG
    Gamma1

1961 TCAAGGCAGCC TGAACAGCGG CGTGCACACC TCCCGGCTTG
    AGTCCCCGGG ACTGTCGCCG GCACGTGGCT AAGGGGCGAGC
    Gamma1

2001 TCCTACAGTC CTCAAGACTCT TACCTCCTCA GCAGCGTGGT
    AGGATGTCAG GAGTCGATCC AATAGGGGAST GTCCGACCCA
    Gamma1

2041 GACCGTGCCC TCCAGCACTCT TGGGCAACCA GACCTACTCT
    CTGCAGCGGG AGGTCTCCGA ACCCGTGGGT CTGGATGTAG
    Gamma1

2081 TGCAACGTGGA ATCACAAGCC CACCAACACC AAGGTGACCA
    ACGTTGCACT TATGTTTCCG GTCTGTGTTG TTCCACCTGT
    Gamma1

2121 AGAAAGGTTGA CCCCAAATCTTT TGTCGACAAAA CTCAACCATAG
    TCTTCACACT CGGGTTTAGA ACACTGTGTG GAGTGTGTAC
Gamma1

2161 CCCACCGTGC CCAGCACCTG AACTCCTTGGG GGGACCGTCA
GGGTGGCACG GTGCGTGGAC TTGAGGACCC CCCTGGCAGT
Gamma1

2201 GTCTTTCCTCT TCCCCAAAAC ACAAGGAGAC AGCCCTCATGA
CAGAAGGAGA AGGGGGGTTTT TGGGTTCTTG TGGAAGTACT
Gamma1

2241 TCTCCCCGAC CCCTGAGGTC ACATGCGTGG TGGTGGACGT
AGAGGCGCTT GGGACCTCCAG TGTAACGACC ACCACCTGCA
Gamma1

2281 GAGCCACGAA CACCTCTAGG TCGAGGTCAA CTGATACCTG
CTCGGTGCTT CTGGGACTCC AGGTCAAGTT GACCATGCAC
Gamma1

2321 GACGCGCTTG AGGTTGATTAA TGCCAGACAA AAGCCCGCGCC
CTGCCGCAACC TCCACGTATT ACGGTTCTGT TCGGCGGCC
Gamma1

2361 AAGAGCAGTA CAACAGCACG TACCGGTCTG TGACGGTCCT
TCCGCGTCAT GTGCGTCGTC ATGGCACACC AGTCGCGAGA
Gamma1

2401 CACCGGTCCTG CACCGAGACT GGCTGAATGG AAGGAGGACTG
GTGGGAGGAG CTTGACTCTTA CCGACTTACC GTCCTCTAGT
Gamma1

2441 AAGTGCAAGG TCTCCACCAA AGCCCTCCCAG CCCCCCATCC
TTCAGTTTCC AGAGGTGTTT TCGGAGGAGG CGGGGTAGGC
Gamma1

2481 AGAAAAACCT CTCCAAGGCC AAAGGCGAGC CCCGAGAACCC
TCTTTTGGTA GAGGTTCCGG TTCCCCGTCG GGGCTCTTGG
Gamma1

BsrGI SmaI

2521 ACAGGCTGTC ACCTCGGCCCC ATCCCAGGGGA TGAGGCTGACC
TGTCACATG TGGGACGGGG GTAGGGCCCT ACTCGACTGG
Gamma1
2561 AAGAACCAGG TCAGCTGGCT CGTCCCTGGTCT AAGGCTTTCTTTCT
TTCTTGGTCTC AGTCCGGACTC GACGGACCCAG TTTCCGAAGA
Gamma1

2601 ATCCCCAGCGA CATCGGCGGCTG GAGTGAGGAGA CAAATTGGGCA
TAGGGTGCTCT GTACCGGCTG CAGCTCCTCT CTCACTCCTCT GTTACCGGGT
Gamma1

2641 GCCGGAGAAGA AACTACAAGA CCAAGCCTCC CGTGGCTGGAC
CGGCTCGTACG TGTGATGTTCT GTTGCGGGAGG GCACGACCTG
Gamma1

2681 TCCGACGGCT CCTTCTCTCT CTACAGCAAG CTCACGCTGG
AGGCTTGGCCGA GGAAGAGGGA GATGTCATTC GAGTGGCACC
Gamma1

2721 ACAAGAACAG GTGGCAAGCAG GGGAACTGCT TCTCATGCTC
TGTTCTCGTC CACCGCTCCT CTGCCAGGA AGATACGAG
Gamma1

2761 CGTGATGCAT GAGGCTCTGC ACAAACACTA CAGCAGGAG
GCACTACGTA CTCCGAGACG TGTTGTTGAT GTGCGTCCTCT
Gamma1

-----------------------------------
EcoRV

---------
BclI

2801 AGCCTCCCTCC TGTCTCCCGGG TAAATGAGGA TCCAATTCGA
TCCGAGAGG AGAGAGGCC ATTTAACTCT ATGGTACAGCT
2841 CTGATGGCAAC GGTACCGGCA AACCTAAGACT GGAATCGGTGA
GACTACGTTG CCAATGGGG ATGGATCTGA CCAAGACACT
BGH

---------

2881 CAACATGCGG CGGTGATATAC TACGTTAGAT CAGGCTCAGAC
GGTGACGCCG CTACACTAGT ATGCAATACTA GGGACATCCTG
BGH

---------

2921 TGTGCCCTCT GATGGGCCCAG CATCTGTTGCT GTGCGGGCTCC
ACAGGAGAGA CTCAACGCGCAT GCTAGACAAA AACGAGGAGG
BGH

---------

2961 CCCGTGCTTT CCGTCACCT GGAAGGTTGGC ACTCCCCACGT
GGCGACGGAA GGAAGCTGGGA CCTCCACGG TGAGGCTGAC
BGH

3001 TCTTCTCTA ATAAATTGAG GAAATTGCAT CGCATTTGCT
AGGAAAGGAT TATTTTACTC CTTTAACGTA GCGTACACAG

BGH

3041 GAGTAGGTTG CATTCTATTC TGGGGGTGGG GGTGGGGGCA
CTCATCCACA GTAAGATAAG ACCTCCCCAC CCACCCCCGT

BGH

3081 GACAGCAAGG GGAGGATTTG GGAAGACAAAT AGCAGGGCATG
CTGTCGTTCC CCTCCTCATT CTTCTGTTA TCGTCGGTAC

BGH

3121 CTGGGGAATGC GGGGCTCCTCT ATGCTTCTTG AGGGCGGAAG
GACCCCTACG CCACCCGAGA TACCGAAGAC TCCGCTTTTC

Beta

Eco47III SpeI

3161 AACCAGCTGG GGTCTCGACG GCCTGCGCTG GCCTCGACTA
TTGGTCGACC CGAGCTGTC GCCTACGCTAG CGGAGCTGAT

Beta

SpeI

3201 GTAGCTTTTG CTTCTCAATTT CTTATTGCA TAATGAGAAA
CATGCAAAACG AAGAGTTAAA GAATAAACGGT ATTACTCCTTT

Beta

3241 AAAAGGAATAA TTTAATTTTT CACCAATTCA GTAGTTGATT
TTCTCCTTTTT CATTTAATTG GTGTAAAGTT CATCAACTAA

Beta

3281 GAGCAATATGC GTGCGCAAAA AGGATGCTTT AGAGACAGTG
CTGTTTACCG CAACGGTTTT TCCTACGAAA TCTCTGTCAC

Beta

3321 TTCTCTCGAC AGATAAGGAG AACAAAAATT CAGAGGGAGT
AAGAGACGTG TCTATTCTTT TTTGTAATAA GTCTCCCTCA

Beta

3361 ACCCAGAGCT GAGACTCCCTA AGGCCAGTGAG TGGCACAGCA
TGGGGTCGCA CTCTCAGGAT TCGTCACTC ACCGCTCAGT
Beta

3401 TCCAGGGACA AATATGCTTG TCATCACCGA AGGCTGATTCC AGGTTCCCTCT TTATACTGAAAC AGTACGTTGCT TCGGACTAAG
Beta

3441 CGTAGAGCAGCA CACCTCGGTA AAGGCACAATC TGCTCACAAC GACATCTCGG GTGGGACCATA TCCCCTGTAG ACAGATGTGT
Beta

3481 GGATAGAGAG GGCAGGAGCC AGGCGAGAGC ATATAAAGGTG CCTATCTCTC CGTGCTCGGAG TCCCCGTCTCG TATATTCCAC
Beta

3521 AGGTAGGATTC AGTGGCTTCTT CACATTTGCT TCTGACATAG TCCCCATCTAG TCAACGAGGA GTGTAACGAA AGACCTGTATC
Beta

3561 TTGTTGGTGG ACGCTGCTGATA GCCAAAAGGG GGGACAGTTC AACACAAACCC TCGAACATTAT CGAACCAACCC CCCCCTGTGAG
3601 AGGGCTCGCGA TTTCGCGGCAA AACTTGAAGG CAATCCTAGC TCCCCGACCGT AAAGCGCGGT TTGAACGTGCC GTAGGATCAG DHFR

3641 GTGAAGGCTG GTAGGATTTT ATCCCCGGCTG CCACTCATGGT CACTTCCGAC CATCCTAAAA TAGGGGCGAC GTAGTACCA DHFR

3681 TCGACCATTG AACTGCATCG TCCTCGGTGTC CCAAAATATG AGCTGGTAAC TTGACGTAGC AGCGGCACAG GGTCTTTATTAC DHFR

3721 GGGATTGGCA AGAAGCAGGAGA CCTACCCCTGGA CCTCCTCGCTCA CCCTAACCCTG TCTTGCCCTCT GGAATGGGACC GGAGGCAGGT DHFR

3761 GGAACGAGGGT CAAGTACCTTC CAAAGAATGAA CCAACACCTCC CCTGCTCAGA GTTCTGAAAG GTTTCTTACT GGTGTTGGAG DHFR

3801 TTCAGTGGGAA GGTAAACAGGA ATCTGGTGAT TATGGGTAGG AAGTACCCCTC CCAATTTCTCT TAGACCACCTA ATACCACATCC DHFR
3841 AAAACCTGGT TCTCCATTCC TGAGAAGAAT CGACCTTTAA
TTTGGAGCCA AGAGGTAAGG ACTCTTCTTA GCTGGAAATT
DHFR

3881 AGGACAGAAT TAATATAGTT CTCAGTAGAG AACTCAAAGA
TCCTGTCTTA ATTTATACCA GAGTCATCTC TTGAGTTTCT
DHFR

SacI

3921 ACCACCACGA GGAGCTCATT TTCTTGCCAA AAGTTTGGAT
TGTTGGTGTG CTCTGAGTAA AAGAACGGTT TTCAAACCTA
DHFR

AfIII

3961 GATGCCCTAA GACTTTTGGGA ACAACCCGAA TTGGCAAGTA
CTACGGAATT CTGAATAACT TGTTGGCCTT AACCGTTCCAT
DHFR

4001 AAGTGAGACAT GTTTGGGATA GTCGGAGGCA GTTCGTTTTA
TTCATCTGTA CCAAACCTAT CAGGCTCCGT CAAGACAAAT
DHFR

4041 CCAGGAAGCCC ACTGATCAAAC CAGGCCACCT CAGACTCTTCTT
GGTCTTTCCG TACTTTAGTTG GTCCGGTGGA GTCTGAAGAA
DHFR

4081 GTGACAAAGGA TCAAGCAGGA AATTGAAAGT GACAGGTTCCT
CAGCTTTTCT AGTACGTCTT TAAACTTTCA CTGTGCAAAA
DHFR

4121 TCCCCAAGAAA TGGATTGGGG AAATATAAAC TTCTCCCAGA
AGGGTCTTTA ACTAAACCCC TTATATTTG AAGAGGTTCT
DHFR

4161 ATACCGAGGC GTCTCTTCCTG AGGTCCAGGA GGAAGAAAGGC
TATGGTGCCG CAGGAGAAGC TCCAGGTCCT CTTTTTCCG
DHFR

4201 ATCAAGTATA AGTTTGAAAGT CTACGAGAAG AAAGACCTAAC
TAGTCTATAT TCACACACTCA GATGCTCTTC TTTTGATTG
4241 AGGAAGATGC TTCTAGGTTT TCTGCTCCCA TCCCTAAAGCT
TCCTCTACG AAAGTTCAAG AGACGAGGGG AGGATTTCGA
ATGCATTTTT ATAAAGCCAT GGGACTTTTG CTGGCTTTAG 
TACGTAAAA TATTCTGGTA CCCTGAAAAC GACCCGAATC 
BGH

ATCAGCCTCG ACTGTGCCTT CTAGTTGCAA GCCATCTGTT 
TAGTCGGGAC TGACACGGA AATCAACGGT CGGTAGACAA 
BGH

GTTTGCCCCCT CCCCCGTCGCG TTCTCTGACC CTGGAAGGTTG 
CAAACCGGGA GGGGCCACCG AAGGAACTG GACCTTCCAC 
BGH

CCACTCCAC TGTCCTTTCC TAAATAAAATG AGGAAATTGC 
GGTGAGGGTG ACAGGAAAGG ATTTATTTAC TCCCTTAACG 
BGH

ATCCGATTTGT CTCGATAGGT GTCATCTCAT TCTGGGGGGTT 
TAGCGTAAACA GACTCATCCA CAGTAAGATA AGACCCCCCA 
BGH

GGGGTGGGGC AGGACAGGAA GGGGGAGGAT TGGGAGGACA 
CCCCACCCCG TTCTGCTGTT CCCCTCTCCT ACCCTTCTGT 
BGH

ATAGCAGGCA TGCTGGGGAT GGGTGGGGCT CTATGGCTTC 
TATCGTCCGT ACGACCCCTA CGCCACCCGA GATACCGAAG 
BspEI

---

TGAGGCGGAA AGAACAGGCT GGGGCTCGAA GCGGCCGCTC 
ACTCCGCCCTT TCTTGGTCCA CCCCGAGCTT CGCCGGCGAG 
BspEI

---

CGGATCGAGG CCCGCTACTAA CCTCTCCCTC CTCCTCCTTTT 
GCCTAGCTCC GCGATGATT GAGAGAGGAG GAGGAAAAAA 
N2

MscI

---

CCTGCAAGGAC GAGGCAGCGG CGCTATCGTG GCTGGCCCCAG 
GGACGTCTCTG CTCCGTCCGC CCGATAGCAC CGACC GGGTGC 
N2

---

ACGGGGCGTTT CGTGCAGCAGC TGTGGCTCAGC GTGGTCACTG 
TGCCCGCAAG GAACGGGTCAG ACACGAGCTG CAACAGTGA
N2

4721 AAGCGGGGAG GGACTGGGCTG CTATTGGGCG AAGTGCGCGG
TTGCCTCCTC CCTGACGGAC GATAACCGGC TTCAGGCCCC
N2

4761 GCAGGATCTC CTGTGATCTC ACCTTGTCTC TGCCGAGAAA
CGTCCATGAG GACAGTGAAG TGGAGCCAGG ACGGCTCCTTT
N2

4801 GTATCCATCA TGGCTGATGC AATGCAGCCGG CTGCACTACGC
CATAGGTGTG ACCGACTACGC TTACGCCCGC GACGTATGCG
N2

4841 TTGATACCGG TACCTGCCCA TTGCAGCCACC AAGCGAAACA
AATAGGGCCG ATGAGCCGGG TAACTGGTTG TTCGCTTGT
N2

4881 TCCGATGAG CGGAGCGCTA CTCGAGATGA AGCCGCTCTT
AGCGTACGTC GCTCCTGCTG GAGCTACCT TCCGCCAGAA
N2

4921 GTGCAATGAG ATGATCTGGA CGAAGACCAT CAGGGGCCTCG
GAGCTAGTTC TACTACCGCT GCTTTTCTGTA GTCCCCGAGC
N2

4961 CGCCAAGCGA ACTGTTGGCG AGGCTCAAAG CGGCCGATGCG
GCGGTGGGCT TGAAGAAGCGG TCCGAGTTCC GCGGCTACGG
N2

5001 CGACGCGCGG GATCCTGGTG TGACGCGATG CGAGCTGCTGC
GCTGCCTGTC CTAGAGCCACG ACTGGGTACG GCTACGCGACG
N2

5041 TTGCAGAATA TCGATGTCGA AAATGGCCCG TTTTCTGAGAT
AACGGCTTAT AGTACCACCT TTTACCGCGG AAAAGATCTTA
N2

---------------------------------------------
RsrII
---------------------------------------------

5081 TCATGACTG TGGCCGCGCTG GGTGTCCGGG ACCGCTATCA
AGTAGGTGAC ACCGCGCGAC CCACACCGCC TGCCGATAGT
N2

---------------------------------------------
GGACATAGCG TTTGGCTACC CCCTGATATGC TGAAGACCTT
CCTGATCCGA AAACGATGGG CACTATAACG ACTTCTCGAA
N2

GGCGGCCGAT AGGCTGACCG CTTCCTCGTG CTTTACGGTA
CCGCGCGCTTA CCGACTGGC GAAGGAGCAC GAAATGCCC
N2

TCGCCGCCTCC CGATTCGCAG CGCATCGCCT TCTATCGGCT
AGCGGCGAGG GCTAAGCGTC GCGTAGCCGA AGATAGCGGA
N2

EstBI

TCTTGAGCGAG TTTCTTCGAG CGGGACTCTG GGGTTGCAAA
AGAATCGCTC AAAGAGACTC GCCCTGAGAC CCCAAGCCTT
TGACCGACCA AGCGGCGCC AACTCTGCAAT CAGAGGTATT
ACTGGCTGGT TCGCTGGCGG TTGGACGGTA GTGCCTAAA

CGATTCGACCC GCCGCTCTCT AGGAAAGGTT GGGCTTCGGA
GCTAGGTGGC CGCCGGGAAGA TACCTGCAAA CCCGAAGCCT

ATGATTAAAA GAGACGGCGG CTGGAAGTCT CCACACGGCG
TAGAAAAAGG GACCTGACCC GAGGTGCGGC

GGATCTCAT GCTGAGGTTC TTGCCCAAC CGACCTCTAG AAGCGGGTGG GTTGAACAAA

ATTCAGCCCTT AATATGTTTA CAATAAAAGC AATAGCATCA
TACGTGCGAA TATTTACATT GTTTATTCG TTATGCTAGT

BsmI

CAAATTTTAC AAATAAACG CTAGTTTTTCC AGCATTCTAG
GTTTAAAGTG TTTTTTCTGT AAAAAAAGTG ACGTAAGATC

TTGTTGGTTTG CTTTAAACTCA TCAATGTATC TTATACGTAC
AACAACAAAC AGGTGAGTG AGTTACATAG AATAAGTACAG

TGGATCGCGGA CCGCGCGGCA CGCCGGTGGA GTTTTAATA
ACCTAGCGGC GCGCGCGCGG GCGCCACCT CGAAATTAAT
AscJ
-------

5601  AGCCGCGCCA  GCTCCAGCTT  TTGTTCCCTT  TAGTGAGGGT
tccggccggt  ccaggtcgaa  aacaagggaa  atcacctcca
5641  TAATTTCGAG  CTTGGCGTAA  TCCATGGTAC  AGCTGTTTCC
attaaagcctc  gaacccgatt  agtaccagta  tcgacaagg
5681  TGTGTGAAAT  TGTTATCCGC  TCACAAATCC  ACAACAAAATA
acacactttta  acaatagggcg  agtgtaaggg  tgtgttgtat
5721  CGAGCCCGGAA  GCATAAAGTG  TAAAGCCTTG  GGTGCCCTAT
ctcgcattt  gtaattccac  atttcgggacc  ccaggatat
5761  GAGTGAGCTA  ACTCATACTA  ATTGGCTTGCC  GCTCAGCTGCG
ctcactcgat  tgagtgtat  tcagcagaag  cagtgagccgg
5801  CGCTTTTCGAG  TCGGGAACCC  TGTCGATGCC  GCTGCAATTAA
ccaaagagggc  agccctttgg  acagacaggg  ccagctaatat
5841  TGATTCGAGC  AAACCGCGGG  GAGGGACGGT  TTGCTGTATAG
acttagcggc  ttgccccggc  ctctccgcac  aacgcataac
5881  GGCCTTCTCC  CCGCTCTCCG  CTCACTGACTG  CGCTGGCTC
cccgagaagaa  cgggagggag  gtagcagtgg  cgcacgaggg
5921  GGTCGTTCGG  CTGCGGCGAG  CGGTATGCTG  TCACTCAAAG
ccagcaagcc  gacgccgctc  gcatagctgc  agtgagtttcc
5961  GGCTTAATAC  GGTATCCAC  AGAATCAGGG  GATACCCGAG
cccatcattg  ccaataatgg  tcttacttcctc  cattgcgcgct

BspLUIII
-------

6001  GAAAGAACATG  GTGAGCAAAA  GCCGAGCAAA  AGGCGAGGAA
cttctgtgta  cactgtggtt  cgggtgtggttt  cgggtgtcctt
6041  CCGTAAAAAG  GCCGCGTTGC  TGCGGTTTTT  CCATAGGCTC
GGCATTTTTC  CCAGGCAACGG  ACCGAAAAAA  GGTATGGCAG
6081  CGCCCCCTCCTG  AGCACATCTCA  CAAAAATCTGA  CGCTCAAGTC
GGGGGGGGAGC  TGCTCGTAGT  TGTTTAGCTG  CGGATGTCG
6121  AGAGGTGGGG  AAACCCGACA  GGACTATAAA  GATACCCGACG
TTCACCCGCG  TTGGGCTGTT  CCTGATATTT  CTATGGTGGC
6161  GGTACCCCGCG  GAAAGCGCTC  TCGTGCGTCC  TGCTGTTCGG
CAAAGGGGGA  CTCCTGAGGG  AGCAGCCGAG  AGGACAPAGGC
6201  ACCCTGCGCG  TTACCGGATA  CCGTGGCGGC  TTTCCTCCCT
TGGGACGGCG  AATGCGCTAT  GGACAGGCGG  AAAGAGGGAA
6241  CGGTAAACTG  CGCGTCTTCG  CATAGCTCAC  CCGTGTTGAGA
GCCCTTCGCA  CCGCGAAAGA  GTATGAGTGG  GACATCCAT
6281  TCTTCGTTCG  GTGTAGGGCG  TTCGCTCCAA  GCTGGCTGCTG
AGACTCAGGC  TCCATCCAGC  AAGGCAAGGT  CGAACCACACA
6321  GTGCACGACG  CCCCCGTTCGC  GCCGACCAGC  TGGCCTTATAT
CAGTGCTTGG  GGGGGCAAGT  CGGGTGGCAG  AGGGGGATAG
6361  CCGTAACTCA  TCGTCTTGGG  TCCACCCCGG  TAAGAACAAGC
GCCCCATGGAT  AGCAGAATCTC  AGGTGGGGCC  ATTCTGTGCT
6401 CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTA GC
GAATAGCGG GTACCCGTAC GGTGACCATT GTGCTATA TG
6441 AGAGCCAGGG ATGTAGCAGG TGCTACAGAG TTTCTGAATG
TCTCGCTCCA TACATCCGGC ACCATGGTTC AAGAACTTCA
6481 GGTGCGCTAA CTACGGCTAC ACTAGAAGAA CAGTATTTTG
CGCCACGGATT GATGCGGATG TGATCTCCTT GTCATAAACCC
6521 TATGCTGCTT GTGCTGAAAGG CAGTTAACCTT CGGAAAAAGA
CATAGACCGGA GACGACTTCG GTCAATGGGA GCTTTTTTCT
6561 GTGTGTAAGCT CTGTGATCGGA CAACAAGACC ACCGCTCGTA
CAACATCGCA GAAGCTCGCC GTTTTGTTTG TGCGGACCAG
6601 GCCTGCTGTTT TTTGCGTCTC AGCAGCTAAG TATGCGCGTCA
CGCCACAAAG AAACAACACAG TCGTGCCTCT AACAGCGGATC
6641 AAAAAAAAGGA TCTTCAAGAA ATGCTTTGAT CTATCTTACG
TTTTTTTCTT AGAGTTCCTC TAGGAAACTA GAAAGATGC
6681 GGGTCTGACG GATCGAGGAA GCAGAACTCA GTTAAAGGA
CCCAGACTGC GAGTCACCTT GCTTTTGGAT GCAATTCCCT
6721 TTGTGTTGAT AGATTTATCG AAAAAAGATCT TCACCTAGAT
AAAAACCTGA CTTAATAGTG TTTTCTTGA AGTGTGATCTA
6761 CTTTTTAAAT TAAAATGGA GTTTTITTAA ATCTAAATGC
GGAAAAATTTA ATTTTTACTT CAAAATTTAG TTGATTTTCA
6801 ATATATAGGT AAACATTTGC TGACAGTTAC CAAAGCTTAA
TATATACCTCA TTTGAAACCAG ACTGTCAATG GTTACGATT

----------------------------------

Amp

6841 TCAGTGAGGC ACCTAATTCA CCGATCTGTC TATTTCGTTC
AGTCACTTCCG TGGATAGAGT CGCTAGACAG ATAAAGCAAG
----------------------------------

Amp

6881 ATCCCATAGTT GCCTGACTCC CCCGTCTGTA GATAAATACG
TAGGTATCAA CGGACTGAGG GGCAGCACAT CTATGATGC
----------------------------------

Amp

6921 ATACGGGGAGG GGTTACCACCT TGGCCCCAGT GCTGCAAAGA
TATGCCCCTCC CAAATGGTAG ACGGGGTCA CGACGTACT
----------------------------------

Amp

6961 TACCGCCGGA CCCCAGCTCA CCGGCTCCAG ATTTATCAAC
ATGGCGCTCTC CGGTGCGAGT CGCCGAGGTC TAAATATCG
----------------------------------

Amp

7001 AATAAACACG CAGGCCGAGA CGGCGAGGCG CAGAAGTGCT
TTATTTGGTC GGTGCGCCCT CCAGGCTCCG GTCTTACCA
7041 CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATGGTT
GGACGTTGAA ATAGGCGGA GAGGTCAGA TAATTAACAA

-------------------------

Amp

7081 GCCGGGAAGC TAGAGTAAGT AGTTCGCCAG TTAATAGTTT
CGGCCCTTCG ATCTCATTCA TCAAGCCGTC AATGATCAA

-------------------------

Amp

7121 GCGCAGAGTT TGGCCCATTT CTACAGGAC ATGTGTTGCA
CGCGTTGCAA CAACGGAATAC GATGTCCGTA GCACACAGT

-------------------------

Amp

7161 CGCTCTCGCT GTGGTAGTCC TTTACATTAC TCCGGTTCCCC
GCGAGGACGA AACCATAACG AAGTAAGTCG AGGCCAACGG

-------------------------

Amp

7201 AACGATCAAG GCCGAGTTACA TGATCCCCCA TTGTGTCGAA
TTGCTAGTTC CGCTCAATGT ACTAGGGGGT ACAACACGTT

-------------------------

Amp

7241 AAAAGCGGGTT AGCTCTCCAG TGCTTCCGAT CTGGTTCAGA
TTTTCCGCAC GCGAGGAAAC CAGGAGGCTA GCAACAGTCT

-------------------------

Amp

7281 AGTAAGTTGG CCCGAGTGTT ATCACTCATG ATTAAGGCAG
TCATCAACC GCCTCGCACA TAGTGAGTAC CAATACCGTC

-------------------------

Amp

7321 CACTGCATAA TTCTCTTACT GTCACTCCAT CGAGTCAGTG
GACGAGTTAT AAAAGAAATCA CAGTAAGGTA GCACATTCTAC

-------------------------

Amp

7361 CTTTTCTGTG ACTGTTGAGT ACTCAACCAA GTCAATTCTGA
GAAAAGACTAC TGACACTACA TGAGTTGTTT CAGTAAAGACT

-------------------------

Amp

7401 GAATAGTGTA TGCGCGCCACG GAGTGTGTCTCT TGGCCCGCGT
CTTATCAGAC ACTCAGCTGGA CTCAGACGGAG ACGGGCGCGA

-------------------------

Amp

7441 CAATACGGGA TAATACCCCGC CCACATAGCA GAACCTTTAAA
GTTATGCGCT ATTATGCCCG GGGTATCGT CTGGAATTTT

-------------------------

Amp
AGTGCTCATC ATTGAAAAAC GTTCTTCGGG GCGAAAAACTC
TCACGAGTAG TAACCTTTTG CAAGAAGCCG CGCTTTTGAG

Amp

TCAAAGATCT TAACGCTTGT GAGATCCAGT TCGATGTAAC
AGTTCTCTAAG ATGCGGACAA CTCTAGGTGA AGCTACATTTG

Amp

CCACCTCGTGCC ACCCAACTGA TCTTCAGCAT CTTTTAATTTT
GGTGACACGG TGGTTGACT AGAAGTCGTA GAAAATGAAA

Amp

CACCACGCTTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT
GTTGTCGCAA AGACCCACTC GTTTTTGTCC TTCCGTTTTA

Amp

GGGCAAAAAA AGGGATAAAG GGGGACACCG AAATGTTGAA
CGGCGTTTTT TCCCTTATTT CCGCGTGGCC TTTCACACTT

Amp

SspI

TACTCATACCT TTTCCCTTTTT CAATATTATT GAAGGATTTA
ATGAGTATGA GAGGAAAATAA GTTATAATAA CTTGGTAAAT

Amp

TCAGGGTTAT TGTCACATGA GCGGATACAT ATTTGAAATGT
AGTCCAATA ACAGAGTACT CGCTATGTAG TAAACTTACA

Amp

ATTTAGAAAA ATAAAACAAAT AGGGTTCCCG GCACATTGTC
TAAGATCCTTT TATTTTGTTTA TCCCCAAGGC GCGTGAAGA

Amp

CCGAAAAGT GCCAC
GGGCTTTTCA CGGTG
HuCC49 Gly/Ser in Polycis 2

[Diagram of gene structure with various labels and restriction sites]
CGTACGCCGCC CGTACGCCGCC GCATTAAGCG CGCGGGTGTC
GACTGCAGCC GACATCGCCG CGTAATTGCG GCCGCCACA
GTTGGTTACG CGCAGGCTGA CCGGTACACT TGGCAGGC
CCACAAAAGG GCAGTCGCCACT GCGATGTA ACGGTCCGGC
CTAGCCGCCG CTCTTTTCGC TTCTTTCTCG TCCTCTCCTG
GATCGCGGGCC AGGAAGAACG CGAAGAAGGA AGGAAAGAGC
CCACGTTCGG CCAGTTTCCC CTCAAGGCTC TAAATCGGGG
GCTGCAAGCG GCCGAAAGGG GCAGTTCCAG ATTTAGCCCC
GCTCCTTTTA GGGTCGGAT TTAGTGGTTT ACGGACCTC
CGAGGGAAT ATCAAGCGAA TGGCCGTGGAG
GCCACCAAAA AACCTTCTATTA AGGAGATGGT TCACGTAGGG
CTGGGGTTTT TTGAAACTAAT CCCACTACCA AGTGCACTCC

---

TGCAGACGTAA CGCGGCCCCC CCTCGATTAA TTATCAGACG
AGCGCTGCAAT GGCGCGGGGG GGAGCTAATT AATTAGCTCG

Beta

TACTAGCTTT GCTTCTCAAT TTCTTTATTTG CATAATGAGA
ATGATCGAAA CGAGAGTTA AAGAATAAACC GTATTACTCT

Beta

AAAGAGGAAG AAATTAATTTT AACACCAATT CAGTAGTTGA
TTTTTTCTTT TTATTTAAAA TTGTGGTTAA GTCATCAACT

Beta

TTGACAAATTC CGGTGGCCAA AAAGGATGCT TTAGAGACAG
AAGCCTGGTTA CGCAAGGGTT TTTCTACGA ATTCTCTGTC

Beta

TGTTCTCTGC ACAGATAAGG ACAAACAITTA TTCAGAGGGA
ACAAGAGAGCG TGCTTATTCG TGTTTTGTAAT AAGTCTCCCT

Beta

GTACCAGAG CTGAGACTCC TAAGCCAGTG AGTGCCACAG
CATGGGTCTCC GACTCTGAGG ATTCGGTCAC TCACCGTGTC

Beta

CATCCAGGGA GAATATGCT TGTCATCACC GAAGCCTGAT
GTAGGTCCCT TTTTATAAGA ACAGTAGTGG CTTCGGACTA

Beta

TCCGTAGAGC CACACCCCTGG TAAGGGCCAA TCTGCTCACA
AGGCAATCTCG GTGTTGGACC ATTCGGGGTT AGACAGGATGT
Beta
561 CAGGATAGAG AGGGCCAGGAG CGGGCCGACA GCATATAAGG
GTCCTATCTC TCCCGTCTCTC GGTCCCGTCT CGTATATTCC
Beta
601 TGAGGTAGGAG TCAGTTGCTC CTCACTATTG CTTCTGACAT
ACTCCCATCCT AGTCAACGAG GAGTGTAAAC GAAGACTGTA
N1
--------
ClaI
-------
641 AGTTTGTCGGA GCAAGGAAGGA ATCGATCTCTC CATGCTTGGAA
TCAACACGGT CTAGAAGGAG GTACGAACTT
N1
--------
681 CAAGATGGAT TGGAGCGAGG TCTCTCGGCC GCTGGGGTGGC
GTTCTACCTA AGCGGCGGTC AAGAGGCCGG CGAACCACC
N1
--------
721 AGAGGCTATT CCCTATGAC TGCCAGACAAC AGACAATCCGG
TCTCCGATAA GCCGATACTG AACCGTGTTG TCTGTTAGCC
N1
--------
761 CTGCTCTGAG GCCCGGCGGTT GCCGTCTCTC AGGCCAGGGG
GACGAGACTA CGGCAGCCACA AGGCCGACAG TCAGCTCCCC
N1
--------
801 CGCCCGGTTC TTTTTGCTAA GACCGACCTG TCCGGTGCCC
GCGGGCCAG AAAAACAGTT CTGGGTGGAC AGGCCACGGG
SVO
--------
StuI
------
841 TGAATGAACT CGAGGTAAGT CCGCCCGCTC TAGGCCTCCA
ACTTACTTGA CGTCCATTCA CGCGGCGAG ATCCGGAGGT
SVO
--------
SfiI
------
881 AAAAAAGCGCTC CTCACTACTT CTGGAATAGC TCAGAGGCAGC
TTTTTCCGAG GAGTGATGAA GCACCTATCG AGTCTCCGGC
SVO
Sfii
--------
921  AGGGCCGCTC GGCCTCTGCA TAAATAAAAA AAATTTAGTCA
     TCCGCCCAGG CCGAGAGCTT AATTATTTTT TTTAATCAGT
     SVO

961  GCCATGGCATG GGGCGCGAGAA TGGGGCGGAGC TGGGCGGAGT
     CGGTACGTAC CCCGCTCTTT ACCCGGCTTG ACCCGCTCTCA
     SVO

1001 TAGGGGGCGGG ATGGGCGGGG TTAGGGGCGG GACTATGTTT
     ATCCCCGCCG TACCCGGCCTC AATCCCGGACC CTGATACCA
     SVO

1041  GCTGACTAAT TGAGATGCAT GCTTTGCAAT CTTCTGCTCT
     CGACCTGATTA ACTCTAGTGT CGAAGACGTAT GAAGACGGAC
     SVO

1081  CTGGGGAGCCC TGGGGACTTT CATACCTGGG TTGGCTGACTA
     GACCCCCCTGG ACCCCCTGAAA GGTGTGGACC AACGACTGAT
     SVO

1121  ATGAGATGCG ATGCTTTTGGC TACTTCTGCC TGCTGGGGAG
     TAACTCTACG TACGAAACGT ATGAAGACGG ACGACCCTCT
     SVO

1161  CCTGGGGACT TTCCACACCC TAATCTGACAC ACATTCCCACA
     GGACCCCCGA AAGGTGTGGG ATTTGACTGG TGATAAGGGT
     SVO

     CMV

1201 GAAATTTATC CCGTCTTGGT TAATAAGTAAT CAATTTACGG
     CTTAATTAAG GGGATCAATA ATTTACATTA GTAATGCCC
     CMV

1241  GTCAATTAGTT CATAGGCCCAT ATATGGAGTT CCGCGTATACA
     CAGTAATCAAA GTATCGGGGTAA TATACCTCAAA GCGCAATGT
     CMV

1281  TAATCTACGG TAAAATGGCCC GCCTGGCTGA CCCGCCAAACG
     ATTGAATGCC ATTTACCAGGG CGACCGACT GCGGGTTG
     CMV

1321 ACCCCCCGCCC ATTTGAGTCA ATAAATGACT ATGTTCCCAT
     TGGGGGGCGGG TACTGCACT TATTTACTGCC TACAAGGGTA
CMV

1361 AGTAAACGCCA ATAGGGACCTT TCCATTGACC TCAATGGGTG
TCATTGGGTT TATCCCCGAA AGGTAACTGC AGTTACCACG
CMV

1401 GAGTATTTAC GGTAATCGGC CCACTTGGCA GTACATCAAG
CTCATAATAG CCAATTTGCA GGTGAACCGT CATGTAGTTC
CMV

       NdeI

1441 TGTATCATAT GCCAAGTACG CCCCCATTTG ACGTCAATGA
ACATAGTATA CGGTTCAATGC GGGGATAAAC TGCAAGTACT
CMV

1481 CGGTAAATGG CCCGCTTGCC ATTATGCCCA GTACATGACC
GCCATTACC GGGGGGACGG TAATACGGTT CATGTACTGG
CMV

1521 TTATGGGACT TTCTACTTGT GCAGTACATC TACGTAATTAG
AATACCCGGA AAGGATGAAC CGTCAATGAG ATGCATAATC
CMV

1561 TCTACGCTAT TACCATGGTG ATGCGGTGGTG GCCAGTACAT
AGTAGCGATA ATGGTACCAC TACGCCAAAA CCGTCATGTA
CMV

1601 CAAATTGGCGT GGTACAGCAG TTGACTCAAG GGAATTCCA
GTACCCGCCA CCTATCCGCA AACTGAGTTG CCCTAAAGGT
CMV

1641 AGTCTCCACC CCATTGACGT CAATGGGAGT TTGTTTTGGG
TCAGAGGTGG GGTAAACTGCA GTACCCTCAG AAACAAACCC
CMV

1681 ACCAAAAATCA ACCGGACTTT CCAAAATGTC GTAACAACCTC
TGTTTTTGT TGCCCTGAAA GGTTTACAG CATTTGTTAG
CMV

1721 CGC CCCATTTG ACGCAAATGG GCGTGACG GGTAGACGTG
GCGGGGTAAC TGCCCTTACC CGCCCATGGC ACATGCCACC
CMV
1761 GAGGTCTATA TAAGCAGAGC TGGGTACGTG AACCGTCAGA
CTCCAGATAT ATTCGTCCTCG ACCCATGCAC TTGGCAGTCT
  PmeI

  --------
  BglII

1801 TCGCCTGAG ACGCCATCAC AGATCTGTTT AAACGCTAGC
AGCGGACCTC TGCGGTAGTG TCTAGACAAA TTTGCGATCG
  L

1841 CTCTCACCAC GAGATGCCAG GCCCAGGTGC TCATGCTCCT
GAGAGTGGA CTTATCGGTC CGGGTCACAG AGTACGAGGA
  VL

  L

1881 GCTGCTGTTG GTGAGCCGCA CATGCCGCCA CATCGTGATC
CGACGACACC CACTCGCCGT GTACGCCGCT GTAGCACTAC
  VL

1921 AGCCAGTCTC CAGACTCCCT GGCGGTGCTCC TGGCGGAGA
TCGGTCAGAG GTCTGAGGGA CGGGCACAGG GACCGCGCTCT
  VL

1961 GGGTGACTCT GAAATTGCAAG TCCAGCCAGT CCCTGCTCTTA
CCCACTGAGA CTATACGTTC AGGTCGCTCA GGGACGAGAT
  VL

2001 TAGCGGAAT CAGGAAGACT ATCTCCTCCTG GTATCAGCGA
ATCGCCTTTTA GTCTTTCTTGA TAGAGCGGAC CATAGTGCTC
  VL

2041 AAACCAGGCC AGAGCCCTAA ACTGCTGATT TACTGGGCAT
TTTGCTCCCAG TCTCGGGATT TGACGACTAA ATGACCGGTA
  VL

2081 CCGCTAGGGA ATCCGCGCGT GCTGATCCTG TCAGCGCGA
GGCGATCCCT TAGGCGCGAC GGACTAGCGA AGTCGCCGTC
  VL

2121 CGGATCTGGG ACAAGACTTCA CTTGAACAAT CAGCAGCGTG
GCCTAGACCC TGTCGTGAAT GAGACTGTGA GTCGTCGCAC
  VL
2161 CAGGCAGAAG ACCTGGCAGT CTATTATTGT CAGCAAGTATT
GTCGCTCTTC TGCAACCCTGCA GATAAATAACA GTGCTCATAAA
VL

2201 ATAGCTATCC CCTCATACTGC GCGCTGGA CCAAGCTGGGA
TATCGATAGGG GAGGTGTAAG CCGCACCCTG GGTTCGACCT
Kappa

VL

BsiWI

2241 ACTGAAACGT ACCGTTGGCTG CACCATCTGT GTCATCTTTC
TGACTTTTCA TGCCACGAC GCTGGTACCA GAAAGTGAAG
Kappa

2281 CCGCCATCTG ATGAGCAGTT GAATCTGGTA AACTGCCTCCTG
GGCAGTAGAC TACTGTGCAA ACTCATGACCT TGACGGAGAC
Kappa

2321 TTGTGTGCTC GTGAAATAAC TTCTATCCCA GAGAGGCCAA
AACACACGGA CGACTTATTG AAGATAGGAT CTCTCCGGTT
Kappa

2361 AGTACAGTGG AAGGTGGGATA ACGCCCTCCA ATCGGGTAAAC
TCATGTCAACC TTCCACCTAT TGGGGAGATT TAGCCATTG
Kappa

2401 TCCCCAGGAGA GTGTCACAGA GCAGGCACGC AAGACAGACA
AGGGTGCTCT CACAGTGTCT CGCTCTGTGC TTCCTGTGCT
Kappa

2441 CTCACACGCT CAGCCAGCACC CTGACGCTGA GCAAAAGCAGA
GGAATGTGGA GTGTCGCTGG GACCTCGACGCT GTTTCGTCCT
Kappa

2481 CTACGAGAAA CACAAAGTCTC ACGCCCTGGGA ACGTCACCCAT
GATGCCCTTT CGGTTCTCAGA TGGCCAGCCGTC TGGCTGGTA
Kappa

2521 CAGGCCCTGTA GCTCGCCCGTG CACAAAGAGGC TTCAACAGGG
GTCGCCGACT CGAGCGGGCA GTGTCTCTCG AAGTTGTCCTC
Kappa

IRES
EcoRI

--------
2561 GAGAGTTTGG AATTCCCTCCC TCCCCCCCC CTAACGTTTAC
CTCTCACAAAC TTAAGGAGGG AGGGGGGGGG GATTGCAATG
IRES

2601 TGGCCGAAGGC CGCTTGGGAAT AAGGCCGGTG TGGGTTCGTC
ACGGGCTTCC GCGAACCCTTA TTCCGCCAC ACGCAAACAG
IRES

2641 TATATGTAT TTTCCACCAT ATTGCCGTCT TTTGGCAATG
ATATAACTA AAAGGTGCTA TAAAGCGCAA AAACGTGTTAC
IRES

2681 TGAGGGCCCG GAAACCTGGC CCTGTCTTCT TGACGACGCAT
ACTCCCAGGC CTGGGAGGCG GACAGAAGA ACTGCTCGTA
IRES

AvrII

--------
2721 TCGTGGGGGT TTTGTCCCCTC TCGCCTAAAGG AATGCAAAGGT
AGGATCCCCA GAAAGGGGAG AGCGGTTCGCA TTAGTTTCCA
IRES

HindIII

--------
2761 CTGTTGAATG TCGTAAGGA AGCAGTTCTT CTGGAAGCCTT
GACAACTTAC AGCACTTCTT TCGCAGAGA GACCTTGCAA
IRES

2801 CTTGAAAGACA AAACACGTCT GTAGGACCCC TTTGCAGGCA
GAACTTCTGT TTGTGCAGA CATCGCTGGG AAACGCTCGT
IRES

2841 GCAGAACCCTT CACCTGGCAG ACAGGTCGCT TCGGCGCCCA
CGCCTGGGGG GTGGGACGCG TGTCCACGGA ACGGCCGTTT
IRES

PmlII

--------
2881 AAGCCCAAGTG TATAAGATAAC ACCTGCAAAAG GCGGCACAAAC
TTGGTTGAC ATATTCTATG TGGAGGTTC CGCCGTGTTG
IRES
2921 CCCAGTGCCA CGTTGTGAGT TGGATAGTTG TGGAAAGAGT
GGTCAAGGTT GCAACACTCA ACCTATCAAC ACCTTTCTCA
IRES

2961 CAAAAGGCCG TCTCAGACCC TATTCAACAA GGGCTGAAG
GTTTACCGAG AGGAGTTGCG ATAAAGTTGTT CCCGACTTC
IRES

KpnI

3001 GATGCCCCGA AGGTACCCCA TTGTATGGGA TCTGATCTGG
CTACGGGTCT TTCCATGGGT AACATACCTT AGACTAGACC
IRES

3041 GGCCTCGGTG CACATGCTTTT ACATGTGTTTT AGTGAGAAGTT
CCGGAGCCAC GTGTCAGAAA TGTAACAAAA TCAGCTCCAA
IRES

3081 AAAAAAGTTC TAGGCCCCCA GAACCACGGG GACGTTGTTT
TTTTTGCGAG ATCCGGGGGG CTTGGTGCCC CTGCACCAAAL

IRES

3121 TCCTTTGAAA AACAGGATGA TAATATGGGT TGGAGCCTCA
AGGAAAACCTTT TTGTGTACTCT ATTATACCAA ACCTCAGAGT
L

MluI

3161 TCTTGTCTTT CCTTGTGCGT GTTGCTAGCG GTGTCTGTTC
AGAAGGAGAAG GGAACAGCGCA CACAGGATCG CACAGGACAG
L

VH

3201 CCAGGTCCAG CTGGTCGAGT CCGCGCTGTA GGTGCTGAAA
GTCCAGGTCG ACCAGCTCA GCGCGGACT CCACCACCTTT
VH

3241 CCTGGGCGTTC CCTGGAAGAT TTCTGGCAAG GCAAGCAGGCT
GGACCCCGAA GCCACTTCTTA AAGGAGGTTCC CTTGGCGCGA
VH
3281 ACACCTTCAC TGATCAGCA ATCCACTGGG TGAAACAGAA
TGTGGAAGTG ACTAGTGCGT TAGTGACC CGTCTTTGCTT
VH

3321 TCCCTGACAG CGCCCTGGAGT GGAATGGATA TTTCTCTCCC
AGGACACGTC GCGGACCTCA CTTAACTCAT AAAGAGAGGG
VH

3361 GGAAACGATG ATTTTGAAGA CAATGAGAGG TTCAAGGGA
CCTTTGCTAC TAAAATTCAT GTTACTCTCC AAAGTTCCGT
VH

3401 AGGCCACACT GACTGCAGAC ACATCTGCCA GCACCTGCTA
TCCGGTGCTGA CTGACGCTCTG TGTAGACGGGT CGTGACGGAT
VH

3441 CGTGGAGCTCT TCCAGGCTGGA GATCCGAGGA TACTGCAAGT
GCACCTCGAG AGGTCCGACT CTAGGCCTCT ATAGACGGTCAC
VH

3481 TACTTCTGCA CAAGATCTCC TGAATATGGCC TACTGGGGAC
ATGAAGACGT GTTCTAGGGA CTTATTCCGG ATGACCCCTG
G1 Gly/Ser DelCH2

VH

3521 AGGGACCCTT GGTCACCGTC TCCAGGCCTA GCACCAAGGG
TCCCTTGGA CCAAGGCGAG AGGTCCGCGAT GTGCTTTCCC
G1 Gly/Ser DelCH2

3561 CCCATCGGTC TTCCCCCCTGG CACCCCTCCTC CAAGAGCACCC
GGTGAGCCAG AAGGGGGGAC GTGGGAGGAG GTTCTCCTGG
G1 Gly/Ser DelCH2

3601 TCTGGGGGCA CAGCCGCGCT CAGGGTCGCTG GTCAAGGACT
AGACCCCCGT GTGCCGCGGA CCCGACGGAC CAGTTCTCTGA
G1 Gly/Ser DelCH2

AgeI

3641 ACTTCCCCGA ACCGGTGACG GTGTCTGGGA ACTCAGGCGC
TGAAGGGGCT TGCCACTGCA CACAGCAACT TGAGTCCGCG
G1 Gly/Ser DelCH2

----------------------------------------
3681  CCTGACCAGC  GGGCTGCAAC  CCTTCCCGGC  TGTCCTACAG
       GGACTGGTCG  CCGCAGCTGT  GGAAGGGCCG  ACAGGATGTC
       G1 Gly/Ser DelCH2

3721  TCCTCAGGAC  TCTACTCCCT  CAGCAGCGTG  GTGACCGTGAC
       AGGAGTCCTG  AGATGAGGGA  GTGCTGCAC  CACTGCGACG
       G1 Gly/Ser DelCH2

3761  CCTCCAGCAG  CTTGGGCACC  CAGACCTACA  TCTGCAACGT
       GGAGGTGCCT  GAAACCGTGG  GTCTGGATGT  AGACGTGGCA
       G1 Gly/Ser DelCH2

3801  GAATCACAAG  CCCAGCAACA  CCAAGGTGGA  CAAGAAAGTT
       CTTAGTGTTT  GGCTGTCGTG  GTTCACCTT  GTCTTTCTAA
       G1 Gly/Ser DelCH2

3841  GAGCCCCAAT  CCTGTGACAA  AACTCACA  TGCCACCCGT
       CTCGGCTTTA  GAACACTGTT  TTGATGTGCT  ACGGTTGGCA
       G1 Gly/Ser DelCH2

3881  GCCCAGACCC  TGAACCTCGT  GGGGGACCGT  CAGTCTCTCC
       CGGGTCCTGG  ACTTGAAGAC  CCCCTGGCA  GTCAGAAGGA
       G1 Gly/Ser DelCH2

3921  CTTCCCCCCA  AAACCAAGG  ACACCCCTCAT  GATCTCCCCG
       GAAGGGGGGT  TTTGGGTCTC  TGTGGGAGTA  CTAAGGGGCC
       G1 Gly/Ser DelCH2

3961  ACCCCTGAGG  TCACATCGGT  GGTGGTGGAC  GTGAGGCACG
       TGGGGACTCC  AGTGTACGCA  CCCACGCTG  CACTGGTGTC
       G1 Gly/Ser DelCH2

4001  AAGACCTCTG  GTGCAAGTTC  AACTGGTACG  TGGAGCGGGT
       TTCTGGGACT  CCAGTCGAAG  TTGACCATGC  ACCTGCGCGA
       G1 Gly/Ser DelCH2

4041  GGAGGTGCAT  AATGCCAAGA  CAAAGCCCGG  GGAGGAGCGA
       CCTCCACGTA  TTACGTTTCT  GGTCCGCAC  CCTCCTCGTC
       G1 Gly/Ser DelCH2

4081  TACACAGCAG  CGTGACGCTG  GGTCAGCCTG  CTCAGGCTCC
       ATGGTGTGCTG  GCATGGCAAC  CCAGTCGCAAG  GAGTGGCAGG
       G1 Gly/Ser DelCH2
4121 TGCACCAGGA CTGGCTGAAT  GGCAAGGAGT ACAAGTGCAA  ACGTGGTCCT  GACCGACTTA  CCGTTCTCTCA  TGGTCACGGT  G1 Gly/Ser DelCH2

4161 GGTTCCTACAA AAAGCCCTCC  CAGCCCCCAT  CGAGAAACCC  CCAGAGGGTG  TTTCGGGAGG  GTCCGGGTA  GCTCTTTTGG  G1 Gly/Ser DelCH2

BsrGI

4201 ATCTCCAAAG CCAAAGGGGA  GCCCCCCAGAA  CCACAGGTGT  TAGAGGTTCG  GGTTCCTCGG  CCGGGCTATT  GGTGTCCACA  G1 Gly/Ser DelCH2

BsrGI  Smal

4241 ACACCCCTGCC  CCACATCCCGG  GATGAGCTGA  CCAAGAACCA  TGTGGGACGG  GGGTAGGGCC  CTACTCGACT  GGTTCCTGGG  G1 Gly/Ser DelCH2

4281 GGTCAGCGCTC  ACCTGCGGTC  TCAAAGCCTG  CTATCCCAAGC  CCAGTCGGAC  TGGACGGACC  AGTTTCCGAA  GATAGGGTCCG  G1 Gly/Ser DelCH2

4321 GACATCGCCG  TGGAGTGCGGA  GAGCAATGGG  CAGCCGGAGA  CTGTAGCGGCG  ACCTCACCCCT  CTCGTTACTCC  GTCGCGCTCTCT  G1 Gly/Ser DelCH2

4361 ACAACTACAA  GACCACGCCT  CCCGTGCTGG  ACTCCGACGG  TGTTGATGGT  CTGGTCGGGA  GGGACAGACC  TGAGGCTGCC  G1 Gly/Ser DelCH2

4401 CTCTTTCTTC  CTCTACAGCA  AGCTCACCGT  GGACAAGAGC  GAGGAAGGAAG  GAGATGCTGT  TCGAGTGGCA  CCGTCTTCTCG  G1 Gly/Ser DelCH2

4441 AGGTGGCAGC  AGGGGAACGT  CTCTCTCATGC  TCGTGATGC  TCCAGCGTGC  TCCCCCTTGCA  GAAGAGTACG  AGGCACTACG  G1 Gly/Ser DelCH2

4481 ATGAGGCTCTG  GCACAACCCAC  TACAGCGAAG  AGAGCCCTCCT  TACTCGGAGA  CGGTGTGGTG  ATGTGCCTCT  TCTCGGAGAG  G1 Gly/Ser DelCH2
BamHI

4521 CCTGTCTCCG GGTAATGAG GATCCAAATTG GACTGATGCA
    GGACAGAGGC CCAATTACTC CATGGTTAAG CTGACTACGT
4561 AGCGTTACCA ACTACCTAGA CTGGATTGC GACAACATGC
    TGCCATAGGT TGATGGATCT GACCTAAGCA CTGTTGTCAG
    BGH

EcoRV

4601 GGCCGTGATA TCTACGTATG ATCAGCCTCG ACTGTGCGTT
    CCGGCACCTAT AGATGCTAC TAGTCCGAGC TGACACGGAA
    BGH
4641 CTAGTTGCCA GCCATCTGTG TTGTGCCCCC CCCCCGTGCC
    GATCAACGGT CGTGAACCAA CAAACGGGA GGGGCACGG
    BGH
4681 TTCCCTGACC CTGGAAAGGTG CCACTCCACC TGTCCTTTCC
    AAGGAACGCG GACCTTCCAC GGTAGGGTG ACAGGAAAAG
    BGH
4721 TAATAAAATG AGGAAAATTGC ATCGCATTGT CTGAGTAGGT
    ATTATTTTAC TCCTTTAACG TAGCGTAACA GACTCATCCA
    BGH
4761 GTCATTCTAT TCTGGGGGGGT GGGTGTTGCG AGGACAGCAA
    CAGTAAGATA AGACCCCCCA CCCCCCCCCG TCCTCTCGTT
    BGH
4801 GGGGGAGGAT TGGAAGACCA ATAGCAGGGA TGCTTGGGGAT
    CCCCCCTCTTA ACCCTTCTGT TAGCTCCGGT ACGACCCCTA
    BGH
4841 GCAGTGGGGCT CTATGGCTTC TGAGGCGGAA AGAACCAGCCT
    CGCCACCDDA GATACCGGAAG ACTCCCGCCTT TCTTGGTGCA
    Beta
4881 GGGGCTCGAC AGCGCTGCGA TCGCCTCGAC TAGTAGCTTT
    CCCGAGCTG TCGCGAGCCT AGCGGAGCTG ATCATCGAAA
    Beta
4921 GCTTCTCAAT TTCTTATTTG CATAATGAGA AAAAAAGGAA  
CGAAGAGTTA AAGAATAAAC GTATTACTCT TTTTTTCCCT
  Beta

4961 AATTAATTTT AACAACAAAT CAGTAGTTGA TTGAGCAAT  
TTAATTTAAA TTGTGGTTAA GTCATCAACT AACTCGTTTA
  Beta

5001 GCGTTGCCAA AAAGGATGCT TTAGAGACAG TGTTCTCTGC  
CGCAACCGTT TTTCTCAGTA AAATCTCTGTC ACAAGAGACG
  Beta

5041 ACAGATAAGG ACAAAACATTA TCCAGAGGGGA GTACCAGAG  
TGCTCTATCC TTGTGTGAAT AAGTCTCCCT CATGGGTCTC
  Beta

5081 CTGAGACTCC TAAGCCAGTG AGTGGCAGACG CATCCAGGGA  
GACTCTGAGG ATTCGGTTCG TCACCGTGTC TAGGTCCCT
  Beta

5121 GAAATATGCT TGTCATCACG GAAGCTGTAT TCAAGTAGAC  
CTTTATACGA ACAGTAGCTG CTTGGACTA AGGCATCTCG
  Beta

5161 CACACCCCTGG TAAGGGCCAA TCTGCTCACA CAGGATAAGG  
GTGTTGGGACC ATTCGGGGTT AGACGAGTGT GTCTCTATTC
  Beta

5201 AGGGCAGGAG CCAGGGCAGA GCATATAAGG TGAGGTAGGA  
TTCCGGTCCT GCACGGTGTC GTATATATCC ACTCCATCCT
  Beta

5241 TCAGTTGCTC TCTACATTTC TTCTGACATG AGTGTGGTTG  
AGTCAACGAG AGAGTTAACG AAGACTGTA TCAACACAAC

5281 GGAGCTTGGGA TAGCTTGGGG GGGGCAGAGC TGAGGGCTGC  
CCTCGAACCT ATTCGGACCC CCCCCCGTGC AGTCCGGACG

5321 GATTTTCGGCC CAACACTTGAC GGCAATTCTA GCGTGAAGGC  
CTAAAGCCCG GTTGAACACT CGTGTAGATT CGCAGCTCCG
  DHFR

5361 TGATTTAGTT TTATCCCCGC TGCCATCATG GTTGACCAT  
ACCACCTTAA AATAGGGCCG ACGTGATGAC CAAGCTGGTA
  DHFR
5401 TGAACCTGCA TGCTGCGGTG TCCCCAAATA TGGGGATTGG
    ACTTGGACTA GCAGCGGCCA AGGGTTATTAT ACCCCCTAACC
    DHFR

5441 CAAGAACGGGA GACCCTACCT GGGCTCCGCT CAGGAAGGAG
    GTTCTTGCCT CTGGATGGGA CCGGAGGCGA GTCCCTGTCT
    DHFR

5481 TTCAAGTACT TCCAAAGAAT GACCAACAACC TCTTCAGTGG
    AAGTTCATGA AGGGTTCTTA CTGGGTGTGG AGAAAGTCACC
    DHFR

5521 AAGGTAAAAC GAATCTGGTG ATTATGGGTA GGAAGACCTG
    TTCCATTTGT CCTAGACCAT TAATACCCAT CCTTTTGAGAC
    DHFR

5561 GTTCTCCATT CCTGAGAAGA ATCGACCTTT AAAGGACAGA
    CAAGAGTTAA GGACTCTTCT TAGCTGAAAA TTTCCGTCT
    DHFR

5601 ATTAATATAG TTCTCAGTATG AGAAGTCAA AAGCCACCAC
    TAATTATATC AAGAGTCATC TCTTGAGTGT CTTGGTGTGG
    DHFR

5641 GAGGAGCTCA TTTCTTGCCT AAAAGTTTGG ATGATGGCTTT
    CTCCTCGAGT AAAAGAACGG TTTTCAAACC TACTACGGAA
    DHFR

5681 AAGACTTATT GAACAAACCGG AATTGGCAAG TAAAGTAGAC
    TTCTGAGTAA CTTGTGGGAC TTAACCGGTC ATTTGATCTG
    DHFR

5721 ATGGTTTGGGA TAGTCCGGAGG CAGTTCCTGG TACACAGGAAG
    TACCAAAACCT ATCAGCCCTCC GCTAGACAA ATGGTCTCTC
    DHFR

5761 CCATGAATCA ACCAGGCGAC CTCAGACTCT TTGTGACAAG
    GGTACTTATG TGGTCGGGTG GAGTCTGAGA AACACTGTTC
    DHFR

AflIII

AflIII

AflIII
5801 GATCATGCAG GAATTTGAAA GTGACACGGT TTTCCAGGA
CTAGTACGTC CTTAAACTTT CACTGTGCAA AAAGGGTCTT
DHFR

5841 ATTGATTGG GGAATATAAA ACTTCTCCCA GAATACCAG
TAACTAAACC CTTTTATAATT TGAAGAGGGT CTTATGGGTC
DHFR

5881 GCGTCTCCTCT TGAGGTCCAG GAGGAAAAG GCATCAAGTA
CGCAGGAGAG ACTCCAGGTC CTCTTTTTTC GCTAGTTCAT
DHFR

5921 TAAGTGTGAA GTCTAGCAGA AGAAGACTA ACAGGAGAT
ATTCAAACCT CAGATGCTCT TCTTTCTGAT TGTCCTCTTA

5961 GCTTTCAAGT TCTCTGCTCC CTCCTAAGG CTATGCATTT
CGAAAGTTCA AGAGACGAGG GAGGATTTTC GATAGTAAA

6001 TTATAAGACC ATGGGACCTT TGCTGGCTTT AGATCAGCCT
AATATTCTGG TACCCTGAAA ACGACCGAAA TCTAGTCCGA
BGH

6041 CGACTGTGCC TTCTAGTTGC CAGCCATCTG TTTGTGCCCC
GCTGACACGG AAGATCAACG GTCGGTAGAC AACAAAACGGG
BGH

6081 CTCCCCCGTG CCGTCCCTTA AACCAAGGGG CACCATCTTT
GAGGGGACAC GGAAGGAACG GGGACCTTCC ACAGGAGGG
BGH

6121 ACTGCTCCTTT CCTAATAAAA TGAGGAAATT GCATCGCATT
TGACAGGAAA GGATTATTTT ACTCCTTTAA CGTAGCGTAA
BGH

6161 GTCTGAGTAG GTGTCATTCT ATTTCTGGGG GTGGGGTGCGG
CAGACTCATC CACAGTAAAG TAAGACCCCC CACCCCCC
BGH

6201 GCAGGACAGC AAGGGGAGAG ATGGAAGAGA CAATAGCCAG
CGTCTGCTCG TTCCTCCTCC TAAACCTTCT GTATATGCTG
BGH

6241 CATGCTGGGG ATGCGGTGGG CTCTATGGCT TCTGAGGCGG
GTACGCCCTC TACGCCACCC GAGATACCCGA AGACTCCGCC
BspEI
N2

6721 AGGATCTCGT CTTGACCCTAT GGCGATGCCT GCTTGCCGAA
TCCTAGAGCA GCACCTGGGT ACGCTACGGG CGAACCGCTT
N2

6761 TATCATGCTGT GAAAATGGCC GCTTTTCTGG ATTCATCGAC
ATAGTACCAC CTTTTACCGG CGAAAAAGACC TAATGCTCTG
N2

RsrII

6801 TGTGGCCCGG CGGGTGTGCG GCACGCTTAT CAGGACATAG
ACACCGGCGG ACCACACCGG CCTGGCGATA GTCTGTATTC
N2

6841 CGTTGCGCTAC CGTGATATT GCCTGAGAGG TTTGGCGCCGA
GCAACCGATG GGCACCTAAA CGACTTCTCG ACGGCGCCGT
N2

6881 ATGGCGCTAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT
TACCCGACTG GCGAAGGAGC ACGAATGCC ATACGGCGGA
N2

6921 CCCGATTTCGG AGCGGATCCG CTTCTATCGC CTTCTTGAAG
GGGCTAAAGCG TCGCGTACCG GAAGATAGCG GAAGAATGCG
N2

BstBI

6961 AGTTCTTCTCT AGCGGGAACCTC TGGGGTTCGA AATGACCGAC
TCAAGAAGAC TCCGCCCTGAG ACCCAAAGCT TTACTGGCTG
7001 CAAGCGACGC CCAACCTGACC ATCACGAGAT TCGATTCCA
GTCGCTGAGC GCGGACGAGC TAGTGCTCTA AAGCTAAGGT
7041 CGGCCGCCCTT CTATGAAAGG TTTGGCTTCGA GAATCGTTTT
GCGGGCGGAA GTAATTCTCC ACGCGAAGC CTTAGCAAAA
7081 CCGGGACGCC GGCTGGAAGA TCCCTCAAGCG CGGGAATCTC
GGCCCTGCGG CCGACCTACT AGGAGTTCGG GCCCCTAGAG
SV

7121 ATGCTGGAGT TCTTCCGGCC CAACACTTGT TTATTCGAGC
TACGACCTCA AGAAGCCGGG TGGTTGACAA AATAACGCTCG
SV
7161 TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTTC
AATATTACCA ATGTTTATTT CGTATCGTA GTGTTAAAG
SV

7201 ACAAAATAAG CATTTTTTTC ACTGCAATTCT AGTTTGTGGTT
TGTTTTATTC GTAATAAAAG TGAGTGTAAGA TCAACACCAA
SV

7241 TGCTCCAAACT CATCAATGTA TCTTATCATG TCTGGATCGC
ACAGGTTTGA GTAGTACTAC AGAATAGTAC AGACCTAGCG
Asci

7281 GGCCGCGCGCG CACCCGGCTTG GAGCTTTAAT TAAAGGCGCCG
CGACGGGCGGC GTGGCGGCCAC CTGGAATTA ATTCACGGCGG
Asci

7321 CAGCTCCAGC TTTTTTTCCC TTTAGTGAGG GTTAAATTTCG
GTCGAGGTCG AAAACAAAGGG AAACTACTCC CAATTAAAGC

7361 AGCTTGGCTG AATCATGTTG TATGCTGGTTT CCGGCTGGAA
TCGAGCCGCA TTAGTACAGC TATCGACAAA GGACACACTT

7401 ATGTTATTCG GCTCAACATT CCCAACAACA TACCGAGCGG
TAACACATTG GAGGTGTAAA GGTGTGTTGT ATGTCGTCGG

7441 AAGCATAAGG TGTAAAGCCTT GGGTGCTCCT ATGAGTGAGC
TTCTGTATTTC ACATTTCCGA CCCCAGGAT TACTACCTGG

7481 TAACCTCAGAT TAATTTGCTT GGGCTACTGC CCCGCTTTCC
ATTTAGTTGA ATTAACCGCAA CGCGAGTGCAG GGGCGAAAGG

7521 ATCTGGGAAA CCTGTCGTGC CAGCTGCATT AATGAACTGG
TCAGCCCTTTT GACAGCCACGC TGAGCTGAA TTACTTGACC

7561 CCAACGCGCG GGGAGAGCGG GTTGTGCTAT TGGGGCGTCT
GGTTGCGCCG CCCCTTCCCG CAACACGATA ACCGGAGAGA

7601 TCCGTCTCTC GCTCAGCTGA CTGCTGTGCG TGCGTCGTTC
AGCGGAAGGA GGCAAGTACT GAGCGACCGG AGCCAGCAAG

7641 GGCTCGGGCG AGGCGTATCA GCTCAGCTCA AAGGCGTAAAG
CCGACGGGCC TCAGCCATTGT CGAGGTAGTTT TCCGGCATTA

7681 ACAGGTTATCC ACAGAAATCG GGATATAGAG AGGAAAGAAC
TGCAATATAG TGTCTTAGTC CCTATTGCG TCTTTTCTTG

7721 ATGTGAGCAA AAGGCGCCAA AAGGGCCAGG AAACGGAAAT
TACACGAGTTT TGGCGGTGCT TGGCGGTCC TGGGCAATTAT

7761 AGCCGGCGTT GCTGCGGTGT TCTCATACGC TGGGGGGGCG
TGGCGCGCAA CGACGCCAAA AGGATATCGC AAGGGGGGCG

7801 TGACGGAGCAT CACAAAGATGC GACGCTCAAG TCAGAGTGAG
ACTGCTGTAG GTGGTTTATT CTGCGAGTTC AGTCTGCCAAC

7841 CGAAACCGGCA CAGAGACCTA AAGATAACCG GGGTTGTCGC
GCTTTGGGCAT GCTGCTGATAT TTTCTATGGTC CGCAAGGGG
7881 CTGGAAGCTC CTCGTCGGGC TCTCCTGTTC CGACCCTGCC
GACCTTCGAG GAGACAGCGG AGAGAACAGG GCTGGAGACG
7921 GCTTCAGGGA TACTGTTCCG CTTTCTCCG TTGGGAGACC
CGAATGCGCT ATGGACAGGC GGAAGAGGG AGGCCCTTCG
7961 GTGGCCGCTTT CCTCATAGCTC AGCTGTTAGG TATCTCAGTT
CACCCGAAAG AGTATCAGG TGCCGACATCC ATAGAGTCAA
8001 CGGTGTAGGT CGTTTCGCTCC AAGCTGGGCT GTGTCAGCGA
GCCACATCCA GCAAGCGGAG TTTGACCGCA CACAGTCGCT
8041 ACACCCCGTGT CAGCGCGACC GCTGCGCCTT ATCCGGAAC
TGGGGGGCAA GTCGGGCTGG CGACGCGGAA TAGGCCATTG
8081 TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC
ATAGCAGAAC TCAGGTGTTGG GCCATCTGTTG CTGAAATAGCC
8121 CACTGGCAGC AGCCACTGTT AACAGGATTA GCAGAGCGAG
GTGACCGTGCG GCGTTGACCA TTGTCTCAATC CGTCTCCTTG
8161 GTATGTAGGC GGTGCTACAG AGTTCCTGAA GTGGTGGCCT
CATACATCGG CCAAGATGTC TCAAGAACCT CACCCCGGAG
8201 AACATCGGCT ACACTAGAAG AACGATATTG GTGATCTCGG
TTGATGCGGA TGTGATCTTC TTGTCATAAA CATAGACGC
8241 CCTCTGCTGAA GCCAGGTACC TTGGGAAAG AAAGTTGTAG
GAGACGACTT CGTCAATGGA AGCCCTTTTT TCTAACATCC
8281 CTCTTGTAGCC GAACCAACAA CCAGGGCTGG TACGGGTTGT
GAGAACTAGG CGGTGGTTTG GGGGCGGACC ATGGCACCAG
8321 TTTTTTTTTG CCAGACGACAT GATACGACC AGAAGAAAGG
AAAAAACAACG CGTTGTCTGT CTAATGGCGG TCTTTTTTCC
8361 GATCTCAAGAA AGATCTTTTG ATCTTCTTCA CGGGGTCTGA
CTAGAGTTCT TCTAGGAAAAC TAGAAAAGAT GCCCCAGACT
8401 CGCTCAGTGG AACGAACACT CACGTTAAGG GATTTTGTCG
GGGAGTCAACC TTGCTTTTGA GTGCATAATCC CAATAACCGAG
8441 ATGAGATTAT CAAAAGAGAT CTTCCGCTAG ATCCTTTTAA
TACTCTAAAT GTTTTTCTCA GAAATGAGTC TAGAAATATT
8481 ATTAAAAATT AAGTTTTAAA TCAATCTAAA GATATATAG
TAAATTTTACT TCTAAATTTT AGTGAAGATT CATAATACCT
8521 GTAAACTTGG TCTGACAGTT ACCAATGCTT AATCAGTGA
CATTGGAACC AGACTGTCAGA TTGGTACGAA TTAGTACCTC

-----------------

Amp

8561 GCACCTATCT CAGCGATCTG TCTATTTTGCT CTATCCATAG
CGTGGATAGA GTGCGTACAGC AGATAAAGCA AGTAGATATC

-----------------

Amp

8601 TTGCGCTAGT CCCGTCGCTG TAGATAACTA CGATACGGGA
AAGGAGTCTG GGGCCGACAC ATCTATTGAT GCTATGCCCT

-----------------
8641  GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA
       CCCGAATGCT AGACCGGGGT CACGACGTAA CTAATGGCGCT

Amp

8681  GACCCACGCT CACCAGGTCC AGATTTATCA GCAATAAAC
       CTGGGGTCCGA GTGGCCCGAGG TCTAAATAGT CGTTATTTGG

Amp

8721  AGCCAGGCGG AAGGGCGCGG CGCGAGGTAG GTCCTGCAAC
       TCGTGGCCGA TTCCCGGCTC GGGTCTCACT CAGGACGTCTG

Amp

8761  TTTATCCCGT TTCATCCAGT CATATATTG AGGCTGGGAA
       AAATAGGCGG AGGTTAGGCTA GATAATTAAC AAGGGCCTTT

Amp

8801  GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGGCGCAACG
       CGATCTCATT CATCAAGCGG TCAATTATCA AACCGCGTGC

Amp

8841  TTGTTGCGAT TGCTACAGGC ATCGTGGGTG CACGCTGTC
       AACAAACGGTA ACGATGTCCG TAGCACCACA GTGCGGACCAG

Amp

8881  GTTGGGTATG GCTTCATTTAG GCTCCCGTTC CCAACGATCA
       CAAACCATAAC CGAAAGTAAA GTCGGCACAAG GGGTGTAGTG

Amp

8921  AGGCGAGTTC CAGTATCCCC CATGTTGGTC AAAAAGCCCG
       TCCGCTCAAT GTACTAAGGG GTACAACAGG TTTTTTCCGC

Amp

8961  TTAGCTCTTT TTGGTCTCCCG ATCGTTGTCG GAAATGATTT
       AATCGAGGAA GCAGGAGGAC TAGCAACAGT TTTTCTCAA

Amp

9001  GGCtGCGATTT TTTATCCTCA TGTTATGAGC AGCACGCTGC
       CCGGCGGTCAC ATAGTGAAGT ACCAATACCG TGGTGACGTA

Amp

9041  AATTCCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTTG
       TTAAGGAAAT GACAGTGACG TAGGATCTCT ACGAAAAAGAC

Amp
9081 TGACTGGTGA GTACTGAAACC AAGTÇATTCT GAGAATAGTG
ACTGACCACCT CATGAGTTGG TTCAGTAAAGA CTCTTATCAC

Amp

9121 TATGGCGGCA CCGAGTTGCT CTTGCCCGGC GTCAATACGG
ATACGCGCCT GCTCAGAAGA GAACGGGCCG CAGTTATGCC

Amp

9161 GATAATACCG CGCCACATAG CAGAACTTTTA AAAGTGCTCA
CTATTATGGC GCGGTGTATTC GTCTTGAATAT TTTACAGGAT

Amp

9201 TCATTGGAAA ACGTTCTCAG GGGCGAAAAC TCTCAAGGAT
AGTAACCTTT TGCAAGAAGC CCCCCTTTTG AGAGTTCCCTA

Amp

9241 CTTACCGCCTG TTGAGATCCA GTCGCCAGTA ACCCACTCTGT
GAATGGCCAC AACTCTAGGT CAAGCTACAT TGGGTGAGCA

Amp

9281 GCACCCCAACT GATCTCCAGC ATCTTTTACT TTCACGAGCG
CGTGGGTGTA CTAGAAGTCG TAAAAATGA AAGTGCGCCG

Amp

9321 TTTCTGGGTTG AGCAAAAAACA GGAAGGCCAA ATGCCGCAA
AAAGACCCAC TCGTTTTTTT CCTCCGGTTT TAGGGCGTTT

Amp

9361 AAAGGGGAATA AGGGCGACAC GGAAATGTTG AATACTCATA
TTTCCCCITTAT TCCCGCTGTG CTTTACAAC TTATGAGTAT

Amp

SspI

9401 CTCTTCTTTT TTCAATATTA TTGAAGCATT TATCAGGTTG
GAGAAGGAAA AAGTTATAAT AACTTGTAA ATAGTCCCAA

9441 ATATGCTCAT GACGGGATAC ATATTTGAAT GATATTTAGAA
TAACAGAGTA CTCCGGCTATG TATAACTATA CATAAATCTT

9481 AAATAAACCAA ATAGGGGTTC CGCGCACATT TCCCCGAAAA
TTTATTTGTT TATCCCCAAAG CGCGGCTGAA AGGGCGTTTT

9521 GTGCCAC
CACGGTG