US 20020061567A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2002/0061567 A1

Tang et al. (43) Pub. Date: May 23, 2002

(54) NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(76) Inventors: **Y. Tom Tang**, San Jose, CA (US); **Chenghua Liu**, San Jose, CA (US); **Ping Zhou**, Cupertino, CA (US); **Vinod Asundi**, Foster City, CA (US); **Feiyan Ren**, Cupertino, CA (US); **Jie Zhang**, Campbell, CA (US); **Qing A. Zhao**, San Jose, CA (US); **Aidong J. Xue**, Sunnyvale, CA (US); **Ryle Goodrich**, San Jose, CA (US); **Tom Wehrman**, Stanford, CA (US); **Radoje T. Drmanac**, Palo Alto, CA (US)

Correspondence Address:
Ivor R. Elrifi
Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C
One Financial Center
Boston, MA 02111 (US)

(21) Appl. No.: **09/728,711**

(22) Filed: **Nov. 30, 2000**

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/649,167, filed on Aug. 23, 2000, which is a continuation-in-part of application No. 09/540,217, filed on Mar. 31, 2000.

Publication Classification

(51) **Int. Cl.**[7] **C12N 9/00**; C07H 21/04; C12P 21/02; C12N 5/06

(52) **U.S. Cl.** **435/183**; 435/69.1; 435/325; 435/320.1; 536/23.2

(57) **ABSTRACT**

The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part application of U.S. application Ser. No. 09/649,167, filed Aug. 23, 2000, Attorney Docket No. 790CIP, which in turn is a continuation-in-part application of U.S. application Ser. No. 09/540,217, filed Mar. 31, 2000, Attorney Docket No. 790, both which are incorporated herein by reference in their entirety.

BACKGROUND OF THE INVENTION

[0002] 1. Technical Field

[0003] The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

[0004] 2. Background

[0005] Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

[0006] Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

SUMMARY OF THE INVENTION

[0007] The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

[0008] The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

[0009] The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-10 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

[0010] The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-10 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-10. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-10 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

[0011] The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-10. The sequence information can be a segment of any one of SEQ ID NO: 1-10 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-10.

[0012] A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

[0013] This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

[0014] In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-10 or novel segments or parts

of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-10 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

[0015] The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-10; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-10; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-10. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-10; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

[0016] The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-10; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

[0017] The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

[0018] The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

[0019] The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

[0020] Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, e.g., in situ hybridization.

[0021] In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

[0022] The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

[0023] Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

[0024] In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

[0025] The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

[0026] The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for

evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

[0027] The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound the binds to a polypeptide of the invention is identified.

[0028] The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

[0029] The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

DETAILED DESCRIPTION OF THE INVENTION

[0030] DEFINITIONS

[0031] It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

[0032] The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

[0033] The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

[0034] The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

[0035] The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

[0036] The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

[0037] As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

[0038] The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonculeotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

[0039] The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-10.

[0040] Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P. S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F. M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York N.Y., both of which are incorporated herein by reference in their entirety.

[0041] The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-10. The sequence information can be a segment of any one of SEQ ID NOs: 1-10 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-10. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

[0042] Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3×25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

[0043] The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

[0044] The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

[0045] The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

[0046] The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

[0047] The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

[0048] The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

[0049] The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

[0050] The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

[0051] The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

[0052] Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

[0053] Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, i.e., conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

[0054] Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

[0055] The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, e.g., polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

[0056] The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

[0057] The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (e.g., microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli,* will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

[0058] The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

[0059] The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic

gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

[0060] The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (e.g., soluble proteins) or partially (e.g., receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P. A. and Young, P. R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W. P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

[0061] Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

[0062] The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65° C., and washing in 0.1×SSC/0.1% SDS at 68° C.), and moderately stringent conditions (i.e., washing in 0.2×SSC/0.1% SDS at 42° C.). Other exemplary hybridization conditions are described herein in the examples.

[0063] In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6×SSC/0.05% sodium pyrophosphate at 37° C. (for 14-base oligonucleotides), 48° C. (for 17-base oligos), 55° C. (for 20-base oligonucleotides), and 60° C. (for 23-base oligonucleotides).

[0064] As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (i.e., the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, e.g., mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, e.g., mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity. Substantially equivalent nucleotide sequences of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (e.g., via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

[0065] The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

[0066] The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

[0067] As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UME to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

[0068] Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

[0069] NUCLEIC ACIDS OF THE INVENTION

[0070] Nucleotide sequences of the invention are set forth in the Sequence Listing.

[0071] The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-10; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 1-10; and a polynucleotide comprising the nucleotide sequence encod-

ing the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-10. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1-10; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 1-10. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

[0072] The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

[0073] The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-10 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-10 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-10 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

[0074] The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

[0075] The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, e.g., at least about 65%, at least about 70%, at least about 75%, at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above.

[0076] Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-10, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

[0077] The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-10, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1-10 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

[0078] The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1-10, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S. F. J Mol. Evol. 36 290-300 (1993) and Altschul S. F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

[0079] Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

[0080] The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

[0081] The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations

will typically be modified in series, e.g., by substituting first with conservative choices (e.g., hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (e.g., hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

[0082] In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

[0083] A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology,* Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

[0084] Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

[0085] The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

[0086] In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-10, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

[0087] A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

[0088] The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1-10 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1-10 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

[0089] The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General meth-

ods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

[0090] Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lac, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

[0091] As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, Wis., USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

[0092] Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

[0093] HOSTS

[0094] The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

[0095] Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

[0096] The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

[0097] Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells,

Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y. (1989), the disclosure of which is hereby incorporated by reference.

[**0098**] Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from in vitro culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

[**0099**] Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae, Schizosaccharomyces pombe,* Kluyveromyces strains, Candida, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium,* or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

[**0100**] In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

[**0101**] The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

[**0102**] The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Pat. No. 5,272,071 to Chappel; U.S. Pat. No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

[**0103**] POLYPEPTIDES OF THE INVENTION

[**0104**] The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 1-10 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1-10 or the corresponding full

length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NOs: 1-10 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 1-10 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 1-10 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, typically at least about 95%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 1-10.

[0105] Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

[0106] The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

[0107] Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

[0108] The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

[0109] A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

[0110] The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

[0111] The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

[0112] In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, e.g., Scopes, *Protein Purification: Principles and Practice,* Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: A Laboratory Manual; Ausubel et al., *Current Protocols in Molecular Biology.* Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

[0113] The purified polypeptides can be used in in vitro binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in in vivo tissue

culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

[0114] In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 1-10.

[0115] The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

[0116] The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

[0117] Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

[0118] The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, e.g., Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

[0119] The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (i.e., from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

[0120] Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

[0121] Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

[0122] The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

[0123] DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

[0124] Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in

computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, Wis.), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S. F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S. F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, Md. 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

[0125] GENE THERAPY

[0126] Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected ex vivo, in situ, or in vivo by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or ex vivo by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

[0127] Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

[0128] The present invention still further provides cells genetically engineered in vivo to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

[0129] Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

[0130] In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

[0131] The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the

use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

[0132] The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Pat. No. 5,272,071 to Chappel; U.S. Pat. No.5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

[0133] TRANSGENIC ANIMALS

[0134] In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Pat. No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Pat. No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

[0135] Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

[0136] The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the in vivo activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

[0137] In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Pat. No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Pat. No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

[0138] Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

[0139] USES AND BIOLOGICAL ACTIVITY

[0140] The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

[0141] The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

[0142] RESEARCH USES AND UTILITIES

[0143] The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

[0144] The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

[0145] Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

[0146] Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

[0147] NUTRITIONAL USES

[0148] Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

[0149] CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

[0150] A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

[0151] Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. B. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

[0152] Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

[0153] Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin

6—Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11—Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9—Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

[0154] Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

[0155] STEM CELL GROWTH FACTOR ACTIVITY

[0156] A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells in vivo or ex vivo is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of biopharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

[0157] It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CFS, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

[0158] Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Pat. No. 5,690,926).

[0159] Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

[0160] Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

[0161] Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

[0162] In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

[0163] HEMATOPOIESIS REGULATING ACTIVITY

[0164] A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelosuppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

[0165] Therapeutic compositions of the invention can be used in the following:

[0166] Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

[0167] Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

[0168] Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

[0169] TISSUE GROWTH ACTIVITY

[0170] A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

[0171] A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

[0172] A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

[0173] Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in

cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

[0174] The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

[0175] Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

[0176] Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

[0177] A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

[0178] A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

[0179] Therapeutic compositions of the invention can be used in the following:

[0180] Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

[0181] Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dennatol 71:382-84 (1978).

[0182] IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

[0183] A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

[0184] Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

[0185] Using the proteins of the invention it may also be possible to modulate immune responses, in a number of

ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

[0186] Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

[0187] The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

[0188] Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

[0189] Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

[0190] Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

[0191] A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MEC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

[0192] The activity of a protein of the invention may, among other means, be measured by the following methods:

[0193] Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in:

Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

[0194] Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

[0195] Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

[0196] Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

[0197] Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

[0198] Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

[0199] ACTIVIN/INHIBIN ACTIVITY

[0200] A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

[0201] The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

[0202] Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

[0203] CHEMOTACTIC/CHEMOKINETIC ACTIVITY

[0204] A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

[0205] A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

[0206] Therapeutic compositions of the invention can be used in the following:

[0207] Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of

assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

[0208] HEMOSTATIC AND THROMBOLYTIC ACTIVITY

[0209] A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

[0210] Therapeutic compositions of the invention can be used in the following:

[0211] Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

[0212] CANCER DIAGNOSIS AND THERAPY

[0213] Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

[0214] Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

[0215] Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

[0216] The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

[0217] In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals

with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

[0218] In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These in vitro models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, N.Y. Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

[0219] RECEPTOR/LIGAND ACTIVITY

[0220] A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

[0221] The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

[0222] Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

[0223] By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

[0224] Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, calorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of calorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

[0225] DRUG SCREENING

[0226] This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

[0227] Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

[0228] Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

[0229] The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

[0230] Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-

707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol,* 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol,* 1(1):114-19 (1997Dorner et al., *Bioorg Med Chem,* 4(5):709-15 (1996) (alkylated dipeptides).

[0231] Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in in vivo tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

[0232] The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

[0233] ASSAY FOR RECEPTOR ACTIVITY

[0234] The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (i.e., increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

[0235] The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

[0236] ANTI-INFLAMMATORY ACTIVITY

[0237] Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

[0238] LEUKEMIAS

[0239] Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J. B. Lippincott Co., Pa.).

[0240] NERVOUS SYSTEM DISORDERS

[0241] Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be

treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

[0242] (i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

[0243] (ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

[0244] (iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

[0245] (iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

[0246] (v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

[0247] (vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

[0248] (vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

[0249] (viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

[0250] Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

[0251] (i) increased survival time of neurons in culture;

[0252] (ii) increased sprouting of neurons in culture or in vivo;

[0253] (iii) increased production of a neuron-associated molecule in culture or in vivo, e.g., choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

[0254] (iv) decreased symptoms of neuron dysfunction in vivo.

[0255] Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, etc., depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, e.g., weakness, motor neuron conduction velocity, or functional disability.

[0256] In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

[0257] OTHER ACTIVITIES

[0258] A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act

as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

[0259] IDENTIFICATION OF POLYMORPHISMS

[0260] The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

[0261] Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

[0262] Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

[0263] ARTHRITIS AND INFLAMMATION

[0264] The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

[0265] The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

[0266] THERAPEUTIC METHODS

[0267] The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

EXAMPLE

[0268] One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 μg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

[0269] PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

[0270] A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the

effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, EL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

[0271] The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

[0272] As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, Pa., latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

[0273] In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

[0274] ROUTES OF ADMINISTRATION

[0275] Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

[0276] Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

[0277] The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

[0278] COMPOSITIONS/FORMULATIONS

[0279] Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, e.g., by means of conventional

mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

[0280] When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

[0281] For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

[0282] Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

[0283] For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

[0284] Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in

powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

[0285] The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

[0286] A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

[0287] The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

[0288] The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

[0289] The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Pat. Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

[0290] The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or

injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

[0291] The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

[0292] A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

[0293] The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, e.g., amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (e.g., bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

[0294] Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

[0295] EFFECTIVE DOSAGE

[0296] Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (i.e., the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

[0297] A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be deter-

mined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from in vitro data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

[0298] Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

[0299] An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

[0300] The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

[0301] PACKAGING

[0302] The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

[0303] ANTIBODIES

[0304] Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies include monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, and complementary determining region (CDR)-grafted antibodies, including compounds which include CDR and/or antigen-binding sequences, which specifically recognize a polypeptide of the invention. Preferred antibodies of the invention are human antibodies which are produced and identified according to methods described in WO93/11236, published Jun. 20, 1993, which is incorporated herein by reference in its entirety. Antibody fragments, including Fab, Fab', $F(ab')_2$, and F_v, are also provided by the invention. The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (i.e., able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, N.Y. (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins. Antibodies of the invention can be produced using any method well known and routinely practiced in the art.

[0305] Non-human antibodies may be humanized by any methods known in the art. In one method, the non-human CDRs are inserted into a human antibody or consensus antibody framework sequence. Further changes can then be introduced into the antibody framework to modulate affinity or immunogenicity.

[0306] Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

[0307] Polypeptides of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the carboxyl terminus, and are conjugated to a hapten such as

keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R. P. Merrifield, J. Amer. Chem. Soc. 85, 2149-2154 (1963); J. L. Krstenansky, et al., FEBS Lett. 211, 10 (1987).

[0308] Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., J. Inmunol. 35:1-21 (1990); Kohler and Milstein, Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today 4:72 (1983); Cole et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985), pp. 77-96).

[0309] Any animal (mouse, rabbit, etc.) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

[0310] For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, Western blot analysis, or radioimmunoassay (Lutz et al., Exp. Cell Research. 175:109-124 (1988)). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

[0311] For polyclonal antibodies, antibody-containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The present invention further provides the above-described antibodies in delectably labeled form. Antibodies can be delectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example, see (Sternberger, L. A. et al., J. Histochem. Cytochem. 18:315 (1970); Bayer, E. A. et al., Meth. Enzym. 62:308 (1979); Engval, E. et al., Immunol. 109:129 (1972); Goding, J. W. J. Immunol. Meth. 13:215 (1976)).

[0312] The labeled antibodies of the present invention can be used for in vitro, in vivo, and in situ assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for in vitro, in vivo, and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

[0313] COMPUTER READABLE SEQUENCES

[0314] In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

[0315] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be repre-

sented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[0316] By providing any of the nucleotide sequences SEQ ID NOs: 1-10 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NOs: 1-10 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

[0317] As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

[0318] As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[0319] As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

[0320] TRIPLE HELIX FORMATION

[0321] In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix—see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense—Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, Fla. (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

[0322] DIAGNOSTIC ASSAYS AND KITS

[0323] The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

[0324] In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that

if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

[0325] In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

[0326] In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

[0327] Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, Fla. Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

[0328] In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

[0329] In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

[0330] MEDICAL IMAGING

[0331] The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. No. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide in vivo at the target site.

[0332] SCREENING ASSAYS

[0333] Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NOs: 1-10, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

[0334] (a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

[0335] (b) determining whether the agent binds to said protein or said nucleic acid.

[0336] In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

[0337] Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

[0338] Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

[0339] Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

[0340] The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

[0341] For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

[0342] In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

[0343] Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix—see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense—Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, Fla. (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

[0344] Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

[0345] USE OF NUCLEIC ACIDS AS PROBES

[0346] Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1-10. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NOs: 1-10 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

[0347] Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in U.S. Pat. Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

[0348] Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York N.Y.

[0349] Fluorescent in situ hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265: 1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject

invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

[0350] PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

[0351] Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

[0352] Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata et al., 1985; Dahlen et al., 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller et al., 1988; 1989); all references being specifically incorporated herein.

[0353] Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude et al. (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, Calif.).

[0354] Nunc Laboratories (Naperville, Ill.) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen et al., (1991) Anal. Biochem. 198(1) 138-42).

[0355] The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). in this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

[0356] More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95° C. and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm$_7$), is then added to a final concentration of 10 mM 1-MeIm$_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

[0357] Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm$_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50° C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50° C.).

[0358] It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

[0359] An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor et al. (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness et al. (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

[0360] To link an oligonucleotide to a nylon support, as described by Van Ness et al. (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

[0361] One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease et al., (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected N-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

[0362] PREPARATION OF NUCLEIC ACID FRAGMENTS

[0363] The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook et al. (1989) describes

three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

[0364] DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multi-well plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

[0365] The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook et al. (1989), shearing by ultrasound and NaOH treatment.

[0366] Low pressure shearing is also appropriate, as described by Schriefer et al. (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

[0367] One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, CviJI, described by Fitzgerald et al. (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

[0368] The restriction endonuclease CviJI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (CviJI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald et al. (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a CviJI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that CviJI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

[0369] As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

[0370] Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90° C. The solution is then cooled quickly to 2° C. to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

[0371] PREPARATION OF DNA ARRAYS

[0372] Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in $1mm^2$, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8×12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

[0373] Another approach is to use membranes or plates (available from NUNC, Naperville, Ill.) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

[0374] The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

[0375] All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5.0	EXAMPLES
5.1	EXAMPLE 1

[0376] Novel Nucleic Acid Sequences Obtained From Various Libraries

[0377] A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived

from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7 mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

[0378] In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

Example 2

[0379] Novel Nucleic Acids

[0380] The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

[0381] Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 120, gb pri 120, UniGene version 120, Genpept release 120). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1-10.

[0382] Table 1 shows the various tissue sources of SEQ ID NO: 1-10.

[0383] The homology for SEQ ID NO: 1-10 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 120 and the amino acid version of Geneseq released on Oct. 26, 2000, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1-10 from Genpept. The homologues with identifiable functions for SEQ ID NO: 1-10 are shown in Table 2 below.

[0384] Using eMatrix software package (Stanford University, Stanford, Calif.) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

[0385] Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

[0386] The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et as reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

TABLE 1

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	ABD003	5-6
adult brain	Clontech	ABR008	5-6
cultured preadipocytes	Strategene	ADP001	5-6
adult kidney	GIBCO	AKD001	2
adult liver	Invitrogen	ALV002	5-6
adult ovary	Invitrogen	AOV001	5-6 10
adult spleen	GIBCO	ASP001	5-6
bone marrow	Clontech	BMD002	2 5-6
adult cervix	BioChain	CVX001	8
fetal brain	Clontech	FBR006	2-3
fetal brain	Invitrogen	FBT002	3 5-7
fetal kidney	Clontech	FKD002	5-6
fetal liver-spleen	Columbia University	FLS001	5-6
fetal liver-spleen	Columbia University	FLS002	2 5-6
fetal liver-spleen	Columbia University	FLS003	5-6
fetal muscle	Invitrogen	FMS002	5-6
fetal skin	Invitrogen	FSK001	10
umbilical cord	BioChain	FUC001	3
infant brain	Columbia University	IB2002	8
infant brain	Columbia University	IB2003	4
infant brain	Columbia University	IBM002	2
lung tumor	Invitrogen	LGT002	5-6 8
lymphocytes	ATCC	LPC001	5-6
mammary gland	Invitrogen	MMG001	1 4-6
induced neuron cells	Strategene	NTD001	9
rectum	Invitrogen	REC001	8
salivary gland	Clontech	SAL001	10
skeletal muscle	Clontech	SKM001	1
thyroid gland	Clontech	THR001	2
trachea	Clontech	TRC001	2

[0387]

TABLE 2

SEQ ID NO:	CORRESPONDING SEQ ID NO. IN U.S.S.N 09/649,167	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	6138	AL136365	*Homo sapiens* bA143M15.2 (novel DM DNA binding domain containing protein)	1715	100
2	6221	AE003742	*Drosophila melanogaster* CG13827 gene product	189	28
3	8371				
4	11110 11111 11112	AF231024	*Homo sapiens* protocadherin Flamingo 2	15953	100
5	15219 15220 15221	AC016795	*Arabidopsis thaliana* putative ubiquitin carboxyl-terminal hydrolase	449	37
6	15219 15220 15221	AC016795	*Arabidopsis thaliana* putative ubiquitin carboxyl-terminal hydrolase	449	37
7	16198 16199 16200 16201	D29822	*Callithrix jacchus* Cytochrome P-450	267	70
8	16629 16630	Z29332_aa1	Human receptor tyrosine kinase coding region.	2189	99
9	21383	Y36307	Human secreted protein encoded by gene 84	1053	100
10	23036	AF242769	*Homo sapiens* mesenchymal stem cell protein DSC54	1521	100

[0388]

TABLE 3

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
1	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803I 15.63 8.226e-07 210-246
2	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519C 9.73 7.202e-06 130-145
3	DM00522	499 kw TRYPSIN KINASE KUNITZ PANCREATIC.	DM00522B 9.43 7.923e-06 50-64
4	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517E 10.98 9.270e-08 88-105
5	PF00602	Influenza RNA-dependant RNA polymerase subunit PB1.	PF00602F 11.46 8.275e-07 1118-1173
6	PF00602	Influenza RNA-dependant RNA polymerase subunit PB1.	PF00602F 11.46 8.275e-07 1140-1195
7	BL01095	Chitinases family 18 proteins.	BL01095C 10.76 1.000e-05 82-94
9	PR00669	INHIBIN ALPHA CHAIN SIGNATURE	PR00669B 8.27 9.910e-06 164-181

TABLE 3-continued

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
10	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513D 11.06 2.991e-06 54-72

*Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence.

[0389]

TABLE 4

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
1	DM-domain	DM DNA binding domain	4.2e-25	96.8
4	7tm_2	7 transmembrane receptor (Secretin family)	6.4e-90	312.2
5	UCH-2	Ubiquitin carboxyl-terminal hydrolase family 2	0.0003	17.5
6	UCH-2	Ubiquitin carboxyl-terminal hydrolase family 2	0.0003	17.5

TABLE 4-continued

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
7	p450	Cytochrome P450	0.0014	15.4
8	pkinase	Eukaryotic protein kinase domain	8.6e-39	142.3

[0390]

TABLE 5

SEQ ID NO:	POSITION OF SIGNAL IN AMINO ACID SEQUENCE	maxS (MAXIMUM SCORE)	meanS (MEAN SCORE)
4	1-20	0.971	0.941
7	1-31	0.936	0.729
9	1-30	0.970	0.574

[0391]

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 10

<210> SEQ ID NO 1
<211> LENGTH: 2157
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(1020)

<400> SEQUENCE: 1

```
atg aac ggc tac ggc tcc ccc tac ctg tac atg ggc ggc ccg gtg tcg       48
Met Asn Gly Tyr Gly Ser Pro Tyr Leu Tyr Met Gly Gly Pro Val Ser
  1               5                  10                  15

cag ccg cca cgg gcg ccc ctg cag cgc acg ccc aag tgc gcg cgc tgc       96
Gln Pro Pro Arg Ala Pro Leu Gln Arg Thr Pro Lys Cys Ala Arg Cys
             20                  25                  30

cgc aac cat ggc gtc ctg tcc tgg ctc aag ggc cac aag cgt tac tgc      144
Arg Asn His Gly Val Leu Ser Trp Leu Lys Gly His Lys Arg Tyr Cys
         35                  40                  45

cgc ttc aag gac tgc acc tgc gag aag tgc atc ctc atc atc gag cgg      192
Arg Phe Lys Asp Cys Thr Cys Glu Lys Cys Ile Leu Ile Ile Glu Arg
     50                  55                  60

cag cgg gtc atg gct gcg cag gtg gcg ctg cgc cgg cag cag gcc aac      240
Gln Arg Val Met Ala Ala Gln Val Ala Leu Arg Arg Gln Gln Ala Asn
 65                  70                  75                  80

gag agc ttg gag agc ctc atc ccc gac tcg ctg cgc gct ctg cca ggg      288
Glu Ser Leu Glu Ser Leu Ile Pro Asp Ser Leu Arg Ala Leu Pro Gly
                 85                  90                  95

ccc ccg ccg ccg ggg gac gcc gtc gcc gcc ccg cag ccg ccg cca gcc      336
Pro Pro Pro Pro Gly Asp Ala Val Ala Ala Pro Gln Pro Pro Pro Ala
            100                 105                 110

tct cag ccg tcg cag ccg cag ccg ccg cgc cct gct gcc gag ttg gcc      384
Ser Gln Pro Ser Gln Pro Gln Pro Pro Arg Pro Ala Ala Glu Leu Ala
        115                 120                 125

gcg gcc gcc gcg ctg cgt tgg act gcc gag ccg cag ccc ggg gct ctg      432
Ala Ala Ala Ala Leu Arg Trp Thr Ala Glu Pro Gln Pro Gly Ala Leu
    130                 135                 140

cag gcg cag ctc gcc aag cca gat ttg act gaa gaa cga ctt gga gac      480
Gln Ala Gln Leu Ala Lys Pro Asp Leu Thr Glu Glu Arg Leu Gly Asp
145                 150                 155                 160

ggc aag tcg gca gac aat aca gag gtc ttc agt gac aaa gac act gac      528
Gly Lys Ser Ala Asp Asn Thr Glu Val Phe Ser Asp Lys Asp Thr Asp
                165                 170                 175

cag agg agt tcc cca gat gtg gca aag agt aag ggc tgc ttc acc cct      576
```

-continued

```
Gln Arg Ser Ser Pro Asp Val Ala Lys Ser Lys Gly Cys Phe Thr Pro
            180                 185                 190

gag agc cct gag ata gtg tcc gtg gag gaa ggg gga tac gct gtc cag      624
Glu Ser Pro Glu Ile Val Ser Val Glu Glu Gly Gly Tyr Ala Val Gln
        195                 200                 205

aaa aac gga ggc aac ccc gag agc cgc cct gac agc ccc aag tgt cac      672
Lys Asn Gly Gly Asn Pro Glu Ser Arg Pro Asp Ser Pro Lys Cys His
    210                 215                 220

gcg gag cag aat cac ctc ctg att gag ggc ccc tcg ggg act gtt tct      720
Ala Glu Gln Asn His Leu Leu Ile Glu Gly Pro Ser Gly Thr Val Ser
225                 230                 235                 240

ctg ccc ttc agc ttg aaa gcc aac aga ccg ccg ctt gaa gtg tta aaa      768
Leu Pro Phe Ser Leu Lys Ala Asn Arg Pro Pro Leu Glu Val Leu Lys
                245                 250                 255

aag ata ttc ccc aac cag aag cca acg gtg ctt gag ctc atc ctc aag      816
Lys Ile Phe Pro Asn Gln Lys Pro Thr Val Leu Glu Leu Ile Leu Lys
            260                 265                 270

ggc tgt ggc ggg gac ctg gtg agc gcc gtg gaa gtc ctt ctg tcc agc      864
Gly Cys Gly Gly Asp Leu Val Ser Ala Val Glu Val Leu Leu Ser Ser
        275                 280                 285

cga tcc tca gtc acg gga gca gag cga act tcc gca gaa cct gag agt      912
Arg Ser Ser Val Thr Gly Ala Glu Arg Thr Ser Ala Glu Pro Glu Ser
    290                 295                 300

cta gcg ttg ccc tcc aat ggg cac atc ttt gaa cac acc ttg agc tcc      960
Leu Ala Leu Pro Ser Asn Gly His Ile Phe Glu His Thr Leu Ser Ser
305                 310                 315                 320

tac ccc atc tcg tct tcc aat ggt ctg tgg gat cag cct ttc gag tcc     1008
Tyr Pro Ile Ser Ser Ser Asn Gly Leu Trp Asp Gln Pro Phe Glu Ser
                325                 330                 335

cag aca cgt tga ggt tttctgccga ctctagcaac gttgtcccca gtcccttggc     1063
Gln Thr Arg  *
            340

tgggcctctg cagccccctt tcccccagcc accccggtac ccgctgatgc tgaggaatac   1123

tttggcgaga agccagtcga gccccttttt gcccaatgat gtcaccctgt ggaacaccat   1183

gacgctgcag cagcagtatc agctgaggtc ccagtatgtc agtcctttcc ccagtaactc   1243

taccagcgtc ttcagaagct cgcccgtcct tcctgcccgc gccacggaag accctcggat   1303

ttccatccct gatgatgggt gtccatttgt gtcaaagcag tccatttaca ccgaggacga   1363

ctatgacgag aggtctgact cctcagactc tagaacactc aacacatcat cttaaagtgg   1423

tgctggatgg gtggtggcca ggtgacattt tctgtgcgtt ttgaccctga ggcatctgag   1483

gagaggccac atcttgtgta tgccctttcc ttctgtttga caaggtgact gtgcttgatt   1543

ctatacatta gcaataaaaa cataacttat ttaacttctt gcacttcact ggaaaatgcc   1603

aaatagctct gttctgtggc tttagtgctg aatgtttatt gtaaaagaga gtctaatgtt   1663

aagaatagtc ttgggaaggc tgggtccgtg gaagatttat ttggggatgt aaagctgaag   1723

gtcagccttg cacctaaacc caacctggaa tgttaaatga aataatatac ttgaatgcaa   1783

ttttgtaaaa gtggattcct caggatatgt gaaacctaaa ggaagtggtt ctgttgcaaa   1843

tggactataa acagggacat tatattctta cactaaaaat ccttgcattt taaagagaga   1903

tgcacttaag aatagagtga actgctcata tgcttattta agcttggaca gttttcagag   1963

acaaacccca ttaagaatta ctcttttcac atggctgaat cgaaacatgt gtaatgtcaa   2023

tgtaaaacca atcacagctg tgaactgcat gaaatgtatt gtgaaacgaa cacaagatta   2083
```

-continued

```
agctttgtca ggttaatgta gcatgttaag gactctagaa aaaaataaac taaggagacg   2143

agaaaaaaaa aaaa                                                    2157


<210> SEQ ID NO 2
<211> LENGTH: 1330
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (478)..(1101)

<400> SEQUENCE: 2

tggttcttcc tttattctta attcttaact gtgatggatc cttgcaatga aagaattaaa     60

aagggcgggt gcggtggggc acgcctgtaa ttccagcact ttggaaggtc gacgtgggac    120

gattgcttaa gcccagggct caaggccagc ctgggtaaca tgaacaattc ttaacgtcca    180

aagataacgt ctttccttaa cgtcgaaaga tgccttgttg tgatattaca gcagtcattc    240

ctgctcacta cgggcaggca gacagtgcag aaaaggtgaa gcacaaaaac taaaaaggaa    300

aatctctcct tgaaccactg cgaacagatc cgagtgctgg ggtactgctg ccagctggtt    360

ggtggagttc tggttgaaca gtgtcccgcc aggtccgaag tggggacacg tctgttggtg    420

gtgtccaccc aactcagcca ctgcaggacc atcttgcgac tctttgatga cctggcc       477

atg ttt gtc tac act aag caa tat ggc ctg ggg gca cag gag gag gac      525

Met Phe Val Tyr Thr Lys Gln Tyr Gly Leu Gly Ala Gln Glu Glu Asp
  1               5                  10                  15

gcc ttt gtc cgc tgt gtc tcc gtc cta ggg aac ctg gct gac cag ctc      573
Ala Phe Val Arg Cys Val Ser Val Leu Gly Asn Leu Ala Asp Gln Leu
             20                  25                  30

tac tac ccc tgt gag cac gtg gcc tgg gcg gct gat gcc cgg gtc ctc      621
Tyr Tyr Pro Cys Glu His Val Ala Trp Ala Ala Asp Ala Arg Val Leu
         35                  40                  45

cac gtg gac tct tct cgg tgg tgg acg ctg agt aca acc ctg tgg gcc      669
His Val Asp Ser Ser Arg Trp Trp Thr Leu Ser Thr Thr Leu Trp Ala
     50                  55                  60

ctc tct ctg ctc ctg ggg gtt gcc agg tcc ctg tgg atg ctg ctg aaa      717
Leu Ser Leu Leu Leu Gly Val Ala Arg Ser Leu Trp Met Leu Leu Lys
 65                  70                  75                  80

ctg aga cag agg ctg cgg agc ccc acg gcg ccc ttc acc agc ccg ctg      765
Leu Arg Gln Arg Leu Arg Ser Pro Thr Ala Pro Phe Thr Ser Pro Leu
                 85                  90                  95

ccc cgg ggc aag cgg agg gcc atg gag gcg cag atg cag tcg gag gcg      813
Pro Arg Gly Lys Arg Arg Ala Met Glu Ala Gln Met Gln Ser Glu Ala
            100                 105                 110

ctg tca ctt ctc agc aac ctg gcc gac ctg gcc aac gcc gtg cac tgg      861
Leu Ser Leu Leu Ser Asn Leu Ala Asp Leu Ala Asn Ala Val His Trp
        115                 120                 125

ctg ccc cgg ggc gtg ctg tgg gcc ggc cgc ttc ccg ccg tgg cta gtg      909
Leu Pro Arg Gly Val Leu Trp Ala Gly Arg Phe Pro Pro Trp Leu Val
    130                 135                 140

ggc ctc atg ggc acc acc tcc tgc ctg cca gca tgt acc agg cgg ccc      957
Gly Leu Met Gly Thr Thr Ser Cys Leu Pro Ala Cys Thr Arg Arg Pro
145                 150                 155                 160

ggg ccg gcg gcc agg ccg agg cca cta ccc cct gac act gcc gga aga     1005
Gly Pro Ala Ala Arg Pro Arg Pro Leu Pro Pro Asp Thr Ala Gly Arg
                165                 170                 175
```

-continued

```
gca cag gga cac agc cag agc cca cgg agg gcc ctt ccc gca aag cag     1053
Ala Gln Gly His Ser Gln Ser Pro Arg Arg Ala Leu Pro Ala Lys Gln
            180                 185                 190

aag ccg cca ggg cag ggg ccg ggg ctc tca tgg agc tgc ctg tgg tga     1101
Lys Pro Pro Gly Gln Gly Pro Gly Leu Ser Trp Ser Cys Leu Trp  *
        195                 200                 205

aacccatggg cagggtgggt ggagggactg atgcgaggtc ccccaggcct cagcctggag   1161

ctcctgagcc gtgacagttt tcaggggccg tgcagggctt cagcagcacc tcctgcctgc   1221

cagctggcgt cttgtatcca aatcaagatg aggaagaggg ggcctgtcgt gccttgagaa   1281

agctggaacg ggaatcaatt aaacattgtg gtgctggaaa aaaaaaaaa               1330


<210> SEQ ID NO 3
<211> LENGTH: 992
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (314)..(691)

<400> SEQUENCE: 3

tggtctctgc ccctcctcag atggcctgtt aggtaactca cctgtgtgtg tgtgcatggt     60

ccttcttgag taatggctgc acaggaagga aattacctag tggccactgg aacaggagcc    120

tccggcagcc atctttatgt ccttgggctg agactccagg ctgcccagca gaggataagg    180

tggcctctga gaagcctctt ttttgtcctc tgaccttggg actccacact cctgcttcca    240

tggatttgaa ggttccaaca tccccttct ccacggtctc cagcaggcac agcccaggtc    300

tgacggggag gtg   atg ccc acg ctg gac atg gcc ttg ttc gac tgg acc     349
                Met Pro Thr Leu Asp Met Ala Leu Phe Asp Trp Thr
                  1               5                  10

gat tat gaa gac tta aaa cct gat ggt tgg ccc tct gca aag aag aaa      397
Asp Tyr Glu Asp Leu Lys Pro Asp Gly Trp Pro Ser Ala Lys Lys Lys
         15                  20                  25

gag aaa cac cgc ggt aaa ctc tcc agt gat ggt aac gaa aca tca cca      445
Glu Lys His Arg Gly Lys Leu Ser Ser Asp Gly Asn Glu Thr Ser Pro
     30                  35                  40

gcc gaa ggg gaa cca tgc gac cat cac caa tac tgc ctg aca ggg act      493
Ala Glu Gly Glu Pro Cys Asp His His Gln Tyr Cys Leu Thr Gly Thr
 45                  50                  55                  60

tgc tgc gac ctg cgg gag cat ttc tgc aca ccc cac aac cga ggc ctg      541
Cys Cys Asp Leu Arg Glu His Phe Cys Thr Pro His Asn Arg Gly Leu
                 65                  70                  75

aac aac aaa tgc ttc gat gac tgc atg tgt gtg gaa ggg ctg cgc tgc      589
Asn Asn Lys Cys Phe Asp Asp Cys Met Cys Val Glu Gly Leu Arg Cys
             80                  85                  90

tat gcc aaa ttc cac cgg aac cgc agg gtt aca cgg agg aaa ggg cgc      637
Tyr Ala Lys Phe His Arg Asn Arg Arg Val Thr Arg Arg Lys Gly Arg
         95                  100                 105

tgt gtg gag ccc gag acg gcc aac ggc gac cag gga tcc ttc atc aac      685
Cys Val Glu Pro Glu Thr Ala Asn Gly Asp Gln Gly Ser Phe Ile Asn
    110                 115                 120

gtc tag cggccccgcg ggactgggga ctgagcccag gaggtttgca caagccgggc       741
Val  *
125

gatttgtttg taactagcag tgggagatca agttggggaa cagatggctg aggctgcaga    801
```

-continued

```
ctcaggccca ggacactcaa ccccaggagg ggagccgctc ggcgaatgag ctgggtgggt    861

gcccaggagc cggcccgcag cacctgcaca cacgaagtcc ggacccacgc agcctccatc    921

ccgcgtgtct tgctctccgc gatggcaatg ccgagagtgc cctctactgt ccgactccag    981

cactgcaaca g                                                        992


<210> SEQ ID NO 4
<211> LENGTH: 11389
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(9045)

<400> SEQUENCE: 4

atg gcg ccg ccg ccg ccg ccc gtg ctg ccc gtg ctg ctg ctc ctg gcc       48
Met Ala Pro Pro Pro Pro Pro Val Leu Pro Val Leu Leu Leu Leu Ala
  1               5                  10                  15

gcc gcc gcc gcc ctg ccg gcg atg ggg ctg cga gcg gcc gcc tgg gag       96
Ala Ala Ala Ala Leu Pro Ala Met Gly Leu Arg Ala Ala Ala Trp Glu
             20                  25                  30

ccg cgc gta ccc ggc ggg acc cgc gcc ttc gcc ctc cgg ccc ggc tgt      144
Pro Arg Val Pro Gly Gly Thr Arg Ala Phe Ala Leu Arg Pro Gly Cys
         35                  40                  45

acc tac gcg gtg ggc gcc gct tgc acg ccc cgg gcg ccg cgg gag ctg      192
Thr Tyr Ala Val Gly Ala Ala Cys Thr Pro Arg Ala Pro Arg Glu Leu
     50                  55                  60

ctg gac gtg ggc cgc gat ggg cgg ctg gca gga cgt cgg cgc gtc tcg      240
Leu Asp Val Gly Arg Asp Gly Arg Leu Ala Gly Arg Arg Arg Val Ser
 65                  70                  75                  80

ggc gcg ggg cgc ccg ctg ccg ctg caa gtc cgc ttg gtg gcc cgc agt      288
Gly Ala Gly Arg Pro Leu Pro Leu Gln Val Arg Leu Val Ala Arg Ser
                 85                  90                  95

gcc ccg acg gcg ctg agc cgc cgc ctg cgg gcg cgc acg cac ctt ccc      336
Ala Pro Thr Ala Leu Ser Arg Arg Leu Arg Ala Arg Thr His Leu Pro
            100                 105                 110

ggc tgc gga gcc cgt gcc cgg ctc tgc gga acc ggt gcc cgg ctc tgc      384
Gly Cys Gly Ala Arg Ala Arg Leu Cys Gly Thr Gly Ala Arg Leu Cys
        115                 120                 125

ggg gcg ctc tgc ttc ccc gtc ccc ggc ggc tgc gcg gcc gcg cag cat      432
Gly Ala Leu Cys Phe Pro Val Pro Gly Gly Cys Ala Ala Ala Gln His
    130                 135                 140

tcg gcg ctc gca gct ccg acc acc tta ccc gcc tgc cgc tgc ccg ccg      480
Ser Ala Leu Ala Ala Pro Thr Thr Leu Pro Ala Cys Arg Cys Pro Pro
145                 150                 155                 160

cgc ccc agg ccc cgc tgt ccc ggc cgt ccc atc tgc ctg ccg ccg ggc      528
Arg Pro Arg Pro Arg Cys Pro Gly Arg Pro Ile Cys Leu Pro Pro Gly
                165                 170                 175

ggc tcg gtc cgc ctg cgt ctg ctg tgc gcc ctg cgg cgc gcg gct ggc      576
Gly Ser Val Arg Leu Arg Leu Leu Cys Ala Leu Arg Arg Ala Ala Gly
            180                 185                 190

gcc gtc cgg gtg gga ctg gcg ctg gag gcc gcc acc gcg ggg acg ccc      624
Ala Val Arg Val Gly Leu Ala Leu Glu Ala Ala Thr Ala Gly Thr Pro
        195                 200                 205

tcc gcg tcg cca tcc cca tcg ccg ccc ctg ccg ccg aac ttg ccc gaa      672
Ser Ala Ser Pro Ser Pro Ser Pro Pro Leu Pro Pro Asn Leu Pro Glu
    210                 215                 220

gcc cgg gcg ggg ccg gcg cga cgg gcc cgg cgg ggc acg agc ggc aga      720
```

-continued

```
Ala Arg Ala Gly Pro Ala Arg Arg Ala Arg Arg Gly Thr Ser Gly Arg
225                 230                 235                 240

ggg agc ctg aag ttt ccg atg ccc aac tac cag gtg gcg ttg ttt gag      768
Gly Ser Leu Lys Phe Pro Met Pro Asn Tyr Gln Val Ala Leu Phe Glu
                245                 250                 255

aac gaa ccg gcg ggc acc ctc atc ctc cag ctg cac gcg cac tac acc      816
Asn Glu Pro Ala Gly Thr Leu Ile Leu Gln Leu His Ala His Tyr Thr
            260                 265                 270

atc gag ggc gag gag gag cgc gtg agc tat tac atg gag ggg ctg ttc      864
Ile Glu Gly Glu Glu Glu Arg Val Ser Tyr Tyr Met Glu Gly Leu Phe
        275                 280                 285

gac gag cgc tcc cgg ggc tac ttc cga atc gac tct gcc acg ggc gcc      912
Asp Glu Arg Ser Arg Gly Tyr Phe Arg Ile Asp Ser Ala Thr Gly Ala
    290                 295                 300

gtg agc acg gac agc gta ctg gac cgc gag acc aag gag acg cac gtc      960
Val Ser Thr Asp Ser Val Leu Asp Arg Glu Thr Lys Glu Thr His Val
305                 310                 315                 320

ctc agg gtg aaa gcc gtg gac tac agt acg ccg ccg cgc tcg gcc acc     1008
Leu Arg Val Lys Ala Val Asp Tyr Ser Thr Pro Pro Arg Ser Ala Thr
                325                 330                 335

acc tac atc act gtc ttg gtc aaa gac acc aac gac cac agc ccg gtc     1056
Thr Tyr Ile Thr Val Leu Val Lys Asp Thr Asn Asp His Ser Pro Val
            340                 345                 350

ttc gag cag tcg gag tac cgc gag cgc gtg cgg gag aac ctg gag gtg     1104
Phe Glu Gln Ser Glu Tyr Arg Glu Arg Val Arg Glu Asn Leu Glu Val
        355                 360                 365

ggc tac gag gtg ctg acc atc cgc gcc agc gac cgc gac tcg ccc atc     1152
Gly Tyr Glu Val Leu Thr Ile Arg Ala Ser Asp Arg Asp Ser Pro Ile
    370                 375                 380

aac gcc aac ttg cgt tac cgc gtg ttg ggg ggc gcg tgg gac gtc ttc     1200
Asn Ala Asn Leu Arg Tyr Arg Val Leu Gly Gly Ala Trp Asp Val Phe
385                 390                 395                 400

cag ctc aac gag agc tct ggc gtg gtg agc aca cgg gcg gtg ctg gac     1248
Gln Leu Asn Glu Ser Ser Gly Val Val Ser Thr Arg Ala Val Leu Asp
                405                 410                 415

cgg gag gag gcg gcc gag tac cag ctc ctg gtg gag gcc aac gac cag     1296
Arg Glu Glu Ala Ala Glu Tyr Gln Leu Leu Val Glu Ala Asn Asp Gln
            420                 425                 430

ggg cgc aat ccg ggc ccg ctc agt gcc acg gcc acc gtg tac atc gag     1344
Gly Arg Asn Pro Gly Pro Leu Ser Ala Thr Ala Thr Val Tyr Ile Glu
        435                 440                 445

gtg gag gac gag aac gac aac tac ccc cag ttc agc gag cag aac tac     1392
Val Glu Asp Glu Asn Asp Asn Tyr Pro Gln Phe Ser Glu Gln Asn Tyr
    450                 455                 460

gtg gtc cag gtg ccc gag gac gtg ggg ctc aac acg gct gtg ctg cga     1440
Val Val Gln Val Pro Glu Asp Val Gly Leu Asn Thr Ala Val Leu Arg
465                 470                 475                 480

gtg cag gcc acg gac cgg gac cag ggc cag aac gcg gcc att cac tac     1488
Val Gln Ala Thr Asp Arg Asp Gln Gly Gln Asn Ala Ala Ile His Tyr
                485                 490                 495

agc atc ctc agc ggg aac gtg gcc ggc cag ttc tac ctg cac tcg ctg     1536
Ser Ile Leu Ser Gly Asn Val Ala Gly Gln Phe Tyr Leu His Ser Leu
            500                 505                 510

agc ggg atc ctg gat gtg atc aac ccc ttg gat ttc gag gat gtc cag     1584
Ser Gly Ile Leu Asp Val Ile Asn Pro Leu Asp Phe Glu Asp Val Gln
        515                 520                 525

aaa tac tcg ctg agc att aag gcc cag gat ggg ggc cgg ccc ccg ctc     1632
```

-continued

```
Lys Tyr Ser Leu Ser Ile Lys Ala Gln Asp Gly Gly Arg Pro Pro Leu
    530                 535                 540

atc aat tct tca ggg gtg gtg tct gtg cag gtg ctg gat gtc aac gac     1680
Ile Asn Ser Ser Gly Val Val Ser Val Gln Val Leu Asp Val Asn Asp
545                 550                 555                 560

aac gag cct atc ttt gtg agc agc ccc ttc cag gcc acg gtg ctg gag     1728
Asn Glu Pro Ile Phe Val Ser Ser Pro Phe Gln Ala Thr Val Leu Glu
                565                 570                 575

aat gtg ccc ctg ggc tac ccc gtg gtg cac att cag gcg gtg gac gcg     1776
Asn Val Pro Leu Gly Tyr Pro Val Val His Ile Gln Ala Val Asp Ala
            580                 585                 590

gac tct gga gag aac gcc cgg ctg cac tat cgc ctg gtg gac acg gcc     1824
Asp Ser Gly Glu Asn Ala Arg Leu His Tyr Arg Leu Val Asp Thr Ala
        595                 600                 605

tcc acc ttt ctg ggg ggc ggc agc gct ggg cct aag aat cct gcc ccc     1872
Ser Thr Phe Leu Gly Gly Gly Ser Ala Gly Pro Lys Asn Pro Ala Pro
    610                 615                 620

acc cct gac ttc ccc ttc cag atc cac aac agc tcc ggt tgg atc aca     1920
Thr Pro Asp Phe Pro Phe Gln Ile His Asn Ser Ser Gly Trp Ile Thr
625                 630                 635                 640

gtg tgt gcc gag ctg gac cgc gag gag gtg gag cac tac agc ttc ggg     1968
Val Cys Ala Glu Leu Asp Arg Glu Glu Val Glu His Tyr Ser Phe Gly
                645                 650                 655

gtg gag gcg gtg gac cac ggc tcg ccc ccc atg agc tcc tcc acc agc     2016
Val Glu Ala Val Asp His Gly Ser Pro Pro Met Ser Ser Ser Thr Ser
            660                 665                 670

gtg tcc atc acg gtg ctg gac gtg aat gac aac gac ccg gtg ttc acg     2064
Val Ser Ile Thr Val Leu Asp Val Asn Asp Asn Asp Pro Val Phe Thr
        675                 680                 685

cag ccc acc tac gag ctt cgt ctg aat gag gat gcg gcc gtg ggg agc     2112
Gln Pro Thr Tyr Glu Leu Arg Leu Asn Glu Asp Ala Ala Val Gly Ser
    690                 695                 700

agc gtg ctg acc ctg cag gcc cgc gac cgt gac gcc aac agt gtg att     2160
Ser Val Leu Thr Leu Gln Ala Arg Asp Arg Asp Ala Asn Ser Val Ile
705                 710                 715                 720

acc tac cag ctc aca ggc ggc aac acc cgg aac cgc ttt gca ctc agc     2208
Thr Tyr Gln Leu Thr Gly Gly Asn Thr Arg Asn Arg Phe Ala Leu Ser
                725                 730                 735

agc cag aga ggg ggc ggc ctc atc acc ctg gcg cta cct ctg gac tac     2256
Ser Gln Arg Gly Gly Gly Leu Ile Thr Leu Ala Leu Pro Leu Asp Tyr
            740                 745                 750

aag cag gag cag cag tac gtg ctg gcg gtg aca gca tcc gac ggc aca     2304
Lys Gln Glu Gln Gln Tyr Val Leu Ala Val Thr Ala Ser Asp Gly Thr
        755                 760                 765

cgg tcg cac act gcg cat gtc cta atc aac gtc act gat gcc aac acc     2352
Arg Ser His Thr Ala His Val Leu Ile Asn Val Thr Asp Ala Asn Thr
    770                 775                 780

cac agg cct gtc ttt cag agc tcc cat tac aca gtg agt gtc agt gag     2400
His Arg Pro Val Phe Gln Ser Ser His Tyr Thr Val Ser Val Ser Glu
785                 790                 795                 800

gac agg cct gtg ggc acc tcc att gct acc ctc agt gcc aac gat gag     2448
Asp Arg Pro Val Gly Thr Ser Ile Ala Thr Leu Ser Ala Asn Asp Glu
                805                 810                 815

gac aca gga gag aat gcc cgc atc acc tac gtg att cag gac ccc gtg     2496
Asp Thr Gly Glu Asn Ala Arg Ile Thr Tyr Val Ile Gln Asp Pro Val
            820                 825                 830

ccg cag ttc cgc att gac ccc gac agt ggc acc atg tac acc atg atg     2544
```

-continued

```
Pro Gln Phe Arg Ile Asp Pro Asp Ser Gly Thr Met Tyr Thr Met Met
        835                 840                 845

gag ctg gac tat gag aac cag gtc gcc tac acg ctg acc atc atg gcc     2592
Glu Leu Asp Tyr Glu Asn Gln Val Ala Tyr Thr Leu Thr Ile Met Ala
    850                 855                 860

cag gac aac ggc atc ccg cag aaa tca gac acc acc acc cta gag atc     2640
Gln Asp Asn Gly Ile Pro Gln Lys Ser Asp Thr Thr Thr Leu Glu Ile
865                 870                 875                 880

ctc atc ctc gat gcc aat gac aat gca ccc cag ttc ctg tgg gat ttc     2688
Leu Ile Leu Asp Ala Asn Asp Asn Ala Pro Gln Phe Leu Trp Asp Phe
                885                 890                 895

tac cag ggt tcc atc ttt gag gat gct cca ccc tcg acc agc atc ctc     2736
Tyr Gln Gly Ser Ile Phe Glu Asp Ala Pro Pro Ser Thr Ser Ile Leu
            900                 905                 910

cag gtc tct gcc acg gac cgg gac tca ggt ccc aat ggg cgt ctg ctg     2784
Gln Val Ser Ala Thr Asp Arg Asp Ser Gly Pro Asn Gly Arg Leu Leu
        915                 920                 925

tac acc ttc cag ggt ggg gac gac ggc gat ggg gac ttc tac atc gag     2832
Tyr Thr Phe Gln Gly Gly Asp Asp Gly Asp Gly Asp Phe Tyr Ile Glu
    930                 935                 940

ccc acg tcc ggt gtg att cgc acc cag cgc cgg ctg gac cgg gag aat     2880
Pro Thr Ser Gly Val Ile Arg Thr Gln Arg Arg Leu Asp Arg Glu Asn
945                 950                 955                 960

gtg gcc gtg tac aac ctt tgg gct ctg gct gtg gat cgg ggc agt ccc     2928
Val Ala Val Tyr Asn Leu Trp Ala Leu Ala Val Asp Arg Gly Ser Pro
                965                 970                 975

act ccc ctt agc gcc tcg gta gaa atc cag gtg acc atc ttg gac att     2976
Thr Pro Leu Ser Ala Ser Val Glu Ile Gln Val Thr Ile Leu Asp Ile
            980                 985                 990

aat gac aat gcc ccc atg ttt gag aag gac gaa ctg gag ctg ttt gtt     3024
Asn Asp Asn Ala Pro Met Phe Glu Lys Asp Glu Leu Glu Leu Phe Val
        995                1000                1005

gag gag aac aac cca gtg ggg tcg gtg gtg gca aag att cgt gct aac     3072
Glu Glu Asn Asn Pro Val Gly Ser Val Val Ala Lys Ile Arg Ala Asn
   1010                1015                1020

 gac cct gat gaa ggc cct aat gcc cag atc atg tat cag att gtg gaa    3120
 Asp Pro Asp Glu Gly Pro Asn Ala Gln Ile Met Tyr Gln Ile Val Glu
1025                1030                1035                1040

ggg gac atg cgg cat ttc ttc cag ctg gac ctg ctc aac ggg gac ctg     3168
Gly Asp Met Arg His Phe Phe Gln Leu Asp Leu Leu Asn Gly Asp Leu
               1045                1050                1055

cgt gcc atg gtg gag ctg gac ttt gag gtc cgg cgg gag tat gtg ctg     3216
Arg Ala Met Val Glu Leu Asp Phe Glu Val Arg Arg Glu Tyr Val Leu
           1060                1065                1070

gtg gtg cag gcc acg tcg gct ccg ctg gtg agc cga gcc acg gtg cac     3264
Val Val Gln Ala Thr Ser Ala Pro Leu Val Ser Arg Ala Thr Val His
       1075                1080                1085

atc ctt ctc gtg gac cag aat gac aac ccg cct gtg ctg ccc gac ttc     3312
Ile Leu Leu Val Asp Gln Asn Asp Asn Pro Pro Val Leu Pro Asp Phe
   1090                1095                1100

 cag atc ctc ttc aac aac tat gtc acc aac aag tcc aac agt ttc ccc    3360
 Gln Ile Leu Phe Asn Asn Tyr Val Thr Asn Lys Ser Asn Ser Phe Pro
1105                1110                1115                1120

acc ggc gtg atc ggc tgc atc ccg gcc cat gac ccc gac gtg tca gac     3408
Thr Gly Val Ile Gly Cys Ile Pro Ala His Asp Pro Asp Val Ser Asp
               1125                1130                1135

agc ctc aac tac acc ttc gtg cag ggc aac gag ctg cgc ctg ttg ctg     3456
```

-continued

```
Ser Leu Asn Tyr Thr Phe Val Gln Gly Asn Glu Leu Arg Leu Leu Leu
          1140                1145                1150

ctg gac ccc gcc acg ggc gaa ctg cag ctc agc cgc gac ctg gac aac     3504
Leu Asp Pro Ala Thr Gly Glu Leu Gln Leu Ser Arg Asp Leu Asp Asn
       1155                1160                1165

aac cgg ccg ctg gag gcg ctc atg gag gtg tct gtg tct gat ggc atc     3552
Asn Arg Pro Leu Glu Ala Leu Met Glu Val Ser Val Ser Asp Gly Ile
   1170                1175                1180

 cac agc gtc acg gcc ttc tgc acc ctg cgt gtc acc atc atc acg gac    3600
 His Ser Val Thr Ala Phe Cys Thr Leu Arg Val Thr Ile Ile Thr Asp
1185                1190                1195                1200

gac atg ctg acc aac agc atc act gtc cgc ctg gag aac atg tcc cag     3648
Asp Met Leu Thr Asn Ser Ile Thr Val Arg Leu Glu Asn Met Ser Gln
               1205                1210                1215

gag aag ttc ctg tcc ccg ctg ctg gcc ctc ttc gtg gag ggg gtg gcc     3696
Glu Lys Phe Leu Ser Pro Leu Leu Ala Leu Phe Val Glu Gly Val Ala
          1220                1225                1230

gcc gtg ctg tcc acc acc aag gac gac gtc ttc gtc ttc aac gtc cag     3744
Ala Val Leu Ser Thr Thr Lys Asp Asp Val Phe Val Phe Asn Val Gln
       1235                1240                1245

aac gac acc gac gtc agc tcc aac atc ctg aac gtg acc ttc tcg gcg     3792
Asn Asp Thr Asp Val Ser Ser Asn Ile Leu Asn Val Thr Phe Ser Ala
   1250                1255                1260

 ctg ctg cct ggc ggc gtc cgc ggc cag ttc ttc ccg tcg gag gac ctg    3840
 Leu Leu Pro Gly Gly Val Arg Gly Gln Phe Phe Pro Ser Glu Asp Leu
1265                1270                1275                1280

cag gag cag atc tac ctg aat cgg acg ctg ctg acc acc atc tcc acg     3888
Gln Glu Gln Ile Tyr Leu Asn Arg Thr Leu Leu Thr Thr Ile Ser Thr
               1285                1290                1295

cag cgc gtg ctg ccc ttc gac gac aac atc tgc ctg cgc gag ccc tgc     3936
Gln Arg Val Leu Pro Phe Asp Asp Asn Ile Cys Leu Arg Glu Pro Cys
          1300                1305                1310

gag aac tac atg aag tgc gtg tcc gtt ctg cga ttc gac agc tcc gcg     3984
Glu Asn Tyr Met Lys Cys Val Ser Val Leu Arg Phe Asp Ser Ser Ala
       1315                1320                1325

ccc ttc ctc agc tcc acc acc gtg ctc ttc cgg ccc atc cac ccc atc     4032
Pro Phe Leu Ser Ser Thr Thr Val Leu Phe Arg Pro Ile His Pro Ile
   1330                1335                1340

 aac ggc ctg cgc tgc cgc tgc ccg ccc ggc ttc acc ggc gac tac tgc    4080
 Asn Gly Leu Arg Cys Arg Cys Pro Pro Gly Phe Thr Gly Asp Tyr Cys
1345                1350                1355                1360

gag acg gag atc gac ctc tgc tac tcc gac ccg tgc ggc gcc aac ggc     4128
Glu Thr Glu Ile Asp Leu Cys Tyr Ser Asp Pro Cys Gly Ala Asn Gly
               1365                1370                1375

cgc tgc cgc agc cgc gag ggc ggc tac acc tgc gag tgc ttc gag gac     4176
Arg Cys Arg Ser Arg Glu Gly Gly Tyr Thr Cys Glu Cys Phe Glu Asp
          1380                1385                1390

ttc act gga gag cac tgt gag gtg gat gcc cgc tca ggc cgc tgt gcc     4224
Phe Thr Gly Glu His Cys Glu Val Asp Ala Arg Ser Gly Arg Cys Ala
       1395                1400                1405

aac ggg gtg tgc aag aac ggg ggc acc tgc gtg aac ctg ctc atc ggc     4272
Asn Gly Val Cys Lys Asn Gly Gly Thr Cys Val Asn Leu Leu Ile Gly
   1410                1415                1420

 ggc ttc cac tgc gtg tgt cct cct ggc gag tat gag agg ccc tac tgt    4320
 Gly Phe His Cys Val Cys Pro Pro Gly Glu Tyr Glu Arg Pro Tyr Cys
1425                1430                1435                1440

gag gtg acc acc agg agc ttc ccg ccc cag tcc ttc gtc acc ttc cgg     4368
```

-continued

```
Glu Val Thr Thr Arg Ser Phe Pro Pro Gln Ser Phe Val Thr Phe Arg
             1445                1450                1455

ggc ctg aga cag cgc ttc cac ttc acc atc tcc ctc acg ttt gcc act     4416
Gly Leu Arg Gln Arg Phe His Phe Thr Ile Ser Leu Thr Phe Ala Thr
          1460                1465                1470

cag gaa agg aac ggc ttg ctt ctc tac aac ggc cgc ttc aat gag aag     4464
Gln Glu Arg Asn Gly Leu Leu Leu Tyr Asn Gly Arg Phe Asn Glu Lys
       1475                1480                1485

cac gac ttc atc gcc ctg gag atc gtg gac gag cag gtg cag ctc acc     4512
His Asp Phe Ile Ala Leu Glu Ile Val Asp Glu Gln Val Gln Leu Thr
   1490                1495                1500

 ttc tct gca ggc gag aca aca acg acc gtg gca ccg aag gtt ccc agt    4560
 Phe Ser Ala Gly Glu Thr Thr Thr Thr Val Ala Pro Lys Val Pro Ser
1505                1510                1515                1520

ggt gtg agt gac ggg cgg tgg cac tct gtg cag gtg cag tac tac aac     4608
Gly Val Ser Asp Gly Arg Trp His Ser Val Gln Val Gln Tyr Tyr Asn
             1525                1530                1535

aag ccc aat att ggc cac ctg ggc ctg ccc cat ggg ccg tcc ggg gaa     4656
Lys Pro Asn Ile Gly His Leu Gly Leu Pro His Gly Pro Ser Gly Glu
          1540                1545                1550

aag atg gcc gtg gtg aca gtg gat gat tgt gac aca acc atg gct gtg     4704
Lys Met Ala Val Val Thr Val Asp Asp Cys Asp Thr Thr Met Ala Val
       1555                1560                1565

cgc ttt gga aag gac atc ggg aac tac agc tgc gct gcc cag ggc act     4752
Arg Phe Gly Lys Asp Ile Gly Asn Tyr Ser Cys Ala Ala Gln Gly Thr
   1570                1575                1580

 cag acc ggc tcc aag aag tcc ctg gat ctg acc ggc cct cta ctc ctg    4800
 Gln Thr Gly Ser Lys Lys Ser Leu Asp Leu Thr Gly Pro Leu Leu Leu
1585                1590                1595                1600

ggg ggt gtc ccc aac ctg cca gaa gac ttc cca gtg cac aac cgg cag     4848
Gly Gly Val Pro Asn Leu Pro Glu Asp Phe Pro Val His Asn Arg Gln
             1605                1610                1615

ttc gtg ggc tgc atg cgg aac ctg tca gtc gac ggc aaa aat gtg gac     4896
Phe Val Gly Cys Met Arg Asn Leu Ser Val Asp Gly Lys Asn Val Asp
          1620                1625                1630

atg gcc gga ttc atc gcc aac aat ggc acc cgg gaa ggc tgc gct gct     4944
Met Ala Gly Phe Ile Ala Asn Asn Gly Thr Arg Glu Gly Cys Ala Ala
       1635                1640                1645

cgg agg aac ttc tgc gat ggg agg cgg tgt cag aat gga ggc acc tgt     4992
Arg Arg Asn Phe Cys Asp Gly Arg Arg Cys Gln Asn Gly Gly Thr Cys
   1650                1655                1660

 gtc aac agg tgg aat atg tat ctg tgt gag tgt cca ctc cga ttc ggc    5040
 Val Asn Arg Trp Asn Met Tyr Leu Cys Glu Cys Pro Leu Arg Phe Gly
1665                1670                1675                1680

ggg aag aac tgt gag caa gcc atg cct cac ccc cag ctc ttc agc ggt     5088
Gly Lys Asn Cys Glu Gln Ala Met Pro His Pro Gln Leu Phe Ser Gly
             1685                1690                1695

gag agc gtc gtg tcc tgg agt gac ctg aac atc atc atc tct gtg ccc     5136
Glu Ser Val Val Ser Trp Ser Asp Leu Asn Ile Ile Ile Ser Val Pro
          1700                1705                1710

tgg tac ctg ggg ctc atg ttc cgg acc cgg aag gag gac agc gtt ctg     5184
Trp Tyr Leu Gly Leu Met Phe Arg Thr Arg Lys Glu Asp Ser Val Leu
       1715                1720                1725

atg gag gcc acc agt ggt ggg ccc acc agc ttt cgc ctc cag atc ctg     5232
Met Glu Ala Thr Ser Gly Gly Pro Thr Ser Phe Arg Leu Gln Ile Leu
   1730                1735                1740

 aac aac tac ctc cag ttt gag gtg tcc cac ggc ccc tcc gat gtg gag    5280
```

-continued

```
 Asn Asn Tyr Leu Gln Phe Glu Val Ser His Gly Pro Ser Asp Val Glu
1745                1750                1755                1760

tcc gtg atg ctg tcc ggg ttg cgg gtg acc gac ggg gag tgg cac cac     5328
Ser Val Met Leu Ser Gly Leu Arg Val Thr Asp Gly Glu Trp His His
                1765                1770                1775

ctg ctg atc gag ctg aag aat gtt aag gag gac agt gag atg aag cac     5376
Leu Leu Ile Glu Leu Lys Asn Val Lys Glu Asp Ser Glu Met Lys His
            1780                1785                1790

ctg gtc acc atg acc ttg gac tat ggg atg gac cag aac aag gca gat     5424
Leu Val Thr Met Thr Leu Asp Tyr Gly Met Asp Gln Asn Lys Ala Asp
        1795                1800                1805

atc ggg ggc atg ctt ccc ggg ctg acg gta agg agc gtg gtg gtc gga     5472
Ile Gly Gly Met Leu Pro Gly Leu Thr Val Arg Ser Val Val Val Gly
    1810                1815                1820

 ggc gcc tct gaa gac aag gtc tcc gtg cgc cgt gga ttc cga ggc tgc    5520
 Gly Ala Ser Glu Asp Lys Val Ser Val Arg Arg Gly Phe Arg Gly Cys
1825                1830                1835                1840

atg cag gga gtg agg atg ggg ggg acg ccc acc aac gtc gcc acc ctg     5568
Met Gln Gly Val Arg Met Gly Gly Thr Pro Thr Asn Val Ala Thr Leu
                1845                1850                1855

aac atg aac aac gca ctc aag gtc agg gtg aag gac ggc tgt gat gtg     5616
Asn Met Asn Asn Ala Leu Lys Val Arg Val Lys Asp Gly Cys Asp Val
            1860                1865                1870

gac gac ccc tgt acc tcg agc ccc tgt ccc ccc aat agc cgc tgc cac     5664
Asp Asp Pro Cys Thr Ser Ser Pro Cys Pro Pro Asn Ser Arg Cys His
        1875                1880                1885

gac gcc tgg gag gac tac agc tgc gtc tgt gac aaa ggg tac ctt gga     5712
Asp Ala Trp Glu Asp Tyr Ser Cys Val Cys Asp Lys Gly Tyr Leu Gly
    1890                1895                1900

 ata aac tgt gtg gat gcc tgt cac ctg aac ccc tgc gag aac atg ggg    5760
 Ile Asn Cys Val Asp Ala Cys His Leu Asn Pro Cys Glu Asn Met Gly
1905                1910                1915                1920

gcc tgc gtg cgc tcc ccc ggc tcc ccg cag ggc tac gtg tgc gag tgt     5808
Ala Cys Val Arg Ser Pro Gly Ser Pro Gln Gly Tyr Val Cys Glu Cys
                1925                1930                1935

ggg ccc agt cac tac ggg ccg tac tgt gag aac aaa ctc gac ctt ccg     5856
Gly Pro Ser His Tyr Gly Pro Tyr Cys Glu Asn Lys Leu Asp Leu Pro
            1940                1945                1950

tgc ccc aga ggc tgg tgg ggg aac ccc gtc tgt gga ccc tgc cac tgt     5904
Cys Pro Arg Gly Trp Trp Gly Asn Pro Val Cys Gly Pro Cys His Cys
        1955                1960                1965

gcc gtc agc aaa ggc ttt gat ccc gac tgt aat aag acc aac ggc cag     5952
Ala Val Ser Lys Gly Phe Asp Pro Asp Cys Asn Lys Thr Asn Gly Gln
    1970                1975                1980

 tgc caa tgc aag gag aat tac tac aag ctc cta gcc cag gac acc tgt    6000
 Cys Gln Cys Lys Glu Asn Tyr Tyr Lys Leu Leu Ala Gln Asp Thr Cys
1985                1990                1995                2000

ctg ccc tgc gac tgc ttc ccc cat ggc tcc cac agc cgc act tgc gac     6048
Leu Pro Cys Asp Cys Phe Pro His Gly Ser His Ser Arg Thr Cys Asp
                2005                2010                2015

atg gcc acc ggg cag tgt gcc tgc aag ccc ggc gtc atc ggc cgc cag     6096
Met Ala Thr Gly Gln Cys Ala Cys Lys Pro Gly Val Ile Gly Arg Gln
            2020                2025                2030

tgc aac cgc tgc gac aac ccg ttt gcc gag gtc acc acg ctc ggc tgt     6144
Cys Asn Arg Cys Asp Asn Pro Phe Ala Glu Val Thr Thr Leu Gly Cys
        2035                2040                2045

gaa gtg atc tac aat ggc tgt ccc aaa gca ttt gag gcc ggc atc tgg     6192
```

-continued

```
Glu Val Ile Tyr Asn Gly Cys Pro Lys Ala Phe Glu Ala Gly Ile Trp
   2050                2055                2060

 tgg cca cag acc aag ttc ggg cag ccg gct gcg gtg cca tgc cct aag    6240
 Trp Pro Gln Thr Lys Phe Gly Gln Pro Ala Ala Val Pro Cys Pro Lys
2065                2070                2075                2080

gga tcc gtt gga aat gcg gtc cga cac tgc agc ggg gag aag ggc tgg    6288
Gly Ser Val Gly Asn Ala Val Arg His Cys Ser Gly Glu Lys Gly Trp
               2085                2090                2095

ctg ccc cca gag ctc ttt aac tgt acc acc atc tcc ttc gtg gac ctc    6336
Leu Pro Pro Glu Leu Phe Asn Cys Thr Thr Ile Ser Phe Val Asp Leu
          2100                2105                2110

agg gcc atg aat gag aag ctg agc cgc aat gag acg cag gtg gac ggc    6384
Arg Ala Met Asn Glu Lys Leu Ser Arg Asn Glu Thr Gln Val Asp Gly
       2115                2120                2125

gcc agg gcc ctg cag ctg gtg agg gcg ctg cgc agt gct aca cag cac    6432
Ala Arg Ala Leu Gln Leu Val Arg Ala Leu Arg Ser Ala Thr Gln His
   2130                2135                2140

 acg ggc acg ctc ttt ggc aat gac gtg cgc acg gcc tac cag ctg ctg    6480
 Thr Gly Thr Leu Phe Gly Asn Asp Val Arg Thr Ala Tyr Gln Leu Leu
2145                2150                2155                2160

ggc cac gtc ctt cag cac gag agc tgg cag cag ggc ttc gac ctg gca    6528
Gly His Val Leu Gln His Glu Ser Trp Gln Gln Gly Phe Asp Leu Ala
               2165                2170                2175

gcc acg cag gac gcc gac ttt cac gag gac gtc atc cac tcg ggc agc    6576
Ala Thr Gln Asp Ala Asp Phe His Glu Asp Val Ile His Ser Gly Ser
          2180                2185                2190

gcc ctc ctg gcc cca gcc acc agg gcg gcg tgg gag cag atc cag cgg    6624
Ala Leu Leu Ala Pro Ala Thr Arg Ala Ala Trp Glu Gln Ile Gln Arg
       2195                2200                2205

agc gag ggc ggc acg gca cag ctg ctc cgg cgc ctc gag ggc tac ttc    6672
Ser Glu Gly Gly Thr Ala Gln Leu Leu Arg Arg Leu Glu Gly Tyr Phe
   2210                2215                2220

 agc aac gtg gca cgc aac gtg cgg cgg acg tac ctg cgg ccc ttc gtc    6720
 Ser Asn Val Ala Arg Asn Val Arg Arg Thr Tyr Leu Arg Pro Phe Val
2225                2230                2235                2240

atc gtc acc gcc aac atg att ctt gct gtc gac atc ttt gac aag ttc    6768
Ile Val Thr Ala Asn Met Ile Leu Ala Val Asp Ile Phe Asp Lys Phe
               2245                2250                2255

aac ttt acg gga gcc agg gtc ccg cga ttc gac acc atc cat gaa gag    6816
Asn Phe Thr Gly Ala Arg Val Pro Arg Phe Asp Thr Ile His Glu Glu
          2260                2265                2270

ttc ccc agg gag ctg gag tcc tcc gtc tcc ttc cca gcc gac ttc ttc    6864
Phe Pro Arg Glu Leu Glu Ser Ser Val Ser Phe Pro Ala Asp Phe Phe
       2275                2280                2285

aga cca cct gaa gaa aaa gaa ggc ccc ctg ctg agg ccg gct ggc cgg    6912
Arg Pro Pro Glu Glu Lys Glu Gly Pro Leu Leu Arg Pro Ala Gly Arg
   2290                2295                2300

 agg acc acc ccg cag acc acg cgc ccg ggg cct ggc acc gag agg gag    6960
 Arg Thr Thr Pro Gln Thr Thr Arg Pro Gly Pro Gly Thr Glu Arg Glu
2305                2310                2315                2320

gcc ccg atc agc agg cgg agg cga cac cct gat gac gct ggc cag ttc    7008
Ala Pro Ile Ser Arg Arg Arg Arg His Pro Asp Asp Ala Gly Gln Phe
               2325                2330                2335

gcc gtc gct ctg gtc atc att tac cgc acc ctg ggg cag ctc ctg ccc    7056
Ala Val Ala Leu Val Ile Ile Tyr Arg Thr Leu Gly Gln Leu Leu Pro
          2340                2345                2350

gag cgc tac gac ccc gac cgt cgc agc ctc cgg ttg cct cac cgg ccc    7104
```

-continued

```
Glu Arg Tyr Asp Pro Asp Arg Arg Ser Leu Arg Leu Pro His Arg Pro
       2355                2360                2365

atc att aat acc ccg atg gtg agc acg ctg gtg tac agc gag ggg gct     7152
Ile Ile Asn Thr Pro Met Val Ser Thr Leu Val Tyr Ser Glu Gly Ala
   2370                2375                2380

 ccg ctc ccg aga ccc ctg gag agg ccc gtc ctg gtg gag ttc gcc ctg    7200
 Pro Leu Pro Arg Pro Leu Glu Arg Pro Val Leu Val Glu Phe Ala Leu
2385                2390                2395                2400

ctg gag gtg gag gag cga acc aag cct gtc tgc gtg ttc tgg aac cac     7248
Leu Glu Val Glu Glu Arg Thr Lys Pro Val Cys Val Phe Trp Asn His
               2405                2410                2415

tcc ctg gcc gtt ggt ggg acg gga ggg tgg tct gcc cgg ggc tgc gag     7296
Ser Leu Ala Val Gly Gly Thr Gly Gly Trp Ser Ala Arg Gly Cys Glu
           2420                2425                2430

ctc ctg tcc agg aac cgg aca cat gtc gcc tgc cag tgc agc cac aca     7344
Leu Leu Ser Arg Asn Arg Thr His Val Ala Cys Gln Cys Ser His Thr
       2435                2440                2445

gcc agc ttt gcg gtg ctc atg gat atc tcc agg cgt gag aac ggg gag     7392
Ala Ser Phe Ala Val Leu Met Asp Ile Ser Arg Arg Glu Asn Gly Glu
   2450                2455                2460

 gtc ctg cct ctg aag att gtc acc tat gcc gct gtg tcc ttg tca ctg    7440
 Val Leu Pro Leu Lys Ile Val Thr Tyr Ala Ala Val Ser Leu Ser Leu
2465                2470                2475                2480

gca gcc ctg ctg gtg gcc ttc gtc ctc ctg agc ctg gtc cgc atg ctg     7488
Ala Ala Leu Leu Val Ala Phe Val Leu Leu Ser Leu Val Arg Met Leu
               2485                2490                2495

cgc tcc aac ctg cac agc att cac aag cac ctc gcc gtg gcg ctc ttc     7536
Arg Ser Asn Leu His Ser Ile His Lys His Leu Ala Val Ala Leu Phe
           2500                2505                2510

ctc tct cag ctg gtg ttc gtg att ggg atc aac cag acg gaa aac ccg     7584
Leu Ser Gln Leu Val Phe Val Ile Gly Ile Asn Gln Thr Glu Asn Pro
       2515                2520                2525

ttt ctg tgc aca gtg gtt gcc atc ctc ctc cac tac atc tac atg agc     7632
Phe Leu Cys Thr Val Val Ala Ile Leu Leu His Tyr Ile Tyr Met Ser
   2530                2535                2540

 acc ttt gcc tgg acc ctc gtg gag agc ctg cat gtc tac cgc atg ctg    7680
 Thr Phe Ala Trp Thr Leu Val Glu Ser Leu His Val Tyr Arg Met Leu
2545                2550                2555                2560

acc gag gtg cgc aac atc gac acg ggg ccc atg cgg ttc tac tac gtc     7728
Thr Glu Val Arg Asn Ile Asp Thr Gly Pro Met Arg Phe Tyr Tyr Val
               2565                2570                2575

gtg ggc tgg ggc atc ccg gcc att gtc aca gga ctg gcg gtc ggc ctg     7776
Val Gly Trp Gly Ile Pro Ala Ile Val Thr Gly Leu Ala Val Gly Leu
           2580                2585                2590

gac ccc cag ggc tac ggg aac ccc gac ttc tgc tgg ctg tcg ctt caa     7824
Asp Pro Gln Gly Tyr Gly Asn Pro Asp Phe Cys Trp Leu Ser Leu Gln
       2595                2600                2605

gac acc ctg att tgg agc ttt gcg ggg ccc atc gga gct gtt ata atc     7872
Asp Thr Leu Ile Trp Ser Phe Ala Gly Pro Ile Gly Ala Val Ile Ile
   2610                2615                2620

 atc aac aca gtc act tct gtc cta tct gca aag gtt tcc tgc caa aga    7920
 Ile Asn Thr Val Thr Ser Val Leu Ser Ala Lys Val Ser Cys Gln Arg
2625                2630                2635                2640

aag cac cat tat tat ggg aaa aaa ggg atc gtc tcc ctg ctg agg acc     7968
Lys His His Tyr Tyr Gly Lys Lys Gly Ile Val Ser Leu Leu Arg Thr
               2645                2650                2655

gca ttc ctc ctg ctg ctg ctc atc agc gcc acc tgg ctg ctg ggg ctg     8016
```

-continued

```
Ala Phe Leu Leu Leu Leu Leu Ile Ser Ala Thr Trp Leu Leu Gly Leu
          2660                2665                2670

ctg gct gtg aac cgc gat gca ctg agc ttt cac tac ctc ttc gcc atc     8064
Leu Ala Val Asn Arg Asp Ala Leu Ser Phe His Tyr Leu Phe Ala Ile
       2675                2680                2685

ttc agc ggc tta cag ggc ccc ttc gtc ctc ctt ttc cac tgc gtg ctc     8112
Phe Ser Gly Leu Gln Gly Pro Phe Val Leu Leu Phe His Cys Val Leu
   2690                2695                2700

 aac cag gag gtc cgg aag cac ctg aag ggc gtg ctc ggc ggg agg aag    8160
 Asn Gln Glu Val Arg Lys His Leu Lys Gly Val Leu Gly Gly Arg Lys
2705                2710                2715                2720

ctg cac ctg gag gac tcc gcc acc acc agg gcc acc ctg ctg acg cgc     8208
Leu His Leu Glu Asp Ser Ala Thr Thr Arg Ala Thr Leu Leu Thr Arg
              2725                2730                2735

tcc ctc aac tgc aac acc acc ttc ggt gac ggg cct gac atg ctg cgc     8256
Ser Leu Asn Cys Asn Thr Thr Phe Gly Asp Gly Pro Asp Met Leu Arg
          2740                2745                2750

aca gac ttg ggc gag tcc acc gcc tcg ctg gac agc atc gtc agg gat     8304
Thr Asp Leu Gly Glu Ser Thr Ala Ser Leu Asp Ser Ile Val Arg Asp
       2755                2760                2765

gaa ggg atc cag aag ctc ggc gtg tcc tct ggg ctg gtg agg ggc agc     8352
Glu Gly Ile Gln Lys Leu Gly Val Ser Ser Gly Leu Val Arg Gly Ser
   2770                2775                2780

 cac gga gag cca gac gcg tcc ctc atg ccc agg agc tgc aag gat ccc    8400
 His Gly Glu Pro Asp Ala Ser Leu Met Pro Arg Ser Cys Lys Asp Pro
2785                2790                2795                2800

cct ggc cac gat tcc gac tca gat agc gag ctg tcc ctg gat gag cag     8448
Pro Gly His Asp Ser Asp Ser Asp Ser Glu Leu Ser Leu Asp Glu Gln
              2805                2810                2815

agc agc tct tac gcc tcc tca cac tcg tca gac agc gag gac gat ggg     8496
Ser Ser Ser Tyr Ala Ser Ser His Ser Ser Asp Ser Glu Asp Asp Gly
          2820                2825                2830

gtg gga gct gag gaa aaa tgg gac ccg gcc agg ggc gcc gtc cac agc     8544
Val Gly Ala Glu Glu Lys Trp Asp Pro Ala Arg Gly Ala Val His Ser
       2835                2840                2845

acc ccc aaa ggg gac gct gtg gcc aac cac gtt ccg gcc ggc tgg ccc     8592
Thr Pro Lys Gly Asp Ala Val Ala Asn His Val Pro Ala Gly Trp Pro
   2850                2855                2860

 gac cag agc ctg gct gag agt gac agt gag gac ccc agc ggc aag ccc    8640
 Asp Gln Ser Leu Ala Glu Ser Asp Ser Glu Asp Pro Ser Gly Lys Pro
2865                2870                2875                2880

cgc ctg aag gtg gag acc aag gtc agc gtg gag ctg cac cgc gag gag     8688
Arg Leu Lys Val Glu Thr Lys Val Ser Val Glu Leu His Arg Glu Glu
              2885                2890                2895

cag ggc agt cac cgt gga gag tac ccc ccg gac cag gag agc ggg ggc     8736
Gln Gly Ser His Arg Gly Glu Tyr Pro Pro Asp Gln Glu Ser Gly Gly
          2900                2905                2910

gca gcc agg ctt gct agc agc cag ccc cca gag cag agg aaa ggc atc     8784
Ala Ala Arg Leu Ala Ser Ser Gln Pro Pro Glu Gln Arg Lys Gly Ile
       2915                2920                2925

ttg aaa aat aaa gtc acc tac ccg ccg ccg ctg acg ctg acg gag cag     8832
Leu Lys Asn Lys Val Thr Tyr Pro Pro Pro Leu Thr Leu Thr Glu Gln
   2930                2935                2940

 acg ctg aag ggc cgg ctc cgg gag aag ctg gcc gac tgt gag cag agc    8880
 Thr Leu Lys Gly Arg Leu Arg Glu Lys Leu Ala Asp Cys Glu Gln Ser
2945                2950                2955                2960

ccc aca tcc tcg cgc acg tct tcc ctg ggc tct ggc ggc ccc gac tgc     8928
```

-continued

```
Pro Thr Ser Ser Arg Thr Ser Ser Leu Gly Ser Gly Gly Pro Asp Cys
               2965                2970                2975

gcc atc aca gtc aag agc cct ggg agg gag ccg ggg cgt gac cac ctc     8976
Ala Ile Thr Val Lys Ser Pro Gly Arg Glu Pro Gly Arg Asp His Leu
           2980                2985                2990

aac ggg gtg gcc atg aat gtg cgc act ggg agc gcc cag gcc gat ggc     9024
Asn Gly Val Ala Met Asn Val Arg Thr Gly Ser Ala Gln Ala Asp Gly
       2995                3000                3005

tcc gac tct gag aaa ccg tga gg caagcccgtc accccacaca ggctgcggca     9077
Ser Asp Ser Glu Lys Pro  *
   3010                3015

tcaccctcag accttggagc ccaaggggcc actgcccttg aagtggagtg ggcccagagt   9137

gtggcggtcc ccatggtggc agccccccga ctgatcatcc agacacaaag gtcttggttc   9197

tcccaggagc tcagggcctg tcagacctgg tgacaagtgc caaaggccac aggcatgagg   9257

gaggcgtgga ccactgggcc agcaccgctg agtcctaaga ctgcagtcaa agccagaact   9317

gagaggggac cccagactgg gcccagaggc tggccagagt tcaggaacgc cgggcacaga   9377

ccaaagaccg cggtccagcc ccgcccaggc gggcatctca tggcagtgcg gacccgtggc   9437

tggcagcccg ggcagtcctt tgcaaaggca ccccttgtct taaaatcact tcgctatgtg   9497

ggaaaggtgg agatactttt atatatttgt atgggactct gaggaggtgc aacctgtata   9557

tatattgcat tcgtgctgac tttgttatcc cgagagatcc atgcaatgat ctcttgctgt   9617

cttctctgtc aagattgcac agttgtactt gaatctggca tgtgttgacg aaactggtgc   9677

cccagcagat caaaggtggg aaatacgtca gcagtggggc taaaaccaag cggctagaag   9737

ccctacagct gccttcggcc aggaagtgag gatggtgtgg gccctccccg ccggcccccc   9797

gggtccccag tgttcgctgt gtgtgcgttt gtcctctgct gccatctgcc ccggctgtgt   9857

gaattcaaga cagggcagtg cagcactagg caggtgtgag gagccctgct gaggtcactg   9917

tggggcacgg ttgccacacg gctgtcattt ttcacctggt cattctgtga ccaccacccc   9977

ctcccctcac cgcctcccag gtggcccggg agctgcaggt ggggatggct ttgtcctttg  10037

ctcctgctcc ccgtgggacc tgggacctta aagcgttgca ggttcctgat ttggacagag  10097

gtgtggggcc ttccaggccg ttacatacct cctgccaatt ctctaactct ctgagactgc  10157

gaggatctcc aggcagggtt ctcccctctg gagtctgacc aattacttca ttttgcttca  10217

aatggccaat tgtgcagagg gacaaagcca cagccacact cttcaacggt taccaaactg  10277

tttttggaaa ttcacaccaa ggtcgggccc actgcaggca gctggcacag cgtggcccga  10337

ggggctgtgg aacgggtccc ggaactgtca gacatgtttg attttagcgt ttcctttgtt  10397

cttcaaatca ggtgcccaaa taagtgatca gcacagctgc ttccaaatag gagaaaccat  10457

aaaataggat gaaaatcaag taaaatgcaa agatgtccac actgttttaa acttgaccct  10517

gatgaaaatg tgagcactgt tagcagatgc ctatgggaga ggaaaagcgt atctgaaaat  10577

ggtccaggac aggaggatga aatgagatcc cagagtcctc acacctgaat gaattataca  10637

tgtgccttac caggtgagtg gtctttcgaa gataaaaaac tctagtccct ttaaacgttt  10697

gcccctggcg tttcctaagt acgaaaaggt ttttaagtct tcgaacagtc tcctttcatg  10757

actttaacag gattctgccc cctgaggtgt aatttttttg ttctattttt ttccacgtac  10817

tccacagcca acatcacgag gtgtaatttt taatttgatc agaactgtta ccaaaaaaca  10877

actgtcagtt ttattgagat gggaaaaatg taaacctatt tttattactt aagactttat  10937
```

-continued

```
gggagagatt agacactgga ggtttttaac agaacgtgta tttattaatg ttcaaaacac  10997

tggaattaca aatgagaaga gtctacaata aattaagatt tttgaatttg tacttctgcg  11057

gtgctggttt ttctccacaa acacccccgc ccctccccat gcccagggtg gccgtggaag  11117

ggacggttta cggacgtgca gctgagctgt ccgtgtccca tgctccctca gccagtggaa  11177

cgtgccggaa ctttttgtcc attccctagt aggcctgcca cagcctagat gggcagtttt  11237

tgtctttcac caaatttgag gacttttttt ttttgccatt atttcttcag ttttcttttc  11297

ttgcactgat ctttctcctc tccttctgtg actccagtga ctcagacgtt agacctcttg  11357

atgttttccc actggtccct gaggctctgt tc                               11389


<210> SEQ ID NO 5
<211> LENGTH: 4087
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (113)..(4087)

<400> SEQUENCE: 5

aattcccggg tcgacgattt cgtggcggcg cccagttggg gcgggttcgt tcgcttcgcg     60

ttttggccag ggcgggggtc tgggctttag gcaggtagta tttagtttca ca    atg     115
                                                             Met
                                                               1

ttt ggg gac ctg ttt gaa gag gag tat tcc act gtg tct aat aat cag      163
Phe Gly Asp Leu Phe Glu Glu Glu Tyr Ser Thr Val Ser Asn Asn Gln
              5                  10                  15

tat gga aaa ggg aag aaa tta aag act aaa gct ttg gag cca cct gct      211
Tyr Gly Lys Gly Lys Lys Leu Lys Thr Lys Ala Leu Glu Pro Pro Ala
         20                  25                  30

cct aga gaa ttc acc aat tta agc gga atc aga aat cag ggt gga acc      259
Pro Arg Glu Phe Thr Asn Leu Ser Gly Ile Arg Asn Gln Gly Gly Thr
     35                  40                  45

tgt tac ctc aat tcc ctt ctt cag act ctt cat ttc aca cct gaa ttc      307
Cys Tyr Leu Asn Ser Leu Leu Gln Thr Leu His Phe Thr Pro Glu Phe
 50                  55                  60                  65

aga gaa gct cta ttt tct ctt ggc cca gaa gag ctt ggt ttg ttt gaa      355
Arg Glu Ala Leu Phe Ser Leu Gly Pro Glu Glu Leu Gly Leu Phe Glu
                 70                  75                  80

gat aag gat aaa ccc gat gca aag gtt cga atc atc cct tta cag tta      403
Asp Lys Asp Lys Pro Asp Ala Lys Val Arg Ile Ile Pro Leu Gln Leu
             85                  90                  95

cag cgc ttg ttt gct cag ctt ctg ctc tta gac cag gaa gct gca tcc      451
Gln Arg Leu Phe Ala Gln Leu Leu Leu Leu Asp Gln Glu Ala Ala Ser
        100                 105                 110

aca gca gac ctc act gac agc ttt ggg tgg acc agt aat gag gaa atg      499
Thr Ala Asp Leu Thr Asp Ser Phe Gly Trp Thr Ser Asn Glu Glu Met
    115                 120                 125

agg caa cat gat gtg cag gaa ctg aat cga atc ctc ttc agc gct ttg      547
Arg Gln His Asp Val Gln Glu Leu Asn Arg Ile Leu Phe Ser Ala Leu
130                 135                 140                 145

gaa act tct tta gtt ggg acc tcc ggt cat gac ctc atc tat cgt ctg      595
Glu Thr Ser Leu Val Gly Thr Ser Gly His Asp Leu Ile Tyr Arg Leu
                150                 155                 160

tac cat gga acc att gtt aac cag att gtt tgt aaa gaa tgt aag aac      643
Tyr His Gly Thr Ile Val Asn Gln Ile Val Cys Lys Glu Cys Lys Asn
```

-continued

```
           165                 170                 175

gtt agc gag agg cag gaa gac ttc tta gat cta aca gta gca gtc aaa      691
Val Ser Glu Arg Gln Glu Asp Phe Leu Asp Leu Thr Val Ala Val Lys
        180                 185                 190

aat gta tcc ggt ttg gaa gat gct ctc tgg aac atg tat gta gaa gag      739
Asn Val Ser Gly Leu Glu Asp Ala Leu Trp Asn Met Tyr Val Glu Glu
    195                 200                 205

gaa gtt ttt gat tgt gac aac ttg tac cac tgt gga act tgt gac agg      787
Glu Val Phe Asp Cys Asp Asn Leu Tyr His Cys Gly Thr Cys Asp Arg
210                 215                 220                 225

ctg gtt aaa gca gca aag tcg gcc aaa tta cgt aag ctg cct cct ttt      835
Leu Val Lys Ala Ala Lys Ser Ala Lys Leu Arg Lys Leu Pro Pro Phe
                230                 235                 240

ctt act gtt tca tta cta aga ttt aat ttt gat ttt gtg aaa tgc gaa      883
Leu Thr Val Ser Leu Leu Arg Phe Asn Phe Asp Phe Val Lys Cys Glu
            245                 250                 255

cgc tac aag gaa act agc tgt tat aca ttc cct ctc cgg att aat ctc      931
Arg Tyr Lys Glu Thr Ser Cys Tyr Thr Phe Pro Leu Arg Ile Asn Leu
        260                 265                 270

aag ccc ttt tgt gaa cag agt gaa ttg gat gac tta gaa tat ata tat      979
Lys Pro Phe Cys Glu Gln Ser Glu Leu Asp Asp Leu Glu Tyr Ile Tyr
    275                 280                 285

gac ctc ttc tca gtt att ata cac aaa ggt ggc tgc tac gga ggc cat     1027
Asp Leu Phe Ser Val Ile Ile His Lys Gly Gly Cys Tyr Gly Gly His
290                 295                 300                 305

tac cat gta tat att aaa gat gtt gat cat ttg gga aac tgg cag ttt     1075
Tyr His Val Tyr Ile Lys Asp Val Asp His Leu Gly Asn Trp Gln Phe
                310                 315                 320

caa gag gaa aaa agt aaa cca gat gtg aat ctg aaa gat ctc cag agt     1123
Gln Glu Glu Lys Ser Lys Pro Asp Val Asn Leu Lys Asp Leu Gln Ser
            325                 330                 335

gaa gaa gag att gat cat cca ctg atg att cta aaa gca atc tta tta     1171
Glu Glu Glu Ile Asp His Pro Leu Met Ile Leu Lys Ala Ile Leu Leu
        340                 345                 350

gag gag aat aat cta att cct gtt gat cag ctg ggc cag aaa ctt ttg     1219
Glu Glu Asn Asn Leu Ile Pro Val Asp Gln Leu Gly Gln Lys Leu Leu
    355                 360                 365

aaa aag ata gga ata tct tgg aac aag aag tac aga aaa cag cat gga     1267
Lys Lys Ile Gly Ile Ser Trp Asn Lys Lys Tyr Arg Lys Gln His Gly
370                 375                 380                 385

cca ttg cgg aag ttc tta cag ctc cat tct cag ata ttt cta ctc agt     1315
Pro Leu Arg Lys Phe Leu Gln Leu His Ser Gln Ile Phe Leu Leu Ser
                390                 395                 400

tca gat gaa agt aca gtt cgt ctc ttg aag aat agt tct ctc cag gct     1363
Ser Asp Glu Ser Thr Val Arg Leu Leu Lys Asn Ser Ser Leu Gln Ala
            405                 410                 415

gag tct gat ttc caa agg aat gac cag caa att ttc aag atg ctt cct     1411
Glu Ser Asp Phe Gln Arg Asn Asp Gln Gln Ile Phe Lys Met Leu Pro
        420                 425                 430

cca gaa tcc cca ggt tta aac aat agc atc tcc tgt ccc cac tgg ttt     1459
Pro Glu Ser Pro Gly Leu Asn Asn Ser Ile Ser Cys Pro His Trp Phe
    435                 440                 445

gat ata aat gat tct aaa gtc cag cca atc agg gaa aag gat att gaa     1507
Asp Ile Asn Asp Ser Lys Val Gln Pro Ile Arg Glu Lys Asp Ile Glu
450                 455                 460                 465

cag caa ttt cag ggt aaa gaa agt gcc tac atg ttg ttt tat cgg aaa     1555
Gln Gln Phe Gln Gly Lys Glu Ser Ala Tyr Met Leu Phe Tyr Arg Lys
```

-continued

```
               470                 475                 480

tcc cag ttg cag aga ccc cct gaa gct cga gct aat cca aga tat ggg     1603
Ser Gln Leu Gln Arg Pro Pro Glu Ala Arg Ala Asn Pro Arg Tyr Gly
            485                 490                 495

gtt cca tgt cat tta ctg aat gaa atg gat gca gct aac att gaa ctg     1651
Val Pro Cys His Leu Leu Asn Glu Met Asp Ala Ala Asn Ile Glu Leu
        500                 505                 510

caa acc aaa agg gca gaa tgt gat tct gca aac aat act ttt gaa ttg     1699
Gln Thr Lys Arg Ala Glu Cys Asp Ser Ala Asn Asn Thr Phe Glu Leu
    515                 520                 525

cat ctt cac ctg ggc cct cag tat cat ttc ttc aat ggg gct ctg cac     1747
His Leu His Leu Gly Pro Gln Tyr His Phe Phe Asn Gly Ala Leu His
530                 535                 540                 545

cca gta gtc tct caa aca gaa agc gtg tgg gat ttg acc ttt gat aaa     1795
Pro Val Val Ser Gln Thr Glu Ser Val Trp Asp Leu Thr Phe Asp Lys
                550                 555                 560

aga aaa act tta gga gat ctc cgg cag tca ata ttt cag ctg tta gaa     1843
Arg Lys Thr Leu Gly Asp Leu Arg Gln Ser Ile Phe Gln Leu Leu Glu
            565                 570                 575

ttt tgg gaa gga gac atg gtt ctt agt gtt gca aag ctt gta cca gca     1891
Phe Trp Glu Gly Asp Met Val Leu Ser Val Ala Lys Leu Val Pro Ala
        580                 585                 590

gga ctt cac att tac cag tca ctt ggc ggg gat gaa ctg aca ctg tgt     1939
Gly Leu His Ile Tyr Gln Ser Leu Gly Gly Asp Glu Leu Thr Leu Cys
    595                 600                 605

gaa act gaa att gct gat ggg gaa gac atc ttt gtg tgg aat ggg gtg     1987
Glu Thr Glu Ile Ala Asp Gly Glu Asp Ile Phe Val Trp Asn Gly Val
610                 615                 620                 625

gag gtt ggt gga gtc cac att caa act ggt att gac tgc gaa cct cta     2035
Glu Val Gly Gly Val His Ile Gln Thr Gly Ile Asp Cys Glu Pro Leu
                630                 635                 640

ctt tta aat gtt ctt cat cta gac aca agc agt gat gga gaa aag tgt     2083
Leu Leu Asn Val Leu His Leu Asp Thr Ser Ser Asp Gly Glu Lys Cys
            645                 650                 655

tgt cag gtg ata gaa tct cca cat gtc ttt cca gct aat gca gaa gtg     2131
Cys Gln Val Ile Glu Ser Pro His Val Phe Pro Ala Asn Ala Glu Val
        660                 665                 670

ggc act gtc ctc aca gcc tta gca atc cca gca ggt gtc atc ttc atc     2179
Gly Thr Val Leu Thr Ala Leu Ala Ile Pro Ala Gly Val Ile Phe Ile
    675                 680                 685

aac agt gct gga tgt cca ggt ggg gag ggt tgg acg gcc atc ccc aag     2227
Asn Ser Ala Gly Cys Pro Gly Gly Glu Gly Trp Thr Ala Ile Pro Lys
690                 695                 700                 705

gaa gac atg agg aag acg ttc agg gag caa ggg ctc aga aat gga agc     2275
Glu Asp Met Arg Lys Thr Phe Arg Glu Gln Gly Leu Arg Asn Gly Ser
                710                 715                 720

tca att tta att cag gat tct cat gat gat aac agc ttg ttg acc aag     2323
Ser Ile Leu Ile Gln Asp Ser His Asp Asp Asn Ser Leu Leu Thr Lys
            725                 730                 735

gaa gag aaa tgg gtc act agt atg aat gag att gac tgg ctc cac gtt     2371
Glu Glu Lys Trp Val Thr Ser Met Asn Glu Ile Asp Trp Leu His Val
        740                 745                 750

aaa aat tta tgc cag tta gaa tct gaa gag aag caa gtt aaa ata tca     2419
Lys Asn Leu Cys Gln Leu Glu Ser Glu Glu Lys Gln Val Lys Ile Ser
    755                 760                 765

gca act gtt aac aca atg gtg ttt gat att cga att aaa gcc ata aag     2467
Ala Thr Val Asn Thr Met Val Phe Asp Ile Arg Ile Lys Ala Ile Lys
```

-continued

```
770                 775                 780                 785

gaa tta aaa tta atg aag gaa cta gct gac aac agc tgt ttg aga cct     2515
Glu Leu Lys Leu Met Lys Glu Leu Ala Asp Asn Ser Cys Leu Arg Pro
                790                 795                 800

att gat aga aat ggg aag ctt ctt tgt cca gtg ccg gac agc tat act     2563
Ile Asp Arg Asn Gly Lys Leu Leu Cys Pro Val Pro Asp Ser Tyr Thr
            805                 810                 815

ttg aag gaa gca gaa ttg aag atg gga agt tca ttg gga ctg tgt ctt     2611
Leu Lys Glu Ala Glu Leu Lys Met Gly Ser Ser Leu Gly Leu Cys Leu
        820                 825                 830

gga aaa gca cca agt tcg tct cag ttg ttc ctg ttt ttt gca atg ggg     2659
Gly Lys Ala Pro Ser Ser Ser Gln Leu Phe Leu Phe Phe Ala Met Gly
    835                 840                 845

agt gac gtt caa cct ggg aca gaa atg gaa atc gta gta gaa gaa aca     2707
Ser Asp Val Gln Pro Gly Thr Glu Met Glu Ile Val Val Glu Glu Thr
850                 855                 860                 865

ata tct gtg aga gat tgt tta aag tta atg ctg aag aaa tct ggc cta     2755
Ile Ser Val Arg Asp Cys Leu Lys Leu Met Leu Lys Lys Ser Gly Leu
                870                 875                 880

caa gga gat gcc tgg cat tta cga aaa atg gat tgg tgc tat gaa gct     2803
Gln Gly Asp Ala Trp His Leu Arg Lys Met Asp Trp Cys Tyr Glu Ala
            885                 890                 895

gga gag cct tta tgt gaa gaa gat gca aca ctg aaa gaa ctt ctg ata     2851
Gly Glu Pro Leu Cys Glu Glu Asp Ala Thr Leu Lys Glu Leu Leu Ile
        900                 905                 910

tgt tct gga gat act ttg ctt tta att gaa gga caa ctt cct cct ctg     2899
Cys Ser Gly Asp Thr Leu Leu Leu Ile Glu Gly Gln Leu Pro Pro Leu
    915                 920                 925

ggt ttc ctg aag gtg ccc atc tgg tgg tac cag ctt cag ggt ccc tca     2947
Gly Phe Leu Lys Val Pro Ile Trp Trp Tyr Gln Leu Gln Gly Pro Ser
930                 935                 940                 945

gga cac tgg gag agt cat cag gac cag acc aac tgt act tcg tct tgg     2995
Gly His Trp Glu Ser His Gln Asp Gln Thr Asn Cys Thr Ser Ser Trp
                950                 955                 960

ggc aga gtt tgg aga gcc act tcc agc caa ggt gct tct ggg aac gag     3043
Gly Arg Val Trp Arg Ala Thr Ser Ser Gln Gly Ala Ser Gly Asn Glu
            965                 970                 975

cct gcg caa gtt tct ctc ctc tac ttg gga gac ata gag atc tca gaa     3091
Pro Ala Gln Val Ser Leu Leu Tyr Leu Gly Asp Ile Glu Ile Ser Glu
        980                 985                 990

gat gcc acg ctg gcg gag ctg aag tct cag gcc atg acc ttg cct cct     3139
Asp Ala Thr Leu Ala Glu Leu Lys Ser Gln Ala Met Thr Leu Pro Pro
    995                1000                1005

 ttc ctg gag ttc ggt gtc ccg tcc cca gcc cac ctc aga gcc tgg acg    3187
 Phe Leu Glu Phe Gly Val Pro Ser Pro Ala His Leu Arg Ala Trp Thr
1010                1015                1020                1025

gtg gag agg aag cgc cca ggc agg ctt tta cga act gac cgg cag cca     3235
Val Glu Arg Lys Arg Pro Gly Arg Leu Leu Arg Thr Asp Arg Gln Pro
               1030                1035                1040

ctc agc ccc cag gac gtg ctg ctg agg aca cag gtg cgc atc cct ggt     3283
Leu Ser Pro Gln Asp Val Leu Leu Arg Thr Gln Val Arg Ile Pro Gly
           1045                1050                1055

gag agg acc tat gcc cct gcc ctg gac ctg gtg tgg aac gcg gcc cag     3331
Glu Arg Thr Tyr Ala Pro Ala Leu Asp Leu Val Trp Asn Ala Ala Gln
       1060                1065                1070

ggt ggg act gcc ggc tcc ctg agg cag aga gtt gcc gat ttc tat cgt     3379
Gly Gly Thr Ala Gly Ser Leu Arg Gln Arg Val Ala Asp Phe Tyr Arg
```

-continued

```
   1075                1080                1085

 ctt ccc gtg gag aag att gaa att gcc aaa tac ttt ccc gaa aag ttc    3427
 Leu Pro Val Glu Lys Ile Glu Ile Ala Lys Tyr Phe Pro Glu Lys Phe
1090                1095                1100                1105

gag tgg ctt ccg ata tct agc tgg aac caa caa ata acc aag agg aaa    3475
Glu Trp Leu Pro Ile Ser Ser Trp Asn Gln Gln Ile Thr Lys Arg Lys
               1110                1115                1120

aag aaa aaa aaa caa gat tat ttg caa ggg gca ccg tat tac ttg aaa    3523
Lys Lys Lys Lys Gln Asp Tyr Leu Gln Gly Ala Pro Tyr Tyr Leu Lys
           1125                1130                1135

gac gga gat act att ggt gtt aag ccc ttt aaa gtc ccc ctg ttt ggc    3571
Asp Gly Asp Thr Ile Gly Val Lys Pro Phe Lys Val Pro Leu Phe Gly
       1140                1145                1150

cac tcg cag tgg aag ctt ctg gag ctg aag gga aag aat ctc ctg att    3619
His Ser Gln Trp Lys Leu Leu Glu Leu Lys Gly Lys Asn Leu Leu Ile
   1155                1160                1165

 gac gac gat gat gat ttc agt aca atc aga gat gac act gga aaa gaa    3667
 Asp Asp Asp Asp Asp Phe Ser Thr Ile Arg Asp Asp Thr Gly Lys Glu
1170                1175                1180                1185

aag cag aaa caa cgg gcc ctg ggg aga agg aaa agt atc ttg gct gcc    3715
Lys Gln Lys Gln Arg Ala Leu Gly Arg Arg Lys Ser Ile Leu Ala Ala
               1190                1195                1200

cgt gtg cca ggg act gca tgc cct ctg cca ctg ggg gcg tct ggg gaa    3763
Arg Val Pro Gly Thr Ala Cys Pro Leu Pro Leu Gly Ala Ser Gly Glu
           1205                1210                1215

gag agg tcc cag gtg gag gtg gca ggg cag ctc tgc agc cct cag gtc    3811
Glu Arg Ser Gln Val Glu Val Ala Gly Gln Leu Cys Ser Pro Gln Val
       1220                1225                1230

ctg gtg ccc tca gag acc tgg gat atg gga gag ggg ctg cag cgc tgg    3859
Leu Val Pro Ser Glu Thr Trp Asp Met Gly Glu Gly Leu Gln Arg Trp
   1235                1240                1245

 ctg ctt cgg ggg ttt tcc tgt tta aag aaa atg cac aac gtg tgc gtg    3907
 Leu Leu Arg Gly Phe Ser Cys Leu Lys Lys Met His Asn Val Cys Val
1250                1255                1260                1265

aac cgc agg tat gga ggc agc ggc atg ccg ttg ctc cgc tgt ggg aga    3955
Asn Arg Arg Tyr Gly Gly Ser Gly Met Pro Leu Leu Arg Cys Gly Arg
               1270                1275                1280

cgg ttg act ttc cca gac agc agc tta gcg gaa agc agc atc ccg aag    4003
Arg Leu Thr Phe Pro Asp Ser Ser Leu Ala Glu Ser Ser Ile Pro Lys
           1285                1290                1295

acg gtg tcg cat gtc ctg agc ctg ctt gtc acc tgg gcc tca cgg ttt    4051
Thr Val Ser His Val Leu Ser Leu Leu Val Thr Trp Ala Ser Arg Phe
       1300                1305                1310

ctc gtc agc ccc gtc act gat ggg aag cag cac tga                    4087
Leu Val Ser Pro Val Thr Asp Gly Lys Gln His  *
   1315                1320                1325


<210> SEQ ID NO 6
<211> LENGTH: 4153
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (113)..(4153)

<400> SEQUENCE: 6

aattcccggg tcgacgattt cgtggcggcg cccagttggg gcgggttcgt tcgcttcgcg     60

ttttggccag ggcgggggtc tgggctttag gcaggtagta tttagtttca ca    atg     115
```

-continued

```
                                                                  Met
                                                                   1

ttt ggg gac ctg ttt gaa gag gag tat tcc act gtg tct aat aat cag      163
Phe Gly Asp Leu Phe Glu Glu Glu Tyr Ser Thr Val Ser Asn Asn Gln
             5                  10                  15

tat gga aaa ggg aag aaa tta aag act aaa gct ttg gag cca cct gct      211
Tyr Gly Lys Gly Lys Lys Leu Lys Thr Lys Ala Leu Glu Pro Pro Ala
         20                  25                  30

cct aga gaa ttc acc aat tta agc gga atc aga aat cag ggt gga acc      259
Pro Arg Glu Phe Thr Asn Leu Ser Gly Ile Arg Asn Gln Gly Gly Thr
     35                  40                  45

tgt tac ctc aat tcc ctt ctt cag act ctt cat ttc aca cct gaa ttc      307
Cys Tyr Leu Asn Ser Leu Leu Gln Thr Leu His Phe Thr Pro Glu Phe
 50                  55                  60                  65

aga gaa gct cta ttt tct ctt ggc cca gaa gag ctt ggt ttg ttt gaa      355
Arg Glu Ala Leu Phe Ser Leu Gly Pro Glu Glu Leu Gly Leu Phe Glu
                 70                  75                  80

gat aag gat aaa ccc gat gca aag gtt cga atc atc cct tta cag tta      403
Asp Lys Asp Lys Pro Asp Ala Lys Val Arg Ile Ile Pro Leu Gln Leu
             85                  90                  95

cag cgc ttg ttt gct cag ctt ctg ctc tta gac cag gaa gct gca tcc      451
Gln Arg Leu Phe Ala Gln Leu Leu Leu Leu Asp Gln Glu Ala Ala Ser
         100                 105                 110

aca gca gac ctc act gac agc ttt ggg tgg acc agt aat gag gaa atg      499
Thr Ala Asp Leu Thr Asp Ser Phe Gly Trp Thr Ser Asn Glu Glu Met
     115                 120                 125

agg caa cat gat gtg cag gaa ctg aat cga atc ctc ttc agc gct ttg      547
Arg Gln His Asp Val Gln Glu Leu Asn Arg Ile Leu Phe Ser Ala Leu
130                 135                 140                 145

gaa act tct tta gtt ggg acc tcc ggt cat gac ctc atc tat cgt ctg      595
Glu Thr Ser Leu Val Gly Thr Ser Gly His Asp Leu Ile Tyr Arg Leu
                150                 155                 160

tac cat gga acc att gtt aac cag att gtt tgt aaa gaa tgt aag aac      643
Tyr His Gly Thr Ile Val Asn Gln Ile Val Cys Lys Glu Cys Lys Asn
            165                 170                 175

gtt agc gag agg cag gaa gac ttc tta gat cta aca gta gca gtc aaa      691
Val Ser Glu Arg Gln Glu Asp Phe Leu Asp Leu Thr Val Ala Val Lys
        180                 185                 190

aat gta tcc ggt ttg gaa gat gct ctc tgg aac atg tat gta gaa gag      739
Asn Val Ser Gly Leu Glu Asp Ala Leu Trp Asn Met Tyr Val Glu Glu
    195                 200                 205

gaa gtt ttt gat tgt gac aac ttg tac cac tgt gga act tgt gac agg      787
Glu Val Phe Asp Cys Asp Asn Leu Tyr His Cys Gly Thr Cys Asp Arg
210                 215                 220                 225

ctg gtt aaa gca gca aag tcg gcc aaa tta cgt aag ctg cct cct ttt      835
Leu Val Lys Ala Ala Lys Ser Ala Lys Leu Arg Lys Leu Pro Pro Phe
                230                 235                 240

ctt act gtt tca tta cta aga ttt aat ttt gat ttt gtg aaa tgc gaa      883
Leu Thr Val Ser Leu Leu Arg Phe Asn Phe Asp Phe Val Lys Cys Glu
            245                 250                 255

cgc tac aag gaa act agc tgt tat aca ttc cct ctc cgg att aat ctc      931
Arg Tyr Lys Glu Thr Ser Cys Tyr Thr Phe Pro Leu Arg Ile Asn Leu
        260                 265                 270

aag ccc ttt tgt gaa cag agt gaa ttg gat gac tta gaa tat ata tat      979
Lys Pro Phe Cys Glu Gln Ser Glu Leu Asp Asp Leu Glu Tyr Ile Tyr
    275                 280                 285

gac ctc ttc tca gtt att ata cac aaa ggt ggc tgc tac gga ggc cat     1027
```

-continued

```
Asp Leu Phe Ser Val Ile Ile His Lys Gly Gly Cys Tyr Gly Gly His
290                 295                 300                 305

tac cat gta tat att aaa gat gtt gat cat ttg gga aac tgg cag ttt     1075
Tyr His Val Tyr Ile Lys Asp Val Asp His Leu Gly Asn Trp Gln Phe
                310                 315                 320

caa gag gaa aaa agt aaa cca gat gtg aat ctg aaa gat ctc cag agt     1123
Gln Glu Glu Lys Ser Lys Pro Asp Val Asn Leu Lys Asp Leu Gln Ser
            325                 330                 335

gaa gaa gag att gat cat cca ctg atg att cta aaa gca atc tta tta     1171
Glu Glu Glu Ile Asp His Pro Leu Met Ile Leu Lys Ala Ile Leu Leu
        340                 345                 350

gag gag aat aat cta att cct gtt gat cag ctg ggc cag aaa ctt ttg     1219
Glu Glu Asn Asn Leu Ile Pro Val Asp Gln Leu Gly Gln Lys Leu Leu
    355                 360                 365

aaa aag ata gga ata tct tgg aac aag aag tac aga aaa cag cat gga     1267
Lys Lys Ile Gly Ile Ser Trp Asn Lys Lys Tyr Arg Lys Gln His Gly
370                 375                 380                 385

cca ttg cgg aag ttc tta cag ctc cat tct cag ata ttt cta ctc agt     1315
Pro Leu Arg Lys Phe Leu Gln Leu His Ser Gln Ile Phe Leu Leu Ser
                390                 395                 400

tca gat gaa agt aca gtt cgt ctc ttg aag aat agt tct ctc cag gct     1363
Ser Asp Glu Ser Thr Val Arg Leu Leu Lys Asn Ser Ser Leu Gln Ala
            405                 410                 415

gag tct gat ttc caa agg aat gac cag caa att ttc aag atg ctt cct     1411
Glu Ser Asp Phe Gln Arg Asn Asp Gln Gln Ile Phe Lys Met Leu Pro
        420                 425                 430

cca gaa tcc cca ggt tta aac aat agc atc tcc tgt ccc cac tgg ttt     1459
Pro Glu Ser Pro Gly Leu Asn Asn Ser Ile Ser Cys Pro His Trp Phe
    435                 440                 445

gat ata aat gat tct aaa gtc cag cca atc agg gaa aag gat att gaa     1507
Asp Ile Asn Asp Ser Lys Val Gln Pro Ile Arg Glu Lys Asp Ile Glu
450                 455                 460                 465

cag caa ttt cag ggt aaa gaa agt gcc tac atg ttg ttt tat cgg aaa     1555
Gln Gln Phe Gln Gly Lys Glu Ser Ala Tyr Met Leu Phe Tyr Arg Lys
                470                 475                 480

tcc cag ttg cag aga ccc cct gaa gct cga gct aat cca aga tat ggg     1603
Ser Gln Leu Gln Arg Pro Pro Glu Ala Arg Ala Asn Pro Arg Tyr Gly
            485                 490                 495

gtt cca tgt cat tta ctg aat gaa atg gat gca gct aac att gaa ctg     1651
Val Pro Cys His Leu Leu Asn Glu Met Asp Ala Ala Asn Ile Glu Leu
        500                 505                 510

caa acc aaa agg gca gaa tgt gat tct gca aac aat act ttt gaa ttg     1699
Gln Thr Lys Arg Ala Glu Cys Asp Ser Ala Asn Asn Thr Phe Glu Leu
    515                 520                 525

cat ctt cac ctg ggc cct cag tat cat ttc ttc aat ggg gct ctg cac     1747
His Leu His Leu Gly Pro Gln Tyr His Phe Phe Asn Gly Ala Leu His
530                 535                 540                 545

cca gta gtc tct caa aca gaa agc gtg tgg gat ttg acc ttt gat aaa     1795
Pro Val Val Ser Gln Thr Glu Ser Val Trp Asp Leu Thr Phe Asp Lys
                550                 555                 560

aga aaa act tta gga gat ctc cgg cag tca ata ttt cag ctg tta gaa     1843
Arg Lys Thr Leu Gly Asp Leu Arg Gln Ser Ile Phe Gln Leu Leu Glu
            565                 570                 575

ttt tgg gaa gga gac atg gtt ctt agt gtt gca aag ctt gta cca gca     1891
Phe Trp Glu Gly Asp Met Val Leu Ser Val Ala Lys Leu Val Pro Ala
        580                 585                 590

gga ctt cac att tac cag tca ctt ggc ggg gat gaa ctg aca ctg tgt     1939
```

-continued

```
Gly Leu His Ile Tyr Gln Ser Leu Gly Gly Asp Glu Leu Thr Leu Cys
    595                 600                 605

gaa act gaa att gct gat ggg gaa gac atc ttt gtg tgg aat ggg gtg     1987
Glu Thr Glu Ile Ala Asp Gly Glu Asp Ile Phe Val Trp Asn Gly Val
610                 615                 620                 625

gag gtt ggt gga gtc cac att caa act ggt att gac tgc gaa cct cta     2035
Glu Val Gly Gly Val His Ile Gln Thr Gly Ile Asp Cys Glu Pro Leu
                630                 635                 640

ctt tta aat gtt ctt cat cta gac aca agc agt gat gga gaa aag tgt     2083
Leu Leu Asn Val Leu His Leu Asp Thr Ser Ser Asp Gly Glu Lys Cys
            645                 650                 655

tgt cag gtg ata gaa tct cca cat gtc ttt cca gct aat gca gaa gtg     2131
Cys Gln Val Ile Glu Ser Pro His Val Phe Pro Ala Asn Ala Glu Val
        660                 665                 670

ggc act gtc ctc aca gcc tta gca atc cca gca ggt gtc atc ttc atc     2179
Gly Thr Val Leu Thr Ala Leu Ala Ile Pro Ala Gly Val Ile Phe Ile
    675                 680                 685

aac agt gct gga tgt cca ggt ggg gag ggt tgg acg gcc atc ccc aag     2227
Asn Ser Ala Gly Cys Pro Gly Gly Glu Gly Trp Thr Ala Ile Pro Lys
690                 695                 700                 705

gaa gac atg agg aag acg ttc agg gag caa ggg ctc aga aat gga agc     2275
Glu Asp Met Arg Lys Thr Phe Arg Glu Gln Gly Leu Arg Asn Gly Ser
                710                 715                 720

tca att tta att cag gat tct cat gat gat aac agc ttg ttg acc aag     2323
Ser Ile Leu Ile Gln Asp Ser His Asp Asp Asn Ser Leu Leu Thr Lys
            725                 730                 735

gaa gag aaa tgg gtc act agt atg aat gag att gac tgg ctc cac gtt     2371
Glu Glu Lys Trp Val Thr Ser Met Asn Glu Ile Asp Trp Leu His Val
        740                 745                 750

aaa aat tta tgc cag tta gaa tct gaa gag aag caa gtt aaa ata tca     2419
Lys Asn Leu Cys Gln Leu Glu Ser Glu Glu Lys Gln Val Lys Ile Ser
    755                 760                 765

gca act gtt aac aca atg gtg ttt gat att cga att aaa gcc ata aag     2467
Ala Thr Val Asn Thr Met Val Phe Asp Ile Arg Ile Lys Ala Ile Lys
770                 775                 780                 785

gaa tta aaa tta atg aag gaa cta gct gac aac agc tgt ttg aga cct     2515
Glu Leu Lys Leu Met Lys Glu Leu Ala Asp Asn Ser Cys Leu Arg Pro
                790                 795                 800

att gat aga aat ggg aag ctt ctt tgt cca gtg ccg gac agc tat act     2563
Ile Asp Arg Asn Gly Lys Leu Leu Cys Pro Val Pro Asp Ser Tyr Thr
            805                 810                 815

ttg aag gaa gca gaa ttg aag atg gga agt tca ttg gga ctg tgt ctt     2611
Leu Lys Glu Ala Glu Leu Lys Met Gly Ser Ser Leu Gly Leu Cys Leu
        820                 825                 830

gga aaa gca cca agt tcg tct cag ttg ttc ctg ttt ttt gca atg ggg     2659
Gly Lys Ala Pro Ser Ser Ser Gln Leu Phe Leu Phe Phe Ala Met Gly
    835                 840                 845

agt gac gtt caa cct ggg aca gaa atg gaa atc gta gta gaa gaa aca     2707
Ser Asp Val Gln Pro Gly Thr Glu Met Glu Ile Val Val Glu Glu Thr
850                 855                 860                 865

ata tct gtg aga gat tgt tta aag tta atg ctg aag aaa tct ggc cta     2755
Ile Ser Val Arg Asp Cys Leu Lys Leu Met Leu Lys Lys Ser Gly Leu
                870                 875                 880

caa gga gat gcc tgg cat tta cga aaa atg gat tgg tgc tat gaa gct     2803
Gln Gly Asp Ala Trp His Leu Arg Lys Met Asp Trp Cys Tyr Glu Ala
            885                 890                 895

gga gag cct tta tgt gaa gaa gat gca aca ctg aaa gaa ctt ctg ata     2851
```

-continued

```
Gly Glu Pro Leu Cys Glu Glu Asp Ala Thr Leu Lys Glu Leu Leu Ile
        900                 905                 910

tgt tct gga gat act ttg ctt tta att gaa gga caa ctt cct cct ctg     2899
Cys Ser Gly Asp Thr Leu Leu Leu Ile Glu Gly Gln Leu Pro Pro Leu
    915                 920                 925

ggt ttc ctg caa gtg ccc atc tgg tgg tac cag ctt cag ggt ccc tca     2947
Gly Phe Leu Gln Val Pro Ile Trp Trp Tyr Gln Leu Gln Gly Pro Ser
930                 935                 940                 945

gga cac tgg gag agt cat cag gac cag acc aac tgt act tcg tct tgg     2995
Gly His Trp Glu Ser His Gln Asp Gln Thr Asn Cys Thr Ser Ser Trp
                950                 955                 960

ggc aga gtt tgg aga gcc act tcc agc caa ggt gct tct ggg aac gag     3043
Gly Arg Val Trp Arg Ala Thr Ser Ser Gln Gly Ala Ser Gly Asn Glu
            965                 970                 975

cct gcg caa gtt tct ctc ctc tac ttg gga gac ata gag atc tca gaa     3091
Pro Ala Gln Val Ser Leu Leu Tyr Leu Gly Asp Ile Glu Ile Ser Glu
        980                 985                 990

gat gcc acg ctg gcg gag ctg aag tct cag gcc atg acc ttg cct cct     3139
Asp Ala Thr Leu Ala Glu Leu Lys Ser Gln Ala Met Thr Leu Pro Pro
    995                1000                1005

 ttc ctg gag ttc ggt gtc ccg tcc cca gcc cac ctc aga gcc tgg acg    3187
 Phe Leu Glu Phe Gly Val Pro Ser Pro Ala His Leu Arg Ala Trp Thr
1010                1015                1020                1025

gtg gag agg aag cgc cca ggc agg ctt tta cga act gac cgg cag cca     3235
Val Glu Arg Lys Arg Pro Gly Arg Leu Leu Arg Thr Asp Arg Gln Pro
               1030                1035                1040

ctc agg gaa tat aaa cta gga cgg aga att gag atc tgc tta gag ccc     3283
Leu Arg Glu Tyr Lys Leu Gly Arg Arg Ile Glu Ile Cys Leu Glu Pro
           1045                1050                1055

ctt cag aaa ggc gaa aac ttg ggc ccc cag gac gtg ctg ctg agg aca     3331
Leu Gln Lys Gly Glu Asn Leu Gly Pro Gln Asp Val Leu Leu Arg Thr
       1060                1065                1070

cag gtg cgc atc cct ggt gag agg acc tat gcc cct gcc ctg gac ctg     3379
Gln Val Arg Ile Pro Gly Glu Arg Thr Tyr Ala Pro Ala Leu Asp Leu
   1075                1080                1085

 gtg tgg aac gcg gcc cag ggt ggg act gcc ggc tcc ctg agg cag aga    3427
 Val Trp Asn Ala Ala Gln Gly Gly Thr Ala Gly Ser Leu Arg Gln Arg
1090                1095                1100                1105

gtt gcc gat ttc tat cgt ctt ccc gtg gag aag att gaa att gcc aaa     3475
Val Ala Asp Phe Tyr Arg Leu Pro Val Glu Lys Ile Glu Ile Ala Lys
               1110                1115                1120

tac ttt ccc gaa aag ttc gag tgg ctt ccg ata tct agc tgg aac caa     3523
Tyr Phe Pro Glu Lys Phe Glu Trp Leu Pro Ile Ser Ser Trp Asn Gln
           1125                1130                1135

caa ata acc aag agg aaa aag aaa aaa aaa caa gat tat ttg caa ggg     3571
Gln Ile Thr Lys Arg Lys Lys Lys Lys Lys Gln Asp Tyr Leu Gln Gly
       1140                1145                1150

gca ccg tat tac ttg aaa gac gga gat act att ggt gtt aag ccc ttt     3619
Ala Pro Tyr Tyr Leu Lys Asp Gly Asp Thr Ile Gly Val Lys Pro Phe
   1155                1160                1165

 aaa gtc ccc ctg ttt ggc cac tcg cag tgg aag ctt ctg gag ctg aag    3667
 Lys Val Pro Leu Phe Gly His Ser Gln Trp Lys Leu Leu Glu Leu Lys
1170                1175                1180                1185

gga aag aat ctc ctg att gac gac gat gat gat ttc agt aca atc aga     3715
Gly Lys Asn Leu Leu Ile Asp Asp Asp Asp Asp Phe Ser Thr Ile Arg
               1190                1195                1200

gat gac act gga aaa gaa aag cag aaa caa cgg gcc ctg ggg aga agg     3763
```

-continued

```
Asp Asp Thr Gly Lys Glu Lys Gln Lys Gln Arg Ala Leu Gly Arg Arg
          1205                1210                1215

aaa agt atc ttg gct gcc cgt gtg cca ggg act gca tgc cct ctg cca     3811
Lys Ser Ile Leu Ala Ala Arg Val Pro Gly Thr Ala Cys Pro Leu Pro
       1220                1225                1230

ctg ggg gcg tct ggg gaa gag agg tcc cag gtg gag gtg gca ggg cag     3859
Leu Gly Ala Ser Gly Glu Glu Arg Ser Gln Val Glu Val Ala Gly Gln
   1235                1240                1245

 ctc tgc agc cct cag gtc ctg gtg ccc tca gag acc tgg gat atg gga    3907
 Leu Cys Ser Pro Gln Val Leu Val Pro Ser Glu Thr Trp Asp Met Gly
1250                1255                1260                1265

gag ggg ctg cag cgc tgg ctg ctt cgg ggg ttt tcc tgt tta aag aaa     3955
Glu Gly Leu Gln Arg Trp Leu Leu Arg Gly Phe Ser Cys Leu Lys Lys
              1270                1275                1280

atg cac aac gtg tgc gtg aac cgc agg tat gga ggc agc ggc atg ccg     4003
Met His Asn Val Cys Val Asn Arg Arg Tyr Gly Gly Ser Gly Met Pro
          1285                1290                1295

ttg ctc cgc tgt ggg aga cgg ttg act ttc cca gac agc agc tta gcg     4051
Leu Leu Arg Cys Gly Arg Arg Leu Thr Phe Pro Asp Ser Ser Leu Ala
       1300                1305                1310

gaa agc agc atc ccg aag acg gtg tcg cat gtc ctg agc ctg ctt gtc     4099
Glu Ser Ser Ile Pro Lys Thr Val Ser His Val Leu Ser Leu Leu Val
   1315                1320                1325

 acc tgg gcc tca cgg ttt ctc gtc agc ccc gtc act gat ggg aag cag    4147
 Thr Trp Ala Ser Arg Phe Leu Val Ser Pro Val Thr Asp Gly Lys Gln
1330                1335                1340                1345

cac tga                                                            4153
His  *


<210> SEQ ID NO 7
<211> LENGTH: 776
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (147)..(446)

<400> SEQUENCE: 7

cagcaggttg gctcacagca gaaggcaaag gccatcatca gctcccttta taagggaacg     60

gtcatgcact gggtgtgctg agagtgtcct gcctggtcct ctgtgcctgg tggggtggag    120

gtgccaggtg tgtccagagg agccca atg ggc agt gag gca gcc atg ggg ctg     173
                            Met Gly Ser Glu Ala Ala Met Gly Leu
                              1               5

gat gca ctg gtg ccc ctg gca gtg aca gtg gcc atc ttc ctg ctc ctg      221
Asp Ala Leu Val Pro Leu Ala Val Thr Val Ala Ile Phe Leu Leu Leu
 10                  15                  20                  25

gtg gac ctg atg cag cag cac caa cgc tgg act gca cgc tac ccg cca      269
Val Asp Leu Met Gln Gln His Gln Arg Trp Thr Ala Arg Tyr Pro Pro
                 30                  35                  40

ggc ccc ctg cca ctg ccc ggg ctg ggc aac ttg ctg cat gtg gac ttc      317
Gly Pro Leu Pro Leu Pro Gly Leu Gly Asn Leu Leu His Val Asp Phe
             45                  50                  55

cag aac ata tac acc ttc aac cag gtg agg gag gag gtc ctg agg atc      365
Gln Asn Ile Tyr Thr Phe Asn Gln Val Arg Glu Glu Val Leu Arg Ile
         60                  65                  70

ccc cac cac cag caa aca tgg gtg gtg ggt gga gcc aca gtc tgg aca      413
Pro His His Gln Gln Thr Trp Val Val Gly Gly Ala Thr Val Trp Thr
     75                  80                  85
```

-continued

```
aga agc cag gct gag aag ggg aag cag att tga gggacttc ctgggggagg      464
Arg Ser Gln Ala Glu Lys Gly Lys Gln Ile  *
 90                  95                 100

gcatttatgc atggcatgaa agatggcatt ttccaaaggc caaggaagag taaggcaagg    524

gcctggaggt ggaactgtac ttggcagtgg gcgtgcacgc ccattggcca tcatatgtta    584

aggaccacaa ggtccctctt gtgacaccag aatgaaaggg cctggcaaat ggaacaacca    644

gccaggggcc ctggttagcc cttttatatc ccacaattat accctggggc tctcttgtcc    704

ctacaccccc gtttccttct ctttgtttct ctacggaatc tcctgtcttg tcgctgtttt    764

ccttttttg gc                                                        776


<210> SEQ ID NO 8
<211> LENGTH: 1621
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (38)..(1414)

<400> SEQUENCE: 8

gatatgtcca agaaaccacc ccctgataac gtctttg atg aca tgc caa gct tgg      55
                                      Met Thr Cys Gln Ala Trp
                                        1               5

cac gag gca gag gag gga cac gga att act cac agc tgt gct gtg tgc      103
His Glu Ala Glu Glu Gly His Gly Ile Thr His Ser Cys Ala Val Cys
            10                  15                  20

att ctc tgt ggg cct agc agg gaa ggg gac agc cct gtg gca atg ggc      151
Ile Leu Cys Gly Pro Ser Arg Glu Gly Asp Ser Pro Val Ala Met Gly
         25                  30                  35

atg aca cgg atg ctc ctg gaa tgc agt ctc agt gac aag ttg tgt gtc      199
Met Thr Arg Met Leu Leu Glu Cys Ser Leu Ser Asp Lys Leu Cys Val
     40                  45                  50

atc cag gag aag cag tat gaa gtg att atc gtc cca act ttg ttg gtt      247
Ile Gln Glu Lys Gln Tyr Glu Val Ile Ile Val Pro Thr Leu Leu Val
 55                  60                  65                  70

act atc ttc ctc atc ctt ctt ggg gtc atc ctg tgg ctt ttt atc aga      295
Thr Ile Phe Leu Ile Leu Leu Gly Val Ile Leu Trp Leu Phe Ile Arg
                75                  80                  85

gaa caa aga act caa cag cag cgt tct gga cct caa ggc att gcc cct      343
Glu Gln Arg Thr Gln Gln Gln Arg Ser Gly Pro Gln Gly Ile Ala Pro
            90                  95                 100

gtt cct cca cct agg gac cta agc tgg gaa gca gga cat gga gga aat      391
Val Pro Pro Pro Arg Asp Leu Ser Trp Glu Ala Gly His Gly Gly Asn
        105                 110                 115

gtg gct ttg cca ctt aag gag aca tcc gtg gaa aac ttt ctg gga gct      439
Val Ala Leu Pro Leu Lys Glu Thr Ser Val Glu Asn Phe Leu Gly Ala
    120                 125                 130

acc aca cct gcc ctg gct aag ctg cag gtg ccg cgg gag caa ctc tct      487
Thr Thr Pro Ala Leu Ala Lys Leu Gln Val Pro Arg Glu Gln Leu Ser
135                 140                 145                 150

gaa gtt ctg gag cag att tgc agt ggt agc tgt ggg ccc atc ttt cga      535
Glu Val Leu Glu Gln Ile Cys Ser Gly Ser Cys Gly Pro Ile Phe Arg
                155                 160                 165

gcc aat atg aac act ggg gac cct tct aag ccc aag agt gtt att ctc      583
Ala Asn Met Asn Thr Gly Asp Pro Ser Lys Pro Lys Ser Val Ile Leu
            170                 175                 180
```

-continued

```
aag gct tta aaa gaa cca gct ggg ctc cat gag gta caa gat ttc tta      631
Lys Ala Leu Lys Glu Pro Ala Gly Leu His Glu Val Gln Asp Phe Leu
        185                 190                 195

ggg cga atc aaa ttc cat caa tac ctg ggg aaa cac aaa aac ctg gtg      679
Gly Arg Ile Lys Phe His Gln Tyr Leu Gly Lys His Lys Asn Leu Val
    200                 205                 210

cag ctg gaa ggc tgc tgc act gaa aag ctg cca ctc tat atg gtg ttg      727
Gln Leu Glu Gly Cys Cys Thr Glu Lys Leu Pro Leu Tyr Met Val Leu
215                 220                 225                 230

gag gat gtg gcc cag ggg gac ctg ctc ggc ttt ctc tgg acc tgt cgg      775
Glu Asp Val Ala Gln Gly Asp Leu Leu Gly Phe Leu Trp Thr Cys Arg
                235                 240                 245

cgg gat gtg atg act atg gat ggt ctt ctc tat gat ctc aca gaa aaa      823
Arg Asp Val Met Thr Met Asp Gly Leu Leu Tyr Asp Leu Thr Glu Lys
            250                 255                 260

caa gta tat cac atc gga aag cag gtc ctt ttg gcg ctg gaa ttc ctg      871
Gln Val Tyr His Ile Gly Lys Gln Val Leu Leu Ala Leu Glu Phe Leu
        265                 270                 275

cag gag aag cat ttg ttc cat ggg gat gtg gca gcc agg aat att ctg      919
Gln Glu Lys His Leu Phe His Gly Asp Val Ala Ala Arg Asn Ile Leu
    280                 285                 290

atg caa agt gat ctc act gct aag ctc tgt gga tta ggc ctg gct tat      967
Met Gln Ser Asp Leu Thr Ala Lys Leu Cys Gly Leu Gly Leu Ala Tyr
295                 300                 305                 310

gaa gtt tac acc cga ggg gcc atc tcc tct act caa acc ata cct ctc     1015
Glu Val Tyr Thr Arg Gly Ala Ile Ser Ser Thr Gln Thr Ile Pro Leu
                315                 320                 325

aag tgg ctt gcc cca gaa cgg ctt ctc ctg aga cct gct agc atc aga     1063
Lys Trp Leu Ala Pro Glu Arg Leu Leu Leu Arg Pro Ala Ser Ile Arg
            330                 335                 340

gca gat gtc tgg tct ttt ggg atc ctg ctc tat gag atg gtg act cta     1111
Ala Asp Val Trp Ser Phe Gly Ile Leu Leu Tyr Glu Met Val Thr Leu
        345                 350                 355

gga gca cca ccg tat cct gaa gtc cct cct acc agc atc cta gag cat     1159
Gly Ala Pro Pro Tyr Pro Glu Val Pro Pro Thr Ser Ile Leu Glu His
    360                 365                 370

ctc caa aga agg aaa atc atg aag aga ccc agt agc tgc aca cat acc     1207
Leu Gln Arg Arg Lys Ile Met Lys Arg Pro Ser Ser Cys Thr His Thr
375                 380                 385                 390

atg tac agt atc atg aag tcc tgc tgg cgc tgg cgt gag gct gac cgc     1255
Met Tyr Ser Ile Met Lys Ser Cys Trp Arg Trp Arg Glu Ala Asp Arg
                395                 400                 405

ccc tca cct aga gag ctg cgc ttg cgc cta gaa gct gcc att aaa act     1303
Pro Ser Pro Arg Glu Leu Arg Leu Arg Leu Glu Ala Ala Ile Lys Thr
            410                 415                 420

gca gat gac gag gct gtg tta caa gta cca gag ttg gtg gta cct gaa     1351
Ala Asp Asp Glu Ala Val Leu Gln Val Pro Glu Leu Val Val Pro Glu
        425                 430                 435

ctg tat gca gct gtg gcc ggc atc aga gtg gag agc ctc ttc tac aac     1399
Leu Tyr Ala Ala Val Ala Gly Ile Arg Val Glu Ser Leu Phe Tyr Asn
    440                 445                 450

tat agc atg ctt tga agagtctcgg gcaagaaaca ttcatgcatg agtatatgtt     1454
Tyr Ser Met Leu  *
455

cttggaatca attcctctaa gaacagagaa tggtctttcc cagggacaca aagggagaaa   1514

tgggacatgg attcttgatc ttcctttaca catttctcgg gaaatctgaa atgatgctgg   1574
```

-continued

```
atgggactct acacatcctg agctaagaca tactgtcagt ctcactt                1621


<210> SEQ ID NO 9
<211> LENGTH: 829
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (70)..(690)

<400> SEQUENCE: 9

gggctgggca gatgggtggg tgagttccct ctccccagag ccatcggcca ggtaccaaag     60

ctcagctgt    atg gat tcc caa cag gag gac ctg cgc ttc cct ggg atg     108
             Met Asp Ser Gln Gln Glu Asp Leu Arg Phe Pro Gly Met
               1               5                  10

tgg gtc tca ttg tac ttt gga atc ctg ggg ctg tgt tct gtg ata act     156
Trp Val Ser Leu Tyr Phe Gly Ile Leu Gly Leu Cys Ser Val Ile Thr
     15                  20                  25

gga ggg tgc att atc ttt ctg cac tgg agg aag aac ttg agg cgg gaa     204
Gly Gly Cys Ile Ile Phe Leu His Trp Arg Lys Asn Leu Arg Arg Glu
 30                  35                  40                  45

gag cat gcc cag cag tgg gtg gag gtg atg aga gct gcc aca ttc acc     252
Glu His Ala Gln Gln Trp Val Glu Val Met Arg Ala Ala Thr Phe Thr
                 50                  55                  60

tac agc cca ttg ttg tac tgg att aac aag cga cgg cgc tac ggc atg     300
Tyr Ser Pro Leu Leu Tyr Trp Ile Asn Lys Arg Arg Arg Tyr Gly Met
             65                  70                  75

aat gca gcc atc aac acg ggc cct gcc cct gct gtc acc aag act gag     348
Asn Ala Ala Ile Asn Thr Gly Pro Ala Pro Ala Val Thr Lys Thr Glu
         80                  85                  90

act gag gtc cag aat cca gat gtt ctg tgg gat ttg gac atc ccc gaa     396
Thr Glu Val Gln Asn Pro Asp Val Leu Trp Asp Leu Asp Ile Pro Glu
     95                  100                 105

ggc agg agc cat gct gac caa gac agc aac ccc aag gcg gaa gcc cct     444
Gly Arg Ser His Ala Asp Gln Asp Ser Asn Pro Lys Ala Glu Ala Pro
110                 115                 120                 125

gct ccc ctg caa cct gca ctg cag ctg gct cca cag cag ccc cag gcc     492
Ala Pro Leu Gln Pro Ala Leu Gln Leu Ala Pro Gln Gln Pro Gln Ala
                130                 135                 140

aga tcc cca ttc cca ctt ccc atc ttt cag gag gtg ccc ttt gcc cca     540
Arg Ser Pro Phe Pro Leu Pro Ile Phe Gln Glu Val Pro Phe Ala Pro
            145                 150                 155

ccc ttg tgc aac cta ccc ccc ctg ctg aac cac tct gtc tcc tat cct     588
Pro Leu Cys Asn Leu Pro Pro Leu Leu Asn His Ser Val Ser Tyr Pro
        160                 165                 170

ttg gcc acc tgt cct gaa agg aat gtt ctc ttc cat tcc ctc ctg aat     636
Leu Ala Thr Cys Pro Glu Arg Asn Val Leu Phe His Ser Leu Leu Asn
    175                 180                 185

ctg gcc cag gaa gac cat agc ttc aat gcc aag cct ttt cct tca gaa     684
Leu Ala Gln Glu Asp His Ser Phe Asn Ala Lys Pro Phe Pro Ser Glu
190                 195                 200                 205

ctg tag cctcctctca ctgaaggtgg gagctgcagg aatcaggtgc agagtaggaa      740
Leu  *

atggaactaa cctcaggaag gtggtattga cagaggtcag gacccacctg gatgtcatgc    800

tatgaaacat taaaagaaaa aaaaaaaaa                                     829


<210> SEQ ID NO 10
```

-continued

<211> LENGTH: 1812
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (358)..(1203)

<400> SEQUENCE: 10

gccgcccggc cgagcgcgga gcgcagccac tcgccgctgc ccagggagcg cccaagatgt 60

gggggggaccg gggcggcagc ggccgtagca gcgccaggga cgggggcacg cagcagcctc 120

cgctcgcccg cctgtcctga cctgcctcgc ttgcccccaa agaatgtcag ccaagtccaa 180

ggggaacccc tcctcgtcct gtccagccga gggaccgccg gcagcctcca aaaccaaggt 240

gaaggaacag atcaagatca tcgtggagga tttggaatta gtcctgggcg acctgaagga 300

cgtggccaag gaacttaagg agatgaagtc ccactctgtt gcccaggcta gagtgca 357

atg gca caa tct ggg ctc act gca acc tct gcc tcc cag gtt caa gct 405
Met Ala Gln Ser Gly Leu Thr Ala Thr Ser Ala Ser Gln Val Gln Ala
1 5 10 15

att ctc ctg cct cag cct gcc tca gtg cgc cac tac gcc tgg gtg gtt 453
Ile Leu Leu Pro Gln Pro Ala Ser Val Arg His Tyr Ala Trp Val Val
20 25 30

gac cag att gac acc ctg acc tct gac cta cag ctg gag gat gag atg 501
Asp Gln Ile Asp Thr Leu Thr Ser Asp Leu Gln Leu Glu Asp Glu Met
35 40 45

act gac agc tcc aaa acg gac acg ctg aat agt agc tca agt ggc aca 549
Thr Asp Ser Ser Lys Thr Asp Thr Leu Asn Ser Ser Ser Ser Gly Thr
50 55 60

aca gcc tcc agc cta gag aag atc aaa gtg cag gct aat gca ccg ctt 597
Thr Ala Ser Ser Leu Glu Lys Ile Lys Val Gln Ala Asn Ala Pro Leu
65 70 75 80

att aaa ccc cca gca cac cca tct gct atc ctc acg gtc ctg aga aag 645
Ile Lys Pro Pro Ala His Pro Ser Ala Ile Leu Thr Val Leu Arg Lys
85 90 95

cca aac cct cca cca cct cct cca agg ttg aca cct gtg aag tgt gaa 693
Pro Asn Pro Pro Pro Pro Pro Arg Leu Thr Pro Val Lys Cys Glu
100 105 110

gac ccc aaa agg gtg gtt cca act gcc aat cct gta aaa acc aat ggc 741
Asp Pro Lys Arg Val Val Pro Thr Ala Asn Pro Val Lys Thr Asn Gly
115 120 125

acc ctt cta cga aat gga ggc tta cca ggt gga cct aac aaa att cca 789
Thr Leu Leu Arg Asn Gly Gly Leu Pro Gly Gly Pro Asn Lys Ile Pro
130 135 140

aat gga gat atc tgc tgc ata ccc aac agt aac ttg gac aag gct cca 837
Asn Gly Asp Ile Cys Cys Ile Pro Asn Ser Asn Leu Asp Lys Ala Pro
145 150 155 160

gtc cag ctt ctg atg cat aga cct gaa aaa gac aga tgt ccc cag gca 885
Val Gln Leu Leu Met His Arg Pro Glu Lys Asp Arg Cys Pro Gln Ala
165 170 175

ggg cct cga gaa cga gtt cgg ttt aat gaa aaa gta cag tac cat ggc 933
Gly Pro Arg Glu Arg Val Arg Phe Asn Glu Lys Val Gln Tyr His Gly
180 185 190

tat tgt cct gac tgt gat acc cgg tat aac ata aaa aac agg gag gtc 981
Tyr Cys Pro Asp Cys Asp Thr Arg Tyr Asn Ile Lys Asn Arg Glu Val
195 200 205

cac tta cac agt gaa cct gtc cac cca ccg gga aag att cct cac caa 1029
His Leu His Ser Glu Pro Val His Pro Pro Gly Lys Ile Pro His Gln
210 215 220

-continued

```
ggc cct ccc ctc cct cct aca ccc cat ctc cct cct ttc cca cta gaa     1077
Gly Pro Pro Leu Pro Pro Thr Pro His Leu Pro Pro Phe Pro Leu Glu
225                 230                 235                 240

aat ggg gga atg gga ata agc cac agt aac agc ttc ccc cct atc aga     1125
Asn Gly Gly Met Gly Ile Ser His Ser Asn Ser Phe Pro Pro Ile Arg
                245                 250                 255

cct gca act gtg cct cct ccc act gca cca aaa cca cag aag acg atc     1173
Pro Ala Thr Val Pro Pro Pro Thr Ala Pro Lys Pro Gln Lys Thr Ile
            260                 265                 270

ttg agg aag tca acc act aca acc gtg tga t gtatgccatt aaaaaaattg     1224
Leu Arg Lys Ser Thr Thr Thr Thr Val  *
        275                 280

ttttttaat tttctatatt ataaacataa aataagtaat gagcactttc tactcaagca   1284

ataaaaagcc caaatatatt aatcctgcat tcagcaaagt ggcataaaaa tcacctggta   1344

agtatgcagc acattgctta tatcctgggt atgcattatt ttaaatgttg tatcattaaa   1404

aacctcagaa tgatgaaaaa tatgaatgat gcattgtttt tgcaattgac ctatgacaaa   1464

ctgtgaacct gcagatttca cctattttga tttactataa gagctgggat ttgattcatt   1524

ttatttatgc ctaagtcatc tatgcattaa catgtcatat tcttaacttt gatctaatgc   1584

tttttactag gaaattttaa tactgaagga ctattttatt atttttttct aaagatgttt   1644

gtcactagtt tttcattatt aaatgctgag gccaatacca agaagtttat tttctatatt   1704

atacaattat gaattacatg ctcagctata tatgtaataa aatactttgg tctgtggaaa   1764

tattgttaaa tcaataaaca atagtaaata atgacaaaaa aaaaaaaa               1812
```

What is claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-10, a mature protein coding portion of SEQ ID NO: 1-10, an active domain of SEQ ID NO: 1-10, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-10.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-10, a mature protein coding portion of SEQ ID NO: 1-10, an active domain of SEQ ID NO: 1-10, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-10, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides from the Sequence Listing, the mature protein portion thereof, or the active domain thereof.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO: 1-10.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or **20** and a pharmaceutically acceptable carrier.

28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or **20** and a pharmaceutically acceptable carrier.

* * * * *