

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/02604 A2

(51) International Patent Classification⁷: **C07K 14/00**

(21) International Application Number: PCT/US01/20999

(22) International Filing Date: 2 July 2001 (02.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/215,446 30 June 2000 (30.06.2000) US

(71) Applicant: **ZYMOGENETICS, INC.** [US/US]; 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(72) Inventors: **THAYER, Edward, C.**; 9701 Sand Point Way NE, #101, Seattle, WA 98115 (US). **SHEPPARD, Paul, O.**; 13532 278th Drive Northeast, Granite Falls, WA 98252 (US). **PRESNELL, Scott, R.**; 2902 N. Puget Sound Avenue, Tacoma, WA 98407 (US).

(74) Agent: **ADAMS, Robyn**; ZymoGenetics, Inc., 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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

(54) Title: LEUCINE-RICH REPEAT PROTEINS, ZLRR7, ZLRR8 AND ZLRR9

(57) Abstract: The present invention provides to polynucleotides and secreted Leucine-Rich Repeat proteins encoded by the polynucleotides. The proteins include binding proteins and fusion proteins operably linked to a second polypeptide. The invention further provides therapeutic and diagnostic methods utilizing the polynucleotides, polypeptides, and antagonists of the polypeptides.



WO 02/02604 A2

Description

5 LEUCINE-RICH REPEAT PROTEINS, ZLRR7, ZLRR8 AND ZLRR9

BACKGROUND OF THE INVENTION

Within the field of genetic engineering, polynucleotides encoding proteins of interest have been identified and cloned by methods that require a detailed
10 knowledge of the structure and/or function of the polynucleotide or the encoded protein. These methods include hybridization screening, polymerase chain reaction (PCR), and expression cloning.

With the more recent advent of large DNA sequence databases and the accompanying data analysis tools, identification of genes of interest is possible through
15 the analysis of raw sequence data. Databases can be "mined" to locate sequences that resemble (are "homologous to") sequences of known function. Alignment of similar sequences can be used to place novel sequences within families of structurally similar sequences. These analytical tools can be combined with structural information obtained from, for example, X-ray crystallography to predict the higher order structure
20 of a novel polypeptide. These analyses also facilitate prediction of polypeptide function. These recent technological advances have greatly increased the pace of gene discovery.

Genetic engineering has made available a number of genes and proteins of pharmaceutical or other economic importance. Such proteins include, for example,
25 tissue plasminogen activator (t-PA) (U.S. Patent No. 4,766,075), coagulation factor VII (U.S. Patent No. 4,784,950), erythropoietin (U.S. Patent No. 4,703,008), platelet derived growth factor (U.S. Patent No. 4,889,919), and various industrial enzymes (e.g., U.S. Patents Nos. 5,965,384; 5,942,431; and 5,922,586).

Although estimates vary as to the amount of the human genome that has
30 been identified to date, there remains a need in the art for further characterization of the human genome and the proteins encoded thereby. Previously unknown genes and proteins will be useful in the treatment and/or prevention of many human diseases, included diseases that have heretofore been refractory to treatment.

Leucine-rich repeats (LRR) are short sequence motifs present in a large
35 number of proteins which appear to be involved in various aspects of protein-protein interaction, such as cell-to-cell communication and signal transduction (for a review, see Kobe and Deisenhofer, *TIBS* 19:415 (1994); Kobe and Deisenhofer, *Curr. Opin.*

Struct. Biol. 5:409 (1995); Kajava, *J. Mol. Biol.* 277:519 (1998)). Proteins that contain an LRR motif include hormone receptors, enzyme subunits, cell adhesion proteins, and ribosome-binding proteins.

A subfamily of the LRR superfamily, referred to as the Small Leucine-Rich Proteoglycan family, illustrates the critical functions fulfilled by proteins containing an LRR motif. Members of this subfamily are believed to play essential biological roles during inflammation and cancer invasion, a regulatory role in collagen fibril formation, suppression of the malignant phenotype of cancer cells, and an inhibition of the growth of certain normal cells (see, for example, Iozzo, *Annu. Rev. Biochem.* 67:609 (1998)).

Kajava, *J. Mol. Biol.* 277:519 (1998), divided the LRR superfamily into subfamilies characterized by different lengths and consensus sequences of the leucine-rich repeats. Based upon this structural analysis, Kajava concluded that LRR proteins of different subfamilies probably emerged independently during evolution, indicating that proteins with the LRR motif provide a unique solution for a wide range of biological functions.

In view of the significant roles played by such proteins, a need exists for the identification of new members of the LRR superfamily, which can provide new tools in detecting and treating alterations in such basic biological functions as cell adhesion and signal transduction.

SUMMARY OF THE INVENTION

Within one aspect the invention provides an isolated polypeptide comprising residues 27 to 234 as shown in SEQ ID NO:2. Within an embodiment, the polypeptide further comprises residues 1 to 234 as shown in SEQ ID NO:2.

Within another aspect the invention provides an isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes an isolated comprising residues 27 to 234 as shown in SEQ ID NO:2. Within an embodiment, is provided an expression vector comprising the following operably linked elements: a transcription promoter; a DNA segment having the isolated polynucleotide; and a transcription terminator. Within another embodiment, the invention provides a cultured cell comprising the expression vector. Within another embodiment is provided a method of producing a polypeptide comprising culturing the cell under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide. Within another embodiment the polypeptide produced by the method is provided

Within another aspect the invention provides an isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:1.

Within another aspect the invention provides an antibody that specifically binds to the isolated protein comprising residues 27 to 234 as shown in
5 SEQ ID NO:2.

Within one aspect the invention provides an isolated polypeptide comprising residues 16 to 279 as shown in SEQ ID NO: 5 or 8. Within an embodiment, the polypeptide further comprises residues 1 to 487 as shown in SEQ ID NO: 5 or 8.

10 Within another aspect the invention provides an isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes an isolated comprising residues 16 to 279 as shown in SEQ ID NO:5 or 8. Within an embodiment, is provided an expression vector comprising the following operably linked elements: a transcription promoter; a DNA segment having the isolated polynucleotide; and a
15 transcription terminator. Within another embodiment, the invention provides a cultured cell comprising the expression vector. Within another embodiment is provided a method of producing a polypeptide comprising culturing the cell under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide. Within another embodiment the polypeptide produced by the method is
20 provided

Within another aspect the invention provides an isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:4 or 7.

Within another aspect the invention provides an antibody that specifically binds to the isolated protein comprising residues 16 to 279 as shown in
25 SEQ ID NO:5 or 8.

Within one aspect the invention provides an isolated polypeptide comprising residues 76 to 363 as shown in SEQ ID NO:11 or 14. Within an embodiment, the polypeptide further comprises residues 1 to 663 as shown in SEQ ID NO:11 or 14.

30 Within another aspect the invention provides an isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes an isolated comprising residues 76 to 363 as shown in SEQ ID NO:11 or 14. Within an embodiment, is provided an expression vector comprising the following operably linked elements: a transcription promoter; a DNA segment having the isolated
35 polynucleotide; and a transcription terminator. Within another embodiment, the invention provides a cultured cell comprising the expression vector. Within another embodiment is provided a method of producing a polypeptide comprising culturing the

cell under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide. Within another embodiment the polypeptide produced by the method is provided

Within another aspect the invention provides an isolated polynucleotide
5 comprising the polynucleotide sequence as shown in SEQ ID NO:10 or 13.

Within another aspect the invention provides an antibody that specifically binds to the isolated protein comprising residues 76 to 363 as shown in SEQ ID NO:11 or 14.

Within one aspect of the invention there is provided an isolated
10 polypeptide comprising fifteen contiguous amino acid residues of a polypeptide as shown in SEQ ID NO:M, wherein M is selected from the group consisting of 2, 5, and 8. Within one embodiment, the at least fifteen contiguous amino acid residues of SEQ ID NO:M are operably linked via a peptide bond or polypeptide linker to a second polypeptide selected from the group consisting of maltose binding protein, an
15 immunoglobulin constant region, a polyhistidine tag, and a peptide as shown in SEQ ID NO:10. Within another embodiment, the polypeptide comprises at least 30 contiguous residues of SEQ ID NO:M. Within a further embodiment, the polypeptide comprises at least 47 contiguous residues of SEQ ID NO:M.

Within another aspect of the invention there are provided
20 polynucleotides encoding the polypeptides disclosed above. Within certain embodiments of the invention the polynucleotides comprise a sequence of nucleotides selected from the group consisting of 1, 4, and 7.

The invention also provides an isolated polynucleotide encoding a
fusion protein, wherein the fusion protein comprises a secretory peptide selected from
25 the group consisting of residues 1 to 22 of SEQ ID NO:2; residues 1 to 18 of SEQ ID NO:5; and residues 1 to 19 of SEQ ID NO:8, and wherein the secretory peptide is operably linked to a second polypeptide.

Within an additional aspect the invention provides a method of
detecting protein secretion from a cell or tissue comprising detecting a mature MSP
30 selected from the group consisting of SEQ ID NO:M, wherein M is selected from the group consisting of 2, 5, and 8.

These and other aspects of the invention will become evident upon reference to the following detailed description of the invention.

DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention in detail, it may be helpful to the understanding thereof to define the following terms:

The term "affinity tag" is used herein to denote a polypeptide segment that can be attached to a second polypeptide to provide for purification of the second polypeptide or provide sites for attachment of the second polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract, protein A (Nilsson et al., *EMBO J.* 4:1075, 1985; Nilsson et al., *Methods Enzymol.* 198:3, 1991), glutathione S transferase (Smith and Johnson, *Gene* 67:31, 1988), Glu-Glu affinity tag (Grussenmeyer et al., *Proc. Natl. Acad. Sci. USA* 82:7952-7954, 1985; see SEQ ID NO:123), substance P, FlagTM peptide (Hopp et al., *Biotechnology* 6:1204-1210, 1988), maltose binding protein (Kellerman and Ferenci, *Methods Enzymol.* 90:459-463, 1982; Guan et al., *Gene* 67:21-30, 1987), streptavidin binding peptide, thioredoxin, ubiquitin, cellulose binding protein, T7 polymerase, immunoglobulin constant domain, or other antigenic epitope or binding domain. See, in general, Ford et al., *Protein Expression and Purification* 2: 95-107, 1991. Affinity tags can be used individually or in combination. DNAs encoding affinity tags and other reagents are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, NJ; Eastman Kodak, New Haven, CT; New England Biolabs, Beverly, MA).

The term "allelic variant" is used herein to denote any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene.

The terms "amino-terminal" and "carboxyl-terminal" are used herein to denote positions within polypeptides. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a polypeptide is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete polypeptide.

A "complement" of a polynucleotide molecule is a polynucleotide molecule having a complementary base sequence and reverse orientation as compared to a reference sequence. For example, the sequence 5' ATGCACGGG 3' is complementary to 5' CCCGTGCAT 3'.

“Corresponding to”, when used in reference to a nucleotide or amino acid sequence, indicates the position in a second sequence that aligns with the reference position when two sequences are optimally aligned.

5 The term “degenerate nucleotide sequence” denotes a sequence of nucleotides that includes one or more degenerate codons (as compared to a reference polynucleotide molecule that encodes a polypeptide). Degenerate codons encompass different triplets of nucleotides, but encode the same amino acid residue (i.e., GAU and GAC triplets each encode Asp).

10 The term “expression vector” is used to denote a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription, wherein said segments are arranged in a way that does not exist naturally. Such additional segments include promoter and terminator sequences, and may also include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc.
15 Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both.

The term “isolated”, when applied to a polynucleotide, denotes that the polynucleotide has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within
20 genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of
25 associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, *Nature* 316:774-78, 1985).

An “isolated” polypeptide or protein is a polypeptide or protein that is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated polypeptide or protein is substantially
30 free of other polypeptides or proteins, particularly other polypeptides or proteins of animal origin. It is preferred to provide the polypeptides or proteins in a highly purified form, i.e. greater than 95% pure, more preferably greater than 99% pure. When used in this context, the term “isolated” does not exclude the presence of the same polypeptide or protein in alternative physical forms, such as dimers or
35 alternatively glycosylated or derivatized forms.

A “mature protein” is a protein that is produced by cellular processing of a primary translation product of a DNA sequence. Such processing may include

removal of a secretory signal peptide, sometimes in combination with a propeptide. Mature sequences can be predicted from full-length sequences using methods known in the art for predicting cleavage sites. See, for example, von Heijne (*Nuc. Acids Res.* 14:4683, 1986). The sequence of a mature protein can be determined experimentally
5 by expressing a DNA sequence of interest in a eukaryotic host cell and determining the amino acid sequence of the final product. For proteins lacking secretory peptides, the primary translation product will be the mature protein.

"Operably linked", when referring to DNA segments, indicates that the segments are arranged so that they function in concert for their intended purposes, e.g.,
10 transcription initiates in the promoter and proceeds through the coding segment to the terminator. When referring to polypeptides, "operably linked" includes both covalently (e.g., by disulfide bonding) and non-covalently (e.g., by hydrogen bonding, hydrophobic interactions, or salt-bridge interactions) linked sequences, wherein the desired function(s) of the sequences are retained.

15 The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

"Paralogs" are distinct but structurally related proteins made by an organism. Paralogs are believed to arise through gene duplication. For example, α -
20 globin, β -globin, and myoglobin are paralogs of each other.

A "polynucleotide" is a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic
25 molecules. Sizes of polynucleotides are expressed as base pairs (abbreviated "bp"), nucleotides ("nt"), or kilobases ("kb"). Where the context allows, the latter two terms may describe polynucleotides that are single-stranded or double-stranded. When the term is applied to double-stranded molecules it is used to denote overall length and will be understood to be equivalent to the term "base pairs". It will be recognized by those
30 skilled in the art that the two strands of a double-stranded polynucleotide may differ slightly in length and that the ends thereof may be staggered as a result of enzymatic cleavage; thus all nucleotides within a double-stranded polynucleotide molecule may not be paired. Such unpaired ends will in general not exceed 20 nt in length.

A "polypeptide" is a polymer of amino acid residues joined by peptide
35 bonds, whether produced naturally or synthetically. Polypeptides of less than about 10 amino acid residues are commonly referred to as "peptides".

The term "promoter" is used herein for its art-recognized meaning to denote a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

5 A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptidic components, such as carbohydrate groups. Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and will vary with the type of cell. Proteins are defined herein in terms of their amino acid backbone structures;
10 substituents such as carbohydrate groups are generally not specified, but may be present nonetheless.

A "secretory signal sequence" is a DNA sequence that encodes a polypeptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized.
15 The larger polypeptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

All references cited herein are incorporated by reference in their entirety.

The present invention is based in part upon the discovery of a group of
20 novel, protein-encoding DNA molecules, designated "MSP" proteins, polypeptides and polynucleotides. Included in this group are secreted proteins herein designated Zlrr7, Zlrr8, and Zlrr9.

A secretory peptide of a protein of the present invention can be used to direct the secretion of other proteins of interest from a host cell. Thus, the present
25 invention provides, *inter alia*, fusions comprising such a secretory peptide of a protein disclosed herein operably linked to another protein of interest. The secretory peptide can be used to direct the secretion of other proteins of interest by joining a polynucleotide sequence encoding it, in the correct reading frame, to the 5' end of a sequence encoding the other protein of interest. Those skilled in the art will recognize
30 that the resulting fused sequence may encode additional residues of a protein of the present invention at the amino terminus of the protein to be secreted. In the extreme case, the fusion may comprise an entire protein of the present invention fused to the amino terminus of a second protein, whereby secretion of the fusion protein is directed by the secretory peptide of the protein of the present invention. It will often be
35 desirable to include a proteolytic cleavage site between the protein of the present invention (or portion thereof) and the other protein of interest. The joined polynucleotide sequences are then introduced into a host cell, which is cultured

according to conventional methods. The protein of interest is then recovered from the culture media. Methods for introducing DNA into host cells, culturing the cells, and isolating recombinant proteins are known in the art. Representative methods are summarized below.

5 Higher order structures of the proteins of the present invention can be predicted by computer analysis using available software (e.g., the Insight II® viewer and homology modeling tools available from MSI, San Diego, CA; and King and Sternberg, *Protein Sci.* 5:2298-310, 1996). In addition, analytical algorithms permit the identification of homologies between newly discovered proteins and known
10 proteins. Such homologies are indicative of related biological functions.

The present invention provides nucleic acid molecules that encode new human polypeptides that are members of the leucine-rich repeat superfamily. Illustrative nucleic acid and polypeptide sequences characterizing polynucleotides and polypeptides of Zlrr7 are shown in SEQ ID NOs:1, 2, and 3. Illustrative nucleic acid
15 and polypeptide sequences characterizing polynucleotides and polypeptides of Zlrr8 are shown in SEQ ID NOs:4, 5, 6, 7, 8, and 9. Illustrative nucleic acid and polypeptide sequences characterizing polynucleotides and polypeptides of Zlrr9 are shown in SEQ ID NOs:10, 11, 12, 13, 14, and 15.

The *Zlrr7* gene described herein encodes a polypeptide of 255 amino
20 acids, as shown in SEQ ID NO:2. The Zlrr7 protein has a signal sequence, N- and C-terminal flanking cysteine-rich regions, and six leucine-rich repeat regions. Table 1 identifies the locations of these structural features and the corresponding nucleotide sequences that encode the polypeptide domains.

Table 1

25

Zlrr7Polypeptide Feature	Amino acid Residues of SEQ ID NO:2	Nucleotides of SEQ ID NO:1
Signal sequence	1-26	1-78
N terminal Cys-Rich Region	27-55	79-165
Leu-Rich Region 1	57-80	169-240
Leu-Rich Region 2	81-104	241-312
Leu-Rich Region 3	105-128	313-384
Leu-Rich Region 4	129-152	385-456
Leu-Rich Region 5	153-176	457-528
Leu-Rich Region 6	177-198	529-594
C terminal Cys-Rich Region	186-234	556-702

The *Zlrr8* gene described herein encodes a polypeptide of 551 amino acids, as shown in SEQ ID NO:8. The *Zlrr8* protein has a signal sequence, N- and C-terminal flanking cysteine-rich regions, and seven leucine-rich repeat regions, an IG domain, and a fibronectin III domain. Table 2 identifies the locations of these structural features and the corresponding nucleotide sequences that encode the polypeptide domains.

Table 2

Zlrr8Polypeptide Feature	Amino acid Residues of SEQ ID NO:11	Nucleotides of SEQ ID NO:10
Signal sequence	1-15	1-45
N terminal Cys-Rich Region	16-47	46-141
Leu-Rich Region 1	49-72	145-216
Leu-Rich Region 2	73-96	217-288
Leu-Rich Region 3	97-120	289-360
Leu-Rich Region 4	121-144	361-432
Leu-Rich Region 5	146-169	436-507
Leu-Rich Region 6	170-193	508-579
Leu-Rich Region 7	194-217	580-651
C terminal Cys-Rich Region	234-279	700-837
IG domain	295-353	883-1059
Fibronectin III domain	404-487	1210-1461

An alternatively spliced variant of the *Zlrr8* gene would encode a polypeptide of 279 amino acids in length from amino acid number 1 to amino acid number 279 of SEQ ID NO:8 or 5. This variant would not include nucleic acid sequence for the IG or Fibronectin domains.

The *Zlrr9* gene described herein encodes a polypeptide of 740 amino acids, as shown in SEQ ID NO:14. The *Zlrr9* protein has a signal sequence, a C-terminal flanking cysteine-rich regions, and 9 leucine-rich repeat regions, and a fibronectin III domain. Table 3 identifies the locations of these structural features and the corresponding nucleotide sequences that encode the polypeptide domains.

Table 3

Zlrr9 Polypeptide Feature	Amino acid Residues of SEQ ID NO:14	Nucleotides of SEQ ID NO:13
Signal sequence	1-18	1-54
Leu-Rich Region 1	76-99	226-297
Leu-Rich Region 2	100-122	298-366
Leu-Rich Region 3	123-146	367-438
Leu-Rich Region 4	147-170	439-510
Leu-Rich Region 5	171-199	511-597
Leu-Rich Region 6	205-228	613-683
Leu-Rich Region 7	229-253	684-759
Leu-Rich Region 8	254-278	760-834
Leu-Rich Region 9	279-302	835-906
C terminal Cys-Rich Region	311-363	931-1089
Fibronectin III domain	580-663	1738-1989

An alternatively spliced variant of the *Zlrr9* gene would encode a polypeptide of 906 amino acids in length from amino acid number 1 to amino acid number 906 of SEQ ID NO:14 or 11. This variant would not include nucleic acid sequence for the IC-terminal Cys-Rich region or Fibronectin domains.

A common leucine-rich motif (LxxLxLxxNxL, where "x" is any amino acid) is present in proteins that are involved in specific protein-protein interaction or cell adhesion (see, for example, Schneider and Schweiger, *Oncogene* 6:1807 (1991), and Kobe and Deisenhofer, *Trends Biochem. Sci.* 19:415 (1994)). The largest subfamily of proteins that contain a leucine-rich domain are extracellular proteins having the following motif: LxxLxxLxLxxNxLxxLPxxOFxx, where "x" is any amino acid and "O" is a non-polar residue (Kajava, *J. Mol. Biol.* 277:519 (1998)).

In certain members of the leucine-rich repeat superfamily, cysteine-rich domains flank the leucine-rich repeat region. Kobe and Deisenhofer, *TIBS* 19:415 (1994), have identified the N-terminal cysteine-rich motif as "CP[about 2x]CxC[about 6x]C," while the C-terminal-like cysteine-rich motif is represented by "PxxCxC[about 20x]C[about 20x]C." The presence of cysteine-rich domains in a leucine-rich repeat protein indicates that the protein is extracellular, and that it may function as an adhesive protein or receptor (Kobe and Deisenhofer, *TIBS* 19:415 (1994); Kajava, *J. Mol. Biol.* 277:519 (1998)). The leucine-rich repeat domains and C- and N-terminal flanking domains of *Zlrr7* are similar to those found in the ALS (insulin-like growth factor binding protein) and SLIT protein.

Certain members of the leucine-rich repeat superfamily that contain flanking cysteine-rich domains appear to play a role in neural differentiation and development (see, for example, Suzuki *et al.*, *J. Biol. Chem.* 271:22522 (1996)). The *Zlrr7* gene appears to be primarily expressed in a retinoblastoma tissue library. The *Zlrr9* gene appears to be in brain, kidney (renal cell adenocarcinoma), and eye tissues and a endometrium, adenocarcinoma cell line.. The *Zlrr8* gene appears to be in several carcinoma tissues including, glioblastoma (pooled); neuroblastoma cells; B-cell, chronic lymphocytic leukemia; 2 pooled kidney tumors (clear cell type); pooled germ cell tumors; lung large cell carcinoma, undifferentiated; breast, pooled: mammary adenocarcinoma, cell line; colon - tumor RER+; colon adenocarcinoma cell line; retinoblastoma; genitourinary tract - 2 pooled high-grade transitional cell tumors; rhabdomyosarcoma; brain, (pineal gland); bone marrow line: lung; lung tumor, squamous cell CA; kidney epithelial transf embryo line; bladder tumor; colon, mixed tissues; esophagus tumor; small intestine, fetal; uterus; endometrium; fetal - 8-9 weeks; pancreas - adenocarcinoma; prostate - adenocarcinoma, cell line; and colon. The *Zlrr8* gene is located on chromosome 11q13.

As described herein, the present invention provides isolated polypeptides having an amino acid sequence that is at least 70%, at least 80%, or at least 90% identical to the amino acid sequence of SEQ ID NO:2, wherein such isolated polypeptides can specifically bind with an antibody that specifically binds with a polypeptide consisting of the amino acid sequence of SEQ ID NO:2. An illustrative polypeptide is a polypeptide that comprises the amino acid sequence of SEQ ID NO:2. Additional exemplary polypeptides include the following: (a) polypeptides comprising a leucine-rich region having the following amino acid residue motif: "Y-x(2)-L-x(2)-L-x-L-x(2)-N-x-L-x(2)-L-P-x(2)-L-F-L," wherein "x" is any amino acid, and wherein the values in parentheses indicate the number of occurrences of "x," (b) polypeptides comprising a leucine-rich region with the recited motif and at least one cysteine-rich region that resides in an N-terminal or C-terminal position relative to the leucine-rich region, wherein the N-terminal cysteine-rich region has the motif of "C-P-x(2)-C-x-C-x(6)-C," and the C-terminal cysteine-rich region has the motif of "P-x(2)-C-x-C-x(24)-C-x(21)-C," and (c) polypeptides comprising a leucine-rich region with the recited motif and both cysteine-rich regions.

The present invention also provides isolated polypeptides comprising at least 15, or at least 30, contiguous amino acid residues of an amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, and SEQ ID NO:14.

Further as a secreted proteins, the polynucleotides of the secretory signal sequences of *Zlrr7*, *Zlrr8*, and *Zlrr9* can be fused to a different protein of interest

to direct the protein through the secretory pathway of the cell. These secretory signal sequences are from residue 1 to residue 26 of SEQ ID NO:2 for Zlrr7; from residue 1 to residue 15 of SEQ ID NO:5 for Zlrr8; and from residue 1 to residue 18 of SEQ ID NO:8 for Zlrr9.

5 A protein of the present invention can be prepared as a fusion protein by joining it to a second polypeptide or a plurality of additional polypeptides. Suitable second polypeptides include amino- or carboxyl-terminal extensions, such as linker peptides of up to about 20-25 residues and extensions that facilitate purification (affinity tags) as disclosed above. A protein of interest can be prepared as a fusion to a
10 dimerizing protein as disclosed in U.S. Patents Nos. 5,155,027 and 5,567,584. Preferred dimerizing proteins in this regard include immunoglobulin constant region domains. Immunoglobulin-polypeptide fusions can be expressed in genetically engineered cells to produce a variety of multimeric analogs of a protein of interest. Fusion proteins can also comprise auxiliary domains that target the protein of interest
15 to specific cells, tissues, or macromolecules (e.g., collagen). For example, a protein of interest can be targeted to a predetermined cell type by fusing it to a ligand that specifically binds to a receptor on the surface of a target cell. In this way, proteins can be targeted for therapeutic or diagnostic purposes. A protein can be fused to two or more moieties, such as an affinity tag for purification and a targeting domain. Protein
20 fusions can also comprise one or more cleavage sites, particularly between domains. See, Tuan et al., *Connective Tissue Research* 34:1-9, 1996. Proteins of the present invention can also be used as targetting moieties within fusion proteins comprising, for example, cytokines, cytotoxins, or other biologically active polypeptide moieties.

Protein fusions of the present invention will usually contain not more
25 than about 1,200 amino acid residues joined to the MSP protein. For example, an MSP protein can be fused to *E. coli* β -galactosidase (1,021 residues; see Casadaban et al., *J. Bacteriol.* 143:971-980, 1980), a 10-residue spacer, and a 4-residue factor Xa cleavage site. Such a protein comprising, for example, MSPZ809335G4P (SEQ ID NO:100), contains 723 amino acid residues. In a second example, an MSP protein can be fused
30 to maltose binding protein (approximately 370 residues), a 4-residue cleavage site, and a 6-residue polyhistidine tag.

As disclosed above, the proteins of the present invention or portions thereof can also be used to direct the secretion of a second protein. When such fusions are designed so that the secreted protein retains a portion of the protein of the present
35 invention, the fusion protein can be purified by means that exploit the properties of the protein of the present invention. Typical of such methods is immunoaffinity chromatography using an antibody directed against a protein of the present invention.

When such a fusion is engineered to contain a cleavage site at the fusion point, the fusion can be cleaved and the protein of interest recovered free of extraneous sequence.

As secreted proteins, MSP can be useful to monitor secretion of proteins in general from cells or tissue. In doing so, cell lines or tissues shown to express the proteins of the present invention can be monitored to detect for example if the cell or tissue has a defective secretory pathway. Thus, cell membrane extracts or conditioned media from the cell or tissue can be tested for the presence or absence of the mature MSP protein. Similarly, when trying to identify if a recombinantly expressed protein is has a functional signal peptide, the molecules of the present invention can be used as a positive control.

The present invention also provides polynucleotide molecules, including DNA and RNA molecules, that encode the proteins disclosed above. Those skilled in the art will readily recognize that, in view of the degeneracy of the genetic code, considerable sequence variation is possible among these polynucleotide molecules. The amino acid sequence information provided herein can be used by one of ordinary skill in the art to generate degenerate sequences comprising all nucleotide sequences encoding a particular polypeptide. Table 4 sets forth the one-letter codes used to denote degenerate nucleotide positions. "Resolutions" are the nucleotides denoted by a code letter. "Complement" indicates the code for the complementary nucleotide(s). For example, the code Y denotes either C or T, and its complement R denotes A or G, A being complementary to T, and G being complementary to C.

TABLE 4

Nucleotide	Resolutions	Complement	Resolutions
A	A	T	T
C	C	G	G
G	G	C	C
T	T	A	A
R	A G	Y	C T
Y	C T	R	A G
M	A C	K	G T
K	G T	M	A C
S	C G	S	C G
W	A T	W	A T
H	A C T	D	A G T
B	C G T	V	A C G
V	A C G	B	C G T
D	A G T	H	A C T
N	A C G T	N	A C G T

5

Degenerate codons encompassing all possible codons for a given amino acid are set forth in Table 5, below.

TABLE 5

Amino Acid	One-Letter Code	Codons	Degenerate Codon
Cys	C	TGC TGT	TGY
Ser	S	AGC AGT TCA TCC TCG TCT	WSN
Thr	T	ACA ACC ACG ACT	CAN
Pro	P	CCA CCC CCG CCT	CCN
Ala	A	GCA GCC GCG GCT	GCN
Gly	G	GGA GGC GGG GGT	GGN
Asn	N	AAC AAT	AAY
Asp	D	GAC GAT	GAY
Glu	E	GAA GAG	GAR
Gln	Q	CAA CAG	CAR
His	H	CAC CAT	CAY
Arg	R	AGA AGG CGA CGC CGG CGT	MGN
Lys	K	AAA AAG	AAR
Met	M	ATG	ATG
Ile	I	ATA ATC ATT	ATH
Leu	L	CTA CTC CTG CTT TTA TTG	YTN
Val	V	GTA GTC GTG GTT	GTN
Phe	F	TTC TTT	TTY
Tyr	Y	TAC TAT	TAY
Trp	W	TGG	TGG
Ter	.	TAA TAG TGA	TRR
Asn Asp	B		RAY
Glu Gln	Z		SAR
Any	X		NNN
Gap	-	---	

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, the degenerate codon for serine (WSN) can, in some circumstances, encode arginine (AGR), and the degenerate codon for arginine (MGN) can, in some circumstances, encode serine (AGY). A similar relationship exists between codons encoding phenylalanine and leucine. Thus, some polynucleotides encompassed by the degenerate sequences may encode variant amino

acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequences disclosed in the accompanying Sequence Listing.

Methods for preparing DNA and RNA are well known in the art.

5 Complementary DNA (cDNA) clones are prepared from RNA that is isolated from a tissue or cell that produces large amounts of the cognate mRNA. Such tissues and cells are identified by methods commonly known in the art, such as Northern blotting (Thomas, *Proc. Natl. Acad. Sci. USA* 77:5201, 1980). Databases of expressed sequence tags (ESTs) can be analyzed to produce an "electronic Northern" wherein

10 sequences are assigned to specific cell or tissue sources on the basis of their abundance within libraries.

A panel of cDNAs from human tissues can be screened for tissue expression using PCR. As an example, such a panel can be made from commercially available first strand cDNA (Clontech laboratories, Inc., Palo Alto CA) and would

15 contain 20 first-strand cDNA samples from the human tissues shown in Table 6. The panel can be set up in a 96-well format that further included a human genomic DNA (obtained from Clontech Laboratories, Inc.) positive control sample and a water-only well as a negative control sample. Each well can contain approximately 0.2-100 pg/ μ l of cDNA, diluted with water to 17.5 μ l. The PCR reactions are set up by adding

20 oligonucleotide primers, DNA polymerase (Ex TaqTM; TAKARA Shuzo Co. Ltd. Biomedicals Group, Japan or AdvantageTM 2 cDNA polymerase mix; Clontech Laboratories, Inc.) with the appropriate supplied buffer, dNTP mix (TAKARA Shuzo Co. Ltd.), and a density increasing agent and tracking dye (RediLoad; Research Genetics, Inc., Huntsville, AL) to each sample on the panel. The amplification of the

25 cDNA can be carried out as follows: incubation at 94°C for 2 minutes; 35 cycles of 94°C for 30 seconds, 60°C for 20 seconds, and 72°C for 30 seconds; followed by incubation at 72°C for 5 minutes. A portion of the PCR reaction product is then subjected to standard agarose gel electrophoresis using a 4% agarose gel.

Total RNA can be prepared using guanidine HCl extraction followed by

30 isolation by centrifugation in a CsCl gradient (Chirgwin et al., *Biochemistry* 18:52-94, 1979). Poly (A)⁺ RNA is prepared from total RNA using the method of Aviv and Leder (*Proc. Natl. Acad. Sci. USA* 69:1408-1412, 1972). Complementary DNA (cDNA) is prepared from poly(A)⁺ RNA using known methods. In the alternative, genomic DNA can be isolated. For some applications (e.g., expression in transgenic

35 animals) it may be preferable to use a genomic clone, or to modify a cDNA clone to include at least one genomic intron. Methods for identifying and isolating cDNA and genomic clones are well known and within the level of ordinary skill in the art, and

include the use of the sequences disclosed herein, sequences complementary thereto, or parts thereof, for probing or priming a library. Such methods include, for example, hybridization or polymerase chain reaction ("PCR", Mullis, U.S. Patent 4,683,202). Expression libraries can be probed with antibodies to a protein of interest, receptor
5 fragments, or other specific binding partners.

The polynucleotides of the present invention can also be prepared by automated synthesis. Synthesis of polynucleotides is within the level of ordinary skill in the art, and suitable equipment and reagents are available from commercial suppliers. See, in general, Glick and Pasternak, Molecular Biotechnology, Principles
10 & Applications of Recombinant DNA, ASM Press, Washington, D.C., 1994; Itakura et al., *Ann. Rev. Biochem.* 53: 323-56, 1984; and Climie et al., *Proc. Natl. Acad. Sci. USA* 87:633-7, 1990.

The present invention further provides antisense polynucleotides that are complementary to a segment of a polynucleotide as set forth in one of SEQ ID
15 NO:N, wherein N is an odd integer from 1 to 327. Such antisense polynucleotides are designed to bind to the corresponding mRNA and inhibit its translation. Antisense polynucleotides are used to inhibit gene expression in cell culture or in a patient, and can be used as probes or primers for research or diagnostic purposes.

Probes and primers of the present invention comprise a suitable
20 fragment, and may comprise up to the complete sequence, of a polynucleotide as shown in SEQ ID NO:N or the complement thereof, wherein N is an odd integer from 1 to 327. Probes will generally be at least 20 nucleotides in length, although somewhat shorter probes (14-17 nucleotides) can be used. PCR primers are at least 5 nucleotides in length, preferably 15 or more nt, more preferably 20-30 nt. Shorter polynucleotide
25 probes and primers are referred to in the art as "oligonucleotides," and can be DNA or RNA. Probes will generally comprise an oligonucleotide linked to a label, such as a radionuclide.

Probes and primers as disclosed herein can be used for cloning allelic, orthologous, and paralogous sequences. Allelic variants of the disclosed sequences can
30 be cloned by probing cDNA or genomic libraries from different individuals according to standard procedures. Orthologous sequences can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a cDNA can be cloned using mRNA obtained from a tissue or cell type that expresses the protein. Suitable sources of mRNA can be
35 identified by probing Northern blots with probes designed from the sequences disclosed herein. A library is then prepared from mRNA of a positive tissue or cell line. A cDNA can then be isolated by a variety of methods, such as by probing with a

complete or partial human cDNA or with one or more sets of degenerate probes based on the disclosed sequences. A cDNA can also be cloned by PCR using primers designed from the sequences disclosed herein. Within an additional method, the cDNA library can be used to transform or transfect host cells, and expression of the cDNA of interest can be detected with an antibody to the encoded protein. Similar techniques can also be applied to the isolation of genomic clones. Orthologous and paralogous sequences can be identified from libraries by probing blots at low stringency and washing the blots at successively higher stringency until background is suitably reduced.

Probes and primers disclosed herein can be used to clone 5' non-coding regions of a corresponding gene. Such promoter elements can thus be used to direct the tissue-specific expression of heterologous genes in, for example, transgenic animals or patients treated with gene therapy. Cloning of 5' flanking sequences also facilitates production of a protein of interest by "gene activation" as disclosed in U.S. Patent No. 5,641,670. Briefly, expression of an endogenous gene in a cell is altered by introducing into its locus a DNA construct comprising at least a targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site. The targeting sequence is a 5' non-coding sequence that permits homologous recombination of the construct with the endogenous locus, whereby the sequences within the construct become operably linked with the endogenous coding sequence. In this way, an endogenous promoter can be replaced or supplemented with other regulatory sequences to provide enhanced, tissue-specific, or otherwise regulated expression.

The polynucleotides of the present invention further include polynucleotides encoding the fusion proteins, including signal peptide fusions, disclosed above.

In studying cell biology it is useful to monitor changing levels of mRNA as a function of cell and tissue development. As a set of cDNA molecules, the polynucleotides of the present invention, for example SEQ ID NO:N, wherein N is an odd integer from 1 to 327, can also be used to screen for levels of mRNA of the polypeptides, i.e., as shown in SEQ ID NO:M, wherein M is selected from the group consisting of 2, 5, 8, 11, and 14. Screening for message level fluctuations of the molecules of the present invention are useful for studying, for example, cell replication, activation, and quiescence.

The present invention further provides a computer-readable medium encoded with a data structure that provides at least one of SEQ ID NO:1 through SEQ ID NO:9. Suitable forms of computer-readable media include magnetic media and optically-readable media. Examples of magnetic media include a hard or fixed drive, a

random access memory (RAM) chip, a floppy disk, digital linear tape (DLT), a disk cache, and a ZIP® disk. Optically readable media are exemplified by compact discs (e.g., CD-read only memory (ROM), CD-rewritable (RW), and CD-recordable), and digital versatile/video discs (DVD) (e.g., DVD-ROM, DVD-RAM, and DVD+RW).

5 The polypeptides of the present invention, including full-length proteins, biologically active fragments, immunogenic fragments, and fusion proteins, can be produced in genetically engineered host cells according to conventional techniques. Suitable host cells are those cell types that can be transformed or transfected with exogenous DNA and grown in culture, and include bacteria, fungal
10 cells, and cultured higher eukaryotic cells. Eukaryotic cells, particularly cultured cells of multicellular organisms, are generally preferred for the production of proteins having higher eukaryotic-type post-translational modifications (e.g., γ -carboxylation) and for making proteins, especially secretory proteins, for pharmaceutical use in humans. Techniques for manipulating cloned DNA molecules and introducing
15 exogenous DNA into a variety of host cells are disclosed by Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and Ausubel et al., eds., *Current Protocols in Molecular Biology*, Green and Wiley and Sons, NY, 1993.

 In general, a DNA sequence encoding a polypeptide of interest is
20 operably linked to other genetic elements required for its expression, generally including a transcription promoter and terminator, within an expression vector. The vector will also commonly contain one or more selectable markers and one or more origins of replication, although those skilled in the art will recognize that within certain systems selectable markers can be provided on separate vectors, and replication of the
25 exogenous DNA can be achieved through integration into the host cell genome. Selection of promoters, terminators, selectable markers, vectors and other elements is a matter of routine design within the level of ordinary skill in the art. Many such elements are described in the literature and are available through commercial suppliers.

 To direct a polypeptide into the secretory pathway of a host cell, a
30 secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) is provided in the expression vector. The secretory signal sequence may be that of the protein of interest, or may be derived from another secreted protein (e.g., t-PA; see U.S. Patent No. 5,641,655) or synthesized *de novo*. The secretory signal sequence is operably linked to the DNA sequence encoding the protein of interest, i.e.,
35 the two sequences are joined in the correct reading frame and positioned to direct the newly synthesized protein into the secretory pathway of the host cell. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the protein of

interest, although certain secretory signal sequences may be positioned elsewhere in the DNA sequence of interest (see, e.g., Welch et al., U.S. Patent No. 5,037,743; Holland et al., U.S. Patent No. 5,143,830).

Cultured mammalian cells are suitable hosts for use within the present invention. Methods for introducing exogenous DNA into mammalian host cells include calcium phosphate-mediated transfection (Wigler et al., *Cell* 14:725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7:603, 1981; Graham and Van der Eb, *Virology* 52:456, 1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), DEAE-dextran mediated transfection (Ausubel et al., *ibid.*), and liposome-mediated transfection (Hawley-Nelson et al., *Focus* 15:73, 1993; Ciccarone et al., *Focus* 15:80, 1993). The production of recombinant polypeptides in cultured mammalian cells is disclosed by, for example, Levinson et al., U.S. Patent No. 4,713,339; Hagen et al., U.S. Patent No. 4,784,950; Palmiter et al., U.S. Patent No. 4,579,821; and Ringold, U.S. Patent No. 4,656,134. Suitable cultured mammalian cells include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK (ATCC No. CRL 1632), BHK 570 (ATCC No. CRL 10314), 293 (ATCC No. CRL 1573; Graham et al., *J. Gen. Virol.* 36:59-72, 1977) and Chinese hamster ovary (e.g. CHO-K1; ATCC No. CCL 61) cell lines. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Manassas, Virginia. In general, strong transcription promoters are preferred, such as promoters from SV-40 or cytomegalovirus. See, e.g., U.S. Patent No. 4,956,288. Other suitable promoters include those from metallothionein genes (U.S. Patent Nos. 4,579,821 and 4,601,978) and the adenovirus major late promoter. Within an alternative embodiment, adenovirus vectors can be employed. See, for example, Garnier et al., *Cytotechnol.* 15:145-55, 1994.

Drug selection is generally used to select for cultured mammalian cells into which foreign DNA has been inserted. Such cells are commonly referred to as "transfectants". Cells that have been cultured in the presence of the selective agent and are able to pass the gene of interest to their progeny are referred to as "stable transfectants." An exemplary selectable marker is a gene encoding resistance to the antibiotic neomycin. Selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems can also be used to increase the expression level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. An exemplary amplifiable selectable marker is dihydrofolate reductase, which confers resistance to

methotrexate. Other drug resistance genes (e.g. hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used.

Insect cells can be infected with recombinant baculovirus, commonly derived from *Autographa californica* nuclear polyhedrosis virus (AcNPV). See, King and Possee, The Baculovirus Expression System: A Laboratory Guide, London, Chapman & Hall; O'Reilly et al., Baculovirus Expression Vectors: A Laboratory Manual, New York, Oxford University Press., 1994; and Richardson, Ed., Baculovirus Expression Protocols. Methods in Molecular Biology, Humana Press, Totowa, NJ, 1995. Recombinant baculovirus can also be produced through the use of a transposon-based system described by Luckow et al. (*J. Virol.* 67:4566-4579, 1993). This system, which utilizes transfer vectors, is commercially available in kit form (Bac-to-Bac™ kit; Life Technologies, Rockville, MD). See also, Hill-Perkins and Possee, *J. Gen. Virol.* 71:971-976, 1990; Bonning et al., *J. Gen. Virol.* 75:1551-1556, 1994; and Chazenbalk and Rapoport, *J. Biol. Chem.* 270:1543-1549, 1995.

For protein production, the recombinant virus is used to infect host cells, typically a cell line derived from the fall armyworm, *Spodoptera frugiperda* (e.g., Sf9 or Sf21 cells) or *Trichoplusia ni* (e.g., High Five™ cells; Invitrogen, Carlsbad, CA). See, in general, Glick and Pasternak, Molecular Biotechnology: Principles and Applications of Recombinant DNA, ASM Press, Washington, D.C., 1994. See also, U.S. Patent No. 5,300,435. Serum-free media are used to grow and maintain the cells. Suitable media formulations are known in the art and can be obtained from commercial suppliers. The cells are grown up from an inoculation density of approximately $2-5 \times 10^5$ cells to a density of $1-2 \times 10^6$ cells, at which time a recombinant viral stock is added at a multiplicity of infection (MOI) of 0.1 to 10, more typically near 3. Procedures used are generally described in available laboratory manuals (e.g., King and Possee, *ibid.*; O'Reilly et al., *ibid.*; Richardson, *ibid.*). See also, Guarino et al., U.S. Patent No. 5,162,222 and WIPO publication WO 94/06463.

Fungal cells, including yeast cells, can also be used within the present invention. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Pichia methanolica*. Methods for transforming *S. cerevisiae* cells with exogenous DNA and producing recombinant polypeptides therefrom are disclosed by, for example, Kawasaki, U.S. Patent No. 4,599,311; Kawasaki et al., U.S. Patent No. 4,931,373; Brake, U.S. Patent No. 4,870,008; Welch et al., U.S. Patent No. 5,037,743; and Murray et al., U.S. Patent No. 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (e.g., leucine). A preferred vector system for use in *Saccharomyces cerevisiae* is the

POT1 vector system disclosed by Kawasaki et al. (U.S. Patent No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, e.g., Kawasaki, U.S. Patent No. 4,599,311; Kingsman et al., U.S. Patent No. 4,615,974; and Bitter, U.S. Patent No. 4,977,092) and alcohol dehydrogenase genes. See also U.S. Patents Nos. 4,990,446; 5,063,154; 5,139,936 and 4,661,454.

Transformation systems for other yeasts, including *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Ustilago maydis*, *Pichia pastoris*, *Pichia methanolica*, *Pichia guilliermondii* and *Candida maltosa* are known in the art. See, for example, Gleeson et al., *J. Gen. Microbiol.* 132:3459-3465, 1986 and Cregg, U.S. Patent No. 4,882,279. Aspergillus cells may be utilized according to the methods of McKnight et al., U.S. Patent No. 4,935,349. Methods for transforming *Acremonium chrysogenum* are disclosed by Sumino et al., U.S. Patent No. 5,162,228. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Patent No. 4,486,533. Production of recombinant proteins in *Pichia methanolica* is disclosed in U.S. Patents No. 5,716,808, 5,736,383, 5,854,039, and 5,888,768; and WIPO publications WO 99/14347 and WO 99/14320.

Other higher eukaryotic cells, including plant cells and avian cells, can also be used as hosts according to methods commonly known in the art. For example, the use of *Agrobacterium rhizogenes* as a vector for expressing genes in plant cells has been reviewed by Sinkar et al., *J. Biosci. (Bangalore)* 11:47-58, 1987.

Prokaryotic host cells, including strains of the bacteria *Escherichia coli*, *Bacillus* and other genera are also useful host cells within the present invention. Techniques for transforming these hosts and expressing foreign DNA sequences cloned therein are well known in the art (see, e.g., Sambrook et al., *ibid.*). When expressing a polypeptide in bacteria such as *E. coli*, the polypeptide may be retained in the cytoplasm, typically as insoluble granules, or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed, and the granules are recovered and denatured using, for example, guanidine isothiocyanate or urea. The denatured polypeptide can then be refolded and dimerized by diluting the denaturant, such as by dialysis against a solution of urea and a combination of reduced and oxidized glutathione, followed by dialysis against a buffered saline solution. In the latter case, the polypeptide can be recovered from the periplasmic space in a soluble and functional form by disrupting the cells (by, for example, sonication or osmotic shock) to release the contents of the periplasmic space and recovering the protein, thereby obviating the need for denaturation and refolding.

Transformed or transfected host cells are cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the chosen host cells. A variety of suitable media, including defined media and complex media, are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals. Media may also contain such components as growth factors or serum, as required. The growth medium will generally select for cells containing the exogenously added DNA by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker carried on the expression vector or co-transfected into the host cell.

It is preferred to purify the polypeptides and proteins of the present invention to $\geq 80\%$ purity, more preferably to $\geq 90\%$ purity, even more preferably $\geq 95\%$ purity, and particularly preferred is a pharmaceutically pure state, that is greater than 99.9% pure with respect to contaminating macromolecules, particularly other proteins and nucleic acids, and free of infectious and pyrogenic agents. Preferably, a purified polypeptide or protein is substantially free of other polypeptides or proteins, particularly those of animal origin.

Expressed recombinant proteins (including single polypeptide chains, chimeric polypeptides, and polypeptide multimers) are purified by conventional protein purification methods, typically by a combination of chromatographic techniques. See, in general, Affinity Chromatography: Principles & Methods, Pharmacia LKB Biotechnology, Uppsala, Sweden, 1988; and Scopes, Protein Purification: Principles and Practice, Springer-Verlag, New York, 1994. Proteins comprising a polyhistidine affinity tag (typically about 6 histidine residues) are purified by affinity chromatography on a nickel chelate resin. See, for example, Houchuli et al., *Bio/Technol.* 6: 1321-1325, 1988. Proteins comprising a glu-glu tag can be purified by immunoaffinity chromatography essentially as disclosed by Grussenmeyer et al., *ibid.* Proteins comprising other affinity tags can be purified by appropriate affinity chromatography methods, which are known in the art.

Proteins of the present invention and fragments thereof can also be prepared through chemical synthesis according to methods known in the art, including exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. See, for example, Merrifield, *J. Am. Chem. Soc.* 85:2149, 1963; Stewart et al., Solid Phase Peptide Synthesis (2nd edition), Pierce Chemical Co., Rockford, IL, 1984; Bayer and Rapp, *Chem. Pept. Prot.* 3:3, 1986; and Atherton et al., Solid Phase Peptide Synthesis: A Practical Approach, IRL Press, Oxford, 1989.

Using methods known in the art, the proteins of the present invention can be prepared in a variety of modified or derivatized forms. For example, the proteins can be prepared glycosylated or non-glycosylated; pegylated or non-pegylated; and may or may not include an initial methionine amino acid residue.

5 Biological activities of the proteins of the present invention can be measured *in vitro* using cultured cells or *in vivo* by administering molecules of the claimed invention to the appropriate animal model. Many such assays and models are known in the art. Guidance in initial assay selection is provided by structural predictions and sequence alignments. However, even if no functional prediction is
10 made, the activity of a protein can be elucidated by known methods, including, for example, screening a variety of target cells for a biological response, other *in vitro* assays, expression in a host animal, or through the use of transgenic and/or "knockout" animals. Through the application of robotics, many *in vitro* assays can be adapted to rapid, high-throughput screening of a large number of samples. Target cells for use in
15 activity assays include, without limitation, vascular cells (especially endothelial cells and smooth muscle cells), hematopoietic (myeloid and lymphoid) cells, liver cells (including hepatocytes, fenestrated endothelial cells, Kupffer cells, and Ito cells), fibroblasts (including human dermal fibroblasts and lung fibroblasts), neurite cells (including astrocytes, glial cells, dendritic cells, and PC-12 cells), fetal lung cells,
20 articular synoviocytes, pericytes, chondrocytes, osteoblasts, adipocytes, and prostate epithelial cells. Endothelial cells and hematopoietic cells are derived from a common ancestral cell, the hemangioblast (Choi et al., *Development* 125:725-732, 1998).

Biological activity can be measured with a silicon-based biosensor microphysiometer that measures the extracellular acidification rate or proton excretion
25 associated with receptor binding and subsequent physiologic cellular responses. An exemplary such device is the Cytosensor™ Microphysiometer manufactured by Molecular Devices, Sunnyvale, CA. A variety of cellular responses, such as cell proliferation, ion transport, energy production, inflammatory response, regulatory and receptor activation, and the like, can be measured by this method. See, for example,
30 McConnell et al., *Science* 257:1906-1912, 1992; Pitchford et al., *Meth. Enzymol.* 228:84-108, 1997; Arimilli et al., *J. Immunol. Meth.* 212:49-59, 1998; and Van Liefde et al., *Eur. J. Pharmacol.* 346:87-95, 1998. The microphysiometer can be used for assaying adherent or non-adherent eukaryotic or prokaryotic cells. By measuring extracellular acidification changes in cell media over time, the microphysiometer
35 directly measures cellular responses to various stimuli, including agonistic and antagonistic stimuli. Preferably, the microphysiometer is used to measure responses of a eukaryotic cell known to be responsive to the protein of interest, compared to a

control eukaryotic cell that does not respond to the protein of interest. Responsive eukaryotic cells comprise cells into which a receptor for the protein of interest has been transfected, as well as naturally responsive cells. Differences in the response of cells exposed to the protein of interest, relative to a control not so exposed, are a direct measurement of protein-modulated cellular responses. Such responses can be assayed under a variety of stimuli. The present invention thus provides methods of identifying agonists and antagonists of proteins of interest, comprising providing cells responsive to a selected protein, culturing a first portion of the cells in the absence of a test compound, culturing a second portion of the cells in the presence of a test compound, and detecting a change in a cellular response of the second portion of the cells as compared to the first portion of the cells. The change in cellular response is shown as a measurable change in extracellular acidification rate. Culturing a third portion of the cells in the presence of the protein of interest and the absence of a test compound provides a positive control and a control to compare the agonist activity of a test compound with that of the protein of interest. Antagonists can be identified by exposing the cells to the protein of interest in the presence and absence of the test compound, whereby a reduction in protein-stimulated activity is indicative of antagonist activity in the test compound.

As expressed secreted proteins the MSP polypeptides disclosed herein can be used to screen for cell metabolism effecting receptors. Thus, the polypeptides of the present invention are useful for identifying new target receptors and drug design. A subset of the polypeptides disclosed herein can include membrane-bound receptors. This subset can include amongst other receptors, homologs of G protein coupled receptors, which are important mediators for cell activities including the synthesis of second messengers. Thus a set of these G protein couple receptors can mediate cell differentiation and proliferation, or hormone expression *in vitro* and *in vivo*.

Assays measuring cell proliferation or differentiation are well known in the art. For example, assays measuring proliferation include such assays as chemosensitivity to neutral red dye (Cavanaugh et al., *Investigational New Drugs* 8:347-354, 1990), incorporation of radiolabelled nucleotides (as disclosed by, e.g., Raines and Ross, *Methods Enzymol.* 109:749-773, 1985; Wahl et al., *Mol. Cell Biol.* 8:5016-5025, 1988; and Cook et al., *Analytical Biochem.* 179:1-7, 1989), incorporation of 5-bromo-2'-deoxyuridine (BrdU) in the DNA of proliferating cells (Porstmann et al., *J. Immunol. Methods* 82:169-179, 1985), and use of tetrazolium salts (Mosmann, *J. Immunol. Methods* 65:55-63, 1983; Alley et al., *Cancer Res.* 48:589-601, 1988; Marshall et al., *Growth Reg.* 5:69-84, 1995; and Scudiero et al., *Cancer Res.* 48:4827-4833, 1988). Differentiation can be assayed using suitable precursor cells that can be

induced to differentiate into a more mature phenotype. Assays measuring differentiation include, for example, measuring cell-surface markers associated with stage-specific expression of a tissue, enzymatic activity, functional activity or morphological changes (Watt, *FASEB*, 5:281-284, 1991; Francis, *Differentiation* 57:63-75, 1994; Raes, *Adv. Anim. Cell Biol. Technol. Bioprocesses*, 161-171, 1989). Effects of a protein on tumor cell growth and metastasis can be analyzed using the Lewis lung carcinoma model, for example as described by Cao et al., *J. Exp. Med.* 182:2069-2077, 1995. Activity of a protein on cells of neural origin can be analyzed using assays that measure effects on neurite growth as disclosed below.

In vitro assays for pro- and anti-inflammatory activity are known in the art. Exemplary activity assays include mitogenesis assays in which IL-1 responsive cells (e.g., D10.N4.M cells) are incubated in the presence of IL-1 or a test protein for 72 hours at 37°C in a 5% CO₂ atmosphere. IL-2 (and optionally IL-4) is added to the culture medium to enhance sensitivity and specificity of the assay. ³H-thymidine is then added, and incubation is continued for six hours. The amount of label incorporated is indicative of agonist activity. See, Hopkins and Humphreys, *J. Immunol. Methods* 120:271-276, 1989; Greenfeder et al., *J. Biol. Chem.* 270:22460-22466, 1995. Stimulation of cell proliferation can also be measured using thymocytes cultured in a test protein in combination with phytohemagglutinin. IL-1 is used as a control. Proliferation is detected as ³H-thymidine incorporation or metabolic breakdown of (MTT) (Mosman, *ibid.*).

Protein activity may also be detected using assays designed to measure induction of one or more growth factors or other macromolecules. Preferred such assays include those for determining the presence of hepatocyte growth factor (HGF), epidermal growth factor (EGF), transforming growth factor alpha (TGFα), interleukin-6 (IL-6), VEGF, acidic fibroblast growth factor (aFGF), angiogenin, and other macromolecules produced by the liver. Suitable assays include mitogenesis assays using target cells responsive to the macromolecule of interest, receptor-binding assays, competition binding assays, immunological assays (e.g., ELISA), and other formats known in the art. Metalloprotease secretion is measured from treated primary human dermal fibroblasts, synoviocytes and chondrocytes. The relative levels of collagenase, gelatinase and stromalysin produced in response to culturing a target cell in the presence of a protein of interest is measured using zymogram gels (Loita and Stetler-Stevenson, *Cancer Biology* 1:96-106, 1990). Procollagen/collagen synthesis by dermal fibroblasts and chondrocytes in response to a test protein is measured using ³H-proline incorporation into nascent secreted collagen. ³H-labeled collagen is visualized by SDS-PAGE followed by autoradiography (Unemori and Amento, *J. Biol. Chem.* 265:

10681-10685, 1990). Glycosaminoglycan (GAG) secretion from dermal fibroblasts and chondrocytes is measured using a 1,9-dimethylmethylene blue dye binding assay (Farndale et al., *Biochim. Biophys. Acta* 883:173-177, 1986). Collagen and GAG assays are also carried out in the presence of IL-1 β or TGF- β to examine the ability of a
5 protein to modify the established responses to these cytokines.

Monocyte activation assays are carried out (1) to look for the ability of a protein of interest to further stimulate monocyte activation, and (2) to examine the ability of a protein of interest to modulate attachment-induced or endotoxin-induced monocyte activation (Fuhlbrigge et al., *J. Immunol.* 138: 3799-3802, 1987). IL-1 β and
10 TNF α levels produced in response to activation are measured by ELISA (Biosource, Inc. Camarillo, CA). Monocyte/macrophage cells, by virtue of CD14 (LPS receptor), are exquisitely sensitive to endotoxin, and proteins with moderate levels of endotoxin-like activity will activate these cells.

Other metabolic effects of proteins can be measured by culturing target
15 cells in the presence and absence of a protein and observing changes in adipogenesis, gluconeogenesis, glycogenolysis, lipogenesis, glucose uptake, or the like. Suitable assays are known in the art.

Hematopoietic activity of proteins can be assayed on various hematopoietic cells in culture. Preferred assays include primary bone marrow colony
20 assays and later stage lineage-restricted colony assays, which are known in the art (e.g., Holly et al., WIPO Publication WO 95/21920). Marrow cells plated on a suitable semi-solid medium (e.g., 50% methylcellulose containing 15% fetal bovine serum, 10% bovine serum albumin, and 0.6% PSN antibiotic mix) are incubated in the presence of test polypeptide, then examined microscopically for colony formation.
25 Known hematopoietic factors are used as controls. Mitogenic activity of a protein of interest on hematopoietic cell lines can be measured as disclosed above.

Cell migration is assayed essentially as disclosed by Kähler et al. (*Arteriosclerosis, Thrombosis, and Vascular Biology* 17:932-939, 1997). A protein is considered to be chemotactic if it induces migration of cells from an area of low
30 protein concentration to an area of high protein concentration. A typical assay is performed using modified Boyden chambers with a polystyrene membrane separating the two chambers (Transwell; Corning Costar Corp.). The test sample, diluted in medium containing 1% BSA, is added to the lower chamber of a 24-well plate containing Transwells. Cells are then placed on the Transwell insert that has been
35 pretreated with 0.2% gelatin. Cell migration is measured after 4 hours of incubation at 37°C. Non-migrating cells are wiped off the top of the Transwell membrane, and cells attached to the lower face of the membrane are fixed and stained with 0.1% crystal

violet. Stained cells are then extracted with 10% acetic acid and absorbance is measured at 600 nm. Migration is then calculated from a standard calibration curve. Cell migration can also be measured using the matrigel method of Grant et al. ("Angiogenesis as a component of epithelial-mesenchymal interactions" in Goldberg and Rosen, *Epithelial-Mesenchymal Interaction in Cancer*, Birkhäuser Verlag, 1995, 235-248; Baatout, *Anticancer Research* 17:451-456, 1997).

Proteins can be assayed for the ability to modulate axon guidance and growth. Suitable assays that detect changes in neuron growth patterns include, for example, those disclosed in Hastings, WIPO Publication WO 97/29189 and Walter et al., *Development* 101:685-96, 1987. Assays to measure the effects on neuron growth are well known in the art. For example, the C assay (e.g., Raper and Kapfhammer, *Neuron* 4:21-9, 1990 and Luo et al., *Cell* 75:217-27, 1993) can be used to determine collapsing activity of a protein of interest on growing neurons. Other methods that can assess protein-induced inhibition of neurite extension or divert such extension are also known. See, Goodman, *Annu. Rev. Neurosci.* 19:341-77, 1996. Conditioned media from cells expressing a protein of interest, or aggregates of such cells, can be placed in a gel matrix near suitable neural cells, such as dorsal root ganglia (DRG) or sympathetic ganglia explants, which have been co-cultured with nerve growth factor. Compared to control cells, protein-induced changes in neuron growth can be measured (as disclosed by, for example, Messersmith et al., *Neuron* 14:949-59, 1995 and Puschel et al., *Neuron* 14:941-8, 1995). Neurite outgrowth can be measured using neuronal cell suspensions grown in the presence of molecules of the present invention. See, for example, O'Shea et al., *Neuron* 7:231-7, 1991 and DeFreitas et al., *Neuron* 15:333-43, 1995.

Cell adhesion activity is assayed essentially as disclosed by LaFleur et al. (*J. Biol. Chem.* 272:32798-32803, 1997). Briefly, microtiter plates are coated with the test protein, non-specific sites are blocked with BSA, and cells (such as smooth muscle cells, leukocytes, or endothelial cells) are plated at a density of approximately 10^4 - 10^5 cells/well. The wells are incubated at 37°C (typically for about 60 minutes), then non-adherent cells are removed by gentle washing. Adhered cells are quantitated by conventional methods (e.g., by staining with crystal violet, lysing the cells, and determining the optical density of the lysate). Control wells are coated with a known adhesive protein, such as fibronectin or vitronectin.

Assays for angiogenic activity are also known in the art. For example, the effect of a protein of interest on primordial endothelial cells in angiogenesis can be assayed in the chick chorioallantoic membrane angiogenesis assay (Leung, *Science* 246:1306-1309, 1989; Ferrara, *Ann. NY Acad. Sci.* 752:246-256, 1995). Briefly, a

small window is cut into the shell of an eight-day old fertilized egg, and a test substance is applied to the chorioallantoic membrane. After 72 hours, the membrane is examined for neovascularization. Other suitable assays include microinjection of early stage quail (*Coturnix coturnix japonica*) embryos as disclosed by Drake et al. (*Proc. Natl. Acad. Sci. USA* 92:7657-7661, 1995); the rodent model of corneal neovascularization disclosed by Muthukkaruppan and Auerbach (*Science* 205:1416-1418, 1979), wherein a test substance is inserted into a pocket in the cornea of an inbred mouse; and the hamster cheek pouch assay (Höckel et al., *Arch. Surg.* 128:423-429, 1993). Induction of vascular permeability, which is indicative of angiogenic activity, is measured in assays designed to detect leakage of protein from the vasculature of a test animal (e.g., mouse or guinea pig) after administration of a test compound (Miles and Miles, *J. Physiol.* 118:228-257, 1952; Feng et al., *J. Exp. Med.* 183:1981-1986, 1996). *In vitro* assays for angiogenic activity include the tridimensional collagen gel matrix model (Pepper et al. *Biochem. Biophys. Res. Comm.* 189:824-831, 1992 and Ferrara et al., *Ann. NY Acad. Sci.* 732:246-256, 1995), which measures the formation of tube-like structures by microvascular endothelial cells; and matrigel models (Grant et al., "Angiogenesis as a component of epithelial-mesenchymal interactions" in Goldberg and Rosen, *Epithelial-Mesenchymal Interaction in Cancer*, Birkhäuser Verlag, 1995, 235-248; Baatout, *Anticancer Research* 17:451-456, 1997), which are used to determine effects on cell migration and tube formation by endothelial cells seeded in matrigel, a basement membrane extract enriched in laminin. It is preferred to carry out angiogenesis assays in the presence and absence of vascular endothelial growth factor (VEGF) to assess possible combinatorial effects. It is also preferred to use VEGF as a control within *in vivo* assays.

Receptor binding can be measured by the competition binding method of Labriola-Tompkins et al., *Proc. Natl. Acad. Sci. USA* 88:11182-11186, 1991. In an exemplary assay for IL-1 receptor binding, membranes prepared from EL-4 thymoma cells (Paganelli et al., *J. Immunol.* 138:2249-2253, 1987) are incubated in the presence of the test protein for 30 minutes at 37°C. Labeled IL-1 α or IL-1 β is then added and the incubation is continued for 60 minutes. The assay is terminated by membrane filtration. The amount of bound label is determined by conventional means (e.g., γ counter). In an alternative assay, the ability of a test protein to compete with labeled IL-1 for binding to cultured human dermal fibroblasts is measured according to the method of Dower et al. (*Nature* 324:266-268, 1986). Briefly, cells are incubated in a round-bottomed, 96-well plate in a suitable culture medium (e.g., RPMI 1640 containing 1% BSA, 0.1% Na azide, and 20 mM HEPES pH 7.4) at 8°C on a rocker platform in the presence of labeled IL-1. Various concentrations of test protein are

added. After the incubation (typically about two hours), cells are separated from unbound label by centrifuging 60- μ l aliquots through 200 μ l of phthalate oils in 400- μ l polyethylene centrifuge tubes and excising the tips of the tubes with a razor blade as disclosed by Segal and Hurwitz, *J. Immunol.* 118:1338-1347, 1977. Receptor binding assays for other cell types are known in the art. See, for example, Bowen-Pope and Ross, *Methods Enzymol.* 109:69-100, 1985.

Receptor binding can also be measured using immobilized receptors or ligand-binding receptor fragments. For example, an immobilized receptor can be exposed to its labeled ligand and unlabeled test protein, whereby a reduction in labeled ligand binding compared to a control is indicative of receptor-binding activity in the test protein. Within another format, a receptor or ligand-binding receptor fragment is immobilized on a biosensor (e.g., BIAcore™, Pharmacia Biosensor, Piscataway, NJ) and binding is determined. Antagonists of the native ligand will exhibit receptor binding but will exhibit essentially no activity in appropriate activity assays or will reduce the ligand-mediated response when combined with the native ligand. In view of the low level of receptor occupancy required to produce a response to some ligands (e.g., IL-1), a large excess of antagonist (typically a 10- to 1000-fold molar excess) may be necessary to neutralize ligand activity.

Receptor activation can be detected in target cells by: (1) measurement of adenylate cyclase activity (Salomon et al., *Anal. Biochem.* 58:541-48, 1974; Alvarez and Daniels, *Anal. Biochem.* 187:98-103, 1990); (2) measurement of change in intracellular cAMP levels using conventional radioimmunoassay methods (Steiner et al., *J. Biol. Chem.* 247:1106-13, 1972; Harper and Brooker, *J. Cyc. Nucl. Res.* 1:207-18, 1975); or (3) through use of a cAMP scintillation proximity assay (SPA) method (such as available from Amersham Corp., Arlington Heights, IL).

Proteins can be tested for serine protease activity or proteinase inhibitory activity using conventional assays. Substrate cleavage is conveniently assayed using a tetrapeptide that mimics the cleavage site of the natural substrate and which is linked, via a peptide bond, to a carboxyl-terminal para-nitro-anilide (pNA) group. The protease hydrolyzes the bond between the fourth amino acid residue and the pNA group, causing the pNA group to undergo a dramatic increase in absorbance at 405 nm. Suitable substrates can be synthesized according to known methods or obtained from commercial suppliers. Inhibitory activity is measured by adding a test sample to a reaction mixture containing enzyme and substrate, and comparing the observed enzyme activity to a control (without the test sample). A variety of such assays are known in the art, including assays measuring inhibition of trypsin, chymotrypsin, plasmin, cathepsin G, and human leukocyte elastase. See, for example,

Petersen et al., *Eur. J. Biochem.* 235:310-316, 1996. In a typical procedure, the inhibitory activity of a test compound is measured by incubating the test compound with the proteinase, then adding an appropriate substrate, typically a chromogenic peptide substrate. See, for example, Norris et al. (*Biol. Chem. Hoppe-Seyler* 371:37-42, 1990). Various concentrations of the inhibitor are incubated in the presence of trypsin, plasmin, and plasma kallikrein in a low-salt buffer at pH 7.4, 25°C. After 30 minutes, the residual enzymatic activity is measured by the addition of a chromogenic substrate (e.g., S2251 (D-Val-Leu-Lys-Nan) or S2302 (D-Pro-Phe-Arg-Nan), available from Kabi, Stockholm, Sweden) and a 30-minute incubation. Inhibition of enzyme activity is indicated by a decrease in absorbance at 405 nm or fluorescence E_m at 460 nm. From the results, the apparent inhibition constant K_i is calculated. When a serine protease is prepared as an active precursor (e.g., comprising N-terminal residues 1-109 of SEQ ID NO:2), it is activated by cleavage with a suitable protease (e.g., furin (Steiner et al., *J. Biol. Chem.* 267:23435-23438, 1992)) prior to assay. Assays of this type are well known in the art. See, for example, Lottenberg et al., *Thrombosis Research* 28:313-332, 1982; Cho et al., *Biochem.* 23:644-650, 1984; Foster et al., *Biochem.* 26:7003-7011, 1987). The inhibition of coagulation factors (e.g., factor VIIa, factor Xa) can be measured using chromogenic substrates or in conventional coagulation assays (e.g., clotting time of normal human plasma; Dennis et al., *J. Biol. Chem.* 270:25411-25417, 1995).

Blood coagulation and chromogenic assays, which can be used to detect both procoagulant, anticoagulant, and thrombolytic activities, are known in the art. For example, pro- and anticoagulant activities can be measured in a one-stage clotting assay using platelet-poor or factor-deficient plasma (Levy and Edgington, *J. Exp. Med.* 151:1232-1243, 1980; Schwartz et al., *J. Clin. Invest.* 67:1650-1658, 1981). As disclosed by Anderson et al. (*Proc. Natl. Acad. Sci. USA* 96:11189-11193, 1999), the effect of a test compound on platelet activation can be determined by a change in turbidity, and the procoagulant activity of activated platelets can be determined in a phospholipid-dependent coagulation assay. Activation of thrombin can be determined by hydrolysis of peptide p-nitroanilide substrates as disclosed by Lottenberg et al. (*Thrombosis Res.* 28:313-332, 1982). Other procoagulant, anticoagulant, and thrombolytic activities can be measured using appropriate chromogenic substrates, a variety of which are available from commercial suppliers. See, for example, Kettner and Shaw, *Methods Enzymol.* 80:826-842, 1981.

Anti-microbial activity of proteins is evaluated by techniques that are known in the art. For example, anti-microbial activity can be assayed by evaluating the sensitivity of microbial cell cultures to test agents and by evaluating the protective

effect of test agents on infected mice. See, for example, Musiek et al., *Antimicrob. Agents Chemother.* 3:40, 1973. Antiviral activity can also be assessed by protection of mammalian cell cultures. Known techniques for evaluating anti-microbial activity include, for example, Barsum et al., *Eur. Respir. J.* 8:709-714, 1995; Sandovsky-Losica et al., *J. Med. Vet. Mycol (England)* 28:279-287, 1990; Mehentee et al., *J. Gen. Microbiol (England)* 135:2181-2188, 1989; and Segal and Savage, *J. Med. Vet. Mycol.* 24:477-479, 1986. Assays specific for anti-viral activity include, for example, those described by Daher et al., *J. Virol.* 60:1068-1074, 1986.

The assays disclosed above can be modified by those skilled in the art to detect the presence of agonists and antagonists of a selected protein of interest.

Expression of a polynucleotide encoding a protein of interest in animals provides models for further study of the biological effects of overproduction or inhibition of protein activity *in vivo*. Polynucleotides and antisense polynucleotides can be introduced into test animals, such as mice, using viral vectors or naked DNA, or transgenic animals can be produced.

One *in vivo* approach for assaying proteins of the present invention utilizes viral delivery systems. Exemplary viruses for this purpose include adenovirus, herpesvirus, retroviruses, vaccinia virus, and adeno-associated virus (AAV). Adenovirus, a double-stranded DNA virus, is currently the best studied gene transfer vector for delivery of heterologous nucleic acids. For review, see Becker et al., *Meth. Cell Biol.* 43:161-89, 1994; and Douglas and Curiel, *Science & Medicine* 4:44-53, 1997. The adenovirus system offers several advantages. Adenovirus can (i) accommodate relatively large DNA inserts; (ii) be grown to high-titer; (iii) infect a broad range of mammalian cell types; and (iv) be used with many different promoters including ubiquitous, tissue specific, and regulatable promoters. Because adenoviruses are stable in the bloodstream, they can be administered by intravenous injection.

By deleting portions of the adenovirus genome, larger inserts (up to 7 kb) of heterologous DNA can be accommodated. These inserts can be incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. In an exemplary system, the essential E1 gene is deleted from the viral vector, and the virus will not replicate unless the E1 gene is provided by the host cell (e.g., the human 293 cell line). When intravenously administered to intact animals, adenovirus primarily targets the liver. If the adenoviral delivery system has an E1 gene deletion, the virus cannot replicate in the host cells. However, the host's tissue (e.g., liver) will express and process (and, if a signal sequence is present, secrete) the heterologous protein. Secreted proteins will enter the circulation in the highly vascularized liver, and effects on the infected animal can be determined.

An alternative method of gene delivery comprises removing cells from the body and introducing a vector into the cells as a naked DNA plasmid. The transformed cells are then re-implanted in the body. Naked DNA vectors are introduced into host cells by methods known in the art, including transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, use of a gene gun, or use of a DNA vector transporter. See, Wu et al., *J. Biol. Chem.* 263:14621-14624, 1988; Wu et al., *J. Biol. Chem.* 267:963-967, 1992; and Johnston and Tang, *Meth. Cell Biol.* 43:353-365, 1994.

Transgenic mice, engineered to express a gene encoding a protein of interest, and mice that exhibit a complete absence of gene function, referred to as "knockout mice" (Snouwaert et al., *Science* 257:1083, 1992), can also be generated (Lowell et al., *Nature* 366:740-742, 1993). These mice can be employed to study the gene of interest and the protein encoded thereby in an *in vivo* system. Transgenic mice are particularly useful for investigating the role of proteins in early development in that they allow the identification of developmental abnormalities or blocks resulting from the over- or underexpression of a specific factor. See also, Maisonpierre et al., *Science* 277:55-60, 1997 and Hanahan, *Science* 277:48-50, 1997. Preferred promoters for transgenic expression include promoters from metallothionein and albumin genes. As disclosed above, the human sequences provided herein can be used to clone orthologous polynucleotides, which may be preferred for use in generating transgenic and knockout animals.

Antisense methodology can be used to inhibit gene transcription to examine the effects of such inhibition *in vivo*. Polynucleotides that are complementary to a segment of a protein-encoding polynucleotide are designed to bind to the encoding mRNA and to inhibit translation of such mRNA. Such antisense oligonucleotides can also be used to inhibit expression of protein-encoding genes in cell culture.

Biological activities of test proteins can also be measured in animal models by administering the test protein, by itself or in combination with other agents, including other proteins. Using such models facilitates the assay of the test protein by itself or as an inhibitor or modulator of another agent, and also facilitates the measurement of combinatorial effects of bioactive compounds.

Anti-inflammatory activity can be tested in animal models of inflammatory disease. For example, animal models of psoriasis include the analysis of histological alterations in adult mouse tail epidermis (Hofbauer et al, *Brit. J. Dermatol.* 118:85-89, 1988; Bladon et al., *Arch Dermatol. Res.* 277:121-125, 1985). In this model, anti-psoriatic activity is indicated by the induction of a granular layer and

orthokeratosis in areas of scale between the hinges of the tail epidermis. Typically, a topical ointment comprising a test compound is applied daily for seven consecutive days, then the animal is sacrificed, and tail skin is examined histologically. An additional model is provided by grafting psoriatic human skin to congenitally athymic (nude) mice (Krueger et al., *J. Invest. Dermatol.* 64:307-312, 1975). Such grafts have been shown to retain the characteristic histology for up to eleven weeks. As in the mouse tail model, the test composition is applied to the skin at predetermined intervals for a period of one to several weeks, at which time the animals are sacrificed and the skin grafts examined histologically. A third model has been disclosed by Fretland et al. (*Inflammation* 14:727-739, 1990). Briefly, inflammation is induced in guinea pig epidermis by topically applying phorbol ester (phorbol-12-myristate-13-acetate; PMA), typically at ca. 2 g/ml in acetone, to one ear and vehicle to the contralateral ear. Test compounds are applied concurrently with the PMA, or may be given orally. Histological analysis is performed at 96 hours after application of PMA. This model duplicates many symptoms of human psoriasis, including edema, inflammatory cell diapedesis and infiltration, high LTB₄ levels and epidermal proliferation.

Cerebral ischemia can be studied in a rat model as disclosed by Relton et al. (*ibid.*) and Loddick et al. (*ibid.*).

The effect of a test protein on primordial endothelial cells in angiogenesis can be assayed in the chick chorioallantoic membrane angiogenesis assay (Leung, *Science* 246:1306-1309, 1989; Ferrara, *Ann. NY Acad. Sci.* 752:246-256, 1995). Briefly, a small window is cut into the shell of an eight-day old fertilized egg, and a test substance is applied to the chorioallantoic membrane. After 72 hours, the membrane is examined for neovascularization. Embryo microinjection of early stage quail (*Coturnix coturnix japonica*) embryos can also be used (Drake et al., *Proc. Natl. Acad. Sci. USA* 92:7657-7661, 1995). Briefly, a solution containing the protein is injected into the interstitial space between the endoderm and the splanchnic mesoderm of early-stage embryos using a micropipette and micromanipulator system. After injection, embryos are placed ventral side down on a nutrient agar medium and incubated for 7 hours at 37°C in a humidified CO₂/air mixture (10%/90%). Vascular development is assessed by microscopy of fixed, whole-mounted embryos and sections.

Stimulation of coronary collateral growth can be measured in known animal models, including a rabbit model of peripheral limb ischemia and hind limb ischemia and a pig model of chronic myocardial ischemia (Ferrara et al., *Endocrine Reviews* 18:4-25, 1997). Test proteins are assayed in the presence and absence of VEGF and basic FGF to test for combinatorial effects. These models can be modified

by the use of adenovirus or naked DNA for gene delivery as disclosed in more detail above, resulting in local expression of the test protein(s).

Angiogenic activity can also be tested in a rodent model of corneal neovascularization as disclosed by Muthukkaruppan and Auerbach, *Science* 205:1416-1418, 1979, wherein a test substance is inserted into a pocket in the cornea of an inbred mouse. For use in this assay, proteins are combined with a solid or semi-solid, biocompatible carrier, such as a polymer pellet. Angiogenesis is followed microscopically. Vascular growth into the corneal stroma can be detected in about 10 days.

Angiogenic activity can also be tested in the hamster cheek pouch assay (Höckel et al., *Arch. Surg.* 128:423-429, 1993). A test substance is injected subcutaneously into the cheek pouch, and after five days the pouch is examined under low magnification to determine the extent of neovascularization. Tissue sections can also be examined histologically.

Induction of vascular permeability is measured in assays designed to detect leakage of protein from the vasculature of a test animal (e.g., mouse or guinea pig) after administration of a test compound (Miles and Miles, *J. Physiol.* 118:228-257, 1952; Feng et al., *J. Exp. Med.* 183:1981-1986, 1996).

Wound-healing models include the linear skin incision model of Mustoe et al. (*Science* 237:1333, 1987). In a typical procedure, a 6-cm incision is made in the dorsal pelt of an adult rat, then closed with wound clips. Test substances and controls (in solution, gel, or powder form) are applied before primary closure. It is preferred to limit administration to a single application, although additional applications can be made on succeeding days by careful injection at several sites under the incision. Wound breaking strength is evaluated between 3 and 21 days post wounding. In a second model, multiple, small, full-thickness excisions are made on the ear of a rabbit. The cartilage in the ear splints the wound, removing the variable of wound contraction from the evaluation of closure. Experimental treatments and controls are applied. The geometry and anatomy of the wound site allow for reliable quantification of cell ingrowth and epithelial migration, as well as quantitative analysis of the biochemistry of the wounds (e.g., collagen content). See, Mustoe et al., *J. Clin. Invest.* 87:694, 1991. The rabbit ear model can be modified to create an ischemic wound environment, which more closely resembles the clinical situation (Ahn et al., *Ann. Plast. Surg.* 24:17, 1990). Within a third model, healing of partial-thickness skin wounds in pigs or guinea pigs is evaluated (LeGrand et al., *Growth Factors* 8:307, 1993). Experimental treatments are applied daily on or under dressings. Seven days after wounding, granulation tissue thickness is determined. This model is preferred for dose-response

studies, as it is more quantitative than other *in vivo* models of wound healing. A full thickness excision model can also be employed. Within this model, the epidermis and dermis are removed down to the panniculus carnosum in rodents or the subcutaneous fat in pigs. Experimental treatments are applied topically on or under a dressing, and can be applied daily if desired. The wound closes by a combination of contraction and cell ingrowth and proliferation. Measurable endpoints include time to wound closure, histologic score, and biochemical parameters of wound tissue. Impaired wound healing models are also known in the art (e.g., Cromack et al., *Surgery* 113:36, 1993; Pierce et al., *Proc. Natl. Acad. Sci. USA* 86:2229, 1989; Greenhalgh et al., *Amer. J. Pathol.* 136:1235, 1990). Delay or prolongation of the wound healing process can be induced pharmacologically by treatment with steroids, irradiation of the wound site, or by concomitant disease states (e.g., diabetes). Linear incisions or full-thickness excisions are most commonly used as the experimental wound. Endpoints are as disclosed above for each type of wound. Subcutaneous implants can be used to assess compounds acting in the early stages of wound healing (Broadley et al., *Lab. Invest.* 61:571, 1985; Sprugel et al., *Amer. J. Pathol.* 129: 601, 1987). Implants are prepared in a porous, relatively non-inflammatory container (e.g., polyethylene sponges or expanded polytetrafluoroethylene implants filled with bovine collagen) and placed subcutaneously in mice or rats. The interior of the implant is empty of cells, producing a "wound space" that is well-defined and separable from the preexisting tissue. This arrangement allows the assessment of cell influx and cell type as well as the measurement of vasculogenesis/angiogenesis and extracellular matrix production.

Inhibition of tumor metastasis can be assessed in mice into which cancerous cells or tumor tissue have been introduced by implantation or injection (e.g., Brown, *Advan. Enzyme Regul.* 35:293-301, 1995; Conway et al., *Clin. Exp. Metastasis* 14:115-124, 1996).

Effects on fibrinolysis can be measured in a rat model wherein the enzyme batroxobin and radiolabeled fibrinogen are administered to test animals. Inhibition of fibrinogen activation by a test compound is seen as a reduction in the circulating level of the label as compared to animals not receiving the test compound. See, Lenfors and Gustafsson, *Semin. Thromb. Hemost.* 22:335-342, 1996.

The invention further provides polypeptides that comprise an epitope-bearing portion of a protein as shown in SEQ ID NO:M, wherein M is selected from the group consisting of 2, 5, and 8. An "epitope" is a region of a protein to which an antibody can bind. See, for example, Geysen et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002, 1984. Epitopes can be linear or conformational, the latter being composed of discontinuous regions of the protein that form an epitope upon folding of

the protein. Linear epitopes are generally at least 6 amino acid residues in length. Relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for example, Sutcliffe et al., *Science* 219:660-666, 1983. Antibodies that recognize short, linear epitopes are particularly useful in analytic and diagnostic applications that employ denatured protein, such as Western blotting (Tobin, *Proc. Natl. Acad. Sci. USA* 76:4350-4356, 1979). Antibodies to short peptides may also recognize proteins in native conformation and will thus be useful for monitoring protein expression and protein isolation, and in detecting proteins in solution, such as by ELISA or in immunoprecipitation studies.

Antigenic, epitope-bearing polypeptides of the present invention are useful for raising antibodies, including monoclonal antibodies, that specifically bind to the corresponding protein. Antigenic, epitope-bearing polypeptides contain a sequence of at least six, preferably at least nine, more preferably from 15 to about 30 contiguous amino acid residues of a protein. Within certain embodiments of the invention, the polypeptides comprise 40, 50, 100, or more contiguous residues of a protein as shown in SEQ ID NO:M, up to the entire predicted mature protein or the primary translation product. It is preferred that the amino acid sequence of the epitope-bearing polypeptide is selected to provide substantial solubility in aqueous solvents, that is the sequence includes relatively hydrophilic residues, and hydrophobic residues are substantially avoided. Table 7 lists preferred hexapeptides for use as antigens. Within Table 3, each the amino termini of the hexapeptides are specified. Those skilled in the art will recognize that longer polypeptides comprising these hexapeptides can also be used and will often be preferred.

Table 3

<u>Protein</u>	<u>Hexapeptide N-termini</u>					
Zlrr7		41	40	2	1	93
Zlrr8		135	134	47	133	44
Zlrr9		19	217	63	216	235

As used herein, the term "antibodies" includes polyclonal antibodies, monoclonal antibodies, antigen-binding fragments thereof such as F(ab')₂ and Fab fragments, single chain antibodies, and the like, including genetically engineered antibodies. Non-human antibodies can be humanized by grafting only non-human CDRs onto human framework and constant regions, or by incorporating the entire non-human variable domains (optionally "cloaking" them with a human-like surface by

replacement of exposed residues, wherein the result is a "veneered" antibody). In some instances, humanized antibodies may retain non-human residues within the human variable region framework domains to enhance proper binding characteristics. Through humanizing antibodies, biological half-life may be increased, and the potential for adverse immune reactions upon administration to humans is reduced. One skilled in the art can generate humanized antibodies with specific and different constant domains (i.e., different Ig subclasses) to facilitate or inhibit various immune functions associated with particular antibody constant domains.

Alternative techniques for generating or selecting antibodies useful herein include *in vitro* exposure of lymphocytes to an immunogenic polypeptide, and selection of antibody display libraries in phage or similar vectors (for instance, through use of an immobilized or labeled polypeptide). Human antibodies can be produced in transgenic, non-human animals that have been engineered to contain human immunoglobulin genes as disclosed in WIPO Publication WO 98/24893. It is preferred that the endogenous immunoglobulin genes in these animals be inactivated or eliminated, such as by homologous recombination.

Antibodies are defined to be specifically binding if they bind to a target polypeptide with an affinity at least 10-fold greater than the binding affinity to control (non-target) polypeptide. It is preferred that the antibodies exhibit a binding affinity (K_a) of 10^6 M^{-1} or greater, preferably 10^7 M^{-1} or greater, more preferably 10^8 M^{-1} or greater, and most preferably 10^9 M^{-1} or greater. The affinity of a monoclonal antibody can be readily determined by one of ordinary skill in the art (see, for example, Scatchard, *Ann. NY Acad. Sci.* 51: 660-672, 1949).

Methods for preparing polyclonal and monoclonal antibodies are well known in the art (see for example, Hurrell, J. G. R., Ed., *Monoclonal Hybridoma Antibodies: Techniques and Applications*, CRC Press, Inc., Boca Raton, FL, 1982). As would be evident to one of ordinary skill in the art, polyclonal antibodies can be generated from a variety of warm-blooded animals such as horses, cows, goats, sheep, dogs, chickens, rabbits, mice, and rats. The immunogenicity of a polypeptide immunogen may be increased through the use of an adjuvant such as alum (aluminum hydroxide) or Freund's complete or incomplete adjuvant. Polypeptides useful for immunization also include fusion polypeptides, such as fusions of a polypeptide of interest or a portion thereof with an immunoglobulin polypeptide or with maltose binding protein. The polypeptide immunogen may be a full-length molecule or a portion thereof. If the polypeptide portion is "hapten-like", such portion may be advantageously joined or linked to a macromolecular carrier (such as keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA) or tetanus toxoid) for immunization.

A variety of assays known to those skilled in the art can be utilized to detect antibodies that specifically bind to a polypeptide of interest. Exemplary assays are described in detail in *Antibodies: A Laboratory Manual*, Harlow and Lane (Eds.), Cold Spring Harbor Laboratory Press, 1988. Representative examples of such assays
5 include concurrent immunoelectrophoresis, radio-immunoassays, radio-immunoprecipitations, enzyme-linked immunosorbent assays (ELISA), dot blot assays, Western blot assays, inhibition or competition assays, and sandwich assays.

Antibodies can be used, for example, to isolate target polypeptides by affinity purification, for diagnostic assays for determining circulating or localized
10 levels of target polypeptides, for tissue typing, for cell sorting, for screening expression libraries; for generating anti-idiotypic antibodies, and as neutralizing antibodies or as antagonists to block protein activity *in vitro* and *in vivo*.

Additionally, antibodies generated to the MSP proteins described herein can be used, either individually or in combination, to monitor these proteins in bodily
15 fluids. In particular, correlating changes in the level of expression of the MSP proteins with changes in metabolism, or progression of disease would be useful in studying metabolism disorders or monitoring the improvement or decline of diseases.

Polynucleotides and polypeptides of the present invention will additionally find use as educational tools as a laboratory practicum kits for courses
20 related to genetics and molecular biology, protein chemistry and antibody production and analysis. Due to their unique polynucleotide and polypeptide sequences molecules of MSP can be used as standards or as "unknowns" for testing purposes. For example, MSP polynucleotides can be used as aids, such as, for example, to teach a student how to prepare expression constructs for bacterial, viral, and/or mammalian expression,
25 including fusion constructs, wherein MSP are the genes to be expressed; for determining the restriction endonuclease cleavage sites of the polynucleotides; determining mRNA and DNA localizations of MSP polynucleotides in tissues (i.e., by Northern and Southern blotting as well as polymerase chain reaction); and for identifying related polynucleotides and polypeptides by nucleic acid hybridization.

30 MSP polypeptides can be used educationally as aids to teach preparation of antibodies; identifying proteins by Western blotting; protein purification; determining the weight of expressed MSP polypeptides as a ratio to total protein expressed; identifying peptide cleavage sites; coupling amino and carboxyl terminal tags; amino acid sequence analysis, as well as, but not limited to monitoring biological
35 activities of both the native and tagged protein (i.e., receptor binding, signal transduction, proliferation, and differentiation) *in vitro* and *in vivo*. MSP polypeptides can also be used to teach analytical skills such as mass spectrometry, circular

dichroism to determine conformation, in particular the locations of the disulfide bonds, x-ray crystallography to determine the three-dimensional structure in atomic detail, nuclear magnetic resonance spectroscopy to reveal the structure of proteins in solution. For example, a kit containing the MSP can be given to the student to analyze. Since the amino acid sequences would be known by the professor, the proteins can be given to the student as a test to determine the skills or develop the skills of the student, the teacher would then know whether or not the student has correctly analyzed the polypeptides. Since every polypeptide is unique, the educational utility of each MSP would be unique unto itself.

The antibodies which bind specifically to a MSP can be used as a teaching aid to instruct students how to prepare affinity chromatography columns to purify MSP, cloning and sequencing the polynucleotide that encodes an antibody and thus as a practicum for teaching a student how to design humanized antibodies. The MSP gene, s polypeptides or antibodies would then be packaged by reagent companies and sold to universities so that the students gain skill in art of molecular biology. Because each gene and protein is unique, each gene and protein creates unique challenges and learning experiences for students in a lab practicum. Such educational kits containing the MSP genes, polypeptides or antibodies are considered within the scope of the present invention.

The present invention also provides reagents for use in diagnostic and therapeutic applications. Such reagents include polynucleotide probes and primers; antibodies, including antibody fragments, single-chain antibodies, and other genetically engineered forms; soluble receptors and other polypeptide binding partners; and the proteins of the invention themselves, including fragments thereof. Those skilled in the art will recognize that diagnostic reagents will commonly be labeled to provide a detectable signal or other second function. Thus, polypeptides, antibodies, receptors, and other binding partners disclosed herein can be directly or indirectly conjugated to drugs, toxins, radionuclides, enzymes, enzyme substrates, cofactors, inhibitors, fluorescent markers, chemiluminescent markers, magnetic particles, and the like, and these conjugates used for *in vivo* diagnostic or therapeutic applications. Cytotoxic molecules, for example, can be directly or indirectly attached to the binding partner (e.g., by chemical coupling or as a fusion protein), and include bacterial or plant toxins (e.g., diphtheria toxin, *Pseudomonas* exotoxin, ricin, saporin, abrin, and the like); therapeutic radionuclides (e.g., iodine-131, rhenium-188 or yttrium-90) which can be directly attached to a polypeptide or antibody or indirectly attached through means of a chelating moiety; and cytotoxic drugs (e.g., adriamycin). Methods for preparing labeled reagents are known in the art. Within an alternative embodiment, the

detectable signal or other function can be provided by a second member of a complement-anticomplement pair, which second member binds to the diagnostic reagent. For example, a first (unlabeled) antibody can be used to bind to a cell-surface polypeptide, after which a second, labeled antibody which binds to the first antibody is added. Other complement-anticomplement pairs are known in the art and include biotin/streptavidin.

Diagnostic reagents as disclosed herein can be used *in vivo* or *in vitro*. *In vitro* diagnostic assays include assays of tissue and fluid samples. Assays for protein in serum, for example, may be used to detect metabolic abnormalities characterized by over- or under-production of the protein, such as cancers, immune system abnormalities, infections, organ failure, metabolic imbalances, inborn errors of metabolism and other disease states. Proteins of the present invention can also be used in the detection of circulating autoantibodies, which are indicative of autoimmune disorders. Those skilled in the art will recognize that conditions related to protein underexpression or overexpression may be amenable to treatment by therapeutic manipulation of the relevant protein level(s). Proteins in serum can be quantitated by known methods known in the art, which include the use of antibodies in a variety of formats. Non-antibody binding partners, such as ligand-binding receptor fragments (commonly referred to as "soluble receptors") can also be used.

In general, diagnostic methods employing oligonucleotide probes or primers comprise the steps of (a) obtaining a genetic sample from a patient; (b) incubating the genetic sample with an oligonucleotide probe or primer as disclosed above, under conditions wherein the probe or primer will hybridize to a complementary polynucleotide sequence, to produce a first reaction product; and (c) comparing the first reaction product to a control reaction product. A difference between the first reaction product and the control reaction product is indicative of a genetic abnormality in the patient. Genetic samples for use within such methods include genomic DNA, cDNA, and RNA. Suitable assay methods in this regard include molecular genetic techniques known to those in the art, such as restriction fragment length polymorphism (RFLP) analysis, short tandem repeat (STR) analysis employing PCR techniques, ligation chain reaction (Barany, *PCR Methods and Applications* 1:5-16, 1991), ribonuclease protection assays, and other genetic linkage analysis techniques known in the art (Sambrook et al., *ibid.*; Ausubel et. al., *ibid.*; A.J. Marian, *Chest* 108:255-65, 1995). Ribonuclease protection assays (see, e.g., Ausubel et al., *ibid.*, ch. 4) comprise the hybridization of an RNA probe to a patient RNA sample, after which the reaction product (RNA-RNA hybrid) is exposed to RNase. Hybridized regions of the RNA are protected from digestion. Within PCR assays, a patient genetic sample is incubated

with a pair of oligonucleotide primers, and the region between the primers is amplified and recovered. Changes in size, amount, or sequence of recovered product are indicative of mutations in the patient. Another PCR-based technique that can be employed is single strand conformational polymorphism (SSCP) analysis (Hayashi, *5 PCR Methods and Applications* 1:34-38, 1991).). Chromosomal localization data can be used to correlate MSP gene locations with known genetic disorders using, for example, the OMIM™ Database, Johns Hopkins University, 2000 (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>).

Polynucleotides of the present invention, including fragments thereof, can also be used for radiation hybrid mapping, a somatic cell genetic technique developed for constructing high-resolution, contiguous maps of mammalian chromosomes (Cox et al., *Science* 250:245-50, 1990). Partial or full knowledge of a gene's sequence allows the design of PCR primers suitable for use with chromosomal radiation hybrid mapping panels. Commercially available radiation hybrid mapping panels which cover the entire human genome, such as the Stanford G3 RH Panel and the GeneBridge 4 RH Panel (Research Genetics, Inc., Huntsville, AL), are available. These panels enable rapid, PCR-based chromosomal localizations and ordering of genes, sequence-tagged sites (STSs), and other nonpolymorphic and polymorphic markers within a region of interest, allowing the establishment of directly proportional physical distances between newly discovered genes of interest and previously mapped markers. The precise knowledge of a gene's position can be useful for a number of purposes, including: 1) determining if a sequence is part of an existing contig and obtaining additional surrounding genetic sequences in various forms, such as YACs, BACs or cDNA clones; 2) providing a possible candidate gene for an inheritable disease which shows linkage to the same chromosomal region; and 3) cross-referencing model organisms, such as mouse, which may aid in determining what function a particular gene might have.

If a mammal has an insufficiency of a protein of interest (due to, for example, a mutated or absent gene), the corresponding wild-type gene can be introduced into the cells of the mammal. In one embodiment, a gene encoding a protein of interest is introduced into the animal using a viral vector. Such vectors include an attenuated or defective DNA virus, such as, but not limited to, herpes simplex virus (HSV), papillomavirus, Epstein Barr virus (EBV), adenovirus, adeno-associated virus (AAV), and the like. Defective viruses, which entirely or almost entirely lack viral genes, are preferred. A defective virus is not infective after introduction into a cell. Use of defective viral vectors allows for administration to cells in a specific, localized area, without concern that the vector can infect other cells.

Examples of particular vectors include, but are not limited to, a defective herpes simplex virus 1 (HSV1) vector (Kaplitt et al., *Molec. Cell. Neurosci.* 2:320-30, 1991); an attenuated adenovirus vector, such as the vector described by Stratford-Perricaudet et al. (*J. Clin. Invest.* 90:626-30, 1992); and a defective adeno-associated virus vector (Samulski et al., *J. Virol.* 61:3096-101, 1987; Samulski et al., *J. Virol.* 63:3822-28, 1989).

Within another embodiment, a gene of interest is introduced into an animal by liposome-mediated transfection ("lipofection") essentially as disclosed above. Lipofection can be used to introduce exogenous genes into specific organs.

10 A gene of interest can also be introduced into an animal for gene therapy as a naked DNA plasmid using the methods disclosed above.

In another embodiment, polypeptide-toxin fusion proteins or antibody/fragment-toxin fusion proteins may be used for targeted cell or tissue inhibition or ablation, such as in cancer therapy. Of particular interest in this regard are
15 conjugates of an MSP protein and a cytotoxin, which can be used to target the cytotoxin to a tumor or other tissue that is undergoing undesired angiogenesis or neovascularization.

In another embodiment, MSP-cytokine fusion proteins or antibody/fragment-cytokine fusion proteins may be used for enhancing *in vitro* cytotoxicity (for instance, that mediated by monoclonal antibodies against tumor
20 targets) and for enhancing *in vivo* killing of target tissues (for example, blood and bone marrow cancers). See, generally, Hornick et al., *Blood* 89:4437-4447, 1997). In general, cytokines are toxic if administered systemically. The described fusion proteins enable targeting of a cytokine to a desired site of action, such as a cell having
25 binding sites for an MSP protein, thereby providing an elevated local concentration of cytokine. Polypeptides, antibodies, or receptors target an undesirable cell or tissue (e.g., a tumor), and the fused cytokine mediates improved target cell lysis by effector cells. Suitable cytokines for this purpose include, for example, interleukin-2 and granulocyte-macrophage colony-stimulating factor (GM-CSF).

30 In another embodiment, polypeptide-toxin fusion proteins or other binding partner-linked toxins may be used for targeted cell or tissue inhibition or ablation (for instance, to treat cancer cells or tissues). Target cells (i.e., those displaying a receptor for a polypeptide of interest) bind the polypeptide-toxin conjugate, which is then internalized, killing the cell. The effects of receptor-specific
35 cell killing (target ablation) are revealed by changes in whole animal physiology or through histological examination. Thus, ligand-dependent, receptor-directed cytotoxicity can be used to enhance understanding of the physiological significance of a

protein ligand. A preferred such toxin is saporin. Mammalian cells have no receptor for saporin, which is non-toxic when it remains extracellular. Alternatively, if the polypeptide of interest has multiple functional domains (i.e., an activation domain or a ligand binding domain, plus a targeting domain), a fusion protein including only the targeting domain may be suitable for directing a detectable molecule, a cytotoxic molecule or a complementary molecule to a cell or tissue type of interest. In instances where the domain-only fusion protein includes a complementary molecule, the anti-complementary molecule can be conjugated to a detectable or cytotoxic molecule. Such domain-complementary molecule fusion proteins thus represent a generic targeting vehicle for cell- or tissue-specific delivery of generic anti-complementary-detectable/cytotoxic molecule conjugates.

The bioactive conjugates described herein can be delivered intravenously, intraarterially or intraductally, or may be introduced locally at the intended site of action.

Due to the particular characteristics of the Zlrr7, Zlrr8, and Zlrr9, polypeptides and polynucleotides of the present invention will find use in diagnosing and treating a variety of disorders related to abnormal cell growth. These include, without limitation, retinoblastoma renal cell adenocarcinoma, endometrial adenocarcinoma, glioblastoma, neuroblastoma, B-cell lymphotic leukemia, kidney tumors (clear cell type), germ cell tumors, lung large cell carcinoma, mammary adenocarcinoma, colon adenocarcinoma, genitourinary tract transitional cell tumors, rhabdomyosarcoma, lung tumor (squamous cell carcinoma), bladder tumor, esophagus tumor, pancreas adenocarcinoma, and prostate adenocarcinoma.

For pharmaceutical use, the proteins of the present invention are formulated according to conventional methods. Routes of delivery include topical, mucosal, and parenteral, the latter including intravenous and subcutaneous delivery. Intravenous administration will be by bolus injection or infusion over a typical period of one to several hours. In general, pharmaceutical formulations will include a protein of the present invention in combination with a pharmaceutically acceptable vehicle, such as saline, buffered saline, 5% dextrose in water or the like. Formulations may further include one or more excipients, diluents, fillers, emulsifiers, preservatives, solubilizers, buffering agents, wetting agents, stabilizers, colorings, penetration enhancers, albumin to prevent protein loss on vial surfaces, etc. Topical formulations are typically provided as liquids, ointments, salves, gels, emulsions and the like. Methods of formulation are well known in the art and are disclosed, for example, in *Remington: The Science and Practice of Pharmacy*, Gennaro, ed., Mack Publishing Co., Easton, PA, 19th ed., 1995. Therapeutic doses will be determined by the clinician

according to accepted standards, taking into account the nature and severity of the condition to be treated, patient traits, etc. Proteins of the present invention will generally be formulated to provide a dose of from 0.01 μ g to 100 mg per kg patient weight per day, more commonly from 0.1 μ g to 10 mg/kg/day, still more commonly
5 from 0.1 μ g to 1.0 mg/kg/day. Determination of dose is within the level of ordinary skill in the art. The proteins may be administered for acute treatment, over one week or less, often over a period of one to three days or may be used in chronic treatment, over several months or years. In general, a therapeutically effective amount is an amount sufficient to produce a clinically significant change in the targeted condition.

10 Within the laboratory research field, the proteins of the present invention can be used as molecular weight standards, or as standards in the analysis of cell phenotype, and as reagents for the study of cells, receptors, and other binding molecules. Such reagents will generally further comprise a second moiety, such as a label, binding partner, or toxin, that facilitates the detection of the protein when bound
15 to its target. Many such systems are known in the art and are summarized above. Receptors and other cell-surface binding sites for proteins of the present invention can be identified by exposing a population of cells to a labeled protein under physiologic conditions, whereby the protein binds to the surface of the cell. Cells bearing receptors for a protein of interest can also be identified using the protein joined to a toxin,
20 whereby receptor-bearing cells are killed by the toxin.

MSP proteins and antagonists thereof can be used as standards in assays of protein and protein inhibitors in both clinical and research settings. Such assays can comprise any of a number of standard formats, include radioreceptor assays and ELISAs. Protein standards can be prepared in labeled form using a radioisotope,
25 enzyme, fluorophore, or other compound that produces a detectable signal. The proteins can be packaged in kit form, such kits comprising one or more vials containing the MSP protein and, optionally, a diluent, an antibody, a labeled binding protein, etc. Assay kits can be used in the research laboratory to detect protein and inhibitor activities produced by cultured cells or test animals.

30 Proteins of the present invention may also be used as protein and amino acid supplements, including hydrolysates. Specific uses in this regard include use as animal feed supplements and as cell culture components. Proteins rich in a particular amino acid can be used as a source of that amino acid.

The invention is further illustrated by the following non-limiting
35 examples.

EXAMPLES

Example 1

A protein of the present invention ("MSP") is produced in *E. coli* using a His₆ tag/maltose binding protein (MBP) double affinity fusion system as generally disclosed by Pryor and Leiting, *Prot. Expr. Pur.* 10:309-319, 1997. A thrombin cleavage site is placed at the junction between the affinity tag and MSP sequences.

The fusion construct is assembled in the vector pTAP98, which comprises sequences for replication and selection in *E. coli* and yeast, the *E. coli* tac promoter, and a unique SmaI site just downstream of the MBP-His₆-thrombin site coding sequences. The MSP cDNA is amplified by PCR using primers each comprising 40 bp of sequence homologous to vector sequence and 25 bp of sequence that anneals to the cDNA. The reaction is run using Taq DNA polymerase (Boehringer Mannheim, Indianapolis, IN) for 30 cycles of 94°C, 30 seconds; 60°C, 60 seconds; and 72°C, 60 seconds. One microgram of the resulting fragment is mixed with 100 ng of SmaI-cut pTAP98, and the mixture is transformed into yeast to assemble the vector by homologous recombination (Oldenburg et al., *Nucl. Acids. Res.* 25:451-452, 1997). Ura⁺ transformants are selected.

Plasmid DNA is prepared from yeast transformants and transformed into *E. coli* MC1061. Pooled plasmid DNA is then prepared from the MC1061 transformants by the miniprep method after scraping an entire plate. Plasmid DNA is analyzed by restriction digestion.

E. coli strain BL21 is used for expression of MSP. Cells are transformed by electroporation and grown on minimal glucose plates containing casamino acids and ampicillin.

Protein expression is analyzed by gel electrophoresis. Cells are grown in liquid glucose media containing casamino acids and ampicillin. After one hour at 37°C, IPTG is added to a final concentration of 1mM, and the cells are grown for an additional 2-3 hours at 37°C. Cells are disrupted using glass beads, and extracts are prepared.

Example 2

Larger scale cultures of MSP transformants are prepared by the method of Pryor and Leiting (*ibid.*). 100-ml cultures in minimal glucose media containing casamino acids and 100 µg/ml ampicillin are grown at 37°C in 500-ml baffled flasks to OD₆₀₀ ≈ 0.5. Cells are harvested by centrifugation and resuspended in 100 ml of the same media at room temperature. After 15 minutes, IPTG is added to 0.5 mM, and cultures are incubated at room temperature (ca. 22.5°C) for 16 to 20 hours with shaking

at 125 rpm. The culture is harvested by centrifugation, and cell pellets are stored at -70°C.

Example 3

5 For larger-scale protein preparation, 500-ml cultures of *E. coli* BL21 expressing the MSP-MBP-His₆ fusion protein are prepared essentially as disclosed in Example 2. Cell pellets are resuspended in 100 ml of binding buffer (20 mM Tris, pH 7.58, 100 mM NaCl, 20 mM NaH₂PO₄, 0.4 mM 4-(2-Aminoethyl)-benzenesulfonyl fluoride hydrochloride [Pefabloc® SC; Boehringer-Mannheim], 2 µg/ml Leupeptin, 2
10 µg/ml Aprotinin). The cells are lysed in a French press at 30,000 psi, and the lysate is centrifuged at 18,000 x g for 45 minutes at 4°C to clarify it. Protein concentration is estimated by gel electrophoresis with a BSA standard.

Recombinant MSP fusion protein is purified from the lysate by affinity chromatography. Immobilized cobalt resin (Talon® resin; Clontech Laboratories, Inc.,
15 Palo Alto, CA) is equilibrated in binding buffer. One ml of packed resin per 50 mg protein is combined with the clarified supernatant in a tube, and the tube is capped and sealed, then placed on a rocker overnight at 4°C. The resin is then pelleted by centrifugation at 4°C and washed three times with binding buffer. Protein is eluted with binding buffer containing 0.2 M imidazole. The resin and elution buffer are
20 mixed for at least one hour at 4°C, the resin is pelleted, and the supernatant is removed. An aliquot is analyzed by gel electrophoresis, and concentration is estimated. Amylose resin is equilibrated in amylose binding buffer (20 mM Tris-HCl, pH 7.0, 100 mM NaCl, 10 mM EDTA) and combined with the supernatant from the Talon resin at a ratio of 2 mg fusion protein per ml of resin. Binding and washing steps are carried out
25 as disclosed above. Protein is eluted with amylose binding buffer containing 10 mM maltose using as small a volume as possible to minimize the need for subsequent concentration. The eluted protein is analyzed by gel electrophoresis and staining with Coomassie blue using a BSA standard, and by Western blotting using an anti-MBP antibody.

30

Example 4

An expression plasmid containing all or part of a polynucleotide encoding MSP is constructed via homologous recombination. An MSP coding sequence comprising the ORF with 5' and 3' ends corresponding to the vector
35 sequences flanking the insertion point is prepared by PCR. The primers for PCR each include from 5' to 3' end: 40 bp of flanking sequence from the vector and 17 bp corresponding to the amino or carboxyl termini from the open reading frame of MSP.

Ten μ l of the 100 μ l PCR reaction mixture is run on a 0.8% low-melting-temperature agarose (SeaPlaque GTG®; FMC BioProducts, Rockland, ME) gel with 1 x TBE buffer for analysis. The remaining 90 μ l of the reaction mixture is precipitated with the addition of 5 μ l 1 M NaCl and 250 μ l of absolute ethanol. The
5 plasmid pZMP6, which has been cut with SmaI, is used for recombination with the PCR fragment. Plasmid pZMP6 is a mammalian expression vector containing an expression cassette having the cytomegalovirus immediate early promoter, multiple restriction sites for insertion of coding sequences, a stop codon, and a human growth hormone terminator; an *E. coli* origin of replication; a mammalian selectable marker
10 expression unit comprising an SV40 promoter, enhancer and origin of replication, a DHFR gene, and the SV40 terminator; and URA3 and CEN-ARS sequences required for selection and replication in *S. cerevisiae*. It was constructed from pZP9 (deposited at the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209, under Accession No. 98668) with the yeast genetic elements taken from
15 pRS316 (available from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA, under Accession No. 77145), an internal ribosome entry site (IRES) element from poliovirus, and the extracellular domain of CD8 truncated at the C-terminal end of the transmembrane domain.

One hundred microliters of competent yeast (*S. cerevisiae*) cells are
20 independently combined with 10 μ l of the various DNA mixtures from above and transferred to a 0.2-cm electroporation cuvette. The yeast/DNA mixtures are electropulsed using power supply (BioRad Laboratories, Hercules, CA) settings of 0.75 kV (5 kV/cm), ∞ ohms, 25 μ F. To each cuvette is added 600 μ l of 1.2 M sorbitol, and the yeast is plated in two 300- μ l aliquots onto two URA-D plates (1.8% agar in 2% D-glucose, 0.67% yeast nitrogen base without amino acids, 0.056% -Ura -Trp -Thr
25 powder [made by combining 4.0 g L-adenine, 3.0 g L-arginine, 5.0 g L-aspartic acid, 2.0 g L-histidine, 6.0 g L-isoleucine, 8.0 g L-leucine, 4.0 g L-lysine, 2.0 g L-methionine, 6.0 g L-phenylalanine, 5.0 g L-serine, 5.0 g L-tyrosine, and 6.0 g L-valine], and 0.5% 200X tryptophan, threonine solution [3.0% L-threonine, 0.8% L-tryptophan in H₂O]) and incubated at 30°C. After about 48 hours, the Ura⁺ yeast
30 transformants from a single plate are resuspended in 1 ml H₂O and spun briefly to pellet the yeast cells. The cell pellet is resuspended in 1 ml of lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA). Five hundred microliters of the lysis mixture is added to an Eppendorf tube containing 300 μ l acid-washed glass beads and 200 μ l phenol-chloroform, vortexed for 1 minute intervals two or three times, and spun for 5 minutes in an Eppendorf centrifuge at maximum speed. Three hundred microliters of the aqueous phase is transferred to a fresh tube, and the

DNA is precipitated with 600 μ l ethanol (EtOH), followed by centrifugation for 10 minutes at 4°C. The DNA pellet is resuspended in 10 μ l H₂O.

Transformation of electrocompetent *E. coli* host cells (Electromax DH10B™ cells; obtained from Life Technologies, Inc., Gaithersburg, MD) is done with 0.5-2 ml yeast DNA prep and 40 μ l of cells. The cells are electropulsed at 1.7 kV, 25 μ F, and 400 ohms. Following electroporation, 1 ml SOC (2% Bacto™ Tryptone (Difco, Detroit, MI), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose) is plated in 250- μ l aliquots on four LB AMP plates (LB broth (Lennox), 1.8% Bacto™ Agar (Difco), 100 mg/L Ampicillin).

Individual clones harboring the correct expression construct for MSP are identified by restriction digest to verify the presence of the MSP insert and to confirm that the various DNA sequences have been joined correctly to one another. The inserts of positive clones are subjected to sequence analysis. Larger scale plasmid DNA is isolated using a commercially available kit (QIAGEN Plasmid Maxi Kit, Qiagen, Valencia, CA) according to manufacturer's instructions. The correct construct is designated pZMP6/MSP.

Recombinant protein is produced in BHK cells transfected with pZMP6/MSP. BHK 570 cells (ATCC CRL-10314) are plated in 10-cm tissue culture dishes and allowed to grow to approximately 50 to 70% confluence overnight at 37°C, 5% CO₂, in DMEM/FBS media (DMEM, Gibco/BRL High Glucose; Life Technologies), 5% fetal bovine serum (Hyclone, Logan, UT), 1 mM L-glutamine (JRH Biosciences, Lenexa, KS), 1 mM sodium pyruvate (Life Technologies). The cells are then transfected with pZMP6/MSP by liposome-mediated transfection using a 3:1 (w/w) liposome formulation of the polycationic lipid 2,3-dioleoyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propaniminium-trifluoroacetate and the neutral lipid dioleoyl phosphatidylethanolamine in membrane-filtered water (Lipofectamine™ Reagent; Life Technologies, Gaithersburg, MD), in serum free (SF) media (DMEM supplemented with 10 mg/ml transferrin, 5 mg/ml insulin, 2 mg/ml fetuin, 1% L-glutamine and 1% sodium pyruvate). The plasmid is diluted into 15-ml tubes to a total final volume of 640 μ l with SF media. 35 μ l of the lipid mixture is mixed with 605 μ l of SF medium, and the resulting mixture is allowed to incubate approximately 30 minutes at room temperature. Five milliliters of SF media is then added to the DNA:lipid mixture. The cells are rinsed once with 5 ml of SF media, aspirated, and the DNA:lipid mixture is added. The cells are incubated at 37°C for five hours, then 6.4 ml of DMEM/10% FBS, 1% PSN media is added to each plate. The plates are incubated at 37°C overnight, and the DNA:lipid mixture is replaced with fresh 5% FBS/DMEM media the next day. On day 5 post-transfection, the cells are

split into T-162 flasks in selection medium (DMEM + 5% FBS, 1% L-Gln, 1% NaPyr, 1 μ M methotrexate). Approximately 10 days post-transfection, two 150-mm culture dishes of methotrexate-resistant colonies from each transfection are trypsinized, and the cells are pooled and plated into a T-162 flask and transferred to large-scale culture.

5

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

We claim:

1. An isolated polypeptide comprising residues 27 to 234 as shown in SEQ ID NO:2.
2. The isolated polypeptide of claim 1 wherein the polypeptide further comprises residues 1 to 234 as shown in SEQ ID NO:2.
3. An isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes the isolated polypeptide of claim 1.
4. An expression vector comprising the following operably linked elements:
 - a transcription promoter;
 - a DNA segment having the isolated polynucleotide of claim 3; and
 - a transcription terminator.
5. A cultured cell comprising the expression vector of claim 4.
6. A method of producing a polypeptide comprising culturing the cell of claim 5 under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide.
7. A polypeptide produced by the method of claim 6.
8. An antibody that specifically binds to the isolated protein of claim 1.
9. An isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:1.
10. An isolated polypeptide comprising residues 16 to 279 as shown in SEQ ID NO:8.
11. The isolated polypeptide of claim 10 wherein the polypeptide further comprises residues 16 to 487 as shown in SEQ ID NO:8.

12. The isolated polypeptide of claim 10 wherein the polypeptide further comprises residues 1 to 487 as shown in SEQ ID NO:8.

13. An isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes the isolated polypeptide of claim 10.

14. An expression vector comprising the following operably linked elements:

a transcription promoter;
a DNA segment having the isolated polynucleotide of claim 13; and
a transcription terminator.

15. A cultured cell comprising the expression vector of claim 14.

16. A method of producing a polypeptide comprising culturing the cell of claim 15 under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide.

17. A polypeptide produced by the method of claim 16.

18. An antibody that specifically binds to the isolated protein of claim 10.

19. An isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:7.

20. An isolated polypeptide comprising residues 16 to 144 as shown in SEQ ID NO:5.

21. The isolated polypeptide of claim 20 wherein the polypeptide further comprises residues 1 to 144 as shown in SEQ ID NO:8.

22. An isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes the isolated polypeptide of claim 20.

23. An expression vector comprising the following operably linked elements:

a transcription promoter;
a DNA segment having the isolated polynucleotide of claim 22; and
a transcription terminator.

24. A cultured cell comprising the expression vector of claim 23.

25. A method of producing a polypeptide comprising culturing the cell of claim 24 under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide.

26. A polypeptide produced by the method of claim 25.

27. An antibody that specifically binds to the isolated protein of claim 20.

28. An isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:4.

29. An isolated polypeptide comprising residues 76 to 363 as shown in SEQ ID NO:14.

30. The isolated polypeptide of claim 29 wherein the polypeptide further comprises residues 76 to 663 as shown in SEQ ID NO:14.

31. The isolated polypeptide of claim 29 wherein the polypeptide further comprises residues 1 to 663 as shown in SEQ ID NO:14.

32. An isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes the isolated polypeptide of claim 29.

33. An expression vector comprising the following operably linked elements:

a transcription promoter;
a DNA segment having the isolated polynucleotide of claim 32; and
a transcription terminator.

34. A cultured cell comprising the expression vector of claim 33.

35. A method of producing a polypeptide comprising culturing the cell of claim 34 under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide.

36. A polypeptide produced by the method of claim 35.

37. An antibody that specifically binds to the isolated protein of claim 29.

38. An isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:13.

39. An isolated polypeptide comprising residues 76 to 278 as shown in SEQ ID NO:11.

40. The isolated polypeptide of claim 39 wherein the polypeptide further comprises residues 1 to 278 as shown in SEQ ID NO:11.

41. An isolated polynucleotide comprising a sequence of nucleotides, wherein the sequence encodes the isolated polypeptide of claim 39.

42. An expression vector comprising the following operably linked elements:

a transcription promoter;

a DNA segment having the isolated polynucleotide of claim 39; and

a transcription terminator.

43. A cultured cell comprising the expression vector of claim 42.

44. A method of producing a polypeptide comprising culturing the cell of claim 43 under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide.

45. A polypeptide produced by the method of claim 44.

46. An antibody that specifically binds to the isolated protein of claim 39.

47. An isolated polynucleotide comprising the polynucleotide sequence as shown in SEQ ID NO:10.

SEQUENCE LISTING

<110> ZymoGenetics, Inc.

<120> Leucine-Rich Repeat Proteins, Zlrr7,
Zlrr8, and Zlrr9

<130> 01-27PC

<150> 60/215,446

<151> 2000-06-30

<160> 15

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 768

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(768)

<400> 1

atg cgc ccc cga gcc cca gcc tgc gcc gcc gcg gcg ctc ggg ctc tgc	48
Met Arg Pro Arg Ala Pro Ala Cys Ala Ala Ala Ala Leu Gly Leu Cys	
1 5 10 15	
agc ctt ctg ctg ctg ctc gcg ccc ggg cac gcg tgc ccc gcg ggc tgc	96
Ser Leu Leu Leu Leu Leu Ala Pro Gly His Ala Cys Pro Ala Gly Cys	
20 25 30	
gcc tgc acc gac ccg cac acc gtg gac tgc cgc gac cgc ggg ctg ccc	144
Ala Cys Thr Asp Pro His Thr Val Asp Cys Arg Asp Arg Gly Leu Pro	
35 40 45	
agc gtg cca gac cct ttc ccc ctg gac gtg cgc aag ctg ctg gtg gcc	192
Ser Val Pro Asp Pro Phe Pro Leu Asp Val Arg Lys Leu Leu Val Ala	
50 55 60	

ggc aac cgc atc cag cgg atc ccc gag gac ttc ttc atc ttc tac ggc	240
Gly Asn Arg Ile Gln Arg Ile Pro Glu Asp Phe Phe Ile Phe Tyr Gly	
65 70 75 80	
gac ctg gtc tac ctg gac ttc agg aac aac tcg ctg cgc tcg ctg gag	288
Asp Leu Val Tyr Leu Asp Phe Arg Asn Asn Ser Leu Arg Ser Leu Glu	
85 90 95	
gag ggc acg ttc agc ggc tcg gcc aag ctc gtg ttc ctc gac ctc agc	336
Glu Gly Thr Phe Ser Gly Ser Ala Lys Leu Val Phe Leu Asp Leu Ser	
100 105 110	
tac aac aac ttg acc cag ctg ggc gcc ggc gcc ttc cgc tcg gcc ggc	384
Tyr Asn Asn Leu Thr Gln Leu Gly Ala Gly Ala Phe Arg Ser Ala Gly	
115 120 125	
agg ctg gtg aag ctt agc ctg gct aac aac aac ctg gtg ggc gtg cac	432
Arg Leu Val Lys Leu Ser Leu Ala Asn Asn Asn Leu Val Gly Val His	
130 135 140	
gag gac gcc ttc gag acc ctg gag tcg ctg cag gtg ctg gag ctc aac	480
Glu Asp Ala Phe Glu Thr Leu Glu Ser Leu Gln Val Leu Glu Leu Asn	
145 150 155 160	
gac aac aac ctg cgc agc ctc agc gtg gcc gcc ctg gcc gcg ctg ccc	528
Asp Asn Asn Leu Arg Ser Leu Ser Val Ala Ala Leu Ala Ala Leu Pro	
165 170 175	
gcg ctg cgc tcc ctg cgt ctg gac ggg aac ccc tgg ctg tgc gac tgt	576
Ala Leu Arg Ser Leu Arg Leu Asp Gly Asn Pro Trp Leu Cys Asp Cys	
180 185 190	
gac ttc gcc cac ctc ttc tcc tgg atc cag gag aac gca tcc aaa ctg	624
Asp Phe Ala His Leu Phe Ser Trp Ile Gln Glu Asn Ala Ser Lys Leu	
195 200 205	
ccc aaa gga ctg gcg ggt gtg gat tac tta tgc gtc cct ggt aag cgg	672
Pro Lys Gly Leu Ala Gly Val Asp Tyr Leu Cys Val Pro Gly Lys Arg	
210 215 220	

aat gca gcc tac tct atg gga aac ggc cgt att ctc agt acc gtg cac 720
 Asn Ala Ala Tyr Ser Met Gly Asn Gly Arg Ile Leu Ser Thr Val His
 225 230 235 240

ggg gag tca gcc agt tcc aag ggc tct cca gca gct tcc cga gcc taa 768
 Gly Glu Ser Ala Ser Ser Lys Gly Ser Pro Ala Ala Ser Arg Ala *
 245 250 255

<210> 2

<211> 255

<212> PRT

<213> Homo sapiens

<400> 2

Met Arg Pro Arg Ala Pro Ala Cys Ala Ala Ala Ala Leu Gly Leu Cys
 1 5 10 15
 Ser Leu Leu Leu Leu Leu Ala Pro Gly His Ala Cys Pro Ala Gly Cys
 20 25 30
 Ala Cys Thr Asp Pro His Thr Val Asp Cys Arg Asp Arg Gly Leu Pro
 35 40 45
 Ser Val Pro Asp Pro Phe Pro Leu Asp Val Arg Lys Leu Leu Val Ala
 50 55 60
 Gly Asn Arg Ile Gln Arg Ile Pro Glu Asp Phe Phe Ile Phe Tyr Gly
 65 70 75 80
 Asp Leu Val Tyr Leu Asp Phe Arg Asn Asn Ser Leu Arg Ser Leu Glu
 85 90 95
 Glu Gly Thr Phe Ser Gly Ser Ala Lys Leu Val Phe Leu Asp Leu Ser
 100 105 110
 Tyr Asn Asn Leu Thr Gln Leu Gly Ala Gly Ala Phe Arg Ser Ala Gly
 115 120 125
 Arg Leu Val Lys Leu Ser Leu Ala Asn Asn Asn Leu Val Gly Val His
 130 135 140
 Glu Asp Ala Phe Glu Thr Leu Glu Ser Leu Gln Val Leu Glu Leu Asn
 145 150 155 160
 Asp Asn Asn Leu Arg Ser Leu Ser Val Ala Ala Leu Ala Ala Leu Pro
 165 170 175
 Ala Leu Arg Ser Leu Arg Leu Asp Gly Asn Pro Trp Leu Cys Asp Cys
 180 185 190
 Asp Phe Ala His Leu Phe Ser Trp Ile Gln Glu Asn Ala Ser Lys Leu
 195 200 205
 Pro Lys Gly Leu Ala Gly Val Asp Tyr Leu Cys Val Pro Gly Lys Arg
 210 215 220

Asn Ala Ala Tyr Ser Met Gly Asn Gly Arg Ile Leu Ser Thr Val His
 225 230 235 240
 Gly Glu Ser Ala Ser Ser Lys Gly Ser Pro Ala Ala Ser Arg Ala
 245 250 255

<210> 3
 <211> 765
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Degenerate polynucleotide sequence

<221> misc_feature
 <222> (1)...(765)
 <223> n = A,T,C or G

<400> 3
 atgmgnccnm gngcncncgc ntgygcngcn gcngcnytn gnytnntyws nytnytnytn 60
 ytnytnngcnc cnggncaygc ntgyccngcn ggntgygcnt gyacngaycc ncayacngtn 120
 gaytgymgng aymnggnytn nccnwsngtn ccngayccnt tyccnytnga ygtnmgnaar 180
 ytnytngtng cnggnaaymg nathcarmgn athccngarg aytttytyat httytayggg 240
 gayytngtnt ayytngaytt ymgnaayaay wsnytnmgw snytnngarga rggnacntty 300
 wsnggnwsng cnaarytngt nttyytngay ytnwsntaya ayaaytnac ncarytnngn 360
 gcngngcncnt tymgnwsngc nggmngnytn gtnaarytnw snytnngcnaa yaayaaytn 420
 gtngnggtnc aygargaygc nttygaracn ytnngarwsny tncargtnyt ngarytnaay 480
 gayaayaay tnmgnwsnytn nwsngtnngcn gcnytnngcng cnytnccngc nytnmgwnsn 540
 ytnmgnytn ayggnaaycc ntggytnytn gaytgygayt tygcncayyt nttywsntgg 600
 athcargara aygcwnsnaa rytncnaar ggnytnngcng gngtnngayta yytnytnytn 660
 ccnggnaarm gnaaygcngc ntaywsnatg ggnaayggm gnathytnws nacngtnay 720
 ggngarwsng cnwsnwsnaa rggwnsnccn gcngcnwsnm gngcn 765

<210> 4
 <211> 432
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)...(432)

<400> 4

5

atg gcc ccg ccg ctc ctg ctg ctg ctg ctg gcc agt gga gcg gcc gcc	48
Met Ala Pro Pro Leu Leu Leu Leu Leu Leu Ala Ser Gly Ala Ala Ala	
1 5 10 15	
tgc ccg ctg ccc tgc gtc tgc cag aac ctg tcc gag tcg ctc agc acc	96
Cys Pro Leu Pro Cys Val Cys Gln Asn Leu Ser Glu Ser Leu Ser Thr	
20 25 30	
ctc tgt gcc cac cga ggc ctg ctg ttt gtg ccg ccc aac gtg gac cgg	144
Leu Cys Ala His Arg Gly Leu Leu Phe Val Pro Pro Asn Val Asp Arg	
35 40 45	
cgc aca gtg gag ctg cgg ctg gct gac aac ttc atc cag gcc ctg ggg	192
Arg Thr Val Glu Leu Arg Leu Ala Asp Asn Phe Ile Gln Ala Leu Gly	
50 55 60	
ccc cct gac ttc cgc aac atg acg gga ctg gtg gac ctg aca ctg tct	240
Pro Pro Asp Phe Arg Asn Met Thr Gly Leu Val Asp Leu Thr Leu Ser	
65 70 75 80	
cgc aat gcc atc acc cgc att ggg gcc cgc gcc ttt ggg gac ctc gag	288
Arg Asn Ala Ile Thr Arg Ile Gly Ala Arg Ala Phe Gly Asp Leu Glu	
85 90 95	
agc ctg cgt tcc ctc cac ctt gac ggc aac agg ctg gtg gag ctg ggc	336
Ser Leu Arg Ser Leu His Leu Asp Gly Asn Arg Leu Val Glu Leu Gly	
100 105 110	
acc ggg agc ctc cgg ggc ccc gtc aat ctg cag cac ctc atc ctc agc	384
Thr Gly Ser Leu Arg Gly Pro Val Asn Leu Gln His Leu Ile Leu Ser	
115 120 125	
ggc aac cag ctg ggc gca tcg cgc cgg gag cct tcg acg act tcc tag	432
Gly Asn Gln Leu Gly Ala Ser Arg Arg Glu Pro Ser Thr Thr Ser *	
130 135 140	

<210> 5

<211> 143

<212> PRT

<213> Homo sapiens

<400> 5

6

```

Met Ala Pro Pro Leu Leu Leu Leu Leu Leu Ala Ser Gly Ala Ala Ala
 1           5           10           15
Cys Pro Leu Pro Cys Val Cys Gln Asn Leu Ser Glu Ser Leu Ser Thr
      20           25           30
Leu Cys Ala His Arg Gly Leu Leu Phe Val Pro Pro Asn Val Asp Arg
      35           40           45
Arg Thr Val Glu Leu Arg Leu Ala Asp Asn Phe Ile Gln Ala Leu Gly
      50           55           60
Pro Pro Asp Phe Arg Asn Met Thr Gly Leu Val Asp Leu Thr Leu Ser
65           70           75           80
Arg Asn Ala Ile Thr Arg Ile Gly Ala Arg Ala Phe Gly Asp Leu Glu
      85           90           95
Ser Leu Arg Ser Leu His Leu Asp Gly Asn Arg Leu Val Glu Leu Gly
      100          105          110
Thr Gly Ser Leu Arg Gly Pro Val Asn Leu Gln His Leu Ile Leu Ser
      115          120          125
Gly Asn Gln Leu Gly Ala Ser Arg Arg Glu Pro Ser Thr Thr Ser
      130          135          140

```

<210> 6

<211> 429

<212> DNA

<213> Artificial Sequence

<220>

<223> Degenerate polynucleotide sequence

<221> misc_feature

<222> (1)...(429)

<223> n = A,T,C or G

<400> 6

```

atggcncnc cnytnytnyt nytnytnytn gcnwsnggng cngcngcntg yccnytnccn      60
tgygtntgyc araaytnws ngarwsnytn wsnacnytn gygcncaymg nggnytnytn      120
ttygtncnc cnaaygtnga ymgnmgnacn gtngarytnm gnytngcnga yaayttyath      180
cargcnytn gncncncnga yttymgnaay atgacnggny tngtngayyt nacnytnwsn      240
mgnaaygcna thacnmgnat hggngcnmgn gcnttyggng ayytngarws nytnmgnwsn      300
ytncayytn ayggnaaymg nytngtngar ytnngnacng gnwsnytnmg nggncncgtn      360
aaytncarc ayytnathyt nwsnggnaay carytnggng cnwsnmgnmg ngarccnwsn      420
acnacnwsn                                     429

```

<210> 7

<211> 1653

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(1653)

<400> 7

atg gcc ccg ccg ctc ctg ctg ctg ctg ctg gcc agt gga gcg gcc gcc	48
Met Ala Pro Pro Leu Leu Leu Leu Leu Leu Ala Ser Gly Ala Ala Ala	
1 5 10 15	
tgc ccg ctg ccc tgc gtc tgc cag aac ctg tcc gag tcg ctc agc acc	96
Cys Pro Leu Pro Cys Val Cys Gln Asn Leu Ser Glu Ser Leu Ser Thr	
20 25 30	
ctc tgt gcc cac cga ggc ctg ctg ttt gtg ccg ccc aac gtg gac cgg	144
Leu Cys Ala His Arg Gly Leu Leu Phe Val Pro Pro Asn Val Asp Arg	
35 40 45	
cgc aca gtg gag ctg cgg ctg gct gac aac ttc atc cag gcc ctg ggg	192
Arg Thr Val Glu Leu Arg Leu Ala Asp Asn Phe Ile Gln Ala Leu Gly	
50 55 60	
ccc cct gac ttc cgc aac atg acg gga ctg gtg gac ctg aca ctg tct	240
Pro Pro Asp Phe Arg Asn Met Thr Gly Leu Val Asp Leu Thr Leu Ser	
65 70 75 80	
cgc aat gcc atc acc cgc att ggg gcc cgc gcc ttt ggg gac ctc gag	288
Arg Asn Ala Ile Thr Arg Ile Gly Ala Arg Ala Phe Gly Asp Leu Glu	
85 90 95	
agc ctg cgt tcc ctc cac ctt gac ggc aac agg ctg gtg gag ctg ggc	336
Ser Leu Arg Ser Leu His Leu Asp Gly Asn Arg Leu Val Glu Leu Gly	
100 105 110	
acc ggg agc ctc cgg ggc ccc gtc aat ctg cag cac ctc atc ctc agc	384
Thr Gly Ser Leu Arg Gly Pro Val Asn Leu Gln His Leu Ile Leu Ser	
115 120 125	
ggc aac cag ctg ggc cgc atc gcg ccg gga gcc ttc gac gac ttc cta	432
Gly Asn Gln Leu Gly Arg Ile Ala Pro Gly Ala Phe Asp Asp Phe Leu	
130 135 140	

gag Glu 145	agc Ser	ctg Leu	gag Glu	gac Asp	ctg Leu	gac Asp	ctg Leu	tcc Ser	tac Tyr	aac Asn	aac Asn	ctc Leu	cgg Arg	cag Gln	gtg Val	480
ccc Pro	tgg Trp	gcc Ala	ggc Gly	atc Ile	ggc Gly	gcc Ala	atg Met	cct Pro	gcc Ala	ctg Leu	cac His	acc Thr	ctc Leu	aac Asn	ctg Leu	528
gac Asp	cat His	aac Asn	ctt Leu	att Ile	gac Asp	gca Ala	ctg Leu	ccc Pro	cca Pro	ggc Gly	gcc Ala	ttc Phe	gcc Ala	cag Gln	ctc Leu	576
ggt Gly	cag Gln	ctc Leu	tcc Ser	cgc Arg	ctg Leu	gac Asp	ctc Leu	acc Thr	tcc Ser	aac Asn	cgc Arg	ctg Leu	gcc Ala	acg Thr	ctg Leu	624
gct Ala	ccg Pro	gac Asp	ccg Pro	ctt Leu	ttc Phe	tct Ser	cgt Arg	ggg Gly	cgt Arg	gat Asp	gca Ala	gag Glu	gcc Ala	tct Ser	ccc Pro	672
gcc Ala	ccc Pro	ctg Leu	gtg Val	ctg Leu	agc Ser	ttt Phe	agc Ser	ggg Gly	aac Asn	ccc Pro	ctg Leu	cac His	tgc Cys	aac Asn	tgt Cys	720
gag Glu	ctg Leu	ctg Leu	tgg Trp	ctg Leu	cgg Arg	cgg Arg	ctg Leu	gcg Ala	cgg Arg	ccg Pro	gac Asp	gac Asp	ctg Leu	gaa Glu	acg Thr	768
tgc Cys	gcc Ala	tcc Ser	ccg Pro	ccc Pro	ggc Gly	ctg Leu	gcc Ala	ggc Gly	cgc Arg	tac Tyr	ttc Phe	tgg Trp	gca Ala	gtg Val	ccc Pro	816
gag Glu	ggc Gly	gag Glu	ttc Phe	tcc Ser	tgt Cys	gag Glu	ccg Pro	ccc Pro	ctc Leu	att Ile	gcc Ala	cgc Arg	cac His	acg Thr	cag Gln	864
cgc Arg	ctc Leu	tgg Trp	gtg Val	ctg Leu	gaa Glu	ggc Gly	cag Gln	cgg Arg	gcc Ala	acg Thr	ctg Leu	cgg Arg	tgc Cys	cgg Arg	gcc Ala	912

ctg ggt gac ccc gcg cct acc atg cac tgg gtc ggt cct gac gac cgg	960
Leu Gly Asp Pro Ala Pro Thr Met His Trp Val Gly Pro Asp Asp Arg	
305 310 315 320	
ttg gtt ggc aac tcc tcc cga gcc cgg gct ttc ccc aac ggg acc tta	1008
Leu Val Gly Asn Ser Ser Arg Ala Arg Ala Phe Pro Asn Gly Thr Leu	
325 330 335	
gag att ggg gtg acc ggc gct ggg gac gct ggg ggc tac acc tgc atc	1056
Glu Ile Gly Val Thr Gly Ala Gly Asp Ala Gly Gly Tyr Thr Cys Ile	
340 345 350	
gcc acc aac cct gct ggt gag gcc aca gcc cga gta gaa ctg cgg gtg	1104
Ala Thr Asn Pro Ala Gly Glu Ala Thr Ala Arg Val Glu Leu Arg Val	
355 360 365	
ctg gcc ttg ccc cat ggt ggg aac agc agt gcc gag ggg ggc cgc ccc	1152
Leu Ala Leu Pro His Gly Gly Asn Ser Ser Ala Glu Gly Gly Arg Pro	
370 375 380	
ggg ccc tcg gac atc gcc gcc tcc gct cgc act gct gcc gag ggt gag	1200
Gly Pro Ser Asp Ile Ala Ala Ser Ala Arg Thr Ala Ala Glu Gly Glu	
385 390 395 400	
ggg acg ctg gag tct gag cca gcc gtg cag gtg acg gag gtg acc gcc	1248
Gly Thr Leu Glu Ser Glu Pro Ala Val Gln Val Thr Glu Val Thr Ala	
405 410 415	
acc tca ggg ctg gtg agc tgg ggt ccc ggg cgg cca gcc gac cca gtg	1296
Thr Ser Gly Leu Val Ser Trp Gly Pro Gly Arg Pro Ala Asp Pro Val	
420 425 430	
tgg atg ttc caa atc cag tac aac agc agc gaa gat gag acc ctc atc	1344
Trp Met Phe Gln Ile Gln Tyr Asn Ser Ser Glu Asp Glu Thr Leu Ile	
435 440 445	
tac cgg att gtc cca gcc tcc agc cac cac ttc ctg ctg aag cac ctc	1392
Tyr Arg Ile Val Pro Ala Ser Ser His His Phe Leu Leu Lys His Leu	
450 455 460	
gtc ccc ggc gct gac tat gac ctc tgc ctg ctg gcc ttg tca ccg gcc	1440
Val Pro Gly Ala Asp Tyr Asp Leu Cys Leu Leu Ala Leu Ser Pro Ala	
465 470 475 480	

10

gct ggg ccc tct gac ctc acg gcc acc agg ctg ctg ggc tgt gcc cat 1488
 Ala Gly Pro Ser Asp Leu Thr Ala Thr Arg Leu Leu Gly Cys Ala His
 485 490 495

ttc tcc acg ctg ccg gcc tcg ccc ctg tgc cac gcc ctg cag gcc cac 1536
 Phe Ser Thr Leu Pro Ala Ser Pro Leu Cys His Ala Leu Gln Ala His
 500 505 510

gtg ctg ggc ggg acc ctg acc gtg gcc gtg ggg ggt gtg ctg gtg gct 1584
 Val Leu Gly Gly Thr Leu Thr Val Ala Val Gly Gly Val Leu Val Ala
 515 520 525

gcg ctt act ggt ctt cac tgt ggc cct tgc tgg ttc ggg gcc ggg ggg 1632
 Ala Leu Thr Gly Leu His Cys Gly Pro Cys Trp Phe Gly Ala Gly Gly
 530 535 540

ccg gaa atg gcc gcc tcc ccc 1653
 Pro Glu Met Ala Ala Ser Pro
 545 550

<210> 8

<211> 551

<212> PRT

<213> Homo sapiens

<400> 8

Met Ala Pro Pro Leu Leu Leu Leu Leu Ala Ser Gly Ala Ala Ala
 1 5 10 15

Cys Pro Leu Pro Cys Val Cys Gln Asn Leu Ser Glu Ser Leu Ser Thr
 20 25 30

Leu Cys Ala His Arg Gly Leu Leu Phe Val Pro Pro Asn Val Asp Arg
 35 40 45

Arg Thr Val Glu Leu Arg Leu Ala Asp Asn Phe Ile Gln Ala Leu Gly
 50 55 60

Pro Pro Asp Phe Arg Asn Met Thr Gly Leu Val Asp Leu Thr Leu Ser
 65 70 75 80

Arg Asn Ala Ile Thr Arg Ile Gly Ala Arg Ala Phe Gly Asp Leu Glu
 85 90 95

Ser Leu Arg Ser Leu His Leu Asp Gly Asn Arg Leu Val Glu Leu Gly
 100 105 110

Thr	Gly	Ser	Leu	Arg	Gly	Pro	Val	Asn	Leu	Gln	His	Leu	Ile	Leu	Ser
	115						120					125			
Gly	Asn	Gln	Leu	Gly	Arg	Ile	Ala	Pro	Gly	Ala	Phe	Asp	Asp	Phe	Leu
	130					135				140					
Glu	Ser	Leu	Glu	Asp	Leu	Asp	Leu	Ser	Tyr	Asn	Asn	Leu	Arg	Gln	Val
145				150					155					160	
Pro	Trp	Ala	Gly	Ile	Gly	Ala	Met	Pro	Ala	Leu	His	Thr	Leu	Asn	Leu
			165					170					175		
Asp	His	Asn	Leu	Ile	Asp	Ala	Leu	Pro	Pro	Gly	Ala	Phe	Ala	Gln	Leu
	180							185				190			
Gly	Gln	Leu	Ser	Arg	Leu	Asp	Leu	Thr	Ser	Asn	Arg	Leu	Ala	Thr	Leu
	195					200					205				
Ala	Pro	Asp	Pro	Leu	Phe	Ser	Arg	Gly	Arg	Asp	Ala	Glu	Ala	Ser	Pro
210					215					220					
Ala	Pro	Leu	Val	Leu	Ser	Phe	Ser	Gly	Asn	Pro	Leu	His	Cys	Asn	Cys
225				230					235					240	
Glu	Leu	Leu	Trp	Leu	Arg	Arg	Leu	Ala	Arg	Pro	Asp	Asp	Leu	Glu	Thr
			245					250					255		
Cys	Ala	Ser	Pro	Pro	Gly	Leu	Ala	Gly	Arg	Tyr	Phe	Trp	Ala	Val	Pro
	260					265						270			
Glu	Gly	Glu	Phe	Ser	Cys	Glu	Pro	Pro	Leu	Ile	Ala	Arg	His	Thr	Gln
	275					280					285				
Arg	Leu	Trp	Val	Leu	Glu	Gly	Gln	Arg	Ala	Thr	Leu	Arg	Cys	Arg	Ala
290					295						300				
Leu	Gly	Asp	Pro	Ala	Pro	Thr	Met	His	Trp	Val	Gly	Pro	Asp	Asp	Arg
305				310					315					320	
Leu	Val	Gly	Asn	Ser	Ser	Arg	Ala	Arg	Ala	Phe	Pro	Asn	Gly	Thr	Leu
			325					330					335		
Glu	Ile	Gly	Val	Thr	Gly	Ala	Gly	Asp	Ala	Gly	Gly	Tyr	Thr	Cys	Ile
	340					345						350			
Ala	Thr	Asn	Pro	Ala	Gly	Glu	Ala	Thr	Ala	Arg	Val	Glu	Leu	Arg	Val
	355					360					365				
Leu	Ala	Leu	Pro	His	Gly	Gly	Asn	Ser	Ser	Ala	Glu	Gly	Gly	Arg	Pro
370					375					380					
Gly	Pro	Ser	Asp	Ile	Ala	Ala	Ser	Ala	Arg	Thr	Ala	Ala	Glu	Gly	Glu
385				390					395					400	
Gly	Thr	Leu	Glu	Ser	Glu	Pro	Ala	Val	Gln	Val	Thr	Glu	Val	Thr	Ala
			405					410					415		
Thr	Ser	Gly	Leu	Val	Ser	Trp	Gly	Pro	Gly	Arg	Pro	Ala	Asp	Pro	Val
	420						425					430			
Trp	Met	Phe	Gln	Ile	Gln	Tyr	Asn	Ser	Ser	Glu	Asp	Glu	Thr	Leu	Ile
	435					440						445			

Tyr Arg Ile Val Pro Ala Ser Ser His His Phe Leu Leu Lys His Leu
 450 455 460
 Val Pro Gly Ala Asp Tyr Asp Leu Cys Leu Leu Ala Leu Ser Pro Ala
 465 470 475 480
 Ala Gly Pro Ser Asp Leu Thr Ala Thr Arg Leu Leu Gly Cys Ala His
 485 490 495
 Phe Ser Thr Leu Pro Ala Ser Pro Leu Cys His Ala Leu Gln Ala His
 500 505 510
 Val Leu Gly Gly Thr Leu Thr Val Ala Val Gly Gly Val Leu Val Ala
 515 520 525
 Ala Leu Thr Gly Leu His Cys Gly Pro Cys Trp Phe Gly Ala Gly Gly
 530 535 540
 Pro Glu Met Ala Ala Ser Pro
 545 550

<210> 9

<211> 1653

<212> DNA

<213> Artificial Sequence

<220>

<223> Degenerate polynucleotide sequence

<221> misc_feature

<222> (1)...(1653)

<223> n = A,T,C or G

<400> 9

atggcncnc	cnytnytnyt	nytnytnytn	gcnwsnggng	cngcngcntg	ycnnytnccn	60
tgygtntgyc	araaytnws	ngarwsnytn	wsnacnytn	gygcncaymg	nggnytnytn	120
ttygtncnc	cnaaygtnga	ymgnmgna	gtngarytnm	gnytngcnga	yaaytthyath	180
cargcnytn	gncncncga	yttymgnaay	atgacnggny	tngtngayyt	nacnytnwsn	240
mgnaaygcna	thacnmgnat	hggngcnmgn	gcnttyggng	ayytngarws	nytnmgnwsn	300
ytncaaytn	ayggnaaymg	nytngtngar	ytnggnacng	gnwsnytnmg	nggncngtn	360
aayytnca	ayytnathyt	nwsnggnaay	carytnngnm	gnathgcnc	nggngcntty	420
gaygaytthy	tngarwsnyt	ngargayytn	gayytnwsnt	ayaayaayyt	nmgncargtn	480
ccntgggng	gnathggngc	natgccngcn	ytncaayacny	tnaayytnga	ycayaayytn	540
athgaygcny	tnccncngg	ngcnttygcn	carytnnggnc	arytnwsnm	nytngayytn	600
acnwsnaaym	gnytngcna	nytngcncn	gayccnytn	tywsnmngg	nmgngaygc	660
gargcnwsnc	cngcncny	ngtnytnwsn	ttywsnggna	ayccnytnca	ytgyaaytgy	720
garytnytn	ggytnmgnm	nytngcnmgn	ccngaygay	tngaracntg	ygcnwsnccn	780
ccnggnytn	cnggmngta	ytytgggcn	gtncngarg	ngarttyws	ntgygarccn	840
ccnytnathg	cnmgncayac	ncarmgnytn	tgggtnytn	arggncarmg	ngcnacnytn	900

```

mgntgymgng cnytnngnga yccngcncn acnatgcayt gggtnngncc ngaygaymgn      960
ytngtnngna aywsnwsnmg ngcnmgngcn ttyccnaayg gnacnytnga rathggngtn      1020
acnggngcng gngaygcngg nggntayacn tgyathgcna cnaayccngc nggngargcn      1080
acngcnmgng tngarytnmg ngtnytngcn ytnccncayg gnggnaayws nwsngcngar      1140
ggnggnmgnc cnggncnws ngayathgcn gcnwsngcnm gnacngcngc ngarggngar      1200
ggnacnytn arwsngarcc ngcngtnar gtnacngarg tnacngcnac nwsnggnytn      1260
gtnwsntggg gnccnggnmg nccngcngay ccngtntgga tgttycarat hcartayaay      1320
wsnwsngarg aygaracny nathtaymgn athgtncng cnwsnwsnca ycaytttytn      1380
ytnaarcayy tngtnccng ngcngaytay gayytnngy tnytnngcny nwsnccngcn      1440
gcnggncnw sngaytnac ngcnacnmgn ytnytngnt gygcncaytt ywsnacnytn      1500
ccngcnwsnc cnytnngyca ygcnytnar gcncaygtny tngnggnac nytnacngtn      1560
gcngtnngng gngtnytngt ngcngcnytn acnggnytn aytyggngcc ntgytggtty      1620
ggngcngng gncngarat ggngcnwsn ccn      1653

```

<210> 10

<211> 864

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(864)

<400> 10

```

atg cgg caa acc cta ccg ctg ctg ctg ctg acg gtg ctg cgc ccc agc      48
Met Arg Gln Thr Leu Pro Leu Leu Leu Leu Thr Val Leu Arg Pro Ser
1           5           10          15

tgg gca gac cct ccc cag gag aag gtc ccg ctc ttc cgg gtc act cag      96
Trp Ala Asp Pro Pro Gln Glu Lys Val Pro Leu Phe Arg Val Thr Gln
          20          25          30

cag ggc ccc tgg ggg agc agt ggc agc aac gcc acc gac tcg ccc tgc      144
Gln Gly Pro Trp Gly Ser Ser Gly Ser Asn Ala Thr Asp Ser Pro Cys
          35          40          45

gag ggg ctg ccc gcc gcg gat gcg acg gcc ttg acc ctg gcg aac cgc      192
Glu Gly Leu Pro Ala Ala Asp Ala Thr Ala Leu Thr Leu Ala Asn Arg
          50          55          60

aac ctg gag cgc ctg ccc ggc tgc cta ccg cgc aca ctg cgc agc ctc      240
Asn Leu Glu Arg Leu Pro Gly Cys Leu Pro Arg Thr Leu Arg Ser Leu
          65          70          75          80

```

gac gcc agc cac aac ctg ctg cgc gcc ctg agc act tcc gag ctc ggc Asp Ala Ser His Asn Leu Leu Arg Ala Leu Ser Thr Ser Glu Leu Gly 85 90 95	288
cac ctg gag cag ctg cag gtg ctg acc ctg cgc cac aac cgc atc gcc His Leu Glu Gln Leu Gln Val Leu Thr Leu Arg His Asn Arg Ile Ala 100 105 110	336
gcg ctg cgc tgg ggc ccg ggt ggg ccg gcg ggg ctg cac acc ctg gac Ala Leu Arg Trp Gly Pro Gly Gly Pro Ala Gly Leu His Thr Leu Asp 115 120 125	384
ctc agc tac aac cag ctg gcc gct ctg ccg ccg tgc acc ggg ccc gcg Leu Ser Tyr Asn Gln Leu Ala Ala Leu Pro Pro Cys Thr Gly Pro Ala 130 135 140	432
ctg agc agc ctc cgc gcc ctg gcg ctc gcc ggg aat ccg ctg cgg gcg Leu Ser Ser Leu Arg Ala Leu Ala Leu Ala Gly Asn Pro Leu Arg Ala 145 150 155 160	480
ctg cag ccc cgg gcc ttc gcc tgc ttc ccc gcg ctg cag ctc ctc aac Leu Gln Pro Arg Ala Phe Ala Cys Phe Pro Ala Leu Gln Leu Leu Asn 165 170 175	528
ctc tcc tgc acc gcg ctg ggt cgc gga gcc cag ggg ggc atc gcc gag Leu Ser Cys Thr Ala Leu Gly Arg Gly Ala Gln Gly Gly Ile Ala Glu 180 185 190	576
gcg gcg ttc gct gga gag gat ggc gcg ccc ctg gtc acg ctc gaa gtc Ala Ala Phe Ala Gly Glu Asp Gly Ala Pro Leu Val Thr Leu Glu Val 195 200 205	624
ctg gat ctc agc ggc acg ttc ctt gaa cgg gtt gag tca ggg tgg atc Leu Asp Leu Ser Gly Thr Phe Leu Glu Arg Val Glu Ser Gly Trp Ile 210 215 220	672
aga gac ctg ccg aag ctc aca tcc ctc tac ctg agg aag atg cct cgg Arg Asp Leu Pro Lys Leu Thr Ser Leu Tyr Leu Arg Lys Met Pro Arg 225 230 235 240	720

15

ctg acg acc ctg gag ggg gac att ttc aag atg acc ccc aac ctg cag 768
 Leu Thr Thr Leu Glu Gly Asp Ile Phe Lys Met Thr Pro Asn Leu Gln
 245 250 255

cag ctg gac tgt cag gac tcc cca gca ctt gct tct gtc gcc aca cac 816
 Gln Leu Asp Cys Gln Asp Ser Pro Ala Leu Ala Ser Val Ala Thr His
 260 265 270

atc ttt caa gat act cca cat cta cag gtc ctt ctg ttc cag aag taa 864
 Ile Phe Gln Asp Thr Pro His Leu Gln Val Leu Leu Phe Gln Lys *
 275 280 285

<210> 11

<211> 287

<212> PRT

<213> Homo sapiens

<400> 11

Met Arg Gln Thr Leu Pro Leu Leu Leu Leu Thr Val Leu Arg Pro Ser
 1 5 10 15
 Trp Ala Asp Pro Pro Gln Glu Lys Val Pro Leu Phe Arg Val Thr Gln
 20 25 30
 Gln Gly Pro Trp Gly Ser Ser Gly Ser Asn Ala Thr Asp Ser Pro Cys
 35 40 45
 Glu Gly Leu Pro Ala Ala Asp Ala Thr Ala Leu Thr Leu Ala Asn Arg
 50 55 60
 Asn Leu Glu Arg Leu Pro Gly Cys Leu Pro Arg Thr Leu Arg Ser Leu
 65 70 75 80
 Asp Ala Ser His Asn Leu Leu Arg Ala Leu Ser Thr Ser Glu Leu Gly
 85 90 95
 His Leu Glu Gln Leu Gln Val Leu Thr Leu Arg His Asn Arg Ile Ala
 100 105 110
 Ala Leu Arg Trp Gly Pro Gly Gly Pro Ala Gly Leu His Thr Leu Asp
 115 120 125
 Leu Ser Tyr Asn Gln Leu Ala Ala Leu Pro Pro Cys Thr Gly Pro Ala
 130 135 140
 Leu Ser Ser Leu Arg Ala Leu Ala Leu Ala Gly Asn Pro Leu Arg Ala
 145 150 155 160
 Leu Gln Pro Arg Ala Phe Ala Cys Phe Pro Ala Leu Gln Leu Leu Asn
 165 170 175
 Leu Ser Cys Thr Ala Leu Gly Arg Gly Ala Gln Gly Gly Ile Ala Glu
 180 185 190

16

Ala Ala Phe Ala Gly Glu Asp Gly Ala Pro Leu Val Thr Leu Glu Val
 195 200 205
 Leu Asp Leu Ser Gly Thr Phe Leu Glu Arg Val Glu Ser Gly Trp Ile
 210 215 220
 Arg Asp Leu Pro Lys Leu Thr Ser Leu Tyr Leu Arg Lys Met Pro Arg
 225 230 235 240
 Leu Thr Thr Leu Glu Gly Asp Ile Phe Lys Met Thr Pro Asn Leu Gln
 245 250 255
 Gln Leu Asp Cys Gln Asp Ser Pro Ala Leu Ala Ser Val Ala Thr His
 260 265 270
 Ile Phe Gln Asp Thr Pro His Leu Gln Val Leu Leu Phe Gln Lys
 275 280 285

<210> 12

<211> 861

<212> DNA

<213> Artificial Sequence

<220>

<223> Degenerate polynucleotide sequence

<221> misc_feature

<222> (1)...(861)

<223> n = A,T,C or G

<400> 12

atgmgn cara	cnytnccnyt	nytnytnytn	acngtnytnm	gnccnwsntg	ggcngayccn	60
ccncargara	argtnccnyt	nttymngn	acncarcarg	gnccntgggg	nwsnwsnggn	120
wsnaaygcna	cngaywsncc	ntgygarggn	ytncncngcng	cngaygcnac	ngcnytnacn	180
ytngcnaaym	gnaayytnga	rmgnytnccn	ggntgyytnc	cnmgnacnyt	nmgwnsnytn	240
gaygcnwsnc	ayaayytnyt	nmgngcnytn	wsnacnwsng	arytnggnca	yytngarcarg	300
ytncargtny	tnacnytnmg	ncayaaymgn	athgcngcny	tnmgntgggg	nccnggnggn	360
ccngcnggny	tncaayacnyt	ngayytwnsn	tayaaycary	tngcngcnyt	nccnccntgy	420
acnggnccng	cnytnwsnws	nytnmgngcn	ytngcnytn	cnggnaaycc	nytnmgngcn	480
ytncarcenm	gngcnttygc	ntgyttyccn	gcnytnrcary	tnytnaayyt	nwsntgyacn	540
gcnytnngnm	gngngncna	rgngngnath	gcngargcng	cnttygcngg	ngargayggn	600
gcncnytn	tnacnytn	rgtnytn	ytwnsnggna	cnttyytnga	rmgngtn	660
wsnggntgga	thmgngayyt	nccnaarytn	acnwsnytn	ayytmgnaa	ratgccnmgn	720
ytncacnny	tnarggnga	yathttyaar	atgacnccna	ayytncarca	rytn	780
cargaywsnc	cngcnytn	ngcngn	acncayath	tycargayac	nccncayyt	840
cargtnytny	tnnttycaraa	r				861

<210> 13

17

<211> 2223
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)...(2223)

<400> 13

atg cgg caa acc cta ccg ctg ctg ctg ctg acg gtg ctg cgc ccc agc	48
Met Arg Gln Thr Leu Pro Leu Leu Leu Leu Thr Val Leu Arg Pro Ser	
1 5 10 15	
tgg gca gac cct ccc cag gag aag gtc ccg ctc ttc cgg gtc act cag	96
Trp Ala Asp Pro Pro Gln Glu Lys Val Pro Leu Phe Arg Val Thr Gln	
20 25 30	
cag ggc ccc tgg ggg agc agt ggc agc aac gcc acc gac tcg ccc tgc	144
Gln Gly Pro Trp Gly Ser Ser Gly Ser Asn Ala Thr Asp Ser Pro Cys	
35 40 45	
gag ggg ctg ccc gcc gcg gat gcg acg gcc ttg acc ctg gcg aac cgc	192
Glu Gly Leu Pro Ala Ala Asp Ala Thr Ala Leu Thr Leu Ala Asn Arg	
50 55 60	
aac ctg gag cgc ctg ccc ggc tgc cta ccg cgc aca ctg cgc agc ctc	240
Asn Leu Glu Arg Leu Pro Gly Cys Leu Pro Arg Thr Leu Arg Ser Leu	
65 70 75 80	
gac gcc agc cac aac ctg ctg cgc gcc ctg agc act tcc gag ctc ggc	288
Asp Ala Ser His Asn Leu Leu Arg Ala Leu Ser Thr Ser Glu Leu Gly	
85 90 95	
cac ctg gag cag ctg cag gtg ctg acc ctg cgc cac aac cgc atc gcc	336
His Leu Glu Gln Leu Gln Val Leu Thr Leu Arg His Asn Arg Ile Ala	
100 105 110	
gcg ctg cgc tgg ggc ccg ggt ggg ccg gcg ggg ctg cac acc ctg gac	384
Ala Leu Arg Trp Gly Pro Gly Gly Pro Ala Gly Leu His Thr Leu Asp	
115 120 125	

ctc agc tac aac cag ctg gcc gct ctg ctg ccg tgc acc ggg ccc gcg Leu Ser Tyr Asn Gln Leu Ala Ala Leu Leu Pro Cys Thr Gly Pro Ala 130 135 140	432
ctg agc agc ctc cgc gcc ctg gcg ctc gcc ggg aat ccg ctg cgg gcg Leu Ser Ser Leu Arg Ala Leu Ala Leu Ala Gly Asn Pro Leu Arg Ala 145 150 155 160	480
ctg cag gcc ccg gcc ttc gcc tgc ttc ccc gcg ctg cag ctc ctc aac Leu Gln Ala Pro Ala Phe Ala Cys Phe Pro Ala Leu Gln Leu Leu Asn 165 170 175	528
ctc tcc tgc acc gcg ctg ggt cgc gga gcc cag ggg ggc atc gcc gag Leu Ser Cys Thr Ala Leu Gly Arg Gly Ala Gln Gly Gly Ile Ala Glu 180 185 190	576
gcg gcg ttc gct gga gag gat ggc gcg ccc ctg gtc acg ctc gaa gtc Ala Ala Phe Ala Gly Glu Asp Gly Ala Pro Leu Val Thr Leu Glu Val 195 200 205	624
ctg gat ctc agc ggc acg ttc ctt gaa cgg gtt gag tca ggg tgg atc Leu Asp Leu Ser Gly Thr Phe Leu Glu Arg Val Glu Ser Gly Trp Ile 210 215 220	672
aga gac ctg ccg aag ctc aca tcc ctc tac ctg agg aag atg cct cgg Arg Asp Leu Pro Lys Leu Thr Ser Leu Tyr Leu Arg Lys Met Pro Arg 225 230 235 240	720
ctg acg acc ctg gag ggg gac att ttc aag atc acc ccc aac ctg cag Leu Thr Thr Leu Glu Gly Asp Ile Phe Lys Ile Thr Pro Asn Leu Gln 245 250 255	768
cag ctg gac tgt cag gac tcc cca gca ctt gct tct gtc gcc aca cac Gln Leu Asp Cys Gln Asp Ser Pro Ala Leu Ala Ser Val Ala Thr His 260 265 270	816
atc ttt caa gat act cca cat cta cag gtc ctt ctg ttc cag aac tgc Ile Phe Gln Asp Thr Pro His Leu Gln Val Leu Leu Phe Gln Asn Cys 275 280 285	864
aac ttg agt tcc ttc cct cct tgg acc ctg gat tcc tcc cag gtc cta Asn Leu Ser Ser Phe Pro Pro Trp Thr Leu Asp Ser Ser Gln Val Leu 290 295 300	912

tcg atc aac ctc ttt ggc aac ccc ctc act tgc agt tgt gac ttg tct	960
Ser Ile Asn Leu Phe Gly Asn Pro Leu Thr Cys Ser Cys Asp Leu Ser	
305 310 315 320	
tgg ctc ctc acg gat gca aag aga act gtc cta agc agg gca gca gac	1008
Trp Leu Leu Thr Asp Ala Lys Arg Thr Val Leu Ser Arg Ala Ala Asp	
325 330 335	
act atg tgc gcg cca gct gcg gga tcc agc ggc ccc ttc tca gcc tcc	1056
Thr Met Cys Ala Pro Ala Ala Gly Ser Ser Gly Pro Phe Ser Ala Ser	
340 345 350	
ctg tca ctc tcc cag ctg ccc gga gtg tgc cag tcc gac caa agc acc	1104
Leu Ser Leu Ser Gln Leu Pro Gly Val Cys Gln Ser Asp Gln Ser Thr	
355 360 365	
act ctc ggg gct tca cac cca cct tgc ttc aac cgc tcc acc tac gca	1152
Thr Leu Gly Ala Ser His Pro Pro Cys Phe Asn Arg Ser Thr Tyr Ala	
370 375 380	
cag ggt acc acc gtc gcg ccc agc gca gcc ccc gcc acc cgg cct gcg	1200
Gln Gly Thr Thr Val Ala Pro Ser Ala Ala Pro Ala Thr Arg Pro Ala	
385 390 395 400	
gga gac cag cag agt gtc tcc aag gcc cct aac gtg ggc tct cgc acg	1248
Gly Asp Gln Gln Ser Val Ser Lys Ala Pro Asn Val Gly Ser Arg Thr	
405 410 415	
ata gct gca tgg ccg cac agc gat gca cgg gag ggg act gcc ccc tcc	1296
Ile Ala Ala Trp Pro His Ser Asp Ala Arg Glu Gly Thr Ala Pro Ser	
420 425 430	
acg acc aac tct gta gca ggt cac agc aac tcc agc gtt ttc ccc agg	1344
Thr Thr Asn Ser Val Ala Gly His Ser Asn Ser Ser Val Phe Pro Arg	
435 440 445	
gct gcc agc acc acc agg acc cag cac cga gga gaa cat gcc ccc gag	1392
Ala Ala Ser Thr Thr Arg Thr Gln His Arg Gly Glu His Ala Pro Glu	
450 455 460	

ctt gtc ctt gag cct gat atc tca gct gcc tcc acc cca ctg gcc agc Leu Val Leu Glu Pro Asp Ile Ser Ala Ala Ser Thr Pro Leu Ala Ser 465 470 475 480	1440
aag ctc ctg ggc ccc ttc cct acc tcg tgg gac cgc agc ata agc tcg Lys Leu Leu Gly Pro Phe Pro Thr Ser Trp Asp Arg Ser Ile Ser Ser 485 490 495	1488
cct cag ccc ggc cag agg aca cac gcc aca ccc caa gcc ccc aac ccg Pro Gln Pro Gly Gln Arg Thr His Ala Thr Pro Gln Ala Pro Asn Pro 500 505 510	1536
agt ctt tcc gag ggc gag att cca gtc ttg ctg ctg gac gac tac agt Ser Leu Ser Glu Gly Glu Ile Pro Val Leu Leu Leu Asp Asp Tyr Ser 515 520 525	1584
gag gag gag gaa ggg agg aag gag gag gtg gga acg cct cac cag gac Glu Glu Glu Glu Gly Arg Lys Glu Glu Val Gly Thr Pro His Gln Asp 530 535 540	1632
gtc ccc tgt gat tac cat ccc tgc aag cac ctg cag acc ccg tgc gcg Val Pro Cys Asp Tyr His Pro Cys Lys His Leu Gln Thr Pro Cys Ala 545 550 555 560	1680
gag ctg cag agg cgg tgg cgg tgc cgg tgc ccc ggc ctc agc ggg gaa Glu Leu Gln Arg Arg Trp Arg Cys Arg Cys Pro Gly Leu Ser Gly Glu 565 570 575	1728
gac acc atc cca gac ccg ccc agg ctg cag ggg gtg acg gag acc acg Asp Thr Ile Pro Asp Pro Pro Arg Leu Gln Gly Val Thr Glu Thr Thr 580 585 590	1776
gac acg tcg gcg ctg gtc cac tgg tgt gcc ccc aac tcg gta gtg cat Asp Thr Ser Ala Leu Val His Trp Cys Ala Pro Asn Ser Val Val His 595 600 605	1824
ggg tac cag atc cgc tac tct gcg gag ggc tgg gcg ggg aac cag tcg Gly Tyr Gln Ile Arg Tyr Ser Ala Glu Gly Trp Ala Gly Asn Gln Ser 610 615 620	1872
gtg gtg ggg gtc atc tac gcc acg gcc cgg cag cac cct ctg tac ggg Val Val Gly Val Ile Tyr Ala Thr Ala Arg Gln His Pro Leu Tyr Gly 625 630 635 640	1920

21

ctg tgc ccg ggc acc acc tac cgc gtg tgc gtg ctg gcg gcc aac agg	1968
Leu Ser Pro Gly Thr Thr Tyr Arg Val Cys Val Leu Ala Ala Asn Arg	
645 650 655	
gcg ggc ttg agc cag cca cgg tct tgc ggc tgg agg agc ccg tgc gcc	2016
Ala Gly Leu Ser Gln Pro Arg Ser Ser Gly Trp Arg Ser Pro Cys Ala	
660 665 670	
gcc ttc acc acc aag ccc agc ttc gcg ctc ctg ctc tct ggg ctg tgc	2064
Ala Phe Thr Thr Lys Pro Ser Phe Ala Leu Leu Leu Ser Gly Leu Cys	
675 680 685	
gcc gcc agc ggc ctg ttg ctc gcc agc acc gtg gtg ctg tcc gca tgt	2112
Ala Ala Ser Gly Leu Leu Leu Ala Ser Thr Val Val Leu Ser Ala Cys	
690 695 700	
ctc tgc agg cgg ggc cag acg ctg ggc ctg cag cgc tgc gac acg cac	2160
Leu Cys Arg Arg Gly Gln Thr Leu Gly Leu Gln Arg Cys Asp Thr His	
705 710 715 720	
ctg gtg gcc tac aaa aac ccg gcc ttt gat gat tac ccg ctg ggg ctc	2208
Leu Val Ala Tyr Lys Asn Pro Ala Phe Asp Asp Tyr Pro Leu Gly Leu	
725 730 735	
cag acc gtc agt tag	2223
Gln Thr Val Ser *	
740	

<210> 14

<211> 740

<212> PRT

<213> Homo sapiens

<400> 14

Met Arg Gln Thr Leu Pro Leu Leu Leu Leu Thr Val Leu Arg Pro Ser	
1 5 10 15	
Trp Ala Asp Pro Pro Gln Glu Lys Val Pro Leu Phe Arg Val Thr Gln	
20 25 30	
Gln Gly Pro Trp Gly Ser Ser Gly Ser Asn Ala Thr Asp Ser Pro Cys	
35 40 45	

Glu Gly Leu Pro Ala Ala Asp Ala Thr Ala Leu Thr Leu Ala Asn Arg
 50 55 60
 Asn Leu Glu Arg Leu Pro Gly Cys Leu Pro Arg Thr Leu Arg Ser Leu
 65 70 75 80
 Asp Ala Ser His Asn Leu Leu Arg Ala Leu Ser Thr Ser Glu Leu Gly
 85 90 95
 His Leu Glu Gln Leu Gln Val Leu Thr Leu Arg His Asn Arg Ile Ala
 100 105 110
 Ala Leu Arg Trp Gly Pro Gly Gly Pro Ala Gly Leu His Thr Leu Asp
 115 120 125
 Leu Ser Tyr Asn Gln Leu Ala Ala Leu Leu Pro Cys Thr Gly Pro Ala
 130 135 140
 Leu Ser Ser Leu Arg Ala Leu Ala Leu Ala Gly Asn Pro Leu Arg Ala
 145 150 155 160
 Leu Gln Ala Pro Ala Phe Ala Cys Phe Pro Ala Leu Gln Leu Leu Asn
 165 170 175
 Leu Ser Cys Thr Ala Leu Gly Arg Gly Ala Gln Gly Gly Ile Ala Glu
 180 185 190
 Ala Ala Phe Ala Gly Glu Asp Gly Ala Pro Leu Val Thr Leu Glu Val
 195 200 205
 Leu Asp Leu Ser Gly Thr Phe Leu Glu Arg Val Glu Ser Gly Trp Ile
 210 215 220
 Arg Asp Leu Pro Lys Leu Thr Ser Leu Tyr Leu Arg Lys Met Pro Arg
 225 230 235 240
 Leu Thr Thr Leu Glu Gly Asp Ile Phe Lys Ile Thr Pro Asn Leu Gln
 245 250 255
 Gln Leu Asp Cys Gln Asp Ser Pro Ala Leu Ala Ser Val Ala Thr His
 260 265 270
 Ile Phe Gln Asp Thr Pro His Leu Gln Val Leu Leu Phe Gln Asn Cys
 275 280 285
 Asn Leu Ser Ser Phe Pro Pro Trp Thr Leu Asp Ser Ser Gln Val Leu
 290 295 300
 Ser Ile Asn Leu Phe Gly Asn Pro Leu Thr Cys Ser Cys Asp Leu Ser
 305 310 315 320
 Trp Leu Leu Thr Asp Ala Lys Arg Thr Val Leu Ser Arg Ala Ala Asp
 325 330 335
 Thr Met Cys Ala Pro Ala Ala Gly Ser Ser Gly Pro Phe Ser Ala Ser
 340 345 350
 Leu Ser Leu Ser Gln Leu Pro Gly Val Cys Gln Ser Asp Gln Ser Thr
 355 360 365
 Thr Leu Gly Ala Ser His Pro Pro Cys Phe Asn Arg Ser Thr Tyr Ala
 370 375 380

Gln	Gly	Thr	Thr	Val	Ala	Pro	Ser	Ala	Ala	Pro	Ala	Thr	Arg	Pro	Ala	
385					390					395					400	
Gly	Asp	Gln	Gln	Ser	Val	Ser	Lys	Ala	Pro	Asn	Val	Gly	Ser	Arg	Thr	
				405					410					415		
Ile	Ala	Ala	Trp	Pro	His	Ser	Asp	Ala	Arg	Glu	Gly	Thr	Ala	Pro	Ser	
			420					425					430			
Thr	Thr	Asn	Ser	Val	Ala	Gly	His	Ser	Asn	Ser	Ser	Val	Phe	Pro	Arg	
		435				440						445				
Ala	Ala	Ser	Thr	Thr	Arg	Thr	Gln	His	Arg	Gly	Glu	His	Ala	Pro	Glu	
	450					455				460						
Leu	Val	Leu	Glu	Pro	Asp	Ile	Ser	Ala	Ala	Ser	Thr	Pro	Leu	Ala	Ser	
465					470					475					480	
Lys	Leu	Leu	Gly	Pro	Phe	Pro	Thr	Ser	Trp	Asp	Arg	Ser	Ile	Ser	Ser	
			485						490					495		
Pro	Gln	Pro	Gly	Gln	Arg	Thr	His	Ala	Thr	Pro	Gln	Ala	Pro	Asn	Pro	
		500						505					510			
Ser	Leu	Ser	Glu	Gly	Glu	Ile	Pro	Val	Leu	Leu	Leu	Asp	Asp	Tyr	Ser	
	515						520					525				
Glu	Glu	Glu	Glu	Gly	Arg	Lys	Glu	Glu	Val	Gly	Thr	Pro	His	Gln	Asp	
	530					535					540					
Val	Pro	Cys	Asp	Tyr	His	Pro	Cys	Lys	His	Leu	Gln	Thr	Pro	Cys	Ala	
545					550					555					560	
Glu	Leu	Gln	Arg	Arg	Trp	Arg	Cys	Arg	Cys	Pro	Gly	Leu	Ser	Gly	Glu	
			565						570					575		
Asp	Thr	Ile	Pro	Asp	Pro	Pro	Arg	Leu	Gln	Gly	Val	Thr	Glu	Thr	Thr	
		580						585					590			
Asp	Thr	Ser	Ala	Leu	Val	His	Trp	Cys	Ala	Pro	Asn	Ser	Val	Val	His	
	595						600					605				
Gly	Tyr	Gln	Ile	Arg	Tyr	Ser	Ala	Glu	Gly	Trp	Ala	Gly	Asn	Gln	Ser	
	610					615					620					
Val	Val	Gly	Val	Ile	Tyr	Ala	Thr	Ala	Arg	Gln	His	Pro	Leu	Tyr	Gly	
625					630					635					640	
Leu	Ser	Pro	Gly	Thr	Thr	Tyr	Arg	Val	Cys	Val	Leu	Ala	Ala	Asn	Arg	
			645						650					655		
Ala	Gly	Leu	Ser	Gln	Pro	Arg	Ser	Ser	Gly	Trp	Arg	Ser	Pro	Cys	Ala	
		660						665					670			
Ala	Phe	Thr	Thr	Lys	Pro	Ser	Phe	Ala	Leu	Leu	Leu	Ser	Gly	Leu	Cys	
	675						680					685				
Ala	Ala	Ser	Gly	Leu	Leu	Leu	Ala	Ser	Thr	Val	Val	Leu	Ser	Ala	Cys	
	690					695					700					
Leu	Cys	Arg	Arg	Gly	Gln	Thr	Leu	Gly	Leu	Gln	Arg	Cys	Asp	Thr	His	
705					710					715					720	

Leu Val Ala Tyr Lys Asn Pro Ala Phe Asp Asp Tyr Pro Leu Gly Leu
 725 730 735
 Gln Thr Val Ser
 740

<210> 15
 <211> 2220
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Artificial polynucleotide sequence

<221> misc_feature
 <222> (1)...(2220)
 <223> n = A,T,C or G

<400> 15
 atgmgncara cnytnccnyt nytnytnytn acngtnytnm gnccnwsntg ggcngayccn 60
 ccncargara argtnccnyt nttymgngtn acncarcarg gnccntgggg nwsnwsnggn 120
 wsnaaygcna cngaywsncc ntgygarggn ytncngcng cngaygcnac ngcnytnacn 180
 ytngcnaaym gnaayytnga rmgnytncn ggntgyytnc cnmgncacnytn nmgnwsnytn 240
 gaygcnwsnc ayaayytnyt nmngncnytn wsnacnwsng arytnngnca yytngarcar 300
 ytncargtny tnacnytnmg ncayaaymgn athgcngcny tnmngntgggg nccngngngn 360
 ccngcnggny tncayacnytn ngaytnwsn tayaaycary tngcngcnytn nytnccntgy 420
 acnggncng cnytnwsnws nytnmgngcn ytngcnytn gngnaaycc nytnmgngcn 480
 ytncargcnc cngcnttygc ntgyttyccn gcnytnary tnytnaayyt nwsntgyacn 540
 gcnytnngnm gngngncna rgngngnath gcngargcng cnttygcng ngargayggn 600
 gcncnytn gtnacnytn rgtnytn gnytnwsnggna cnttyytnga rmngntngar 660
 wsnggntgga thmgngayyt nccnaarytn acnwsnytn ayytnmgnaa ratgccnmgn 720
 ytnacnacny tngarggnga yathttyaar athacncna ayytnarca rytngaytgy 780
 cargaywsnc cngcnytn gcnwsngtngcn acncayath tycargayac nccncayyt 840
 cargtnytny tnttycaraa ytgyaayyt wsnwsnttyc cncnttgac nytn gaywsn 900
 wsncargtny tnwsnathaa yytnnttygg aayccnytna cntgywsntg ygaytnwsn 960
 tggtytnytna cngaygcnaa rmgnacngtn ytwnsmgng cngcngayac natgtgygc 1020
 ccngcngcng gnwsnwsngg nccnttywsn gcwnsytnw snytnwsnca rytncnggn 1080
 gtntgycarw sngaycarws nacnacnytn ggngcnwsnc ayccncntg yttayaaymgn 1140
 wsnacntayg cncarggnac nacngtngcn ccnwsngcng cncngcnac nmgnccngcn 1200
 ggngaycarc arwsngtnws naargcncn aaygtnggnw snmgncacnat hgcngcntgg 1260
 ccncaywsng aygcnmgnga rggnacngcn ccnwsnacna cnaaywsngt ngcnggncay 1320
 wsnaaywsnw sngtnttycc nmngcngcn wsnacnacm gnacncarca ymgngngar 1380
 caygcncng arytngtnyt ngarccngay athwsngcng cnwsnacnc nytn gcnwsn 1440

aarytnytng	gncnttycc	nacnwsntgg	gaymgnwsna	thwsnwsncc	ncarccnggn	1500
carmgnacnc	aygcnaccc	ncargcnccn	aayccnwsny	tnwsngargg	ngarathccn	1560
gtnytnytny	tngaygayta	ywsngargar	gargargggn	gnaargarga	rgtnggnacn	1620
ccncaycarg	aygtncctg	ygaytaycay	ccntgyaarc	ayytncarac	ncctgygcn	1680
garytncarm	gnmgntggmg	ntgymgntgy	ccnggnytnw	snggngarga	yacnathccn	1740
gayccnccnm	gnytncargg	ngtnacngar	acnacngaya	cnwsngcnyt	ngtncaytgg	1800
tgycncnca	aywsngtngt	ncayggntay	carathmgnt	aywsngcnga	rggntgggcn	1860
ggnaaycarw	sngtngtngg	ngtnathtay	gcnacngcnm	gnarcaycc	nytnayggg	1920
ytnwsnccng	gnacnacnta	ymgngtntgy	gtnytngcng	cnaaymgngc	nggnytnwsn	1980
carccnmgnw	snwsnggntg	gmgnwsnccn	tgycngcnt	tyacnacnaa	rccnwsntty	2040
gcnytnytny	tnwsnggnyt	ntgygcngcn	wsnggnytny	tnytngcnws	nacngtngtn	2100
ytnwsngcnt	gyytntgymg	nmgnggnear	acnytnggny	tnarmgntg	ygayacncay	2160
ytngtngcnt	ayaaraaycc	ngcnttygay	gaytayccny	tnggnytnca	racngtnwsn	2220