

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 February 2006 (02.02.2006)

PCT

(10) International Publication Number
WO 2006/012414 A2

(51) International Patent Classification:
C12N 15/85 (2006.01)

(21) International Application Number:
PCT/US2005/025876

(22) International Filing Date: 20 July 2005 (20.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/589,679 20 July 2004 (20.07.2004) US

(71) Applicant (for all designated States except US): **CRITICAL THERAPEUTICS, INC.** [US/US]; 60 Westview Avenue, Lexington, MA 02421 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **O'KEEFE, Theresa** [US/US]; 20 Richgrain Avenue, Waltham, MA 02453 (US).

(74) Agents: **BROOK, David, E.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).

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

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

Published:

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

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

(54) Title: NOVEL POLYADENYLATION SIGNAL FOR USE IN EXPRESSION VECTORS

(57) Abstract: Disclosed are novel polynucleotides and expression vectors which comprise an immunoglobulin gene polyadenylation signal and a DNA sequence encoding a polypeptide wherein the DNA sequence is other than a genomic DNA sequence from the immunoglobulin gene. Also, disclosed is a eukaryotic cell or cell-line comprising the novel expression vectors and a method of producing a desired polypeptide comprising culturing mammalian cells transfected with the novel expression vectors and recovering the desired polypeptide from the culture medium. The novel polynucleotides and expression vectors may comprise an immunoglobulin gene polyadenylation signal from an IgG 1 heavy-chain gene.

WO 2006/012414 A2

- 1 -

NOVEL POLYADENYLATION SIGNAL FOR USE IN EXPRESSION VECTORS

RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application No. 60/589,679,
5 filed on July 20, 2004, the entire teachings of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

The use of expression systems is an important tool for the production of recombinant proteins. Recombinant proteins have been successfully produced in both prokaryotic
10 expression systems (such as *E. coli*) and eukaryotic expression systems (including yeast and mammalian systems) and can have utility as research and/or therapeutic agents. When producing recombinant proteins for therapeutic use, it is often preferred to use mammalian expression systems as these systems are more likely to produce proteins with essential post-translational modifications. The use of mammalian expression systems is, however,
15 associated with a significant shortcoming in that the level of protein expression is often significantly less than that of corresponding prokaryotic expression systems. Several methods have been employed to obtain higher levels of recombinant gene expression; one such method pertains to the use and selection of a polyadenylation signal in the protein construct.

Polyadenylation signals are nucleotide sequences found in nearly all mammalian
20 genes and control the addition of a string of approximately 200 adenosine residues (the poly(A) tail) to the 3' end of the gene transcript (Proudfoot et al., Cell 108: 501-512 (2002)). The function of the poly(A) tail is not fully understood but its presence in mRNA is believed to be essential to mRNA stability as there is evidence that mRNAs lacking the poly(A) tail are rapidly degraded. There is also evidence that the presence of the poly(A) tail positively
25 contributes to the translatability of mRNA by affecting the initiation of translation (Molecular Biology of the Cell, Third Edition by B. Alberts et al., Garland Publishing, 1994).

The polyadenylation signal is that portion of the RNA (and can also refer to the corresponding portion of DNA) that directs binding of the polyadenylation protein complex to mRNA. It is believed that efficient polyadenylation signals comprise at least two regions:

- 2 -

the hexanucleotide sequence, AAUAAA, and either a downstream G/U or simply U-rich region. The AAUAAA sequence is believed to bind cleavage and polyadenylation specificity factor (CPSF) and the CPSF-RNA complex then binds additional proteins, including cleavage stimulatory factor (CStF) and poly(A) polymerase, which are involved in cleavage of the mRNA and addition of the poly(A) tail. The interaction between CStF and the U-rich region is believed to contribute to the stability of the protein complex. Cleavage and polyadenylation ultimately take place 10-35 nucleotides downstream of the AAUAAA sequence.

Several polyadenylation signals have been utilized for recombinant protein expression and include polyadenylation signals from simian virus 40 (SV40) (Schek et al., Mol. Cell Biol. 12(12): 5386-93 (1992)), α -globin (Thein et al., Blood 71(2): 313-19), β -globin (Orkin et al., EMBO J. 4(2): 453-6 (1985)), human collagen, polyoma virus (Batt et al., Mol. Cell Biol. 15(9): 4783-90 (1995)) and bovine growth hormone (bGH) (Woychik et al., Proc. Natl. Acad. Sci. USA 81(13): 3944-8 (1984); U.S. Patent No. 5,122,458). The available polyadenylation signals differ with respect to their effect on the level of protein expression in different cells systems. Although the bGH polyadenylation signal has been shown to result in higher protein expression when compared with other known polyadenylation signals in specific mammalian cells (Pfarr et al., DNA 5(2): 115-22 (1986); Yew et al., Hum. Gen. Ther. 8(5): 575-841 (1997)), there remains a need in the art for the identification and use of new polyadenylation signals which can contribute to efficient protein expression.

20

SUMMARY OF THE INVENTION

It has now been discovered that high protein expression can be achieved using a polyadenylation signal from an immunoglobulin gene. For example, it has been shown that Chinese hamster ovary cell transfectants comprising polyadenylation signal sequences from the human IgG1 gene expressed high levels of recombinant protein (see Example 6 below).

25

The present invention pertains to novel polynucleotides and expression vectors comprising a polyadenylation signal from an immunoglobulin gene, cells comprising the novel expression vectors and novel methods for producing recombinant polypeptides in mammalian cells. In one embodiment, the invention pertains to a novel polynucleotide comprising a polyadenylation signal from an IgG gene. In another embodiment, the invention pertains to a novel expression vector comprising a polyadenylation signal from an IgG gene. In further embodiments, the polyadenylation signal is an IgG heavy-chain gene polyadenylation signal.

30

One embodiment of the present invention is a polynucleotide comprising an IgG heavy-chain gene polyadenylation signal and a DNA sequence encoding a polypeptide. The DNA sequence encoding a polypeptide is other than a genomic DNA sequence from the IgG heavy-chain gene. In addition, the DNA sequence is upstream from, and operably linked, to the IgG heavy-chain gene polyadenylation signal. Another embodiment of the present invention is a novel expression vector comprising an IgG heavy-chain gene polyadenylation signal and a DNA sequence encoding a polypeptide. The DNA sequence encoding a polypeptide is other than a genomic DNA sequence from the IgG heavy-chain gene. In addition, the DNA sequence is upstream from, and operably linked, to the IgG heavy-chain gene polyadenylation signal.

In another embodiment of the present invention, the IgG heavy chain gene polyadenylation signal is a human IgG heavy-chain gene polyadenylation signal. In a further embodiment, the human IgG heavy chain gene polyadenylation signal is an IgG1 heavy-chain, an IgG2 heavy-chain, an IgG3 heavy-chain or an IgG4 heavy-chain gene polyadenylation signal. In a further embodiment, the human IgG heavy-chain gene polyadenylation signal is an IgG1 heavy-chain gene polyadenylation signal. In an additional embodiment, the human IgG1 heavy-chain gene polyadenylation signal comprises a sequence selected from the group consisting of SEQ ID NO:10 and SEQ ID NO:9. In yet another embodiment, the human IgG1 heavy-chain gene polyadenylation signal consists of a sequence selected from the group consisting of SEQ ID NO:10 and SEQ ID NO:9.

Another embodiment of the present invention is a eukaryotic cell comprising the novel expression vector. In a further embodiment, the eukaryotic cell is a mammalian cell. Yet another embodiment of the present invention is a method for producing a desired polypeptide comprising culturing a mammalian cell transfected with the novel expression vector and recovering the desired polypeptide from the culture medium. In certain embodiments, the desired polypeptide is an immunoglobulin fusion protein. In certain other embodiments, the desired polypeptide is a high mobility group protein.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A is the nucleic acid sequence of the human IgG1 heavy-chain constant region (GenBank Accession No. J00228; SEQ ID NO: 1).

FIG. 1B is the nucleic acid sequence of the human IgG2 heavy-chain constant region (GenBank Accession No. J00230; SEQ ID NO: 2).

- 4 -

FIG. 1C is the nucleic acid sequence of the human IgG3 heavy-chain constant region (GenBank Accession No. X03604; SEQ ID NO: 3).

FIG. 1D is the nucleic acid sequence of the human IgG4 heavy-chain constant region (GenBank Accession No. K01316; SEQ ID NO: 4).

5 FIG. 2A is the nucleic acid sequence of the mouse IgG1 heavy-chain constant region (GenBank Accession No. J00453; SEQ ID NO: 5).

FIG. 2B is the nucleic acid sequence of the mouse IgG2a heavy-chain constant region (GenBank Accession No. J00470 SEQ ID NO: 6).

10 FIG. 2C is the nucleic acid sequence of the mouse IgG2b heavy-chain constant region (GenBank Accession No. J00461; SEQ ID NO: 7).

FIG. 2D is the nucleic acid sequence of the mouse IgG3 heavy-chain constant region (nucleotides 6000-7715) (GenBank Accession No. X00915; SEQ ID NO: 8).

FIG. 3 is the nucleic acid sequence of human IgG1 heavy-chain gene polyadenylation signal sequence 1 ("IGPA") (SEQ ID NO: 9). The lower case letters indicate cloning sites.

15 FIG. 4 is the nucleic acid sequence of human IgG1 heavy-chain gene polyadenylation signal sequence 2 ("LIGPA") (SEQ ID NO: 10). The lower case letters indicate cloning sites.

FIG. 5A is the nucleic acid sequence (cDNA) of a RAGE-Ig fusion protein. The sequences in lower case contain restriction enzyme sites and Kozak sequences.

20 FIG. 5B is the amino acid sequence of a RAGE-Ig fusion protein. A hyphen indicates the separation between the RAGE and Ig portions of the protein. The mutated Fc region is underlined.

FIG. 6 is a diagram depicting the construction of IGPA and LIGPA.

FIG. 7A is a diagram showing the construction of mammalian expression vector pCTiTOK18, described in Example 3.

25 FIG. 7B is a diagram showing the construction of mammalian expression vector pCTiTOK28, described in Example 3.

FIG. 7C is a diagram showing mammalian expression vector pCTiTOK29, described in Example 3.

30 FIG. 8 is a graph comparing quantity of RAGE-immunoglobulin fusion protein (ug/ml) produced by CHO cells transfected with vectors comprising the BGH, IGPA or LIGPA polyadenylation signal sequences.

- 5 -

FIG. 9 is a graph comparing the quantity of RAGE-immunoglobulin fusion protein (ug/ml) in the supernatants of different CHO cell clones transfected with the mammalian expression vector pCTiTok29, which contains the LIGPA polyadenylation signal sequence.

5 DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell culture, molecular biology, microbiology, cell biology, and immunology, which are well within the skill of the art. Such techniques are fully explained in the literature. See, e.g., Sambrook et al., 1989, "Molecular Cloning: A Laboratory
10 Manual", Cold Spring Harbor Laboratory Press; Ausubel et al. (1995), "Short Protocols in Molecular Biology", "Molecular Cloning: A Laboratory Manual" by T. Maniatis, et al., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982, John Wiley and Sons; Methods in Enzymology (several volumes); Methods in Cell Biology (several volumes), and Methods in Molecular Biology (several volumes).

15 The present invention encompasses polynucleotides and novel expression vectors which comprise an immunoglobulin gene polyadenylation signal and a DNA sequence encoding a polypeptide wherein the DNA sequence is other than a genomic DNA sequence from the immunoglobulin gene, a eukaryotic cell or cell-line comprising the novel expression vectors, and methods of producing a desired polypeptide comprising culturing the
20 mammalian cells transfected with the novel expression vectors and recovering the desired polypeptide from the culture medium. In an additional embodiment, the immunoglobulin gene polyadenylation signal is a polyadenylation signal from an immunoglobulin gene of the IgG isotype. In a further embodiment, the polyadenylation signal is from an IgG heavy-chain gene. In yet another embodiment, the polyadenylation signal is from a human IgG heavy-
25 chain gene. In an additional embodiment, the human IgG heavy-chain gene polyadenylation signal is selected from the group consisting of an IgG1, IgG2, IgG3 and IgG4 heavy-chain gene polyadenylation signal. In a further embodiment, the novel polynucleotide and expression vector comprise a human IgG1 heavy-chain gene polyadenylation signal.

30 *Polynucleotides and Expression Vectors*

As used herein, the term "polynucleotide" is a polymer of deoxyribonucleic acid or ribonucleic acid linked together by 5'-3' phosphodiester bonds between the sugar (deoxyribose or ribose) and phosphate groups. The polynucleotides are composed of

- 6 -

repeating nucleotide units and, as disclosed herein, are capable of existing in the double helix configuration wherein the paired guanine/cytosine (G/C) and adenine/thymine (A/T) bases form hydrogen bonds stabilizing the double helix.

As used herein, the term "expression vector" is a vector, such as a DNA vector, that
5 comprises a coding sequence and various regulatory sequences which are necessary for the expression of recombinant polypeptides in cells. One type of expression vector is a plasmid, which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. The coding sequence may be any suitable coding sequence including a sequence encoding a desired protein, a desired peptide and/or a desired polypeptide. The
10 coding sequence may be derived from any suitable source, for example, a cDNA-derived nucleic acid fragment or a nucleic acid fragment isolated by polymerase chain reaction (PCR).

Regulatory sequences which may be included in an expression vector include a promoter and/or enhancer operably linked to the DNA sequence to be expressed, 5' or 3'
15 non-translated sequences, such as ribosome binding sites, splice donor and acceptor sites, sequences which control the initiation of transcription (e.g., Kozac sequences, etc.) and sequences which control the termination of transcription and translation, for example, a polyadenylation signal. DNA sequences are operably linked when they are positioned such that they are functionally related to each other. For example, a promoter is operably linked to
20 a DNA coding sequence when it is able to control transcription of the coding sequence. A ribosome binding site is operably linked to a DNA coding sequence when it is positioned so as to assist in translation. A signal peptide is operably linked to a DNA coding sequence when it is expressed as a precursor which participates in the secretion of the polypeptide.

As used herein, an "immunoglobulin gene" is understood to mean a gene from the
25 immunoglobulin gene superfamily. An "immunoglobulin gene polyadenylation signal" is understood to mean a DNA sequence which is a subsequence of the immunoglobulin gene that comprises the AATAAA sequence near the 3' end of the gene wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene. Typically, the subsequence from the immunoglobulin gene comprises about 1000
30 nucleotides. More typically, the immunoglobulin gene polyadenylation signal comprises a subsequence from the immunoglobulin gene of about 700, 600, 500, 400, 300 or fewer nucleotides.

- 7 -

The immunoglobulin gene polyadenylation signal is also a subsequence of an immunoglobulin gene that comprises the AATAAA sequence and a GT/T-rich sequence downstream of the AATAAA sequence wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene. The GT/T-rich sequence (the GT/T-rich sequence corresponds to the mRNA GU/U rich region) is a sequence of nucleotides found approximately 20-50 nucleotides downstream of the AATAAA sequence and which is a T or GT-rich motif. The immunoglobulin gene polyadenylation signal is also a subsequence of the immunoglobulin gene that comprises the AATAAA sequence, the downstream G/T-rich sequence and an additional downstream GC-rich nucleotide sequence of an immunoglobulin gene wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene. The GC-rich sequence is a sequence found downstream of the T or G/T-rich sequence and is a GC-rich motif which may form a hairpin loop structure. In addition, the polyadenylation signal is also a subsequence of the immunoglobulin gene that comprises the AATAAA sequence and a sequence found upstream (5') of the AATAAA sequence in an immunoglobulin gene wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene.

While the AATAAA sequence described above is preferred, it may be substituted with other hexanucleotide sequences with homology to AATAAA which are capable of signaling polyadenylation in natural mRNAs. These homologous hexanucleotide sequences include ATTAAA, AGTAAA, CATAAA, TATAAA, GATAAA, ACTAAA, AATATA, AAGAAA, AATAAT, AAAAAA, AATGAA, AATCAA, AACAAA, AATCAA, AATAAC, AATAGA, AATTAA and AATAAG. Therefore, in one embodiment, the immunoglobulin gene polyadenylation signal is also a subsequence of an immunoglobulin gene that comprises a hexanucleotide sequence selected from the group consisting of AATAAA, ATTAAA, AGTAAA, CATAAA, TATAAA, GATAAA, ACTAAA, AATATA, AAGAAA, AATAAT, AAAAAA, AATGAA, AATCAA, AACAAA, AATCAA, AATAAC, AATAGA, AATTAA, AATAAG and other homologous or substantially identical DNA sequences wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene. The polyadenylation signal is also a subsequence of an immunoglobulin gene that comprises one of the hexanucleotide sequences described above and a G/T-rich sequence found downstream of the hexanucleotide sequence wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene. The polyadenylation signal is also a subsequence of an

immunoglobulin gene that comprises one of the hexanucleotide sequences described above, the downstream G/T-rich sequence and an additional downstream GC-rich nucleotide sequence wherein the subsequence is capable of signaling polyadenylation in the transcription of the immunoglobulin gene.

- 5 The immunoglobulin gene polyadenylation signal may be a subsequence from an immunoglobulin gene of any isotype, including immunoglobulin M (IgM), immunoglobulin D (IgD), immunoglobulin A (IgA1 and IgA2), immunoglobulin E (IgE) and immunoglobulin G (IgG). The polyadenylation signal may be from an immunoglobulin light-chain gene (including λ and κ genes) or from an immunoglobulin heavy-chain gene. In one embodiment
- 10 of the present invention, the polyadenylation signal is not an immunoglobulin κ light-chain gene polyadenylation signal. The polyadenylation signal may be from any mammalian species. Examples of mammalian species include, but are not limited to, human, primate, bovine, ovine, porcine and rodent species. In one embodiment of the present invention, the immunoglobulin gene polyadenylation signal is not a mouse κ gene polyadenylation signal.
- 15 In another embodiment, the polyadenylation signal is not a mouse immunoglobulin light-chain gene polyadenylation signal. In a further embodiment, the polyadenylation signal is not a mouse immunoglobulin gene polyadenylation signal.

- Preferably, the immunoglobulin gene polyadenylation signal is a subsequence from an immunoglobulin gene of an IgG subclass. The IgG gene polyadenylation signal may be
- 20 selected from any appropriate mammalian species. Polyadenylation signals may be selected from IgG genes including, but not limited to, human IgG subclasses IgG1, IgG2, IgG3 and IgG4 (See Takahashi et al. (1982), Cell, 29(2): 671-9), murine IgG subclasses IgG1, IgG2a, IgG2b and IgG3, rat IgG subclasses IgG1, IgG2a, IgG2b, IgG2c and IgG3, bovine and sheep IgG subclasses IgG1 and IgG2. In one embodiment, the polyadenylation signal is an
- 25 IgG heavy-chain gene polyadenylation signal. In another embodiment, the polyadenylation signal is selected from the group consisting of a human and murine IgG heavy-chain gene polyadenylation signal. In a further embodiment, the polyadenylation signal is a human or mouse IgG1 heavy-chain gene polyadenylation signal. In an additional embodiment, the polyadenylation signal is a human IgG1 heavy-chain gene polyadenylation signal. In further
- 30 embodiments, the polyadenylation signal is a nucleotide sequence comprising the sequence of SEQ ID NO:10 (FIG. 4) or a subsequence thereof comprising the sequence of SEQ ID NO:9 (FIG. 3). In other embodiments, the polyadenylation signal is a nucleotide sequence consisting of the sequence of SEQ ID NO: 10 (FIG. 4) or the subsequence thereof, SEQ ID

NO:9 (FIG. 3). The polyadenylation signal is additionally a nucleotide sequence which comprises or consists of a nucleotide sequence with homology or sequence identity to SEQ ID NO:9, SEQ ID NO:10 or any of the other polyadenylation signals sequences described herein. As used herein, two sequences have sequence identity or homology when the

5 nucleotide sequences are homologous or identical by at least about 60%, 70%, 80%, 90%, 95% or more. In some embodiments, the sequence identity can be about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher. Substantial identity also exists when the nucleic acid segments will hybridize under selective hybridization conditions (e.g., highly stringent hybridization conditions), to the complement of the strand. The nucleic acids may

10 be present in whole cells, in a cell lysate, or in a partially purified or substantially pure form.

As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in Current Protocols in Molecular Biology (1989) John Wiley & Sons, N.Y., 6.3.1-6.3.6, which is incorporated by

15 reference. Aqueous and nonaqueous methods are described in that reference and either can be used. For applications that require high selectivity, one will typically desire to employ relatively high stringency conditions to form hybrids. In solutions used for some membrane based hybridizations, addition of an organic solvent such as formamide allows the reaction to occur at a lower temperature. High stringency conditions are, for example, relatively low salt

20 and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50°C to about 70°C. Few, if any, mismatches between the probe or primers and the template or target strand would occur under high stringency conditions. In order to achieve less stringent conditions, the salt concentration may be increased and/or the temperature may be decreased. For example, a medium stringency condition could be

25 achieved at a salt concentration of about 0.1 to 0.25 M NaCl at a temperature of about 37°C to about 55°C, while a low stringency condition could be achieved at a salt concentration of about 0.15 M to about 0.9 M NaCl, at temperatures ranging from about 20°C to about 55°C. Selection of components and conditions for hybridization are well known to those skilled in the art and are reviewed in Ausubel et al. (1997, Short Protocols in Molecular Biology, John

30 Wiley & Sons, New York N.Y., Units 2.8-2.11, 3.18-3.19 and 4-64.9).

Calculations of "homology" or "sequence identity" between two sequences (the terms are used interchangeably herein) are performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a

- 10 -

second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In another embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, or 100%, of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In an additional embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5. The percent identity between two amino acid or nucleotide sequences can also be determined using the algorithm of E. Meyers and W. Miller ((1989) CABIOS, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

It is to be understood that the polyadenylation signals of the present invention may include functional equivalents of a polyadenylation signal from an immunoglobulin gene. A

functional equivalent is a nucleotide sequence which has been modified, e.g., by substitution of one nucleotide for another, but which is able to signal polyadenylation in substantially the same manner as a polyadenylation signal from an immunoglobulin gene and/or which is able to form substantially the same three-dimensional structure as a polyadenylation signal from an immunoglobulin gene, such that the polyadenylation protein complex is able to bind.

The polyadenylation signals described herein may be prepared by procedures well known in the art, including digestion of an immunoglobulin gene nucleotide sequence and/or by chemical synthesis. The DNA sequence comprising the polyadenylation signal can be prepared from an immunoglobulin gene by cleavage with a known restriction endonuclease. Thereafter, the DNA sequence can be used as a polyadenylation signal after cleavage of several additional nucleotides beyond those actually cleaved from the immunoglobulin gene by the restriction endonuclease. The techniques for the use of such restriction endonucleases for accomplishing the cleavage of DNA sequences are well known in the art. Alternatively, additional nucleotides may be added to a polyadenylation signal excised from the immunoglobulin gene to create a larger polyadenylation signal which includes non-genomic sequences. For example, if a genomic immunoglobulin sequence were obtained consisting of 50 nucleotides upstream and 50 nucleotides downstream from the 3' end, a non-genomic DNA compound might be ligated either upstream or downstream thereof (or both) in order to obtain a substantially larger DNA compound comprising an immunoglobulin gene polyadenylation signal in accordance with the present invention. Finally, total or partial chemical synthesis is optionally employed to prepare any immunoglobulin gene polyadenylation signal described above.

Sequences substantially identical, similar or homologous to an immunoglobulin gene polyadenylation signal disclosed herein are also part of this invention. As discussed above, the polyadenylation signal is a nucleotide sequence comprising the AATAAA sequence or other homologous or substantially identical hexanucleotide sequences from an immunoglobulin gene. The polyadenylation signal is also a nucleotide sequence comprising the AATAAA sequence and a downstream G/T-rich sequence of an immunoglobulin gene. Further, as discussed above, the polyadenylation signal is also a nucleotide sequence comprising the AATAAA sequence, a downstream G/T-rich region and a downstream GC-rich region of an immunoglobulin gene which may form a hairpin loop structure.

In the human IgG1 heavy-chain gene described in GenBank Accession Number J00228, an AATAAA sequence is found at or near nucleotides 1903-1908, a GT/T-rich

- 12 -

region is found at or near nucleotide 1941 and a GC-rich region is found at or near nucleotide 1963 to the 3' end of the gene (SEQ ID NO:1). In the human IgG2 heavy-chain gene described in GenBank Accession Number J00230, an AATAAA sequence is found at or near nucleotides 1903-1908, a GT/T-rich region is found at or near nucleotide 1941 and a GC-rich region is found at or about nucleotide 1963 to the 3' end of the gene (SEQ ID NO:2). In the human IgG3 heavy-chain gene described in GenBank Accession Number X03604, an AATAAA sequence is found at or near nucleotides 2484-2489, a GT/T-rich region is found at or near nucleotide 2514 and a GC-rich region is found at or about nucleotides 2535 to the 3' end of the gene (SEQ ID NO:3). In the human IgG4 heavy-chain gene described in GenBank Accession Number K01316, an AATAAA sequence is found at or near nucleotides 1904-1909, a GT/T-rich region is found at or near nucleotide 1941 and a GC-rich region is found at or about nucleotide 1964 to the 3' end (SEQ ID NO:4). In the mouse IgG1 heavy-chain gene described in GenBank Accession Number J00453, an AATAAA sequence is found at or near nucleotides 1716-1721 and a GT/T-rich and GC-rich regions are found downstream of nucleotide 1721 (SEQ ID NO:5). In the mouse IgG2a heavy-chain gene described in GenBank Accession Number J00470, an AATAAA sequence is found at or near nucleotides 1838-1844 and GT/T-rich and GC-rich regions are found downstream of nucleotide 1844 (SEQ ID NO:6). In the mouse IgG2b heavy-chain gene described in GenBank Accession Number J00461, an AATAAA sequence is found at or near nucleotides 1787-1792 and GT/T-rich and GC-rich regions are found downstream of nucleotide 1792 (SEQ ID NO:7). In the mouse IgG3 heavy-chain gene described in GenBank Accession Number X00915, an AATAAA sequence is found at or near nucleotides 6971-6976 and a GT/T-rich and GC-rich regions are found downstream of nucleotide 6976 (SEQ ID NO:8).

In addition to an immunoglobulin gene polyadenylation signal, the novel expression vector comprises a DNA sequence encoding a polypeptide, wherein said DNA sequence is other than a genomic DNA sequence from said immunoglobulin gene. As used herein, a "genomic DNA sequence from an immunoglobulin gene" is a DNA sequence that contains both the exons and the introns from the immunoglobulin gene. The DNA sequence encoding a polypeptide may encode any desired polypeptide with an encoding DNA sequence which is known or can be determined. Such polypeptides include, but are not limited to, proteins such as cytokines (such as TNF- α , IL-1a, IL-1b, IL-4, IL-6, IL-8, IL-10, IL-14, IL-18, IFN- γ , platelet activating factor, macrophage migration inhibitory factor and HMGB1), growth factors (such as granulocyte colony stimulating factor, granulocyte-macrophage colony

- 13 -

stimulating factor, mast cell growth factor, stem cell growth factor, epidermal growth factor and growth hormone), protein receptors (such as receptors for cytokines, growth factors, advanced glycation end-products, neurotransmitters and hormones) and high mobility group proteins. In another embodiment, the DNA sequence encoding a polypeptide, encodes for a high mobility group (HMG) protein or a fragment thereof. Preferably, the HMG protein is an HMGB protein or a fragment thereof. More preferably, the HMGB protein is selected from the group consisting of an HMGB1 protein (for example, as described in GenBank Accession Numbers U51677 and U00431), an HMGB2 protein (for example, as described in GenBank Accession Numbers M83665 and X67668) and an HMGB3 protein (for example, as described in GenBank Accession Numbers NM_005342 and Y10044). Fragments of HMGB proteins include, for example, the A and B box fragments (described for example in U.S. Patent Application Publication 2003/0060410 A1). The DNA sequence encoding an HMGB protein or fragment thereof preferably encodes an HMGB1 protein or fragment thereof.

The DNA sequence encoding a polypeptide may also encode immunoglobulin molecules, or portions thereof, such as antibodies, antibody fragments including Fv, Fab, Fab' and F(ab')₂ fragments. Antibodies include chimeric and humanized antibodies or fragments thereof. The DNA sequence encoding a polypeptide may also encode for various fusion proteins including immunoglobulin fusion proteins. DNA sequences encoding such immunoglobulin fusion proteins may be prepared using standard methods. Methods for the preparation of these DNA sequences are described, for example, in U.S. Patent Numbers 5,116,964, 5,428,130, 5,455,165, 5,514,582, 5,155,027, 5,567,584, 6,018,026, 6,291,646 and 6,323,323.

In addition to a DNA sequence encoding a polypeptide and an immunoglobulin gene polyadenylation signal, the novel expression vector can additionally comprise a suitable promoter which is upstream from and operably linked to the DNA sequence encoding a polypeptide. The promoter utilized in the inventive expression vector can be any promoter which is capable of directing transcription of a DNA sequence encoding a polypeptide when operably linked to said promoter. Commonly utilized promoters which can be used in the inventive expression vectors are derived from Polyoma, Adenovirus, SV40, beta-globin, EF-1a, human cytomegalovirus (CMV), β -actin, α -fetoprotein, γ -globin, β -interferon, Metallothionein II (MT II), amylase, cathepsin, M1 muscarinic receptor, gamma-glutamyl transferase, mouse mammary tumor virus (MMTV) and Rous sarcoma virus (RSV). Suitable CMV promoter sequences can be obtained from the CMV-promoted β -galactosidase

- 14 -

expression vector, CMV.beta. (MacGregor et al. (1989), Nucl. Acids Res. 17:2365). Suitable amylase promoter, especially murine amylase promoter sequences are described by Wu et al. (Molec. Cell. Biol. 11:4423-4430 (1991)). Suitable cathepsin E promoter sequences are described by Azuma et al. (J. Biol. Chem. 267:1609-1614 (1992)). Suitable M1 muscarinic
5 receptor promoter sequences are described by Fraser et al. (Molec. Pharmacol. 36:840-847 (1989)) and by Bonner (Trends Neurosci. 12:148-151 (1989)). Early and late SV40 promoters are described by Fiers et al. (1978) (Nature 273: 113). Preferably, the polynucleotides and expression vectors comprises a CMV or EF-1a promoter.

Optionally, the novel expression vector further comprises transcription enhancers.
10 Enhancers are cis-acting elements of DNA, usually about 10 to 300 nucleotides in size that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes including, for example, globin, elastase, albumin, α -fetoprotein, and insulin. Enhancers from a eukaryotic cell virus can also be utilized. Examples include the SV40 enhancer on the late side of the replication origin, the cytomegalovirus early promoter
15 enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer can be spliced into the vector at a position 5' or 3' to a coding sequence, but is preferably located at a site 5' from the promoter.

The novel expression vectors can additionally comprise a selectable marker which can be utilized to determine whether the recombinant DNA compound has been incorporated into
20 the expressed genetic material of a living cell. The use of such markers and the principles of selectivity therefore are well known and well recognized in the art. The nucleotide sequence which comprises the selectable marker operates essentially independent of the DNA sequence encoding a desired polypeptide, the promoter and the polyadenylation signal. Accordingly, the location of the selectable marker in the vector relative to the other three sequences is not
25 critical. Selectable markers which can be used in the inventive vectors include antibiotic resistance genes which are useful in determining incorporation into those living cells towards which the antibiotic is otherwise lethal. In mammalian cells, for example, the aminoglycoside antibiotic geneticin (G418) results in inhibition of protein synthesis and death of the cell. Accordingly, the G418 resistance gene comprises a suitable selectable marker. Similarly, the
30 dihydrofolate reductase (DHFR) gene provides a useful marker for cells (e.g., Chinese hamster ovary cells) which can be prepared DHFR deficient.

Optionally, another technique for selection which does not require the utilization of a selectable marker is the use of co-transference techniques whereby both the DNA compound

- 15 -

of the present invention absent a marker and a separate DNA molecule containing the marker are co-transferred into a cell. See, for example, U.S. Patent No. 4,399,216. Incorporation into the expressed genetic material of a higher eukaryotic cell is accomplished by known means. Typically, the incorporation will result in the presence of a DNA compound in the

5 chromosomal structure of the cell, but may also be accomplished by an extra-chromosomal incorporation, e.g., utilization of a recombinant bovine papilloma virus. See Hsiung, et al., J. Molecular and Applied Genetics 2:497-506 (1984) for the methodology for such incorporation.

10 *Eukaryotic cells transfected with novel expression vectors and methods of producing a desired polypeptide*

The present invention is also directed to eukaryotic cells or eukaryotic cell-lines transfected with a novel expression vector comprising an immunoglobulin gene polyadenylation signal and a DNA sequence encoding a polypeptide, wherein said DNA

15 sequence is other than a genomic DNA sequence from said immunoglobulin gene and wherein said DNA sequence is upstream from and operably linked to said immunoglobulin gene polyadenylation signal.

One embodiment of the present invention is directed to mammalian cells comprising the inventive expression vectors described above. Any suitable mammalian cell or cell line

20 may be used in accordance with the present invention. Mammalian cell lines that have been used to express recombinant proteins and which can be used in the present invention include, but are not limited to, COS-7 cells (Guzman et al. (1981), Cell 23: 175), CV-1/EBNA cells (ATCC CRL 10478), L cells, Chinese hamster ovary (CHO) cells, HeLa cells, BHK cells, MDCK cells, 323 cells, 293 cells, MCF-7 cells, NSO cells, SP2 cells and HEK-293 cells. In

25 one embodiment, the mammalian cell comprising the novel expression vector is a CHO cell. The cell or cell-line used to express the desired recombinant protein can also be an immortalized mammalian cell line, which may be of lymphoid origin, such as a myeloma, hybridoma, trioma or quadroma cell line. The cell line can also comprise a normal lymphoid cell, such as a B-cell, which has been immortalized by transformation with a virus, such as

30 the Epstein-Barr virus. The immortalized cell line can, for example, be a myeloma cell line or a derivative thereof. Other suitable host cells are known to those skilled in the art.

A cell or cell-line can be transfected using any suitable method now known or later discovered. Several methods of transfection are known in the art and can be utilized for the

- 16 -

practice of the present invention. Known transfection methods include calcium phosphate precipitation, DEAE dextran-mediated gene transfer, liposome-mediated gene transfer, electroporation, microinjection, retroviral transfection and the use of gene guns. The particular transfection method employed will depend on many factors including the eukaryotic cell type utilized and polypeptide to be produced. For example, calcium phosphate precipitation (described for example in Wigler et al. (1980), Proc Natl Acad Sci USA 77: 3567) and liposome-mediated gene transfer are useful for the transfection of most mammalian cells but have been found to be more effective with adherent cells. In contrast, electroporation (described for example in Potter et al. (1988), Proc Natl Acad Sci USA 81: 7161) is primarily utilized for the transfection of cells in suspension. Retroviral transfection methods are useful for transfection of many mammalian cell types (described for example in Mann et al. (1993), Cell 33:153-159; Pear et al., PNAS USA 90(18):8392-6 (1993); Kitamura et al., PNAS USA 92:9146-9150 (1995)). Various techniques for transfection of mammalian cells are further described in Keown et al., 1990 Methods in Enzymology, 185:527-537, Mansour et al. (1998), Nature, 336:348-352 and Sambrook, et al., Molecular Cloning, 2nd Ed., Vol. 3, Chapter 16, particularly sections 68-72 (1989).

The eukaryotic cell or cell line can be transfected with a novel expression vector comprising a polyadenylation signal from an immunoglobulin gene of any isotype as discussed above. In one embodiment, the polyadenylation signal is from an IgG subclass. In another embodiment, the polyadenylation signal is from an IgG1 gene. In a further embodiment, the polyadenylation signal from an IgG1 heavy-chain gene. In an another embodiment, the polyadenylation signal is from a human IgG1 heavy-chain gene. In another embodiment, a mammalian cell or cell line is transfected with a novel expression vector comprising a human IgG1 heavy-chain gene. In yet another embodiment, the mammalian cell or cell line is a CHO cell or cell line.

The present invention is also directed to a method of producing a desired polypeptide by culturing a mammalian cell transfected with a novel expression vector and recovering said desired polypeptide from the culture medium. The polypeptide can be isolated and purified using conventional biochemical methods.

Conditions suitable for growth of the mammalian cell or cell-line and for expression of the vector are well known in the art and are described for example in Sambrook et al., 1989, "Molecular Cloning: A Laboratory Manual." The desired polypeptide can be recovered after culturing cells for a period of time sufficient to allow expression of the desired

- 17 -

polypeptide in the mammalian cell and release of the desired polypeptide into the culture medium where it can be recovered using standard protein purification methods.

The invention is illustrated by the following examples which are not intended to be limiting in any way.

5

EXEMPLIFICATION

Example 1: Construction of human IgG1 heavy-chain gene polyadenylation signal sequence 1 ("IGPA")

A human IgG1 heavy-chain gene polyadenylation signal sequence was assembled by PCR using the overlapping primers designated pIGPA5 and pIGPA3 in Table 1 below.

Briefly, primers were combined in various molar ratios in a cyclor reaction of 1 step at 94°C for 2 minutes, followed by 8 cycles at 94°C for 2 minutes and 5 minutes at 72°C with a 30 second ramp time between each step, followed by a step at 72°C for 10 minutes. The product ("cIGPA") obtained was then used as a template for the final amplification reaction.

The IGPA sequence was amplified using cIGPA as a template. cIGPA was combined with the overlapping primers designated pIGPA5a and pIGPA3a in Table 1.

A standard amplification cycle of 1 step at 94°C for 2 minutes, followed by 30 cycles at 94°C for 1 minute, 55°C for 30 seconds and 72°C for 5 minutes, followed by a step at 72°C for 10 minutes, was used. The complete IGPA (shown below) is 187 base pairs in length and is used in the final expression vector and as a template for the 5' region for IgG1 heavy-chain gene polyadenylation signal sequence 2 (LIGPA).

IGPA: SEQ ID NO:9 (lower case indicates cloning sites)

5' gctctagaGTGCGACGGCCGGCAAGCCCCCGCTCCCCGGGCTCTCGCGGTTCGC
ACGAGGATGCTTGGCACGTACCCCTGTACATACTTCCCGGGCGCCAGCATGGA
AATAAAGCACCCAGCGCTGCCCTGGGCCCCTGCGAGACTGTGATGGTTCTTTCCA
CGGGTCAGgcgccgctaaactat 3'

TABLE 1

Primers used to create polyadenylation signal IGPA

Primer	Sequence	SEQ ID NO.
pIGPA5	5' CGGCCGGCAAGCCCCCGCTCCCCGGGCTCTCG CGGTTCGCACGAGGATGCTTGGCACGTACCCCTG TACATACTTCCCGGGCGCCAGC 3'	SEQ ID NO:10
pIGPA3	5' AAGAACCATCACAGTCTCGCAGGGGCCAGGG CACGGCTGGGTGCTTTATTTCCATGCTGGGCGCC	SEQ ID NO:11

- 18 -

CGGGAAGTATGTAC 3'

pIGPA5a 5' GCTCTAGAGTGCACGGCCGGCAAGCCCCCGC SEQ ID NO:12
 TCCCCGGGCTC 3'

pIGPA3a 5' ATAGTTTAGCGCCGCCTGACCCGTGGAAAGA SEQ ID NO:13
 ACCATCACAGTCTCGCAG 3'

Example 2: Construction of human IgG1 heavy-chain gene polyadenylation signal sequence 2 ("LIGPA")

A second human IgG1 heavy-chain gene polyadenylation signal sequence was
 5 assembled by PCR using the overlapping primers designated as pLIGPA5 and pLIGPA3 in
 Table 2 below.

Briefly, these primers were combined in various molar ratios in a cyler reaction of 1
 step at 94°C for 2 minutes, followed by 8 cycles at 94°C for 2 minutes and 72°C for 5 minutes
 with a 30 second ramp time between each step, followed by a step at 72°C for 10 minutes.
 10 The product ("cLIGPA3") obtained was then used as a template for the amplification reaction
 for the 3' region of LIGPA.

The 5' region of the LIGPA sequence ("LIGPA5i") was amplified using IGPA as a
 template combined with the overlapping primers designated pIGPA5a in Table 1 and
 pLIGPA5b in Table 2 below. The 3' region of LIGPA ("LIGPA3i") was obtained using
 15 cLIGPA3 as a template combined with the overlapping primers designated pLIGPA5a and
 pLIGPA3a in Table 2 below. Standard amplification cycles of 94°C for 2 minutes, followed
 by 30 cycles at 94°C for 1 minute, 55°C for 30 seconds and 72°C for 5 minutes, followed by a
 step at 72°C for 10 minutes, were used.

The products of the above two reactions (LIGPA5i and LIGPA3i) were combined in
 20 various molar ratios in a cyler reaction of 94°C for 2 minutes, followed by 8 cycles at 94°C
 for 2 minutes and 72°C for 5 minutes with a 30 second ramping time before each step,
 followed by a step at 72°C for 10 minutes. The product of this reaction ("cLIGPA") served
 as a template for the final amplification reaction of LIGPA.

The complete LIGPA was amplified using cLIGPA as a template with the primers
 25 designated pLIGPA5a and pLIGPA3a in Table 2 below. A standard amplification cycle of
 one step at 94°C for 2 minutes, followed by 30 cycles at 94°C for 1 minute, 55°C for 30
 seconds and 72°C for 5 minutes, followed by a step at 72°C for 10 minutes, was used. The

- 19 -

product LIGPA (shown directly below) is 347 base pairs in length and is used in the final expression vector.

LIGPA: SEQ ID NO:10 (lower case indicates cloning sites)

```

5  5' gctctagaGTGCGACGGCCGGCAAGCCCCCGCTCCCCGGGCTCTCGCGGTTCGC
   ACGAGGATGCTTGGCACGTACCCCCTGTACATACTTCCCGGGCGCCCAGCATGGA
   AATAAAGCACCCAGCGCTGCCCTGGGCCCCTGCGAGACTGTGATGGTTCTTTCCA
   CGGGTCAGGCCGAGTCTGAGGCCTGAGTGGCATGAGGGAGGCAGAGCGGGTCCC
   ACTGTCCCCACACTGGCCCAGGCTGTGCAGGTGTGCCTGGGCCCCCTAGGGTGGG
10 GCTCAGCCAGGGGCTGCCCTCGGCAGGGTGGGGGATTGTCAGCGTGGCCCTCC
   CTCCAgcgccgctaaactat 3'

```

TABLE 2

Primers used to create polyadenylation signal LIGPA

Primer	Sequence	SEQ ID NO.
pLIGPA5b	5' GCCACTCAGGCCTCAGACTCGGCCTGACCCGTGGAA AGAACCATCACAGTCTCGC 3'	SEQ ID NO:14
pLIGPA5	5' GTGGCATGAGGGAGGCAGAGCGGGTCCCACTGTCCC CACACTGGCCCAGGCTGTGCAGGTGTGCCTGGGCCCCCTAG 3'	SEQ ID NO:15
pLIGPA3	5' GGGAGGGCCACGCTGGCAAATCCCCACCCTGCCGA GGGCAGCCCCTGGCTGAGCCCCACCCTAGGGGGCCCAGGCA CACCTGCACAGC 3'	SEQ ID NO:16
pLIGPA5a	5' GGCCGAGTCTGAGGCCTGAGTGGCATGAGGGAGGC AGAGC 3'	SEQ ID NO:17
pLIGPA3a	5' ATAGTTTtagCGGCCGCTGGAGGGAGGGCCACGCTGG CAAATCC 3'	SEQ ID NO:18

15 *Example 3: Construction of expression vectors comprising either the IgG1 heavy-chain gene polyadenylation signal sequence 1 (pCTiTOK28) or the IgG1 heavy-chain gene polyadenylation signal sequence 2 (pCTiTOK29)*

The mammalian expression vector pcDNA3 (Invitrogen, San Diego, CA) was first altered to replace the CMV promoter with the EF-1a promoter for higher protein expression.

20 The pcDNA3 expression vector was also altered to add a Not I site downstream of the BGH

- 20 -

polyadenylation signal sequence in order to allow the addition of a second expression cassette. This altered vector is designated pCTiTOK18.

The BGH polyadenylation signal sequence of pCTiTOK18, flanked by Xba I and Not I enzyme restriction sites, was removed and replaced by either the IGPA or the LIGPA polyadenylation signal sequences, which were created with flanking Xba I and Not I enzyme restriction sites. These mammalian expression vectors, which contain either the IGPA or LIGPA polyadenylation signal sequence, are designated pCTiTOK28 and pCTiTOK29, respectively.

The vectors pCTiTOK18, pCTiTOK28 and pCTiTOK29 all contain the gene for ampicillin selection in bacteria and geneticin (G418) selection in mammalian cells. The promoter, EF-1a, is separated from the polyadenylation signal sequences by a region of multiple cloning sites into which a desired cDNA may be functionally cloned.

Example 4: RAGE-Ig fusion protein production

The RAGE-Ig fusion protein was created by PCR assembly in which two overlapping cDNA fragments were created, annealed by PCR and used as a template for the final PCR product. The extracellular region of RAGE (e.g., amino acids 1-305 of GenBank Accession No. NP_001127; SEQ ID NO:19) along with its signal sequence was amplified using human splenic cDNA as a template and the primers designated pTOK16a5 and pTOK17aA, described in Table 3 below. The primer pTOK16a5 contains the sequence for an EcoRI restriction site and a Kozak sequence. The primer pTOK17aA contains 35 bases that overlap with the 5' region of the human IgG1 Fc.

The human IgG1 Fc region was amplified from a previous human IgG1 fusion protein in which the Fc receptor region was removed by mutation (L235A and G237A). The sequence used covered the hinge CH2 and CH3 regions (e.g., amino acids 219-447 of GenBank Accession No. J00228) and was amplified using the primers pTOK17aB and pTOK10a3. The primer pTOK17aB contained 35 bases that overlapped with the 3' region of the extracellular region of RAGE and the primer pTOK10a3 contains a stop codon and the sequence for the XbaI restriction site. The PCR cycle involved 1 step at 94°C for 2 minutes, followed by 30 cycles at 94°C for 1 minute, 55°C for 30 seconds and 72°C for 5 minutes, followed by a step of 72°C for 10 minutes. The two fragments were then gel purified and combined in equal molar ratios in a PCR reaction without primers that used a PCR cycle involving one step at 94°C for 2 minutes, followed by 8 cycles of 94°C for 2 minutes and

- 21 -

72°C for 5 minutes with a 30-second ramping time at each step, followed by a step at 72°C for 10 minutes. The product of this reaction then served as a template in another amplification reaction using the primers pTOK16a5 and pTOK10a3 with the same cyclor steps used in the first amplification. This product was TA-cloned and sequenced to find the correct cDNA construct. The product, a RAGE-Ig fusion polypeptide, is described as SEQ ID NO:20.

The three expression vectors pCTiTOK17 (BGHpA), pCTiTOK30 (IGPA) and pCTiTOK31 (LIGPA) containing the RAGE-Ig fusion cDNA were all created by functionally cloning the EcoRI-XbaI fragment containing the RAGE-Ig fusion cDNA into the vectors pCTiTOK18, pCTiTOK28 and pCTiTOK29 respectively. After the sequence confirmed, DNA was maxiprepmed and transfected into CHO cells as described below.

TABLE 3

15 Primers used in the creation of the RAGE-Ig fusion construct

Primer	Sequence	SEQ ID NO.
pTOK16 a5	5' CCGGAATTCCTCACCATGGCAGCCGGAACAGCAGT TGGAG 3'	SEQ ID NO:21
pTOK17aA	5' GTTTTGTCGCTGGAATGGGTGGCCACACAGCTGTA GGTTCCTGGTCCTGAG 3'	SEQ ID NO:22
pTOK17aB	5' GCTGTGTGGCCACCCATTCCAGCGACAAACTCA CACATGCCACCGTGCCAGCACCTG 3'	SEQ ID NO:23
p TOK10a3	5' TGCTCTAGATTATTTACCCGGAGACAGGGAGAGGC CTTCTGCGTGTAGTGGTTGTGCAGAGCCTCATGCAT CACGG 3'	SEQ ID NO:24

20 Example 5: Preparation of CHO cell transfectants

This example describes the transfection of CHO cells with the constructs pCTiTOK18, pCTiTOK28 and pCTiTOK29. CHO-S cells (Gibco/Invitrogen, Carlsbad, CA) were transfected according to the manufacturer's instructions using recommended reagents. Briefly, different concentrations of either of the expression vectors pCTiTOK18, pCTiTOK28 or pCTiTOK29 and the DMRIE-C reagent, a mixture of charged and neutral lipids, were

- 22 -

mixed to form complexes for 30-45 minutes. These complexes were then mixed with the CHO cells in serum-free DMEM in the absence of antibiotics. The cells take up the complexes and incorporate the DNA into their chromosomes over the next 5 hours. The transfection was stopped by the addition of media containing 10% fetal bovine serum. The CHO cells were then allowed to recover and incubated at 37°C overnight in 8% CO₂.

The successfully transfected cells were selected for growth in media containing 10% fetal bovine serum and 50 mg/ml geneticin for 5 days. Single cell clones were then plated into 96-well plates, isolated for the production of human IgG Fc containing protein and selected for an additional 14 days. Single cell clones were then identified and their supernatants were tested for the presence of recombinant protein. Clones determined to be producing the recombinant protein were expanded and adapted to protein-free media (the clones may also be kept in 10% fetal bovine serum, however, the presence of serum increases the difficulty of purification). The CHO-S cells easily adapt to growth in Gibco/Invitrogen's protein-free CD-CHO media and shaker flask culture which allows for rapid expansion on the cells. Cell culture media was then collected and the presence of RAGE-Ig fusion protein was determined as described below.

Example 6: RAGE-Ig fusion protein production by cells transfected with pCTiTok17 (BGHpA), pCTiTok30 (IGPA) and pCTiTok31 (LIGPA)

The presence and quantity of RAGE-Ig fusion protein in the cell culture media of CHO cells transfected with either the expression vector pCTiTok17, pCTiTok30 or pCTiTok31, was determined using a standardized ELISA. Briefly, ELISA plates were coated with 0.4 ug/ml of goat anti-human IgG (Jackson ImmunoResearch Laboratories, West Grove, PA) and incubated overnight at 4°C. The plates were then blocked with PBS containing 1% BSA for one hour at room temperature. After washing, the plates were incubated with the CHO cells supernatants obtained as described above. After 2 hours at room temperature, the plates were washed and incubated with mouse anti-human IgG-HRP (Jackson ImmunoResearch Laboratories, West Grove, PA) diluted at 1:2000 in PBS. After washing, the plates were developed with TMB (Invitrogen, San Diego, CA) and absorbance at 655 nm was measured using a plate reader. The amount of RAGE-Ig fusion protein was quantified by comparison to a standard dilution curve of human IgG (Jackson ImmunoResearch Laboratories, West Grove, PA). The results of this quantification is

- 23 -

presented in FIG. 8. The protein may also be quantified using Western analysis and functional assays.

FIG. 8 shows the results of a preliminary experiment measuring the quantity of RAGE-Ig fusion protein (ug/ml) found in the supernatants of CHO cells transfected with pCTiTOK17, pCTiTOK30 or pCTiTOK31 (comprising the BGH, IGPA or LIGPA polyadenylation signal sequence, respectively). CHO cells transfected with pCTiTOK17 were cloned cells; consequently it was estimated that 100% of the cells were able to produce the RAGE-Ig fusion protein. The cells transfected with pCTiTOK30 and pCTiTOK31 were uncloned cells (i.e., primary pools of transfectants) cultured for 12 and 6 days, respectively. Because the latter cells are uncloned, it is estimated that less than 10% of the cells are able to produce the RAGE-Ig fusion protein. The two bars for the IGPA and LIGPA containing cells represent the results of two separate transfections. A greater quantity of recombinant protein in the supernatants indicates that the protein is properly folded and secreted from the cells, which is suggestive of proper polyadenylation. As shown in FIG. 8, the data suggest that cells transfected with an expression vector comprising either the IGPA or LIGPA polyadenylation signal sequence are able to produce the RAGE-Ig fusion protein

FIG. 9 shows the results of an experiment measuring the quantity of RAGE-Ig fusion protein (ug/ml) in the supernatants from four CHO cell clones transfected with the vector pCTiTOK31 (comprising LIGPA) over a 10 day period. On Day 0, the starting number of cells for each of the five clones was 1.3×10^8 cells in 750 ml CD-CHO/G418. As shown in FIG. 9, the quantity of RAGE-Ig fusion protein produced by the cells increases over time and by the tenth day, the cells produce a quantity of RAGE-Ig fusion protein ranging from about 11 to 21 ug/ml.

The teachings of all publications cited herein are incorporated herein by reference in their entirety.

While this invention has been particularly shown and described with references to the preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

- 5 1. A polynucleotide comprising:
 - a) an IgG heavy-chain gene polyadenylation signal, and
 - b) a DNA sequence encoding a polypeptide, wherein said DNA sequence is other than a genomic DNA sequence from said IgG heavy-chain gene, and wherein said DNA sequence is upstream from and operably linked to said IgG heavy-chain gene polyadenylation signal.
- 10 2. The polynucleotide of claim 1, wherein said IgG heavy-chain gene polyadenylation signal is a human IgG heavy-chain gene polyadenylation signal.
- 15 3. The polynucleotide of claim 2, wherein said human IgG heavy chain gene polyadenylation signal is selected from the group consisting of an IgG1 heavy-chain gene polyadenylation signal, an IgG2 heavy-chain gene polyadenylation signal, an IgG3 heavy-chain gene polyadenylation signal and an IgG4 heavy-chain gene polyadenylation signal.
- 20 4. The polynucleotide of claim 3, wherein said human IgG heavy-chain gene polyadenylation signal is an IgG1 heavy-chain gene polyadenylation signal.
- 25 5. The polynucleotide of claim 4, wherein said IgG1 heavy chain gene polyadenylation signal comprises the nucleotide sequence of SEQ ID NO:9.
6. The polynucleotide of claim 5, wherein said IgG1 heavy chain gene polyadenylation signal comprises the nucleotide sequence of SEQ ID NO:10.
- 30 7. The polynucleotide of claim 1, further comprising a promoter sequence upstream from and operably linked to said DNA sequence encoding a polypeptide.
8. The expression vector of claim 7, wherein said promoter is selected from the group consisting of an EF-1a promoter and a CMV promoter.

- 25 -

9. The expression vector of claim 8, further comprising a selectable marker.
10. The polynucleotide of claim 4, wherein said DNA sequence encodes an immunoglobulin fusion protein.
11. The polynucleotide of claim 4, wherein said DNA sequence encodes a high mobility group protein.
12. The polynucleotide of claim 11, wherein said high mobility group protein is a HMGB protein.
13. The polynucleotide of claim 12, wherein said HMGB protein is selected from the group consisting of HMGB1, HMGB2 and HMGB3.
14. An expression vector comprising:
a) an IgG heavy-chain gene polyadenylation signal, and
b) a DNA sequence encoding a polypeptide, wherein said DNA sequence is other than a genomic DNA sequence from said IgG heavy-chain gene, and wherein said DNA sequence is upstream from and operably linked to said IgG heavy-chain gene polyadenylation signal.
15. The expression vector of claim 14, wherein said IgG heavy-chain gene polyadenylation signal is a human IgG heavy-chain gene polyadenylation signal.
16. The expression vector of claim 15, wherein said human IgG heavy chain gene polyadenylation signal is selected from the group consisting of an IgG1 heavy-chain gene polyadenylation signal, an IgG2 heavy-chain gene polyadenylation signal, an IgG3 heavy-chain gene polyadenylation signal and an IgG4 heavy-chain gene polyadenylation signal.
17. The expression vector of claim 16, wherein said human IgG heavy-chain gene polyadenylation signal is an IgG1 heavy-chain gene polyadenylation signal.

18. The expression vector of claim 17, wherein said IgG1 heavy chain gene polyadenylation signal comprises the nucleotide sequence of SEQ ID NO:9.
- 5 19. The expression vector of claim 17, wherein said IgG1 heavy chain gene polyadenylation signal comprises the nucleotide sequence of SEQ ID NO:10.
20. The expression vector of claim 14, further comprising a promoter sequence upstream from and operably linked to said DNA sequence encoding a polypeptide.
- 10 21. The expression vector of claim 20, wherein said promoter is selected from the group consisting of an EF-1a promoter and a CMV promoter.
22. The expression vector of claim 21, further comprising a selectable marker.
- 15 23. The expression vector of claim 17, further comprising a promoter sequence upstream from and operably linked to said DNA sequence encoding a polypeptide.
24. The expression vector of claim 23, wherein said promoter is selected from the group consisting of an EF-1a promoter or a CMV promoter.
- 20 25. A eukaryotic cell transfected with the expression vector of claim 14.
26. The eukaryotic cell of claim 25, wherein said eukaryotic cell is a mammalian cell.
- 25 27. The mammalian cell of claim 26, wherein said mammalian cell is selected from the group consisting of a CHO cell, a COS cell, a HEK cell, a HeLa cell, a BHK cell, and a CV1 cell.
- 30 28. The mammalian cell of claim 27, wherein said cell is a CHO cell.
29. A method of producing a desired polypeptide, said method comprising:

- 27 -

- 5 a) culturing a mammalian cell transfected with an expression vector, wherein said vector comprises: i) an IgG heavy-chain gene polyadenylation signal, and ii) a DNA sequence encoding a desired polypeptide, wherein said DNA sequence is other than a genomic DNA sequence from said IgG heavy-chain gene, and wherein said DNA sequence is upstream from and operably linked to said IgG heavy-chain gene polyadenylation signal; and
- b) recovering said desired polypeptide.
- 10 30. The method of claim 29, wherein said IgG heavy-chain gene polyadenylation signal is a human IgG heavy-chain gene polyadenylation signal.
- 15 31. The method of claim 30, wherein said human IgG heavy chain gene polyadenylation signal is selected from the group consisting of an IgG1 heavy-chain gene polyadenylation signal, an IgG2 heavy-chain gene polyadenylation signal, an IgG3 heavy-chain gene polyadenylation signal and an IgG4 heavy-chain gene polyadenylation signal.
- 20 32. The method of claim 31, wherein said human IgG heavy-chain gene polyadenylation signal is an IgG1 heavy-chain gene polyadenylation signal.
33. The method of claim 32, wherein said IgG1 heavy chain gene polyadenylation signal comprises the nucleotide sequence of SEQ ID NO:9.
- 25 34. The method of claim 32, wherein said IgG1 heavy chain gene polyadenylation signal comprises the nucleotide sequence of SEQ ID NO: 10.
35. The method of claim 29, further comprising a promoter sequence upstream from and operably linked to said DNA sequence encoding a polypeptide.
- 30 36. The method of claim 29, wherein said mammalian cell is selected from the group consisting of a CHO cell, a HEK cell, a Hela cell, a BHK cell, and a CV1 cell.

- 28 -

37. The method of claim 36, wherein said mammalian cell is a Chinese hamster ovary (CHO) cell.
38. The method of claim 29, wherein said expression vector further comprises a promoter sequence upstream from and operably linked to said DNA sequence encoding a desired polypeptide.
39. The method of claim 27, wherein said promoter is selected from the group consisting of an EF-1a promoter and a CMV promoter.
40. The method of claim 32, wherein said expression vector further comprises a promoter sequence upstream from and operably linked to said DNA sequence encoding a desired polypeptide.
41. The method of claim 40, wherein said promoter is selected from the group consisting of an EF-1a promoter and a CMV promoter.
42. The method of claim 29, wherein said expression vector further comprises a selectable marker.
43. The method of claim 29, wherein said desired protein is an immunoglobulin fusion protein.
44. The method of claim 29, wherein said desired protein is a high mobility group protein.
45. The method of claim 44, wherein said high mobility group protein is an HMGB protein.
46. The method of claim 45, wherein said HMGB protein is selected from the group consisting of HMGB1, HMGB2 and HMGB3.

1 agctttctgg ggcaggccag gcctgacctt ggctttgggg cagggagggg gctaagggtga
61 ggcaggtggc gccagcaggt gcacacccaa tgcccatgag cccagacact ggacgctgaa
121 cctcgcggac agttaagaac ccaggggcct ctgcgcctgg gccagctct gtccacacc
181 gcggtcacat ggcaccacct ctctgcagc ctccaccaag ggcccatcgg tcttccccct
241 ggcacctcc tccaagagca cctctggggg cacagcggcc ctgggctgcc tggtaagga
301 ctactcccc gaaccggtga cgggtgctg gaactcaggc gccctgacca gcggcgtgca
361 caccttcccg gctgtctac agtctcagg actctactcc ctacagcagc tggtagccgt
421 gccctccagc agcttgggca cccagaccta catctgcaac gtgaatcaca agcccagcaa
481 caccaagtg gacaagaaag ttggtgagag gccagcacag ggagggaggg tgtctgctgg
541 aagcaggctc agcgtcctg cctggacgca tcccggctat gcagcccag tccagggag
601 caaggcaggc cccgtctgcc tcttaccctg gacgtctgc cggccact catgtcagg
661 gagagggtct tctggcttt tccaggctc tgggcaggca caggctaggt gccctaacc
721 caggccctgc acacaaagg gcaggtgctg ggctcagacc tgccaagagc catatccggg
781 aggacctgc cctgacctc agcccacccc aaaggccaaa ctctcactc cctcagctc
841 gacacctct ctctcccg attccagtaa ctccaatct tctctctgca gagcccaa
901 cttgtgacaa aactcacaca tgccaccgt gccaggtaa gccagcccag gcctgcct
961 ccagctcaag gcgggacagg tgccctagag tagcctgcat ccagggacag gcccagccg
1021 ggtgctgaca cgtccactc catctctcc tcagcacctg aactcctggg gggaccgtca
1081 gtcttctct tcccccaaa acccaaggac accctcatga tctccggac cctgaggtc
1141 acatgcgtgg tgggtgacgt gagccacgaa gacctgagg tcaagttcaa ctggtacgtg
1201 gacggcgtgg aggtgcataa tgccaagaca aagccgcggg aggagcagta caacagcacg
1261 taccgggtgg tcagcgtcct caccgtcctg caccaggact ggctgaatgg caaggagtac
1321 aagtgcaagg tctccaaca agccctccca gcccctcag agaaaacat ctcaaagcc
1381 aaaggtggga cccgtgggt gcgagggcca catggacaga ggccggctcg gcccacctc
1441 tgccctgaga gtgaccgtg taccaacctc tgcctacag ggcagcccc agaaccacag
1501 gtgtacacc tgccccatc cgggatgag ctgaccaaga accaggctcag cctgacctg
1561 ctggtcaaag gcttctatcc cagcgacatc gccgtggagt gggagagcaa tggcagccg
1621 gagaacaact acaagaccac gcctccgtg ctggactccg acggctcctt ctctctac
1681 agcaagctca cgttgacaa gagcaggtgg cagcagggga acgtcttctc atgctccgtg
1741 atgcatgagg ctctgcacaa cactacacg cagaagagcc tctcctgtc tccgggtaaa
1801 tgagtgcgac ggccggcaag cccgctccc cgggtctcg cggtcgcagc aggatgctt
1861 gcacgtacc cctgtacata ctcccgggc gccagcatg gaaataaagc acccagcgt
1921 gccctgggcc cctgcgagac tgtgatggt cttccacgg gtcaggccga gtctgaggcc
1981 tgagtggcat gagggaggca gacgggtc (SEQ ID NO:1)

FIG. 1A

1 agctttctgg ggcgagccgg gcctgacttt ggctttgggg cagggagtgg gctaagggtga
61 ggcaggtggc gccagccagg tgcacacca atgcccgtga gccagacac tggaccctgc
121 ctggaccctc gtggatagac aagaaccgag gggcctctgc gcctggggcc agctctgtcc
181 cacaccggg tcacatggca ccaactctct tgcagcctcc accaagggcc catcggcttt
241 ccccttggcg ccctgctcca ggagcacctc cgagagcaca gccgccctgg gctgcctggt
301 caaggactac tccccgaac cggtagcggg gtcgtggaac tcaggcgctc tgaccagcgg
361 cgtgcacacc tcccagctg tctacagtc ctcaggactc tactccctca gcagcgtggt
421 gaccgtgccc tccagcaact tcggcaccca gacctacac tgcaacgtag atcacaagcc
481 cagcaacacc aaggtggaca agacagtgg tgagaggcca gctcagggag ggagggtgtc
541 tgctggaagc caggctcagc cctcctgcct ggacgcaccc cggctgtgca gcccagccc
601 agggcagcaa ggcaggcccc atctgtctcc tcaccggag gcctctgccc gcccactca
661 tgctcaggga gagggtcttc tggcttttc caccaggctc caggcaggca caggctgggt
721 gccctaccc caggcccttc acacacaggg gcaggtgctt ggctcagacc tgccaaaagc
781 catatccggg aggacctgc cctgaccta agccgacccc aaaggccaaa ctgtccactc
841 cctcagctcg gacacctct cctctccag atccgagtaa ctccaatct tctcttgca
901 gagcgcaaat gttgtgtcga gtgcccaccg tgcccaggtg agccagccca ggctctgccc
961 tccagctcaa ggcgggacag gtgccctaga gtagcctgca tccagggaca ggccccagct
1021 ggggtgctgac acgtccacct ccactcttc ctcagacca cctgtggcag gaccgtcagt
1081 cttctcttc ccccaaaaac ccaaggacac cctcatgac tcccggaccc ctgaggtcac
1141 gtgcgtgggt gtggacgtga gccacgaaga ccccagggtc cagtcaact ggtacgtgga
1201 cggcgtggag gtgcataatg ccaagacaaa gccacgggag gagcagttca acagcacgt
1261 ccgtgtggtc agcgtcctca ccgttgtgca ccaggactgg ctgaacggca aggagtacaa
1321 gtgcaaggtc tccaacaaag gcctccagc ccccatcgag aaaaccatct ccaaaaccaa
1381 aggtgggacc cgcgggggtat gagggccaca tggacagagg ccggtctggc ccacctctg
1441 ccctgggagt gaccgctgtg ccaacctctg tccctacagg gcagccccga gaaccacagg
1501 tgtacacct gccccatcc cgggaggaga tgaccaagaa ccaggtcagc ctgacctgcc
1561 tggtaaagg cttctacccc agcgacatcg ccgtggagtg ggagagcaat gggcagccgg
1621 agaacaacta caagaccaca cctccatgc tggactccga cggctcttc ttcctctaca
1681 gcaagctcac cgtggacaag agcaggtggc agcaggggaa cgtcttctca tgctccgtga
1741 tgcagtaggc tctgcacaac cactacacgc agaagagcct ctccctgtct ccgggtaaat
1801 gagtgccacg gccggcaagc ccccgtccc caggctctcg gggtcgcgtg aggatgcttg
1861 gcacgtaccc cgtgtacata cttcccaggc acccagcatg gaaataaagc acccagcgt
1921 gccctgggcc cctgcgagac tgtgatggtt cttccgtgg gtcaggccga gtctgaggcc
1981 tgagtggcat gagggaggca gagtgggtc

FIG. 1B

1 agctttctgg ggcaggccag gcctgacttt ggctgggggc agggaggggg ctaaggtgac
61 gcaggtggcg ccagccaggc gcacacccaa tgcccgtgag cccagacact ggacctgcc
121 tggacctcg tggatagaca agaaccgagg ggcctctgcg ccctggggcc agctctgtcc
181 cacaccgcag tcacatggcg ccattctctt tgcagcttcc accaagggcc catcgggtct
241 cccctggcg ccctgtctca ggagcacctc tgggggcaca gcggccctgg gctgcctggt
301 caaggactac tccccgaac cggtgacggt gtcgtggaac tcaggcgcgc tgaccagcgg
361 cgtgcacacc tccccggctg tctacagtc ctcaggactc tactccctca gcagcgtggt
421 gacctgtccc tccagcagct tgggcaccca gacctacacc tgcaactga atcacaagcc
481 cagcaacacc aaggtggaca agagagttgg tgagaggcca gcgcaggag ggagggtgtc
541 tgttggaagc caggctcagc cctctgcct ggacgatcc cggctgtgca gtccagccc
601 agggcagcaa ggcaggcccc gtctgactcc tccccggag cctctgcccg cccactcat
661 gctcagggag aggggtctct ggcttttcc accaggctcc gggcaggcac aggtggatg
721 cccctacccc aggccttca cacacagggg cagggtgtgc gctcagagct gccaaaagcc
781 atatccagga ggacctgcc cctgacctaa gcccaccca aaggccaaac tctctactca
841 ctcagctcag acacttctc tttcccaga tctagtaac tccaatctt ctctctcag
901 agctcaaaac cccacttgt gacacaactc acacatgcc acggtgcca ggtaagccag
961 cccaggactc gccctccagc tcaaggcggg acaagagccc tagagtggcc tgagtccagg
1021 gacaggcccc agcagggtgc tgacgatcc acctccatcc cagatccccg taactccaa
1081 tcttctctc gcagagccca aatctgtga cacactccc cgtgcccac ggtgccagg
1141 taagccagcc caggcctcac cctccagctc aaggcaggac aagagcccta gagtggcctg
1201 agtccaggga caggccccag cagggtgtg acggtccac ctccatcca gatccccgta
1261 actcccaatc ttctctctc agagcccaaa tctgtgaca cacctcccc atgcccacgg
1321 tgcccaggta agccagccca ggcctgccc tccagctcaa ggcgggaca gagccctaga
1381 gtggcctgag tccagggaca ggcccagca ggtgtgtgac gcatccact ccatccaga
1441 tccccgaac tccaatctt ctctctcag agcccaaatc ttgtgacaca cctccccgt
1501 gcccaggtg cccaggtgag ccagcccagg cctgcccctc cagctcaagg caggacaggt
1561 gccctagagt ggctgcatc cagggacagg tccagtcgg gtgtgacac atctgcctcc
1621 atcttctct cagcacctga actcctggga ggacctcag tcttctctt cccccaaaa
1681 cccaaggata ccttatgat tccccggacc cctgaggtca cgtgcgtggt ggtggacgtg
1741 agccacgaag accccgaggt ccagttcaag tggtagctgg acggcgtgga ggtgcataat
1801 gccaaagaca agccgcggga ggagcagtag aacagcacgt tccgtgtggt cagcgtctc
1861 accgtcctgc accaggactg gctgaacggc aaggagtaca agtgcaaggt ctccaacaaa
1921 gccctcccag ccccatcga gaaaaccatc tccaaaacca aaggtgggac ccgcggggta
1981 tgaggggcac atggacagag gccagcttga cccacctct gccctgggag tgaccgctgt
2041 gccaaactct gtccctacag gacagcccc agaaccacag gtgtacacc tgccccatc
2101 ccgggaggag atgaccaaga accaggctcag cctgacctgc ctggtcaaag gcttctacc
2161 cagcgacatc gccgtggagt gggagagcag cgggcagccg gagaacaact acaacaccac
2221 gcctccatg ctggactccg acggctcctt ctctctac agcaagctca ccgtggaca
2281 gagcaggtgg cagcagggga acatctctc atgctccgtg atgcatgagg ctctgcacaa
2341 ccgtttcacg cagaagagcc tctccctgtc tccgggtaaa tgagtgcgac agccggcaag
2401 cccccgtcc ccgggctctc ggggtcgcgc gaggatgctt ggcacgtacc ccgtgtacat
2461 acttcccggg caccagcat ggaaataaag caccagcgc tgccctgggc cctgtgaga
2521 ctgtgatggt ttttccacg ggtcaggccg agtctgaggc ctgagtga tgaggagggc
2581 agagcgggtc (SEQ ID NO:3)

FIG. 1C

1 agctttctgg ggcaggccgg gcctgacttt ggctgggggc agggaggggg ctaaggtgac
61 gcaggtggcg ccagccaggt gcacaccaa tgcccatgag ccagacact ggacctgca
121 tggaccatcg cggatagaca agaaccgagg ggctctgcg cctggggccc agctctgtcc
181 cacaccgagg tcacatggca ccacctctct tgcagcttc accaagggcc catccgtctt
241 cccctggcg cctgtctca ggagcacctc cgagagcaca gccgccctgg gctgctgtg
301 caaggactac ttcccgaac cggtagcggg gtcgtggaac tcaggcgccc tgaccagcgg
361 cgtgcacacc ttcccgctg tctacagtc ctacaggactc tactccctca gcagcgtggt
421 gaccgtgccc tcagcagct tgggcacgaa gacctacacc tgcaacgtag atcacaagcc
481 cagcaacacc aaggtggaca agagagtgg tgagaggcca gcacaggag ggagggtgtc
541 tgctggaagc caggtcagc cctctgct ggacgcacc cggctgtgca gcccagccc
601 agggcagcaa ggcatgcccc atctgtctcc tcaccggag gcctctgacc acccactca
661 tgctcaggga gagggtctt tggattttc caccaggctc ccggcaccac aggctggatg
721 cccctacccc aggcctgcg cctacagggc aggtgtgctg ctacagctg ccaagagcca
781 tatccgggag gacctgccc ctgacctaa cccacccaa agccaaact ctccactccc
841 tcagctcaga cacttctct cctccagat ctgagtaact ccaatcttct tctctgcaga
901 gtccaaatat ggtccccc atgccatcatg ccaggtgaa ccaaccagg cctcgccctc
961 cagctcaagg cgggacaggt gccctagagt agcctgcac caggacagg cccagccgg
1021 gtgtgacgc atccacctc atcttctct cagcacctga gttctgggg ggacctcag
1081 tcttctgtt cccccaaaa ccaaggaca ctctcatgat ctccggacc cctgaggtca
1141 cgtgcgtggt ggtggacgtg agccaggaag accccgaggt ccagttcaac tggtagctg
1201 atggcgtgga ggtgcataat gccaagaca agccgcggga ggagcagttc aacagcacgt
1261 accgtgtggt cagcgtctc accgtcctgc accaggactg gctgaacggc aaggagtaca
1321 agtgcaaggt ctccaacaaa ggctccctg cctccatcga gaaaaccatc tccaaagcca
1381 aaggtgggac ccacggggtg cgaggggccac acggacagag gccagctcgg cccacctct
1441 gccctgggag tgaccgtgt gccaacctct gtccctacag ggcagcccc agagccacag
1501 gtgtacacc tggcccatc ccaggaggag atgaccaaga accaggtcag cctgacctgc
1561 ctggtcaaag gcttctacc cagcgacatc gccgtggagt gggagagcaa tgggcagccg
1621 gagaacaact acaagaccac gcctcccgtg ctggactccg acggctcctt ctctctac
1681 agcaggctaa ccgtggaca gagcagggtg caggagggga atgtcttct atgtccgtg
1741 atgcatgagg ctctgcaca cactacaca cagaagagcc tctccctgtc tctgggtaaa
1801 tgagtgccag ggccggcaag ccccgcctc ccgggctctc ggggtcgcgc gaggatgctt
1861 ggcacgtacc ccgtctacat acttccagg caccagcat ggaataaag caccaccac
1921 tgccctgggc cctgtgaga ctgtgatgt tcttccacg ggtcaggccg agtctgaggc
1981 ctgagtgaca tgagggaggc agagcgggtc cactgtccc cacactgg

FIG. 1D

1 tcggggacat gggaagggtg caaaagtagc ggccttctag aaggtttga cctgtcctgt
61 cctgtccgac agtghtaatca catatacttt ttctgttagc caaaacgaca ccccatctg
121 tctatccact ggcccctgga tctgtgccc aaactaactc catggtgacc ctgggatgcc
181 tgggtcaaggg ctatttcct gagccagtga cagtgtacgt gaactctgga tccctgtcca
241 gcgggtgtgca caccttccca gctgtcctgg agtctgacct ctacactctg agcagctcag
301 tgactgtccc ctccagccct cggcccagcg agaccgtcac ctgcaacgtt gccacccgg
361 ccagcagcac caaggtggac aagaaaattg gtgagaggac atatagggag gaggggttca
421 ctagaagtga ggtcaagcc attagcctgc ctaaaccaac caggctggac agccaaccaa
481 ccaggaaatg gatctcagcc cagaagatca aaagttgttc ttctcccttc tggagatttc
541 tatgtccttt acaactcaat tggttaatat cctgggttgg agtcccacac atcttgacaa
601 acagagacaa atttgagtat caccagccaa aagtcatacc caaaaacagc ctggcatgac
661 cacacaccag actcaaactt accctacctt tatcctggtg gcttctcctc tccagacccc
721 agtaacacat agctttctct ccacagtgc cagggtattgt ggttgtaagc ctgcatatg
841 ggccaggtat tgacctattt ccacctttct tcttctcct tagtcccaga agtatcatct
901 gtcttctct tcccccaaa gccaaggat gtgtcacca ttactctgac tctaagggtc
961 acgtgtgttg tggtagacat cagcaaggat gatcccgagg tccagttcag ctggtttgta
1021 gatgatgtgg aggtgcacac agctcagacg caaccccgagg aggagcagtt caacagcact
1081 ttccgctcag tcagtgaact tcccatcatg caccaggact ggctcaatgg caaggagtgc
1141 aatgcaggg tcaacagtgc agctttccct gcccctcctg agaaaacct ctccaaaacc
1201 aaaggtgaga gctgcagtgt gtgacataga agctgcaata gtcagtccat agacagagct
1261 tggcataaca gaccctgcc ctgttcgtga cctctgtgct gaccaatctc ttaccacc
1321 cacaggcaga ccgaaggctc cacaggtgta caccattcca cctccaagg agcagatggc
1381 caaggataaa gtcagtctga cctgcatgat aacagacttc ttccctgaag acattactgt
1441 ggagtggcag tggaatggc agccagcgga gaactacaag aacactcagc ccatcatgaa
1501 cacgaatggc tcttacttcg tctacagcaa gctcaatgtg cagaagagca actgggaggc
1561 aggaaatact ttcactgct ctgtgttaca tgagggcctg cacaaccacc atactgagaa
1621 gagcctctcc cactctcctg gtaaatgatc ccagtgtcct tggagccctc tggctctaca
1681 ggactctgac acctacctcc accctccct gtataataa agcaccagc actgccttgg
1741 gaccctgcaa taactcctg gtgattctg agatgtagag tctagctagg tcatggaatg
1801 aggggtctcc atggttgag ccc

FIG. 2A

1 ggattcactg ctcccttcta ccagcacaga tatctgaaac acacataata cacaaattca
61 tacatagaaa atataattca tggaaaatta agatcaggat acaccatcaa gaggaggaag
121 gtgcttacc tgctgtctc aaagcaattg aagaagacac agtctttcga gcaaagctaa
181 gaccagagcc tctccaaata tctgaggcca cagataacag aaaagctcac acatcctcct
241 ctcttgacgc caaaacaaca gcccacatgg tctatccact ggcctctgtg tgtggagata
301 caactggctc ctcggtgact ctaggatgcc tggtaaggg ttatttcctt gagccagtga
361 ccttgacctg gaactctgga tccctgtcca gtggtgtgca caccttcca gctgtcctgc
421 agtctgacct ctacacctc agcagctcag tgactgtaac ctgagcacc tggcccagcc
481 agtccatcac ctgcaatgtg gccaccccg caagcagcac caaggtggac aagaaaattg
541 gtgaggaaaa caaggggagt agagggtcac aagtgattag tctaaggcct tagcctagct
601 agaccagcca agatcagcag ccataccaa aaatgggaac ttggcccaga agagaaggag
661 atactgactg ggattccctc ttgaaactt ctaactatga ccactacct tcaagtcct
721 gatcctctag gatagatgtc ctgtcattt ccaggatcat cctgacctaa gccataccc
781 agggacaaag tccctgggtt ggtgccttt ctcttcaaa cttagtaac ccccagcctt
841 ctctctgcag agcccagagg gccacaatc aagcctgtc ctccatgaa atgccagggt
901 aagtcactag accagagctc ccccgggag aatggtagt gctgtaaaca tccctgcact
961 agaggataag ccatgtacag atccatttcc atctctctc atcagcact aacctcttg
1021 gtggaccatc cgtcttcac tccctccaa agatcaagga tgtactcatg atctccctga
1081 gcccatagt cacatgtgtg gtggtggatg tgagcgagga tgaccagat gtccagatca
1141 gctggtttgt gaacaacgtg gaagtacaca cagctcagac acaaaccat agagaggatt
1201 acaacagtac tctccgggtg gtcagtgcct tcccatacca gcaccaggac tggatgagtg
1261 gcaaggagtt caaatgcaag gtcaacaaca aagacctcc agcgcccatc gagagaacca
1321 tctcaaaacc caaaggtgag agctgcagcc tgactgcatg ggggctggga tgggcataag
1381 gataaaggct tgtgtggaca gccttctgt tcagccatga ctttgtgta tgtttctacc
1441 ctacagggt cagtaagagc tccacaggta tatgtcttgc ctccaccaga agaagagatg
1501 actaagaaac aggtcactct gacctgcatg gtcacagact tcatgcctga agacatttac
1561 gtggagtgga ccaacaacgg gaaaacagag ctactataca agaactga accgtcctg
1621 gactctgatg gttcttactt catgtacagc aagctgagag tggaaaagaa gaactgggtg
1681 gaaagaaata gctactcctg ttcagtggct cagcagggtc tgcacaatca ccacacgact
1741 aagagcttct cccggactcc ggtataatga gtcagcacc cacaaaactc tcaggtccaa
1801 agagacaccc acactcatct ccatgcttcc cttgtataaa taaagacccc agcaatgcct
1861 gggaccatgt aaaactgtcc tggttacttt ccaaggtata gagcatagct cacaggctga
1921 tatttctggc cagggttga ggacagcctt gtctatagaa gagaatgagg ttttgcact
1981 gcaggactca gagctc

FIG. 2B

1 aacatggaaa aattaaggtc acagtgaag ctctggaggc ggaaattct gacactgctt
61 gcctcagatc aattgtagaa gacacggttc taagagcaaa gctaagaaca gaatctctcc
121 aaatatccga ggccactgat aagaaaaagc tcacacattc tcctctcttg cagccaaaac
181 aacaccccca tcagtctatc cactggcccc tgggtgtgga gatacaactg gttcctccgt
241 gactctggga tgccctgtca agggctactt cctgagtca gtgactgtga cttggaactc
301 tggatccctg tccagcagtg tgcacacctt cccagctctc ctgcagtctg gactctacac
361 tatgagcagc tcagtactg tccctccag cacttgcca agtcagaccg tcacctgcag
421 cgttctcac ccagccagca gcaccacggt ggacaaaaaa cttggtgaga ggacattcag
481 gggaggaggg attcaccaga gtgaggcaa aggtattagc ctatctaac cagccaggct
541 gggatccatc accaaggagg tgacctagc ccagggaaga gggagatact gtctctgcct
601 cctcctggg aacatctagc tatgaccacc tacactcaag gacatgttcc tctgggatag
661 gtgtgctgt cattccagg atcatcctgg aactaagccc ataccaggac aaactttcct
721 ctctctggtt tgggtcttct ctcttcaaa aaccagtaac atccagcctt ctctctgcag
781 agcccagcgg gccatttca acaatcaacc cctgtcctcc atgcaaggag tgtcacaat
841 gccaggtaa gtcactacca gagctccact cccaggagaa tggtaatgac tgtaaaaac
901 cctgtaatgg aggataagcc atgtacaaat ccaattccat ctctctcat cagctcctaa
961 cctcagggtt ggaccatccg tcttcatctt ccttccaaat atcaaggatg tactcatgat
1021 ctccctgaca cccaagggtc cgtgtgtggt ggtggatgtg agcgaggatg acccagacgt
1081 ccagatcagc tggtttga acaacgtgga agtacacaca gtcagacac aaaccatag
1141 agaggattac aacagtacta tccgggtggt cagcaccctc cccatccagc accaggactg
1201 gatgagtggc aaggagtcca aatgcaagg caacaacaaa gacctccat caccatcga
1261 gagaaccatc taaaaatta aagggtggac ctgcaggaca actgcatggg ggctgggatg
1321 ggcataagaa taaatgtcta tgtggacagc cttccacttc agccatgacc tctatgtatg
1381 ttctaaacc cacagggtca gtcagagctc cacaagtata catcttgccg ccaccagcag
1441 agcagttgtc caggaaagat gtcagtctca cttgcctggt cgtgggcttc aaccctggag
1501 acatcagtggt ggagtggacc agcaatgggc atacagagga gaactacaag gacaccgcac
1561 cagtctaga ctctgacggt tcttacttca tatatagcaa gctcaatatg aaaacaagca
1621 agtgggagaa aacagattcc ttctcatgca acgtgagaca cgagggtctg aaaaattact
1681 acctgaagaa gaccatctcc cggctccgg gtaaatgagc tcagcaccca caaagctctc
1741 aggtcctaag agacactggc acccatatcc atgcatccct tgtataaata aagcatccag
1801 caaagcctgg taccatgtaa aactgtcctg gtctttcca aggtatagag catagctcac
1861 gggctgatag gtctggccag ggccggagaa cagccttgtc tataggaaga gaatgaggt
1921 tctgcctgca t

FIG. 2C

6001 agccaaaaca gtggataggg tcaaaactgc tgtgtttctt ctacatgggtg tctagccctc
6061 agccaggcac tctcccaaga tgggctgcat cacctgcac ctttagatg cagatgagcc
6121 tttaggccag ccctgtctat acagtctctt gaaaaacgtg cttctctgag agaagaggaa
6181 gtttccatgt gtctctaagg ctgagaagge ctcaaaggga actctttggg aaggactgcc
6241 aaagtgact ctaatggcct tcaccaacc agtgcctgc cgtgttttt cacttcttg
6301 aatgtaagat ggggctactt gtgaagacct ctttctctgc catccatgca acatgatgaa
6361 atctcagatg caagatccca tcagacactg ctgatccac agcttgaaga ccagggcact
6421 gttacctcag tggatgacct agagagacac tcagagggcc tgagcttcat agcctctgtt
6481 gggccctagc aggtacagga gccactgagc aaaggctggg caaccagaa aagcccgtag
6541 gagaagagca gaccccggtc agtcccaaag gtgtccagg agcaacatgg ggagagagag
6601 aaagtgatgt taaaagcca ggcaggaatg ggcctttcct gcagcctggc atccggactg
6661 ggccactgac agcagagaat accagttaga gtagatatcc agagagagag cccaggcctg
6721 gggcactgag gctgactgac tgctggcatt ggagaggcac tgtggctgtt tctccattc
6781 atatcaaggc aaatgcaggc cagagtcccc agaaaagaca aaggagaaat ccaggaagaa
6841 aagttgggtg cccggctctt aggggactct aagagatgtt cactgacttg aagcttagct
6901 ttgccagctt tttctttaa tttctggaa cccagctatg tctctaattt cctcttcat
6961 ctttttaat aataaaacat ccttcaacat tgtaaaacac actgtgttct tctctgtag
7021 accaaccctc actctccttg taaaaactc acacacacac acatgcacac acatgcatgc
7081 acatgacac actcacacac acgtacatgc atgcacacaa gcacattcac acacacatat
7141 gcacacacat gcatgcacac acacactcac acacaaatac atgcatcaca caccacacac
7201 cacacatata ccatgctaac attcccaaag caccactcac ccagcattct gctctcagtc
7261 tcaagtcccc tccaactatc atcccataat gaggcaggga tggctgtga gtgggaatca
7321 ctctggattt caggccatgg tgaccacgag ggagaggcag tcaggaagca gaagatagct
7381 aacatccatg tgagaggaac agatgctatg actcaagagt ggattatagc ctgatgactt
7441 gcagtagggc agatagcccc gggaacgtgg cctccacttg cactggtaca gggcgaaagc
7501 ctttctaagg cagatcaggg gactagggaa agaaagatgg agctcagaca tgtcagggtg
7561 ctctgctctc actgccttcc caaggtggag tttgcagca ggaacatcct acaacacggg
7621 ttcaactgtt agaaactgtt tgggtctaca taggccttgg cagaggattg tgttgagac
7681 ttggggctgg actagtcaaa agactcatgc agctg

FIG. 2D

5'
gctctagaGTGCGACGGCCGGCAAGCCCCGCTCCCCGGGCTCTCGCGGTTCGC
ACGAGGATGCTTGGCACGTACCCCTGTACATACTTCCCGGGCGCCCAGC
ATGGAAATAAAGCACCCAGCGCTGCCCTGGGCCCCTGCGAGACTGTGATG
GTTCTTTCCACGGGTCAggcggccgctaaactat 3'

FIG. 3

5'
gctctagaGTGCGACGGCCGGCAAGCCCCGCTCCCCGGGCTCTCGCGGTTCGC
ACGAGGATGCTTGGCACGTACCCCCTGTACATACTTCCCGGGCGCCCAGC
ATGGAAATAAAGCACCCAGCGCTGCCCTGGGCCCCTGCGAGACTGTGATG
GTTCTTTCCACGGGTCAGGCCGAGTCTGAGGCCTGAGTGGCATGAGGGAG
GCAGAGCGGGTCCCAGTGTCCCCACACTGGCCCAGGCTGTGCAGGTGTGC
CTGGGCCCCCTAGGGTGGGGCTCAGCCAGGGGCTGCCCTCGGCAGGGTGG
GGGATTTGCCAGCGTGGCCCTCCCTCCAgcggccgctaaactat 3'

FIG. 4

ccggaattcctcaccATGGCAGCCGGAACAGCAGTTGGAGCCTGGGTGCTGGTCCT
CAGTCTGTGGGGGGCAGTAGTAGGTGCTCAAAACATCACAGCCCGGATTG
GCGAGCCACTGGTGCTGAAGTGTAAGGGGGCCCCCAAGAAACCACCCCA
GCGGCTGGAATGGAACTGAACACAGGCCGGACAGAAGCTTGGAAGGTC
CTGTCTCCCCAGGGAGGAGGCCCTGGGACAGTGTGGCTCGTGTCTTCCC
AACGGCTCCCTCTTCCTTCGGCTGTCTGGGATCCAGGATGAGGGGATTTTC
CGGTGCCAGGCAATGAACAGGAATGGAAAGGAGACCAAGTCCAACCTACC
GAGTCCGTGTCTACCAGATTCCTGGGAAGCCAGAAATTGTAGATTCTGCCT
CTGAACTCACGGCTGGTGTTCCCAATAAGGTGGGGACATGTGTGTCAGAG
GGAAGCTACCCTGCAGGGACTCTTAGCTGGCACTTGGATGGGAAGCCCCT
GGTGCCTAATGAGAAGGGAGTATCTGTGAAGGAACAGACCAGGAGACAC
CCTGAGACAGGGCTCTTCACACTGCAGTCGGAGCTAATGGTGACCCACGC
CCGGGGAGGAGATCCCCGTCCCACCTTCTCCTGTAGCTTCAGCCCAGGCCT
TCCCCGACACCGGGCCTTGCGCACAGCCCCATCCAGCCCCGTGTCTGGG
AGCCTGTGCCTCTGGAGGAGGTCCAATTGGTGGTGGAGCCAGAAGGTGGA
GCAGTAGCTCCTGGTGGAAACCGTAACCCTGACCTGTGAAGTCCCTGCCCA
GCCCTCTCCTCAAATCCACTGGATGAAGGATGGTGTGCCCTTGCCCCCTTC
CCCCAGCCCTGTGCTGATCCTCCCTGAGATAGGGCCTCAGGACCAGGGAA
CCTACAGCTGTGTGGCCACCCATTCCAGCGACAAAACCTCACACATGCCCA
CCGTGCCCAGCACCTGAACTCGCGGGGGCACCGTCAGTCTTCCTCTTCCCC
CCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTACATG
CGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGT
ACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGA
GCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACC
AGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGC
CCTCCCAGCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCC
GAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAG
AACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACAT
CGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACCTACAAGACC
ACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTC
ACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGT
GATGCATGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGT
CTCCGGGTAAAtaatctagagca

FIG. 5A

AVVGAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGG
GPWDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIP
GKPEIVDSASELTAGVPNKVGTCVSEGSYPAGTLSWHLDGKPLVPNEKGVSV
KEQTRRHPETGLFTLQSELMVTPARGGDPRPTFSCSFSPGLPRHRALRTAPIQP
RVWEPVPLEEVQLVVEPEGGAVAPGGTVTLTCEVPAQPSPQIHWMKDGVPLP
LPPSPVLILPEIGPQDQGTYSVATH-
SSDKTHTCPPCPAPELAGAPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPE
V
KFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKV
SNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDI
AVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCFSVM
HEALHNHYTQKSLSLSPGK

FIG. 5B

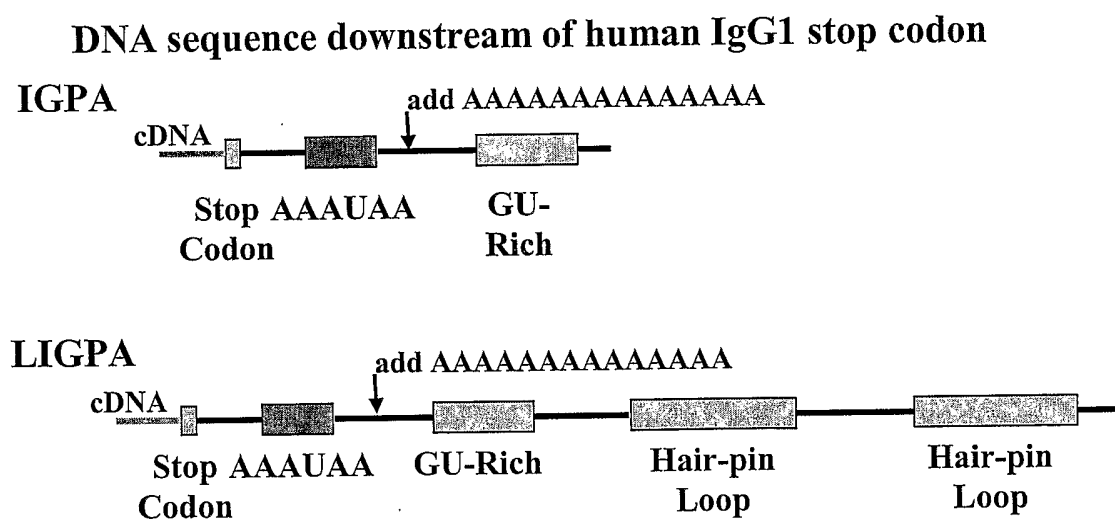


FIG. 6

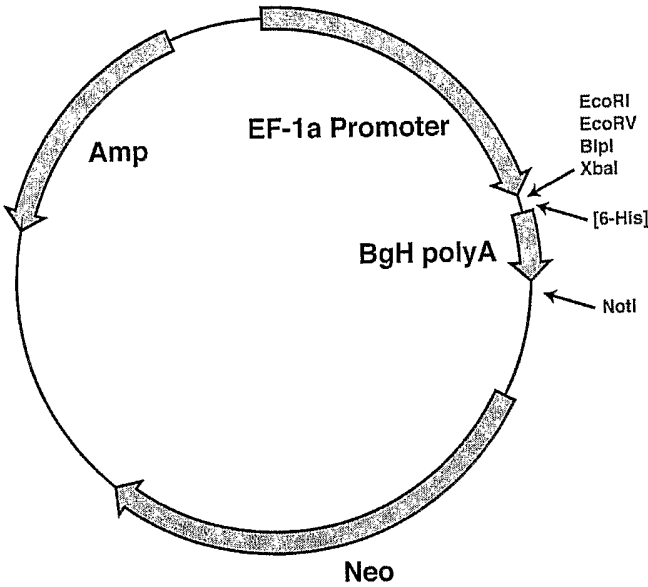


FIG. 7A

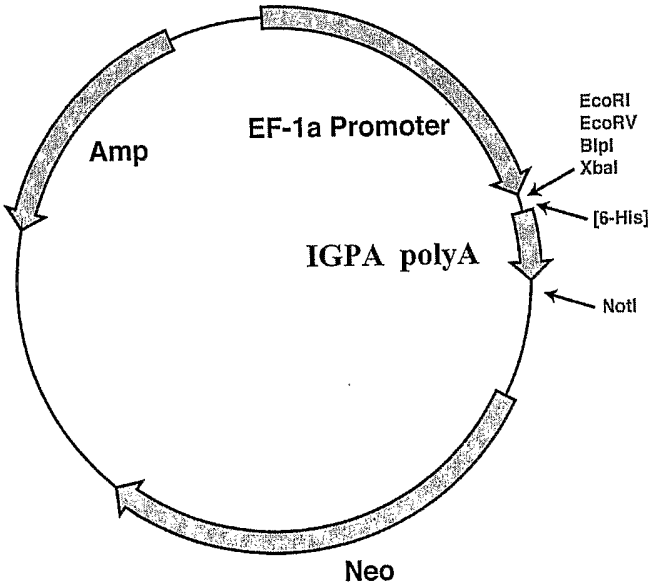


FIG. 7B

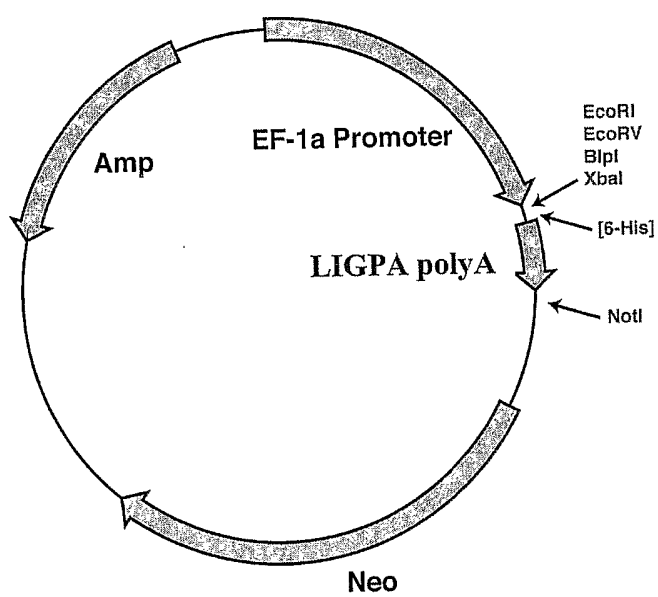


FIG. 7C

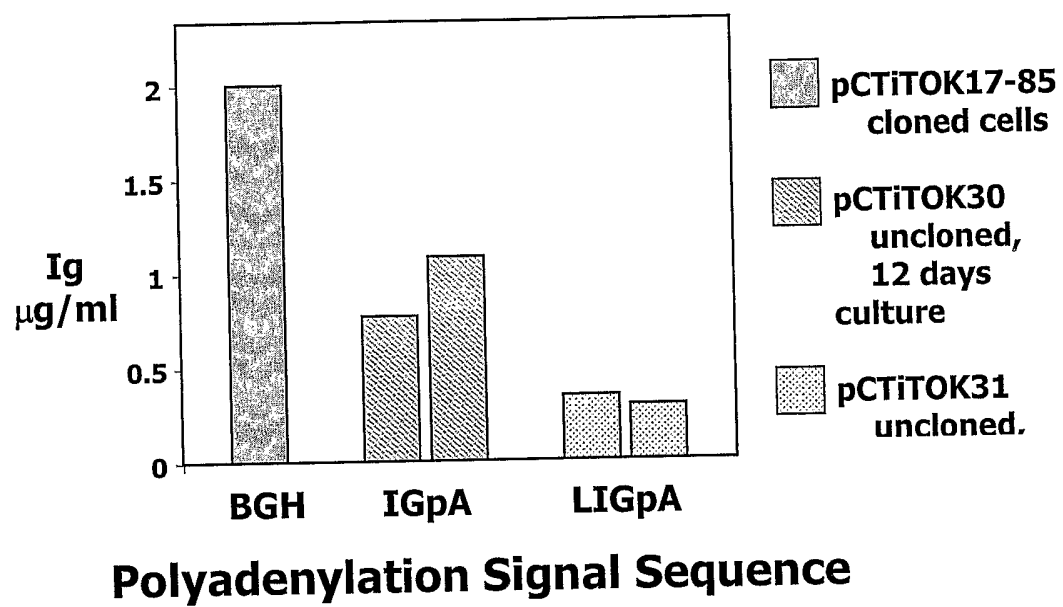


FIG. 8

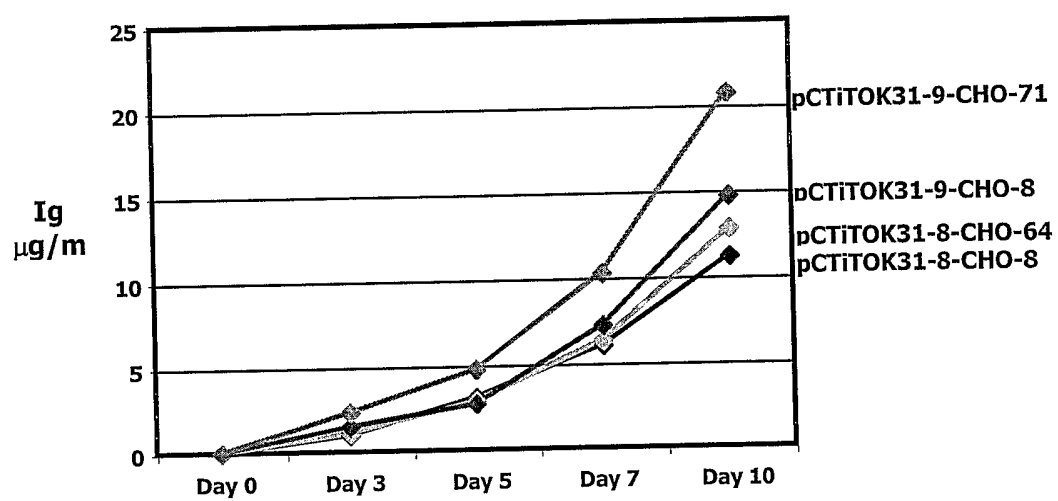


FIG. 9