



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 15/12, C07K 14/47, A61K 38/17, C07K 16/18, G01N 33/50</b>		A2	(11) International Publication Number: <b>WO 00/15797</b>
(21) International Application Number: <b>PCT/US99/21547</b>		(43) International Publication Date: <b>23 March 2000 (23.03.00)</b>	
(22) International Filing Date: <b>15 September 1999 (15.09.99)</b>			
(30) Priority Data: 60/100,858 17 September 1998 (17.09.98) US PCT/US98/19437 17 September 1998 (17.09.98) US			(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, 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, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, 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), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(71) Applicant (for all designated States except US): GENENTECH, INC. [US/US]; 1 DNA Way, South San Francisco, CA 94080-4990 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): FONG, Sherman [US/US]; 19 Basinside Way, Alameda, CA 94502 (US). GODDARD, Audrey [CA/US]; 110 Congo Street, San Francisco, CA 94131 (US). GURNEY, Austin, L. [US/US]; 1 Debbie Lane, Belmont, CA 94002 (US). TUMAS, Daniel [US/US]; 3 Rae Avenue, Orinda, CA 94563 (US). WOOD, William, I. [US/US]; 35 Southdown Court, Hillsborough, CA 94010 (US).			
(74) Agents: SCHWARTZ, Timothy, R. et al.; Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080-4990 (US).			

(54) Title: COMPOSITIONS AND METHODS FOR THE TREATMENT OF IMMUNE RELATED DISEASES

## (57) Abstract

The present invention relates to a composition containing novel proteins and methods for the diagnosis and treatment of immune related diseases.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## COMPOSITIONS AND METHODS FOR THE TREATMENT OF IMMUNE RELATED DISEASES

### Field of the Invention

The present invention relates to compositions and methods for the diagnosis and treatment of immune related diseases.

### Background of the Invention

Immune related and inflammatory diseases are the manifestation or consequence of fairly complex, often multiple interconnected biological pathways which in normal physiology are critical to respond to insult or injury, initiate repair from insult or injury, and mount innate and acquired defense against foreign organisms. Disease or pathology occurs when these normal physiological pathways cause additional insult or injury either as directly related to the intensity of the response, as a consequence of abnormal regulation or excessive stimulation, as a reaction to self, or as a combination of these.

Though the genesis of these diseases often involves multistep pathways and often multiple different biological systems/pathways, intervention at critical points in one or more of these pathways can have an ameliorative or therapeutic effect. Therapeutic intervention can occur by either antagonism of a detrimental process/pathway or stimulation of a beneficial process/pathway.

Many immune related diseases are known and have been extensively studied. Such diseases include immune-mediated inflammatory diseases, non-immune-mediated inflammatory diseases, infectious diseases, immunodeficiency diseases, neoplasia, etc.

T lymphocytes (T cells) are an important component of a mammalian immune response. T cells recognise antigens which are associated with a self-molecule encoded by genes within the major histocompatibility complex (MHC). The antigen may be displayed together with MHC molecules on the surface of antigen presenting cells, virus infected cells, cancer cells, grafts, etc. The T cell system eliminates these altered cells which pose a health threat to the host mammal. T cells include helper T cells and cytotoxic T cells. Helper T cells proliferate extensively following recognition of an antigen-MHC complex on an antigen presenting cell. Helper T cells also secrete a variety of cytokines, i.e. lymphokines, which play a central role in the activation of B cells, cytotoxic T cells and a variety of other cells which participate in the immune response.

A central event in both humoral and cell mediated immune responses is the activation and clonal expansion of helper T cells. Helper T cell activation is initiated by the interaction of the T cell receptor (TCR) - CD3 complex with an antigen-MHC on the surface of an antigen presenting cell.

This interaction mediates a cascade of biochemical events that induce the resting helper T cell to enter a cell cycle (the G0 to G1 transition) and results in the expression of a high affinity receptor for IL-2 and sometimes IL-4. The activated T cell progresses through the cycle proliferating and differentiating into memory cells or effector cells.

In addition to the signals mediated through the TCR, activation of T cells involves additional costimulation induced by cytokines released by the antigen presenting cell or through interactions with membrane bound molecules on the antigen presenting cell and the T cell. The cytokines IL-1 and IL-6 have been shown to provide a costimulatory signal. Also, the interaction between the B7 molecule expressed on the surface of an antigen presenting cell and CD28 and CTLA-4 molecules expressed on the T cell surface effect T cell activation. Activated T cells express an increased number of cellular adhesion molecules, such as ICAM-1, integrins, VLA-4, LFA-1, CD56, etc.

T-cell proliferation in a mixed lymphocyte culture or mixed lymphocyte reaction (MLR) is an established indication of the ability of a compound to stimulate the immune system. In many immune responses, inflammatory cells infiltrate the site of injury or infection. The migrating cells may be neutrophilic, eosinophilic, monocytic or lymphocytic. Histologic examination of the affected tissues provides evidence of an immune stimulating or inhibiting responseCurrent Protocols in Immunology, ed. John E. Coligan, 1994, John Wiley & Sons, Inc.

Immune related diseases can be treated by suppressing the immune response. Using neutralizing antibodies that inhibit molecules having immune stimulatory activity would be beneficial in the treatment of immune-mediated and inflammatory diseases. Molecules which inhibit the immune response can be utilized (proteins directly or via the use of antibody agonists) to inhibit the immune response and thus ameliorate immune related disease.

#### Summary of the Invention

The present invention concerns compositions and methods for the diagnosis and treatment of immune related disease in mammals, including humans. The present invention is based on the identification of proteins (including agonist and antagonist antibodies) which either stimulate or inhibit the immune response in mammals. Immune related diseases can be treated by suppressing or enhancing the immune response. Molecules that enhance the immune response stimulate or potentiate the immune response to an antigen. Molecules which stimulate the immune response can be used therapeutically where enhancement of the immune response would be beneficial. Such stimulatory molecules can also be inhibited where suppression of the immune response would be of value. Neutralizing antibodies are examples of molecules that inhibit molecules having immune stimulatory activity and which would be beneficial in the treatment of immune related and inflammatory diseases. Molecules which inhibit the immune response can also be utilized (proteins directly or via the use of antibody agonists) to inhibit the immune response and thus ameliorate immune related disease.

Accordingly, the proteins of the invention encoded by the genes of the invention are useful for the diagnosis and/or treatment (including prevention) of immune related diseases. Antibodies which bind to stimulatory proteins are useful to suppress the immune system and the immune response. Antibodies which bind to inhibitory proteins are useful to stimulate the immune system and the immune response. The proteins and antibodies of the invention are also useful to prepare medicines and medicaments for the treatment of immune related and inflammatory diseases.

In one embodiment, the present invention concerns an isolated antibody which binds a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. In one aspect, the antibody mimics the activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (an agonist antibody) or conversely the antibody inhibits or neutralizes the activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (an antagonist antibody). In another aspect, the antibody is a monoclonal antibody, which preferably has nonhuman complementarity determining region (CDR) residues and human framework region (FR) residues. The antibody may be labeled and may be immobilized on a solid support. In a further aspect, the antibody is an antibody fragment, a single-chain antibody, or an anti-idiotypic antibody.

In another embodiment, the invention concerns a composition containing a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide or an agonist or antagonist antibody which binds the polypeptide in admixture with a carrier or excipient. In one aspect, the composition contains a therapeutically effective amount of the peptide or antibody. In another aspect, when the composition contains an immune stimulating molecule, the composition is useful for: (a) increasing infiltration of inflammatory cells into a tissue of a mammal in need thereof, (b) stimulating or enhancing an immune response in a mammal in need thereof, or (c) increasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen. In a further aspect, when the composition contains an immune inhibiting molecule, the composition is useful for: (a) decreasing infiltration of inflammatory cells into a tissue of a mammal in need thereof, (b) inhibiting or reducing an immune response in a mammal in need thereof, or (c) decreasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen. In another aspect, the composition contains a further active ingredient, which may, for example, be a further antibody or a cytotoxic or chemotherapeutic agent. Preferably, the composition is sterile.

In another embodiment, the invention concerns the use of the polypeptides and antibodies of the invention to prepare a composition or medicament which has the uses described above.

In a further embodiment, the invention concerns nucleic acid encoding an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody, and vectors and recombinant host cells comprising such nucleic acid. In a still further embodiment, the invention concerns a method for producing such an antibody by culturing a host cell transformed with nucleic acid encoding the antibody under conditions such that the antibody is expressed, and recovering the

antibody from the cell culture.

The invention further concerns antagonists and agonists of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide that inhibit one or more of the functions or activities of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide.

In a further embodiment, the invention concerns isolated nucleic acid molecules that hybridize to the complement of the nucleic acid molecules encoding the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides. The nucleic acid preferably is DNA, and hybridization preferably occurs under stringent conditions. Such nucleic acid molecules can act as antisense molecules of the amplified genes identified herein, which, in turn, can find use in the modulation of the respective amplified genes, or as antisense primers in amplification reactions. Furthermore, such sequences can be used as part of ribozyme and/or triple helix sequence which, in turn, may be used in regulation of the amplified genes.

In another embodiment, the invention concerns a method for determining the presence of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide comprising exposing a cell suspected of containing the polypeptide to an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody and determining binding of the antibody to the cell.

In yet another embodiment, the present invention concerns a method of diagnosing an immune related disease in a mammal, comprising detecting the level of expression of a gene encoding a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (a) in a test sample of tissue cells obtained from the mammal, and (b) in a control sample of known normal tissue cells of the same cell type, wherein a higher expression level in the test sample indicates the presence of immune related disease in the mammal from which the test tissue cells were obtained.

In another embodiment, the present invention concerns a method of diagnosing an immune disease in a mammal, comprising (a) contacting an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody with a test sample of tissue cells obtained from the mammal, and (b) detecting the formation of a complex between the antibody and the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide in the test sample. The detection may be qualitative or quantitative, and may be performed in comparison with monitoring the complex formation in a control sample of known normal tissue cells of the same cell type. A larger quantity of complexes formed in the test sample indicates the presence of tumor in the mammal from which the test tissue cells were obtained. The antibody preferably carries a detectable label. Complex formation can be monitored, for example, by light microscopy, flow cytometry, fluorimetry, or other techniques known in the art. The test sample is usually obtained from an individual suspected of having a deficiency or abnormality of the immune system.

In another embodiment, the present invention concerns a diagnostic kit, containing an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody and a carrier (e.g. a

buffer) in suitable packaging. The kit preferably contains instructions for using the antibody to detect the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide.

In a further embodiment, the invention concerns an article of manufacture, comprising:

- a container;
- a label on the container; and
- a composition comprising an active agent contained within the container; wherein the composition is effective for stimulating or inhibiting an immune response in a mammal, the label on the container indicates that the composition can be used to treat an immune related disease, and the active agent in the composition is an agent stimulating or inhibiting the expression and/or activity of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. In a preferred aspect, the active agent is a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide or an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody.

A further embodiment is a method for identifying a compound capable of inhibiting the expression and/or activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide by contacting a candidate compound with a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide under conditions and for a time sufficient to allow these two components to interact. In a specific aspect, either the candidate compound or the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide is immobilized on a solid support. In another aspect, the non-immobilized component carries a detectable label.

#### Brief Description of the Drawings

Figures 1A-D and Table 4 show hypothetical exemplifications for using the below described method to determine % amino acid sequence identity (Figures 1A-B) and % nucleic acid sequence identity (Figures 1C-D) using the ALIGN-2 sequence comparison computer program, wherein "PRO" represents the amino acid sequence of a hypothetical polypeptide of the invention of interest, "Comparison Protein" represents the amino acid sequence of a polypeptide against which the "PRO" polypeptide of interest is being compared, "PRO-DNA" represents a hypothetical "PRO"-encoding nucleic acid sequence of interest, "Comparison DNA" represents the nucleotide sequence of a nucleic acid molecule against which the "PRO-DNA" nucleic acid molecule of interest is being compared, "X", "Y" and "Z" each represent different hypothetical amino acid residues and "N", "L" and "V" each represent different hypothetical nucleotides.

Figures 2A-P and Table 5 provide the complete source code for the ALIGN-2 sequence comparison computer program. This source code may be routinely compiled for use on a UNIX operating system to provide the ALIGN-2 sequence comparison computer program.

Figure 3 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO245(UNQ219), wherein the nucleotide sequence (SEQ ID NO: 1) is a

clone designated herein as "DNA35638". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 4 and Table 6 show the amino acid sequence (SEQ ID NO: 2) of a native sequence PRO245 polypeptide as derived from the coding sequence of Figure 3. Also shown are the approximate locations of various other important polypeptide domains if known.

Figure 5 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO217(UNQ191), wherein the nucleotide sequence (SEQ ID NO: 3) is a clone designated herein as "DNA33094". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 6 and Table 7 show the amino acid sequence (SEQ ID NO: 4) of a native sequence PRO217 polypeptide as derived from the coding sequence of Figure 5. Also shown are the approximate locations of various other important polypeptide domains if known.

Figure 7 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO301(UNQ264), wherein the nucleotide sequence (SEQ ID NO: 5) is a clone designated herein as "DNA40628". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 8 and Table 8 show the amino acid sequence (SEQ ID NO: 6) of a native sequence PRO301 polypeptide as derived from the coding sequence of Figure 7. Also shown are the approximate locations of various other important polypeptide domains if known.

Figure 9 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO266 (UNQ233), wherein the nucleotide sequence (SEQ ID NO: 7) is a clone designated herein as "DNA37150". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 10 and Table 9 show the amino acid sequence (SEQ ID NO: 8) of a native sequence PRO266 polypeptide as derived from the coding sequence of Figure 9. Also shown are the approximate locations of various other important polypeptide domains if known.

Figure 11 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO335 (UNQ287V), wherein the nucleotide sequence (SEQ ID NO: 9) is a clone designated herein as "DNA41388". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 12 and Table 10 show the amino acid sequence (SEQ ID NO: 10) of a native sequence PRO335 polypeptide as derived from the coding sequence of Figure 11. Also shown are the approximate locations of various other important polypeptide domains if known.

Figure 13 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO331 (UNQ292), wherein the nucleotide sequence (SEQ ID NO: 11) is a

clone designated herein as "DNA40981". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 14 and Table 11 show the amino acid sequence (SEQ ID NO: 12) of a native sequence PRO331 polypeptide as derived from the coding sequence of Figure 13. Also shown are the approximate locations of various other important polypeptide domains if known.

Figure 15 shows the nucleotide sequence of a cDNA containing a nucleotide sequence encoding native sequence PRO326 (UNQ287), wherein the nucleotide sequence (SEQ ID NO: 13) is a clone designated herein as "DNA37140". Also presented in bold font and underlined are the positions of the respective start and stop codons.

Figure 16 and Table 12 show the amino acid sequence (SEQ ID NO: 14) of a native sequence PRO331 polypeptide as derived from the coding sequence of Figure 15. Also shown are the approximate locations of various other important polypeptide domains if known.

#### Detailed Description of the Preferred Embodiments

##### I. Definitions

The term "immune related disease" means a disease in which a component of the immune system of a mammal causes, mediates or otherwise contributes to a morbidity in the mammal. Also included are diseases in which stimulation or intervention of the immune response has an ameliorative effect on progression of the disease. Included within this term are immune-mediated inflammatory diseases, non-immune-mediated inflammatory diseases, infectious diseases, immunodeficiency diseases, neoplasia, etc.

The term "T cell mediated" disease means a disease in which T cells directly or indirectly mediate or otherwise contribute to a morbidity in a mammal. The T cell mediated disease may be associated with cell mediated effects, lymphokine mediated effects, etc., and even effects associated with B cells if the B cells are stimulated, for example, by the lymphokines secreted by T cells.

Examples of immune-related and inflammatory diseases, some of which are immune or T cell mediated, which can be treated according to the invention include systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic arthritis, spondyloarthropathies, systemic sclerosis (scleroderma), idiopathic inflammatory myopathies (dermatomyositis, polymyositis), Sjögren's syndrome, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia (immune pancytopenia, paroxysmal nocturnal hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic purpura, immune-mediated thrombocytopenia), thyroiditis (Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic thyroiditis), diabetes mellitus, immune-mediated renal disease (glomerulonephritis, tubulointerstitial nephritis), demyelinating diseases of the central and peripheral nervous systems such as multiple sclerosis, idiopathic demyelinating polyneuropathy or

Guillain-Barré syndrome, and chronic inflammatory demyelinating polyneuropathy, hepatobiliary diseases such as infectious hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune chronic active hepatitis, primary biliary cirrhosis, granulomatous hepatitis, and sclerosing cholangitis, inflammatory and fibrotic lung diseases such as inflammatory bowel disease (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and Whipple's disease, autoimmune or immune-mediated skin diseases including bullous skin diseases, erythema multiforme and contact dermatitis, psoriasis, allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity and urticaria, immunologic diseases of the lung such as eosinophilic pneumonias, idiopathic pulmonary fibrosis and hypersensitivity pneumonitis, transplantation associated diseases including graft rejection and graft -versus-host-disease. Infectious diseases include AIDS (HIV infection), hepatitis A, B, C, D, and E, bacterial infections, fungal infections, protozoal infections and parasitic infections.

"Treatment" is an intervention performed with the intention of preventing the development or altering the pathology of a disorder. Accordingly, "treatment" refers to both therapeutic treatment and prophylactic or preventative measures. Those in need of treatment include those already with the disorder as well as those in which the disorder is to be prevented. In treatment of an immune related disease, a therapeutic agent may directly decrease or increase the magnitude of response of a component of the immune response, or render the disease more susceptible to treatment by other therapeutic agents, e.g. antibiotics, antifungals, anti-inflammatory agents, chemotherapeutics, etc.

The "pathology" of an immune related disease includes all phenomena that compromise the well-being of the patient. This includes, without limitation, abnormal or uncontrollable cell growth (neutrophilic, eosinophilic, monocytic, lymphocytic cells), antibody production, auto-antibody production, complement production, interference with the normal functioning of neighboring cells, release of cytokines or other secretory products at abnormal levels, suppression or aggravation of any inflammatory or immunological response, infiltration of inflammatory cells (neutrophilic, eosinophilic, monocytic, lymphocytic) into cellular spaces, etc.

"Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, horses, cats, cows, etc. Preferably, the mammal is human.

Administration "in combination with" one or more further therapeutic agents includes simultaneous (concurrent) and consecutive administration in any order.

"Chronic" administration refers to administration of the agent(s) in a continuous mode as opposed to an acute mode, so as to maintain the initial therapeutic effect (activity) for an extended period of time. "Intermittent" administration is treatment that is not consecutively done without interruption, but rather is cyclic in nature.

"Carriers" as used herein include pharmaceutically acceptable carriers, excipients, or stabilizers which are nontoxic to the cell or mammal being exposed thereto at the dosages and concentrations employed. Often the physiologically acceptable carrier is an aqueous pH buffered solution. Examples of physiologically acceptable carriers include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptide; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN™, polyethylene glycol (PEG), and PLURONICSTM.

The term "cytotoxic agent" as used herein refers to a substance that inhibits or prevents the function of cells and/or causes destruction of cells. The term is intended to include radioactive isotopes (e.g. I<sup>131</sup>, I<sup>125</sup>, Y<sup>90</sup> and Re<sup>186</sup>), chemotherapeutic agents, and toxins such as enzymatically active toxins of bacterial, fungal, plant or animal origin, or fragments thereof.

A "chemotherapeutic agent" is a chemical compound useful in the treatment of cancer. Examples of chemotherapeutic agents include adriamycin, doxorubicin, epirubicin, 5-fluorouracil, cytosine arabinoside ("Ara-C"), cyclophosphamide, thiotapec, busulfan, cytoxin, taxoids, e.g. paclitaxel (Taxol, Bristol-Myers Squibb Oncology, Princeton, NJ), and doxetaxel (Taxotere, Rhone-Poulenc Rorer, Antony, France), toxotere, methotrexate, cisplatin, melphalan, vinblastine, bleomycin, etoposide, ifosfamide, mitomycin C, mitoxantrone, vincristine, vinorelbine, carboplatin, teniposide, daunomycin, carminomycin, aminopterin, dactinomycin, mitomycins, esperamicins (see U.S. Pat. No. 4,675,187), melphalan and other related nitrogen mustards. Also included in this definition are hormonal agents that act to regulate or inhibit hormone action on tumors such as tamoxifen and onapristone.

A "growth inhibitory agent" when used herein refers to a compound or composition which inhibits growth of a cell, especially cancer cell overexpressing any of the genes identified herein, either *in vitro* or *in vivo*. Thus, the growth inhibitory agent is one which significantly reduces the percentage of cells overexpressing such genes in S phase. Examples of growth inhibitory agents include agents that block cell cycle progression (at a place other than S phase), such as agents that induce G1 arrest and M-phase arrest. Classical M-phase blockers include the vincas (vincristine and vinblastine), taxol, and topo II inhibitors such as doxorubicin, epirubicin, daunorubicin, etoposide, and bleomycin. Those agents that arrest G1 also spill over into S-phase arrest, for example, DNA alkylating agents such as tamoxifen, prednisone, dacarbazine, mechlorethamine, cisplatin, methotrexate, 5-fluorouracil, and ara-C. Further information can be found in *The Molecular Basis of Cancer*, Mendelsohn and Israel, eds.. Chapter 1, entitled "Cell cycle regulation, oncogens, and

antineoplastic drugs" by Murakami *et al.* (WB Saunders: Philadelphia, 1995), especially p. 13.

The term "cytokine" is a generic term for proteins released by one cell population which act on another cell as intercellular mediators. Examples of such cytokines are lymphokines, monokines, and traditional polypeptide hormones. Included among the cytokines are growth hormone such as human growth hormone, N-methionyl human growth hormone, and bovine growth hormone; parathyroid hormone; thyroxine; insulin; proinsulin; relaxin; prorelaxin; glycoprotein hormones such as follicle stimulating hormone (FSH), thyroid stimulating hormone (TSH), and luteinizing hormone (LH); hepatic growth factor; fibroblast growth factor; prolactin; placental lactogen; tumor necrosis factor-alpha and -beta; mullerian-inhibiting substance; mouse gonadotropin-associated peptide; inhibin; activin; vascular endothelial growth factor; integrin; thrombopoietin (TPO); nerve growth factors such as NGF-beta; platelet-growth factor; transforming growth factors (TGFs) such as TGF-alpha and TGF-beta; insulin-like growth factor-I and -II; erythropoietin (EPO); osteoinductive factors; interferons such as interferon-alpha, -beta, and -gamma; colony stimulating factors (CSFs) such as macrophage-CSF (M-CSF); granulocyte-macrophage-CSF (GM-CSF); and granulocyte-CSF (G-CSF); interleukins (ILs) such as IL-1, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-11, IL-12; a tumor necrosis factor such as TNF-alpha or TNF-beta; and other polypeptide factors including LIF and kit ligand (KL). As used herein, the term cytokine includes proteins from natural sources or from recombinant cell culture and biologically active equivalents of the native sequence cytokines.

As used herein, a "PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide" refers to a native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 having the same amino acid sequence as a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 derived from nature. Such native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be isolated from nature or can be produced by recombinant and/or synthetic means. The term specifically encompasses naturally-occurring truncated or secreted forms (*e.g.*, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. In one embodiment of the invention, the native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is a mature or full-length native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 comprising the amino acid sequences shown in Figures 4, 6, 8, 10, 12, 14 and 16.

The term "polypeptide of the invention" refers to each individual PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. All disclosures in this specification which refer to the "polypeptide of the invention" or to "the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide" refer to each of the polypeptides individually as well as jointly. For example, descriptions of the preparation of, purification of, derivation of, formation of antibodies to or against, administration of, compositions containing, treatment of a disease with, etc.,

pertain to each polypeptide of the invention individually. The term "compound of the invention" includes the polypeptide of the invention, as well as agonist antibodies for and antagonist antibodies to these polypeptide, peptides or small molecules having agonist or antagonist activity developed from the polypeptide, etc.

The term "polypeptide of the invention" also includes variants of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides. A "variant" polypeptide means an active polypeptide as defined below having at least about 80% amino acid sequence identity with the amino acid sequence of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides. Such variant polypeptides include, for instance, polypeptides wherein one or more amino acid residues are added, or deleted, at the N- and/or C-terminus, as well as within one or more internal domains. Ordinarily, a variant polypeptide will have at least about 80% amino acid sequence identity, more preferably at least about 81% amino acid sequence identity, more preferably at least about 82% amino acid sequence identity, more preferably at least about 83% amino acid sequence identity, more preferably at least about 84% amino acid sequence identity, more preferably at least about 85% amino acid sequence identity, more preferably at least about 86% amino acid sequence identity, more preferably at least about 87% amino acid sequence identity, more preferably at least about 88% amino acid sequence identity, more preferably at least about 89% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity, more preferably at least about 91% amino acid sequence identity, more preferably at least about 92% amino acid sequence identity, more preferably at least about 93% amino acid sequence identity, more preferably at least about 94% amino acid sequence identity, more preferably at least about 95% amino acid sequence identity, more preferably at least about 96% amino acid sequence identity, more preferably at least about 97% amino acid sequence identity, more preferably at least about 98% amino acid sequence identity and yet more preferably at least about 99% amino acid sequence identity with the amino acid sequence of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides. Variants do not encompass the native polypeptide sequence.

Ordinarily, variant polypeptides of the invention are at least about 10 amino acids in length, often at least about 20 amino acids in length, more often at least about 30 amino acids in length, more often at least about 40 amino acids in length, more often at least about 50 amino acids in length, more often at least about 60 amino acids in length, more often at least about 70 amino acids in length, more often at least about 80 amino acids in length, more often at least about 90 amino acids in length, more often at least about 100 amino acids in length, more often at least about 150 amino acids in length, more often at least about 200 amino acids in length, more often at least about 300 amino acids in length, or more.

"Percent (%) amino acid sequence identity" with respect to the polypeptide sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are

identical with the amino acid residues in a sequence of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared. For purposes herein, however, % amino acid sequence identity values are obtained as described below by using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Figures 2A-P. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc. and the source code shown in Figures 2A-P has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Figures 2A-P. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

For purposes herein, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

$$100 \text{ times the fraction } X/Y$$

where X is the number of amino acid residues scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A. As examples of % amino acid sequence identity calculations, Figures 1A-B demonstrate how to calculate the % amino acid sequence identity of the amino acid sequence designated "Comparison Protein" to the amino acid sequence designated "PRO".

Unless specifically stated otherwise, all % amino acid sequence identity values used herein are obtained as described above using the ALIGN-2 sequence comparison computer program. However, % amino acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 (Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)). The NCBI-

BLAST2 sequence comparison program may be downloaded from <http://www.ncbi.nlm.nih.gov>. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for amino acid sequence comparisons, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

$$100 \text{ times the fraction } X/Y$$

where X is the number of amino acid residues scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A.

Also included within the term "polypeptides of the invention" are polypeptides which in the context of the amino acid sequence identity comparisons performed as described above, include amino acid residues in the sequences compared that are not only identical, but also those that have similar properties. These polypeptides are termed "positives". Amino acid residues that score a positive value to an amino acid residue of interest are those that are either identical to the amino acid residue of interest or are a preferred substitution (as defined in Table 1 below) of the amino acid residue of interest. For purposes herein, the % value of positives of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % positives to, with, or against a given amino acid sequence B) is calculated as follows:

$$100 \text{ times the fraction } X/Y$$

where X is the number of amino acid residues scoring a positive value as defined above by the sequence alignment program ALIGN-2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % positives of A to B will not equal the % positives of B to A.

An "invention variant polynucleotide" or "invention variant nucleic acid sequence" means a nucleic acid molecule which encodes an active polypeptide of the invention as defined herein and which has at least about 80% nucleic acid sequence identity with the nucleotide acid sequence of DNA35638, DNA33094, DNA40628, DNA37150, DNA41388, DNA40981, or DNA37140 or a specifically derived fragment thereof. Ordinarily, an invention variant polynucleotide will have at least about 80% nucleic acid sequence identity, more preferably at least about 81% nucleic acid sequence identity, more preferably at least about 82% nucleic acid sequence identity, more preferably at least about 83% nucleic acid sequence identity, more preferably at least about 84% nucleic acid sequence identity, more preferably at least about 85% nucleic acid sequence identity, more preferably at least about 86% nucleic acid sequence identity, more preferably at least about 87% nucleic acid sequence identity, more preferably at least about 88% nucleic acid sequence identity, more preferably at least about 89% nucleic acid sequence identity, more preferably at least about 90% nucleic acid sequence identity, more preferably at least about 91% nucleic acid sequence identity, more preferably at least about 92% nucleic acid sequence identity, more preferably at least about 93% nucleic acid sequence identity, more preferably at least about 94% nucleic acid sequence identity, more preferably at least about 95% nucleic acid sequence identity, more preferably at least about 96% nucleic acid sequence identity, more preferably at least about 97% nucleic acid sequence identity, more preferably at least about 98% nucleic acid sequence identity and yet more preferably at least about 99% nucleic acid sequence identity with the nucleic acid sequence of DNA35638, DNA33094, DNA40628, DNA37150, DNA41388, DNA40981, or DNA37140 or a derived fragment thereof. Variants do not encompass the native nucleotide sequence. In this regard, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of invention variant polynucleotides having at least about 80% nucleic acid sequence identity to nucleotides DNA35638, DNA33094, DNA40628, DNA37150, DNA41388, DNA40981, or DNA37140 will encode a polypeptide having an amino acid sequence which is identical to the amino acid sequence of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 or the amino acid sequence encoded by the clones deposited with the ATCC described below.

Ordinarily, invention variant polynucleotides are at least about 30 nucleotides in length, often at least about 60 nucleotides in length, more often at least about 90 nucleotides in length, more often at least about 120 nucleotides in length, more often at least about 150 nucleotides in length, more often at least about 180 nucleotides in length, more often at least about 210 nucleotides in length, more often at least about 240 nucleotides in length, more often at least about 270 nucleotides in length, more often at least about 300 nucleotides in length, more often at least about 450 nucleotides in length, more often at least about 600 nucleotides in length, more often at least about 900 nucleotides in length, or more.

"Percent (%) nucleic acid sequence identity" with respect to the invention polypeptide-encoding nucleic acid sequences identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in an invention polypeptide-encoding sequence of interest, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared. For purposes herein, however, % nucleic acid sequence identity values are obtained as described below by using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Figures 2A-P. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc. and the source code shown in Figures 2A-P has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Figures 2A-P. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

For purposes herein, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

$$100 \text{ times the fraction } W/Z$$

where W is the number of nucleotides scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C. As examples of % nucleic acid sequence identity calculations, Figures 1C-D demonstrate how to calculate the % nucleic acid sequence identity of the nucleic acid sequence designated "Comparison DNA" to the nucleic acid sequence designated "PRO-DNA".

Unless specifically stated otherwise, all % nucleic acid sequence identity values used herein are obtained as described above using the ALIGN-2 sequence comparison computer program. However, % nucleic acid sequence identity may also be determined using the sequence comparison

program NCBI-BLAST2 (Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from <http://www.ncbi.nlm.nih.gov>. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for sequence comparisons, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

$$100 \text{ times the fraction } W/Z$$

where W is the number of nucleotides scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C.

In other embodiments, invention variant polynucleotides are nucleic acid molecules that encode an active polypeptide of the invention and which are capable of hybridizing, preferably under stringent hybridization and wash conditions, to nucleotide sequences encoding the full-length invention polypeptide. Invention variant polypeptides include those that are encoded by an invention variant polynucleotide.

An "isolated" nucleic acid molecule encoding a polypeptide of the invention is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the polypeptide-encoding nucleic acid. An isolated polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated nucleic acid molecules therefore are distinguished from the polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated nucleic acid molecule encoding a polypeptide of the invention includes polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express a polypeptide of the invention where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a

ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 ug/ml), 0.1% SDS, and 10% dextran sulfate at 42C, with washes at 42C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55C.

"Moderately stringent conditions" may be identified as described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (e.g., temperature, ionic strength and %SDS)

less stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/mL denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

The term "epitope tagged" when used herein refers to a chimeric polypeptide comprising a polypeptide of the invention fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody can be made, yet is short enough such that it does not interfere with activity of the polypeptide to which it is fused. The tag polypeptide preferably also is fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 and 50 amino acid residues (preferably, between about 10 and 20 amino acid residues).

"Active" or "activity" in the context of variants of the polypeptide of the invention refers to form(s) of proteins of the invention which retain the biologic and/or immunologic activities of a native or naturally-occurring polypeptide of the invention.

"Biological activity" in the context of an antibody or another molecule that can be identified by the screening assays disclosed herein (e.g. an organic or inorganic small molecule, peptide, etc.) is used to refer to the ability of such molecules to induce or inhibit infiltration of inflammatory cells into a tissue, to stimulate or inhibit T-cell proliferation and to stimulate or inhibit lymphokine release by cells. Another preferred activity is increased vascular permeability or the inhibition thereof.

The term "antagonist" is used in the broadest sense, and includes any molecule that partially or fully blocks, inhibits, or neutralizes a biological activity of a native polypeptide of the invention disclosed herein. In a similar manner, the term "agonist" is used in the broadest sense and includes any molecule that mimics a biological activity of a native polypeptide of the invention disclosed herein. Suitable agonist or antagonist molecules specifically include agonist or antagonist antibodies or antibody fragments, fragments or amino acid sequence variants of native polypeptides of the invention, peptides, small organic molecules, etc.

A "small molecule" is defined herein to have a molecular weight below about 600 daltons.

"Antibodies" (Abs) and "immunoglobulins" (Igs) are glycoproteins having the same structural characteristics. While antibodies exhibit binding specificity to a specific antigen, immunoglobulins include both antibodies and other antibody-like molecules which lack antigen specificity. Polypeptides of the latter kind are, for example, produced at low levels by the lymph system and at increased levels by myelomas. The term "antibody" is used in the broadest sense and specifically covers, without limitation, intact monoclonal antibodies, polyclonal antibodies, multispecific antibodies (e.g. bispecific antibodies) formed from at least two intact antibodies, and antibody

fragments so long as they exhibit the desired biological activity. An anti- PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody is an antibody which immunologically binds to a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. The antibody may bind to any domain of the polypeptide of the invention which may be contacted by the antibody. For example, the antibody may bind to any extracellular domain of the polypeptide and when the entire polypeptide is secreted, to any domain on the polypeptide which is available to the antibody for binding.

"Native antibodies" and "native immunoglobulins" are usually heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light (L) chains and two identical heavy (H) chains. Each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies among the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has at one end a variable domain ( $V_H$ ) followed by a number of constant domains. Each light chain has a variable domain at one end ( $V_L$ ) and a constant domain at its other end; the constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light- chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light- and heavy-chain variable domains.

The term "variable" refers to the fact that certain portions of the variable domains differ extensively in sequence among antibodies and are used in the binding and specificity of each particular antibody for its particular antigen. However, the variability is not evenly distributed throughout the variable domains of antibodies. It is concentrated in three segments called complementarity-determining regions (CDRs) or hypervariable regions both in the light-chain and the heavy-chain variable domains. The more highly conserved portions of variable domains are called the framework (FR). The variable domains of native heavy and light chains each comprise four FR regions, largely adopting a beta-sheet configuration, connected by three CDRs, which form loops connecting, and in some cases forming part of, the beta-sheet structure. The CDRs in each chain are held together in close proximity by the FR regions and, with the CDRs from the other chain, contribute to the formation of the antigen-binding site of antibodies (see Kabat *et al.*, *NIH Publ. No. 91-3242*, Vol. I, pages 647-669 (1991)). The constant domains are not involved directly in binding an antibody to an antigen, but exhibit various effector functions, such as participation of the antibody in antibody-dependent cellular toxicity.

The term "hypervariable region" or "complementarity-determining regions" (CDRs) as used herein define a subregion within the variable region of extreme sequence variability of the antibody, which form the antigen-binding site and are the main determinants of antigen specificity. According to one definition, they can be, for example, residues (Kabat nomenclature) 24-34 (L1), 50-56 (L2) and 89-97 (L3) in the light chain variable region and residues (Kabat nomenclature) 31-35 (H1), 50-65

(H2), 95-102 (H3) in the heavy chain variable region. Kabat *et al.*, *Sequences of Proteins of Immunological Interest*, 5th Ed. Public Health Service, National Institute of Health, Bethesda, MD. [1991]. Alternatively, or in combination with the region defined by Kabat, the hypervariable region can be the "hypervariable loop", comprising, for example, residues (Chothia nomenclature) 26-32 (L1), 50-53 (L2), 91-96 (L3) in the light chain variable region and residue (Chothia nomenclature) 26-32 (H1), 53-55 (L2) and 96-101 (L3); Chothia and Lesk, *J. Mol. Biol.* 196: 901-917 [1987]. "Framework" or "FR" residues are those variable domain residues of relatively low sequence variability which lie in between the CDR regions.

"Antibody fragments" comprise a portion of an intact antibody, preferably the antigen binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')<sub>2</sub>, and Fv fragments; diabodies; linear antibodies (Zapata *et al.*, *Protein Eng.* 8(10):1057-1062 [1995]); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

Papain digestion of antibodies produces two identical antigen- binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, whose PRO245, PRO217, PRO301, Pro266, pro335, pro331 or pro326 reflects its ability to crystallize readily. Pepsin treatment yields an F(ab')<sub>2</sub> fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

"Fv" is the minimum antibody fragment which contains a complete antigen-recognition and - binding site. This region consists of a dimer of one heavy- and one light-chain variable domain in tight, non-covalent association. It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the V<sub>H</sub>-V<sub>L</sub> dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

The Fab fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. F(ab')<sub>2</sub> antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (κ) and lambda (λ), based on the amino acid sequences of their constant domains.

Depending on the amino acid sequence of the constant domain of their heavy chains,

immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), *e.g.*, IgG1, IgG2, IgG3, IgG4, IgA, and IgA2. The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called  $\alpha$ ,  $\delta$ ,  $\varepsilon$ ,  $\gamma$ , and  $\mu$ , respectively. The subunit structures and three-dimensional configurations of different classes of immunoglobulins are well known.

The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. In addition to their specificity, the monoclonal antibodies are advantageous in that they are synthesized by the hybridoma culture, uncontaminated by other immunoglobulins. The modifier "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler *et al.*, Nature, 256:495 [1975], or may be made by recombinant DNA methods (see, *e.g.*, U.S. Patent No. 4,816,567). The "monoclonal antibodies" may also be isolated from phage antibody libraries using the techniques described in Clackson *et al.*, Nature, 352:624-628 [1991] and Marks *et al.*, J. Mol. Biol., 222:581-597 (1991), for example. See also U.S Patent Nos. 5,750,373, 5,571,698, 5,403,484 and 5,223,409 which describe the preparation of antibodies using phagemid and phage vectors.

The monoclonal antibodies herein specifically include "chimeric" antibodies (immunoglobulins) in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity (U.S. Patent No. 4,816,567; Morrison *et al.*, Proc. Natl. Acad. Sci. USA, 81:6851-6855 [1984]).

"Humanized" forms of non-human (*e.g.*, murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human

immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementarity-determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity, and capacity. In some instances, Fv framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and maximize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, see Jones *et al.*, Nature, 321:522-525 (1986); Reichmann *et al.*, Nature, 332:323-329 [1988]; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992). The term "humanized antibody" includes a "primatized" antibody where the antigen-binding region of the antibody is derived from an antibody produced by immunizing macaque monkeys with the antigen of interest. Antibodies containing residues from Old World monkeys are also possible within the invention. See, for example, U.S. Patent Nos. 5,658,570; 5,693,780; 5,681,722; 5,750,105; and 5,756,096.

Antibodies and fragments thereof in this invention also include "affinity matured" antibodies in which an antibody is altered to change the amino acid sequence of one or more of the CDR regions and/or the framework regions to alter the affinity of the antibody or fragment thereof for the antigen to which it binds. Affinity maturation may result in an increase or in a decrease in the affinity of the matured antibody for the antigen relative to the starting antibody. Typically, the starting antibody will be a humanized, human, chimeric or murine antibody and the affinity matured antibody will have a higher affinity than the starting antibody. During the maturation process, one or more of the amino acid residues in the CDRs or in the framework regions are changed to a different residue using any standard method. Suitable methods include point mutations using well known cassette mutagenesis methods (Wells *et al.*, 1985, Gene, 34:315) or oligonucleotide mediated mutagenesis methods (Zoller *et al.*, 1987, Nucleic Acids Res., 10:6487-6504). Affinity maturation may also be performed using known selection methods in which many mutations are produced and mutants having the desired affinity are selected from a pool or library of mutants based on improved affinity for the antigen or ligand. Known phage display techniques can be conveniently used in this approach. See, for example, U.S. 5,750,373; U.S. 5,223,409, etc.

Human antibodies are also within the scope of the antibodies of the invention. Human antibodies can be produced using various techniques known in the art, including phage display

libraries [Hoogenboom and Winter, *J. Mol. Biol.*, **227**:381 (1991); Marks et al., *J. Mol. Biol.*, **222**:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, p. 77 (1985); Boerner et al., *J. Immunol.*, **147**(1):86-95 (1991); U. S. 5,750,373]. Similarly, human antibodies can be made by introducing of human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks *et al.*, *Bio/Technology* **10**, 779-783 (1992); Lonberg *et al.*, *Nature* **368** 856-859 (1994); Morrison, *Nature* **368**, 812-13 (1994); Fishwild *et al.*, *Nature Biotechnology* **14**, 845-51 (1996); Neuberger, *Nature Biotechnology* **14**, 826 (1996); Lonberg and Huszar, *Intern. Rev. Immunol.* **13** 65-93 (1995).

"Single-chain Fv" or "sFv" antibody fragments comprise the  $V_H$  and  $V_L$  domains of antibody, wherein these domains are present in a single polypeptide chain. Preferably, the Fv polypeptide further comprises a polypeptide linker between the  $V_H$  and  $V_L$  domains which enables the sFv to form the desired structure for antigen binding. For a review of sFv see Pluckthun in *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenberg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) in the same polypeptide chain ( $V_H - V_L$ ). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger *et al.*, *Proc. Natl. Acad. Sci. USA*, **90**:6444-6448 (1993).

An "isolated" antibody is one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the compound of the invention will be purified (1) to greater than 95% by weight of the compound as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated compound, e.g. antibody or

polypeptide, includes the compound *in situ* within recombinant cells since at least one component of the compound's natural environment will not be present. Ordinarily, however, isolated compound will be prepared by at least one purification step.

The word "label" when used herein refers to a detectable compound or composition which is conjugated directly or indirectly to the compound, e.g. antibody or polypeptide, so as to generate a "labelled" compound. The label may be detectable by itself (e.g. radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, may catalyze chemical alteration of a substrate compound or composition which is detectable.

By "solid phase" is meant a non-aqueous matrix to which the compound of the present invention can adhere. Examples of solid phases encompassed herein include those formed partially or entirely of glass (e.g., controlled pore glass), polysaccharides (e.g., agarose), polyacrylamides, polystyrene, polyvinyl alcohol and silicones. In certain embodiments, depending on the context, the solid phase can comprise the well of an assay plate; in others it is a purification column (e.g., an affinity chromatography column). This term also includes a discontinuous solid phase of discrete particles, such as those described in U.S. Patent No. 4,275,149.

A "liposome" is a small vesicle composed of various types of lipids, phospholipids and/or surfactant which is useful for delivery of a drug (such as the anti-ErbB2 antibodies disclosed herein and, optionally, a chemotherapeutic agent) to a mammal. The components of the liposome are commonly arranged in a bilayer formation, similar to the lipid arrangement of biological membranes.

As used herein, the term "immunoadhesin" designates antibody-like molecules which combine the binding specificity of a heterologous protein (an "adhesin") with the effector functions of immunoglobulin constant domains. Structurally, the immunoadhesins comprise a fusion of an amino acid sequence with the desired binding specificity which is other than the antigen recognition and binding site of an antibody (i.e., is "heterologous"), and an immunoglobulin constant domain sequence. The adhesin part of an immunoadhesin molecule typically is a contiguous amino acid sequence comprising at least the binding site of a receptor or a ligand. The immunoglobulin constant domain sequence in the immunoadhesin may be obtained from any immunoglobulin, such as IgG-1, IgG-2, IgG-3, or IgG-4 subtypes, IgA (including IgA-1 and IgA-2), IgE, IgD or IgM.

## II. Compositions and Methods of the Invention

### 1. Preparation of the polypeptides of the invention

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 (UNQ219, UNQ191, UNQ264, UNQ233, UNQ287V, UNQ292 or UNQ287 respectively). In particular, cDNA encoding a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide has been identified and isolated, as disclosed in further detail in the

Examples below. It is noted that proteins produced in separate expression rounds may be given different PRO numbers but the UNQ number is unique for any given DNA and the encoded protein, and will not be changed. However, for sake of simplicity, in the present specification the protein encoded by DNA35638, DNA33094, DNA40628, DNA37150, DNA41388, DNA40981 and DNA37140 as well as all further native homologues and variants included in the foregoing definition of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326, will be referred to as PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 or simply as "the polypeptide of the invention", regardless of their origin or mode of preparation.

The description below relates primarily to production of the polypeptide of the invention by culturing cells transformed or transfected with a vector containing nucleic acid which encodes of the polypeptide of the invention. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare of the polypeptide of the invention. For instance, the polypeptide sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of the polypeptide of the invention may be chemically synthesized separately and combined using chemical or enzymatic methods to produce the full-length polypeptide.

In addition to the full-length native sequence polypeptides described herein, it is contemplated that variants can be prepared. Variants can be prepared by introducing appropriate nucleotide changes into the DNA, and/or by synthesis of the desired polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the polypeptide of the invention, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native full-length sequence or in various domains of the polypeptide of the invention described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the polypeptide that results in a change in the amino acid sequence of the polypeptide as compared with the native sequence polypeptide sequence. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the polypeptide of the invention. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the polypeptide of the invention with that of homologous known protein molecules and minimizing the

number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of about 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity exhibited by the full-length or mature native sequence.

Polypeptide fragments of the polypeptides of the invention are also within the scope of the invention. Such fragments may be truncated at the N-terminus or C-terminus, or may lack internal residues, for example, when compared with a full length native protein. Certain fragments lack amino acid residues that are not essential for a desired biological activity of the invention polypeptide.

Polypeptide fragments may be prepared by any of a number of conventional techniques. Desired peptide fragments may be chemically synthesized. An alternative approach involves generating fragments by enzymatic digestion, e.g., by treating the protein with an enzyme known to cleave proteins at sites defined by particular amino acid residues, or by digesting the DNA with suitable restriction enzymes and isolating the desired fragment. Yet another suitable technique involves isolating and amplifying a DNA fragment encoding a desired polypeptide fragment, by polymerase chain reaction (PCR). Oligonucleotides that define the desired termini of the DNA fragment are employed at the 5' and 3' primers in the PCR. Preferably, polypeptide fragments share at least one biological and/or immunological activity with the native invention polypeptide.

In particular embodiments, conservative substitutions of interest are shown in Table 1 under the heading of preferred substitutions. If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions in Table 1, or as further described below in reference to amino acid classes, are introduced and the products screened.

Table 1

<u>Original Residue</u>	<u>Exemplary Substitutions</u>	<u>Preferred Substitutions</u>
Ala (A)	val; leu; ile	val
Arg (R)	lys; gln; asn	lys
Asn (N)	gln; his; lys; arg	gln
Asp (D)	glu	glu
Cys (C)	ser	ser
Gln (Q)	asn	asn
Glu (E)	asp	asp
Gly (G)	pro; ala	ala
His (H)	asn; gln; lys; arg	arg
Ile (I)	leu; val; met; ala; phe; norleucine	leu
Leu (L)	norleucine; ile; val; met; ala; phe	ile
Lys (K)	arg; gln; asn	arg
Met (M)	leu; phe; ile	leu
Phe (F)	leu; val; ile; ala; tyr	leu
Pro (P)	ala	ala
Ser (S)	thr	thr
Thr (T)	ser	ser
Trp (W)	tyr; phe	tyr
Tyr (Y)	trp; phe; thr; ser	phe
Val (V)	ile; leu; met; phe; ala; norleucine	leu

Substantial modifications in function or immunological identity of the invention polypeptide are accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. Naturally occurring residues are divided into groups based on common side-chain properties:

(1) hydrophobic: norleucine, met, ala, val, leu, ile;

- (2) neutral hydrophilic: cys, ser, thr;
- (3) acidic: asp, glu;
- (4) basic: asn, gln, his, lys, arg;
- (5) residues that influence chain orientation: gly, pro; and
- (6) aromatic: trp, tyr, phe.

Non-conservative substitutions will entail exchanging a member of one of these classes for another class. Such substituted residues also may be introduced into the conservative substitution sites or, more preferably, into the remaining (non-conserved) sites.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter et al., *Nucl. Acids Res.*, 13:4331 (1986); Zoller et al., *Nucl. Acids Res.*, 10:6487 (1987)], cassette mutagenesis [Wells et al., *Gene*, 34:315 (1985)], restriction selection mutagenesis [Wells et al., *Philos. Trans. R. Soc. London SerA*, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant [Cunningham and Wells, *Science*, 244: 1081-1085 (1989)]. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, *The Proteins*, (W.H. Freeman & Co., N.Y.); Chothia, *J. Mol. Biol.*, 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

Covalent modifications of polypeptides of the invention are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of an invention polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C-terminal residues of the polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking the invention polypeptide to a water-insoluble support matrix or surface for use in the method for purifying anti-polypeptide antibodies, and vice-versa. Commonly used crosslinking agents include, e.g., 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the -amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the invention polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence polypeptide (either by removing the underlying glycosylation site or by deleting the glycosylation by chemical and/or enzymatic means), and/or adding one or more glycosylation sites that are not present in the native sequence. In addition, the phrase includes qualitative changes in the glycosylation of the native proteins, involving a change in the nature and proportions of the various carbohydrate moieties present.

Addition of glycosylation sites to the polypeptide may be accomplished by altering the amino acid sequence. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence polypeptide (for O-linked glycosylation sites). The amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the polypeptide of the invention is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the polypeptide of the invention may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification comprises linking the invention polypeptide to one of a variety of nonproteinaceous polymers, e.g., polyethylene glycol (PEG), polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

The polypeptide of the present invention may also be modified in a way to form a chimeric molecule comprising the invention polypeptide fused to another, heterologous polypeptide or amino acid sequence.

In one embodiment, such a chimeric molecule comprises a fusion of the invention polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the polypeptide of the invention. The presence of such epitope-tagged forms of the polypeptide of the invention can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the polypeptide of the invention to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., *Mol. Cell. Biol.*, 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., *Molecular and Cellular Biology*, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., *Protein Engineering*, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., *BioTechnology*, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., *Science*, 255:192-194 (1992)]; an -tubulin epitope peptide [Skinner et al., *J. Biol. Chem.*, 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., *Proc. Natl. Acad. Sci. USA*, 87:6393-6397 (1990)].

In an alternative embodiment, the chimeric molecule may comprise a fusion of the polypeptide of the invention with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule (also referred to as an "immunoadhesin"), such a fusion could be to the Fc region of an IgG molecule. The Ig fusions preferably include the substitution of a soluble (transmembrane domain deleted or inactivated) form of an invention polypeptide in place of at least one variable region within an Ig molecule. In a particularly preferred embodiment, the immunoglobulin fusion includes the hinge, CH2 and CH3, or the hinge, CH1, CH2 and CH3 regions of an IgG1 molecule. For the production of immunoglobulin fusions see also US Patent No. 5,428,130 issued June 27, 1995.

i. Isolation of DNA Encoding the Polypeptide of the Invention

DNA encoding the polypeptide of the invention may be obtained from a cDNA library prepared from tissue believed to possess the polypeptide mRNA and to express it at a detectable level. Accordingly, human DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The gene encoding the polypeptide of the invention may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to the polypeptide of the invention

or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding the polypeptide of the invention is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like <sup>32</sup>P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAsstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

#### ii. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for production of the polypeptides of the invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, CaPO<sub>4</sub> and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., *supra*, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., *Gene*, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, *Virology*, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., *J. Bact.*, 130:946 (1977) and Hsiao et al., *Proc. Natl. Acad. Sci. (USA)*, 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, e.g., polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., *Methods in Enzymology*, 185:527-537 (1990) and Mansour et al., *Nature*, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for vectors encoding the polypeptides of the invention. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated polypeptides of the invention are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., *J. Gen Virol.*, 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, *Biol. Reprod.*, 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

iii. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the polypeptides of the invention may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, phagemid or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The polypeptide of the invention may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the DNA encoding the polypeptide of the invention that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* alpha-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2u plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the nucleic acid encoding the polypeptide of the invention,

such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp1* gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the nucleic acid sequence encoding the polypeptide of the invention to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the beta-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the polypeptide of the invention.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytchrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

Transcription of the polypeptide of the invention from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding the polypeptide of the invention by higher eukaryotes may

be increased by inserting an enhancer sequence into the vector. Enhancers are *cis*-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, alpha-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the coding sequence of the polypeptide of the invention, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding the polypeptide of the invention.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of the polypeptide of the invention in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

### iii. Detecting Gene Expression

Gene expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence of the inventive polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to DNA encoding the polypeptide of the invention and encoding a specific antibody epitope.

### iv. Purification of Polypeptide

Forms of the polypeptide of the invention may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (e.g. Triton-X 100) or by enzymatic cleavage. Cells employed in expression of the polypeptide of the invention can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify the polypeptide of the invention from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the polypeptide of the invention. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, *Methods in Enzymology*, 182 (1990); Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular polypeptide of the invention produced.

## 2. Tissue Distribution

The location of tissues expressing the polypeptides of the invention can be identified by determining mRNA expression in various human tissues. The location of such genes provides information about which tissues are most likely to be affected by the stimulating and inhibiting activities of the polypeptides of the invention. The location of a gene in a specific tissue also provides sample tissue for the activity blocking assays discussed below.

As noted before, gene expression in various tissues may be measured by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA (Thomas, *Proc. Natl. Acad. Sci. USA*, 77:5201-5205 [1980]), dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes.

Gene expression in various tissues, alternatively, may be measured by immunological methods, such as immunohistochemical staining of tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence of a polypeptide of the invention or against a synthetic peptide based on the DNA sequences encoding the polypeptide of the invention or against an exogenous sequence fused to a DNA encoding a polypeptide of the invention and encoding a specific antibody epitope. General techniques for

generating antibodies, and special protocols for Northern blotting and *in situ* hybridization are provided below.

### 3. Antibody Binding Studies

The activity of the polypeptides of the invention can be further verified by antibody binding studies, in which the ability of anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibodies to inhibit the effect of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides on tissue cells is tested. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies, the preparation of which will be described hereinbelow.

Antibody binding studies may be carried out in any known assay method, such as competitive binding assays, direct and indirect sandwich assays, and immunoprecipitation assays. Zola, *Monoclonal Antibodies: A Manual of Techniques*, pp.147-158 (CRC Press, Inc., 1987).

Competitive binding assays rely on the ability of a labeled standard to compete with the test sample analyte for binding with a limited amount of antibody. The amount of target protein in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies preferably are insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte which remain unbound.

Sandwich assays involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, of the protein to be detected. In a sandwich assay, the test sample analyte is bound by a first antibody which is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three-part complex. See, e.g., U.S. Pat No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assay). For example, one type of sandwich assay is an ELISA assay, in which case the detectable moiety is an enzyme.

For immunohistochemistry, the tissue sample may be fresh or frozen or may be embedded in paraffin and fixed with a preservative such as formalin, for example.

### 4. Cell-Based Assays

Cell-based assays and animal models for immune related diseases can be used to further understand the relationship between the genes and polypeptides identified herein and the development and pathogenesis of immune related disease.

In a different approach, cells of a cell type known to be involved in a particular immune related disease are transfected with the cDNAs described herein, and the ability of these cDNAs to stimulate or inhibit immune function is analyzed. Suitable cells can be transfected with the desired gene, and monitored for immune function activity. Such transfected cell lines can then be used to test

the ability of poly- or monoclonal antibodies or antibody compositions to inhibit or stimulate immune function, for example to modulate T-cell proliferation or inflammatory cell infiltration. Cells transfected with the coding sequences of the genes identified herein can further be used to identify drug candidates for the treatment of immune related diseases.

In addition, primary cultures derived from transgenic animals (as described below) can be used in the cell-based assays herein, although stable cell lines are preferred. Techniques to derive continuous cell lines from transgenic animals are well known in the art (see, e.g. Small *et al.*, Mol. Cell. Biol. 5, 642-648 [1985]).

One suitable cell based assay is the mixed lymphocyte reaction (MLR). Current Protocols in Immunology, unit 3.12; edited by J. E. Coligan, A. M. Kruisbeek, D. H. Marglies, E. M. Shevach, W. Strober, National Institutes of Health, Published by John Wiley & Sons, Inc. In this assay, the ability of a test compound to stimulate the proliferation of activated T cells is assayed. A suspension of responder T cells is cultured with allogeneic stimulator cells and the proliferation of T cells is measured by uptake of tritiated thymidine. This assay is a general measure of T cell reactivity. Since the majority of T cells respond to and produce IL-2 upon activation, differences in responsiveness in this assay in part reflect differences in IL-2 production by the responding cells. The MLR results can be verified by a standard lymphokine (IL-2) detection assay. Current Protocols in Immunology, above, 3.15, 6.3.

A proliferative T cell response in an MLR assay may be due to direct mitogenic properties of an assayed molecule or to external antigen induced activation. Additional verification of the T cell stimulatory activity of the polypeptides of the invention can be obtained by a costimulation assay. T cell activation requires an antigen specific signal mediated through the T-cell receptor (TCR) and a costimulatory signal mediated through a second ligand binding interaction, for example, the B7(CD80, CD86)/CD28 binding interaction. CD28 crosslinking increases lymphokine secretion by activated T cells. T cell activation has both negative and positive controls through the binding of ligands which have a negative or positive effect. CD28 and CTLA-4 are related glycoproteins in the Ig superfamily which bind to B7. CD28 binding to B7 has a positive costimulation effect of T cell activation; conversely, CTLA-4 binding to B7 has a negative T cell deactivating effect. Chambers, C. A. and Allison, J. P., *Curr. Opin. Immunol.* (1997) 9:396. Schwartz, R. H., *Cell* (1992) 71:1065; Linsey, P. S. and Ledbetter, J. A., *Annu. Rev. Immunol.* (1993) 11:191; June, C. H. et al., *Immunol. Today* (1994) 15:321; Jenkins, M. K., *Immunity* (1994) 1:405. In a costimulation assay, the polypeptides of the invention are assayed for T cell costimulatory or inhibitory activity.

Polypeptides of the invention, as well as other compounds of the invention, which are stimulators (costimulators) of T cell proliferation and agonists, e.g. agonist antibodies, thereto as determined by MLR and costimulation assays, for example, are useful in treating immune related diseases characterized by poor, suboptimal or inadequate immune function. These diseases are treated

by stimulating the proliferation and activation of T cells (and T cell mediated immunity) and enhancing the immune response in a mammal through administration of a stimulatory compound, such as the stimulating polypeptides of the invention. The stimulating polypeptide may, for example, be a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide or an agonist antibody therefor.

Direct use of a stimulating compound as in the invention has been validated in experiments with 4-1BB glycoprotein, a member of the tumor necrosis factor receptor family, which binds to a ligand (4-1BBL) expressed on primed T cells and signals T cell activation and growth. Alderson, M. E. et al., *J. Immunol.* (1994) 24:2219.

The use of an agonist stimulating compound has also been validated experimentally. Activation of 4-1BB by treatment with an agonist anti-4-1BB antibody enhances eradication of tumors. Hellstrom, I. and Hellstrom, K. E., *Crit. Rev. Immunol.* (1998) 18:1. Immunoadjuvant therapy for treatment of tumors, described in more detail below, is another example of the use of the stimulating compounds of the invention.

An immune stimulating or enhancing effect can also be achieved by antagonizing or blocking the activity of a protein which has been found to be inhibiting in the MLR assay. Negating the inhibitory activity of the compound produces a net stimulatory effect. Suitable antagonists/blocking compounds are antibodies or fragments thereof which recognize and bind to the inhibitory protein, thereby blocking the effective interaction of the protein with its receptor and inhibiting signaling through the receptor. This effect has been validated in experiments using anti-CTLA-4 antibodies which enhance T cell proliferation, presumably by removal of the inhibitory signal caused by CTLA-4 binding. Walunas, T. L. et al., *Immunity* (1994) 1:405.

On the other hand, polypeptides of the invention, as well as other compounds of the invention, which are direct inhibitors of T cell proliferation/activation and/or lymphokine secretion, can be directly used to suppress the immune response. These compounds are useful to reduce the degree of the immune response and to treat immune related diseases characterized by a hyperactive, superoptimal, or autoimmune response. This use of the compounds of the invention has been validated by the experiments described above in which CTLA-4 binding to receptor B7 deactivates T cells. The direct inhibitory compounds of the invention function in an analogous manner.

Alternatively, compounds, e.g. antibodies, which bind to stimulating polypeptides of the invention and block the stimulating effect of these molecules produce a net inhibitory effect and can be used to suppress the T cell mediated immune response by inhibiting T cell proliferation/activation and/or lymphokine secretion. Blocking the stimulating effect of the polypeptides suppresses the immune response of the mammal. This use has been validated in experiments using an anti-IL2 antibody. In these experiments, the antibody binds to IL2 and blocks binding of IL2 to its receptor thereby achieving a T cell inhibitory effect.

### 5. Animal Models

The results of the cell based *in vitro* assays can be further verified using *in vivo* animal models and assays for T-cell function. A variety of well known animal models can be used to further understand the role of the genes identified herein in the development and pathogenesis of immune related disease, and to test the efficacy of candidate therapeutic agents, including antibodies, and other antagonists of the native polypeptides, including small molecule antagonists. The *in vivo* nature of such models makes them predictive of responses in human patients. Animal models of immune related diseases include both non-recombinant and recombinant (transgenic) animals. Non-recombinant animal models include, for example, rodent, e.g., murine models. Such models can be generated by introducing cells into syngeneic mice using standard techniques, e.g. subcutaneous injection, tail vein injection, spleen implantation, intraperitoneal implantation, implantation under the renal capsule, etc.

Graft-versus-host disease occurs when immunocompetent cells are transplanted into immunosuppressed or tolerant patients. The donor cells recognize and respond to host antigens. The response can vary from life threatening severe inflammation to mild cases of diarrhea and weight loss.

Graft-versus-host disease models provide a means of assessing T cell reactivity against MHC antigens and minor transplant antigens. A suitable procedure is described in detail in Current Protocols in Immunology, above, unit 4.3.

An animal model for skin allograft rejection is a means of testing the ability of T cells to mediate *in vivo* tissue destruction and a measure of their role in transplant rejection. The most common and accepted models use murine tail-skin grafts. Repeated experiments have shown that skin allograft rejection is mediated by T cells, helper T cells and killer-effector T cells, and not antibodies. Auchincloss, H. Jr. and Sachs, D. H., Fundamental Immunology, 2nd ed., W. E. Paul ed., Raven Press, NY, 1989, 889-992. A suitable procedure is described in detail in Current Protocols in Immunology, above, unit 4.4. Other transplant rejection models which can be used to test the compounds of the invention are the allogeneic heart transplant models described by Tanabe, M. et al., Transplantation (1994) 58:23 and Tinubu, S. A. et al., J. Immunol. (1994) 4330-4338.

Animal models for delayed type hypersensitivity provides an assay of cell mediated immune function as well. Delayed type hypersensitivity reactions are a T cell mediated *in vivo* immune response characterized by inflammation which does not reach a peak until after a period of time has elapsed after challenge with an antigen. These reactions also occur in tissue specific autoimmune diseases such as multiple sclerosis (MS) and experimental autoimmune encephalomyelitis (EAE, a model for MS). A suitable procedure is described in detail in Current Protocols in Immunology, above, unit 4.5.

EAE is a T cell mediated autoimmune disease characterized by T cell and mononuclear cell inflammation and subsequent demyelination of axons in the central nervous system. EAE is generally

considered to be a relevant animal model for MS in humans. Bolton, C., *Multiple Sclerosis* (1995) 1:143. Both acute and relapsing-remitting models have been developed. The compounds of the invention can be tested for T cell stimulatory or inhibitory activity against immune mediated demyelinating disease using the protocol described in *Current Protocols in Immunology*, above, units 15.1 and 15.2. See also the models for myelin disease in which oligodendrocytes or Schwann cells are grafted into the central nervous system as described in Duncan, I. D. et al, *Molec. Med. Today* (1997) 554-561.

Contact hypersensitivity is a simple delayed type hypersensitivity *in vivo* assay of cell mediated immune function. In this procedure, cutaneous exposure to exogenous haptens which gives rise to a delayed type hypersensitivity reaction which is measured and quantitated. Contact sensitivity involves an initial sensitizing phase followed by an elicitation phase. The elicitation phase occurs when the T lymphocytes encounter an antigen to which they have had previous contact. Swelling and inflammation occur, making this an excellent model of human allergic contact dermatitis. A suitable procedure is described in detail in *Current Protocols in Immunology*, Eds. J. E. Cologan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach and W. Strober, John Wiley & Sons, Inc., 1994, unit 4.2. See also Grabbe, S. and Schwarz, T., *Immun. Today* 19(1):37-44 (1998) .

An animal model for arthritis is collagen-induced arthritis. This model shares clinical, histological and immunological characteristics of human autoimmune rheumatoid arthritis and is an acceptable model for human autoimmune arthritis. Mouse and rat models are characterized by synovitis, erosion of cartilage and subchondral bone. The compounds of the invention can be tested for activity against autoimmune arthritis using the protocols described in *Current Protocols in Immunology*, above, units 15.5. See also the model using a monoclonal antibody to CD18 and VLA-4 integrins described in Issekutz, A. C. et al., *Immunology* (1996) 88:569.

A model of asthma has been described in which antigen-induced airway hyper-reactivity, pulmonary eosinophilia and inflammation are induced by sensitizing an animal with ovalbumin and then challenging the animal with the same protein delivered by aerosol. Several animal models (guinea pig, rat, non-human primate) show symptoms similar to atopic asthma in humans upon challenge with aerosol antigens. Murine models have many of the features of human asthma. Suitable procedures to test the compounds of the invention for activity and effectiveness in the treatment of asthma are described by Wolyniec, W. W. et al., *Am. J. Respir. Cell Mol. Biol.* (1998) 18:777 and the references cited therein.

Additionally, the compounds of the invention can be tested on animal models for psoriasis like diseases. Evidence suggests a T cell pathogenesis for psoriasis. The compounds of the invention can be tested in the scid/scid mouse model described by Schon, M. P. et al., *Nat. Med.* (1997) 3:183, in which the mice demonstrate histopathologic skin lesions resembling psoriasis. Another suitable

model is the human skin/scid mouse chimera prepared as described by Nickoloff, B. J. et al., Am. J. Path. (1995) 146:580.

Recombinant (transgenic) animal models can be engineered by introducing the coding portion of the genes identified herein into the genome of animals of interest, using standard techniques for producing transgenic animals. Animals that can serve as a target for transgenic manipulation include, without limitation, mice, rats, rabbits, guinea pigs, sheep, goats, pigs, and non-human primates, e.g. baboons, chimpanzees and monkeys. Techniques known in the art to introduce a transgene into such animals include pronucleic microinjection (Hoppe and Wanger, U.S. Patent No. 4,873,191); retrovirus-mediated gene transfer into germ lines (e.g., Van der Putten *et al.*, Proc. Natl. Acad. Sci. USA 82, 6148-615 [1985]); gene targeting in embryonic stem cells (Thompson *et al.*, Cell 56, 313-321 [1989]); electroporation of embryos (Lo, Mol. Cel. Biol. 3, 1803-1814 [1983]); sperm-mediated gene transfer (Lavitrano *et al.*, Cell 57, 717-73 [1989]). For review, see, for example, U.S. Patent No. 4,736,866.

For the purpose of the present invention, transgenic animals include those that carry the transgene only in part of their cells ("mosaic animals"). The transgene can be integrated either as a single transgene, or in concatamers, e.g., head-to-head or head-to-tail tandems. Selective introduction of a transgene into a particular cell type is also possible by following, for example, the technique of Lasko *et al.*, Proc. Natl. Acad. Sci. USA 89, 6232-636 (1992).

The expression of the transgene in transgenic animals can be monitored by standard techniques. For example, Southern blot analysis or PCR amplification can be used to verify the integration of the transgene. The level of mRNA expression can then be analyzed using techniques such as *in situ* hybridization, Northern blot analysis, PCR, or immunocytochemistry.

The animals may be further examined for signs of immune disease pathology, for example by histological examination to determine infiltration of immune cells into specific tissues. Blocking experiments can also be performed in which the transgenic animals are treated with the compounds of the invention to determine the extent of the T cell proliferation stimulation or inhibition of the compounds. In these experiments, blocking antibodies which bind to the polypeptide of the invention, prepared as described above, are administered to the animal and the effect on immune function is determined.

Alternatively, "knock out" animals can be constructed which have a defective or altered gene encoding a polypeptide identified herein, as a result of homologous recombination between the endogenous gene encoding the polypeptide and altered genomic DNA encoding the same polypeptide introduced into an embryonic cell of the animal. For example, cDNA encoding a particular polypeptide can be used to clone genomic DNA encoding that polypeptide in accordance with established techniques. A portion of the genomic DNA encoding a particular polypeptide can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used

to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, *Cell*, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., *Cell*, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the polypeptide.

#### 6. ImmunoAdjuvant Therapy

In one embodiment, the immunostimulating compounds of the invention can be used in immunoadjuvant therapy for the treatment of tumors (cancer). It is now well established that T cells recognize human tumor specific antigens. One group of tumor antigens, encoded by the MAGE, BAGE and GAGE families of genes, are silent in all adult normal tissues, but are expressed in significant amounts in tumors, such as melanomas, lung tumors, head and neck tumors, and bladder carcinomas. DeSmet, C. et al, (1996) Proc. Natl. Acad. Sci. USA, 93:7149. It has been shown that costimulation of T cells induces tumor regression and an antitumor response both in vitro and in vivo. Melero, I. et al., Nature Medicine (1997) 3:682; Kwon, E. D. et al., Proc. Natl. Acad. Sci. USA (1997) 94:8099; Lynch, D. H. et al., Nature Medicine (1997) 3:625; Finn, O. J. and Lotze, M. T., J. Immunol. (1998) 21:114. The stimulatory compounds of the invention can be administered as adjuvants, alone or together with a growth regulating agent, cytotoxic agent or chemotherapeutic agent, to stimulate T cell proliferation/activation and an antitumor response to tumor antigens. The growth regulating, cytotoxic, or chemotherapeutic agent may be administered in conventional amounts using known administration regimes. Immunostimulating activity by the compounds of the invention allows reduced amounts of the growth regulating, cytotoxic, or chemotherapeutic agents thereby potentially lowering the toxicity to the patient.

#### 7. Screening Assays for Drug Candidates

Screening assays for drug candidates are designed to identify compounds that bind or complex with the polypeptides encoded by the genes identified herein or a biologically active fragment thereof, or otherwise interfere with the interaction of the encoded polypeptides with other cellular proteins. Such screening assays will include assays amenable to high-throughput screening of

chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds, including peptides, preferably soluble peptides, (poly)peptide-immunoglobulin fusions, and, in particular, antibodies including, without limitation, poly- and monoclonal antibodies and antibody fragments, single-chain antibodies, anti-idiotypic antibodies, and chimeric or humanized versions of such antibodies or fragments, as well as human antibodies and antibody fragments. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

All assays are common in that they call for contacting the drug candidate with a polypeptide encoded by a nucleic acid identified herein under conditions and for a time sufficient to allow these two components to interact.

In binding assays, the interaction is binding and the complex formed can be isolated or detected in the reaction mixture. In a particular embodiment, the polypeptide encoded by the gene identified herein or the drug candidate is immobilized on a solid phase, e.g. on a microtiter plate, by covalent or non-covalent attachments. Non-covalent attachment generally is accomplished by coating the solid surface with a solution of the polypeptide and drying. Alternatively, an immobilized antibody, e.g. a monoclonal antibody, specific for the polypeptide to be immobilized can be used to anchor it to a solid surface. The assay is performed by adding the non-immobilized component, which may be labeled by a detectable label, to the immobilized component, e.g. the coated surface containing the anchored component. When the reaction is complete, the non-reacted components are removed, e.g. by washing, and complexes anchored on the solid surface are detected. When the originally non-immobilized component carries a detectable label, the detection of label immobilized on the surface indicates that complexing occurred. Where the originally non-immobilized component does not carry a label, complexing can be detected, for example, by using a labelled antibody specifically binding the immobilized complex.

If the candidate compound interacts with but does not bind to a particular protein encoded by a gene identified herein, its interaction with that protein can be assayed by methods well known for detecting protein-protein interactions. Such assays include traditional approaches, such as, cross-linking, co-immunoprecipitation, and co-purification through gradients or chromatographic columns.

In addition, protein-protein interactions can be monitored by using a yeast-based genetic system described by Fields and co-workers [Fields and Song, *Nature (London)* 340, 245-246 (1989); Chien *et al.*, *Proc. Natl. Acad. Sci. USA* 88, 9578-9582 (1991)] as disclosed by Chevray and Nathans [*Proc. Natl. Acad. Sci. USA* 89, 5789-5793 (1991)]. Many transcriptional activators, such as yeast GAL4, consist of two physically discrete modular domains, one acting as the DNA-binding domain, while the other one functioning as the transcription activation domain. The yeast expression system described in the foregoing publications (generally referred to as the "two-hybrid system") takes advantage of

this property, and employs two hybrid proteins, one in which the target protein is fused to the DNA-binding domain of GAL4, and another, in which candidate activating proteins are fused to the activation domain. The expression of a GAL1-*lacZ* reporter gene under control of a GAL4-activated promoter depends on reconstitution of GAL4 activity via protein-protein interaction. Colonies containing interacting polypeptides are detected with a chromogenic substrate for beta-galactosidase. A complete kit (MATCHMAKER<sup>TM</sup>) for identifying protein-protein interactions between two specific proteins using the two-hybrid technique is commercially available from Clontech. This system can also be extended to map protein domains involved in specific protein interactions as well as to pinpoint amino acid residues that are crucial for these interactions.

In order to find compounds that interfere with the interaction of a gene identified herein and other intra- or extracellular components can be tested, a reaction mixture is usually prepared containing the product of the gene and the intra- or extracellular component under conditions and for a time allowing for the interaction and binding of the two products. To test the ability of a test compound to inhibit binding, the reaction is run in the absence and in the presence of the test compound. In addition, a placebo may be added to a third reaction mixture, to serve as positive control. The binding (complex formation) between the test compound and the intra- or extracellular component present in the mixture is monitored as described above. The formation of a complex in the control reaction(s) but not in the reaction mixture containing the test compound indicates that the test compound interferes with the interaction of the test compound and its reaction partner.

#### 8. Compositions and Methods for the Treatment of Immune Related Diseases

The compositions useful in the treatment of immune related diseases include, without limitation, antibodies, small organic and inorganic molecules, peptides, phosphopeptides, antisense and ribozyme molecules, triple helix molecules, etc. that inhibit or stimulate immune function, for example, T cell proliferation/activation, lymphokine release, or immune cell infiltration.

For example, antisense RNA and RNA molecule act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. When antisense DNA is used, oligodeoxyribonucleotides derived from the translation initiation site, e.g. between about -10 and +10 positions of the target gene nucleotide sequence, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization to the complementary target RNA, followed by endonucleolytic cleavage. Specific ribozyme cleavage sites within a potential RNA target can be identified by known techniques. For further details see, e.g. Rossi, Current Biology 4, 469-471 (1994), and PCT publication No. WO 97/33551 (published September 18, 1997).

Nucleic acid molecules in triple helix formation used to inhibit transcription should be single-stranded and composed of deoxynucleotides. The base composition of these oligonucleotides is designed such that it promotes triple helix formation via Hoogsteen base pairing rules, which

generally require sizeable stretches of purines or pyrimidines on one strand of a duplex. For further details see, e.g. PCT publication No. WO 97/33551, *supra*.

These molecules can be identified by any or any combination of the screening assays discussed above and/or by any other screening techniques well known for those skilled in the art.

#### 9. Antibodies

Some of the most promising drug candidates according to the present invention are antibodies and antibody fragments which may inhibit (antagonists) or stimulate (agonists) T cell proliferation, eosinophil infiltration, etc.

##### i. Polyclonal Antibodies

Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the polypeptide of the invention or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

##### ii. Monoclonal Antibodies

Antibodies which recognize and bind to the polypeptides of the invention or which act as agonist therefor may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the polypeptide of the invention or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, *Monoclonal Antibodies: Principles and Practice*, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture

medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the polypeptide of the invention or having similar activity as the polypeptide of the invention. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxyapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise

produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., *supra*] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies are preferably monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

*In vitro* methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

### iii. Human and Humanized Antibodies

The antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-329 (1988); and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a

humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and coworkers [Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeven et al., *Science*, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, p. 77 (1985); Boerner et al., *J. Immunol.*, 147(1):86-95 (1991); U. S. 5,750,373]. Similarly, human antibodies can be made by introducing of human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks et al., *Bio/Technology* 10, 779-783 (1992); Lonberg et al., *Nature* 368 856-859 (1994); Morrison, *Nature* 368, 812-13 (1994); Fishwild et al., *Nature Biotechnology* 14, 845-51 (1996); Neuberger, *Nature Biotechnology* 14, 826 (1996); Lonberg and Huszar, *Intern. Rev. Immunol.* 13 65-93 (1995).

iv. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities may be for the polypeptide of the invention, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the coexpression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 [1983]). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct

molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are cotransfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

v. Heteroconjugate Antibodies

Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptopbutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

vi. Effector function engineering

It may be desirable to modify the antibody of the invention with respect to effector function, so as to enhance the effectiveness of the antibody in treating an immune related disease, for example. For example cysteine residue(s) may be introduced in the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated may have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med. 176:1191-1195 (1992) and Shope, B. J. Immunol. 148:2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity may also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research 53:2560-2565 (1993). Alternatively, an antibody can be engineered which has dual Fc regions and may thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design 3:219-230 (1989).

vii. Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g. an enzymatically active toxin of bacterial, fungal, plant or animal origin, or fragments thereof), or a radioactive isotope (i.e. a

radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof which can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin and the trichothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$  and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein coupling agents such as N-succinimidyl-3-(2-pyridylthiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta *et al.*, *Science* **238**: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody may be conjugated to a "receptor" (such streptavidin) for utilization in tissue pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g. avidin) which is conjugated to a cytotoxic agent (e.g. a radionucleotide).

#### viii. Immunoliposomes

The proteins, antibodies, etc. disclosed herein may also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein *et al.*, *Proc. Natl. Acad. Sci. USA*, **82**:3688 (1985); Hwang *et al.*, *Proc. Natl. Acad. Sci. USA*, **77**:4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin *et al.*, *J. Biol. Chem.* **257**: 286-288 (1982) via a disulfide interchange reaction. A chemotherapeutic agent (such as doxorubicin) may be optionally

contained within the liposome. See Gabizon *et al.*, *J. National Cancer Inst.* 81(19)1484 (1989).

#### 10. Pharmaceutical Compositions

The active molecules of the invention, polypeptides and antibodies, as well as other molecules identified by the screening assays disclosed above, can be administered for the treatment of immune related diseases, in the form of pharmaceutical compositions.

Therapeutic formulations of the active molecule, preferably a polypeptide or antibody of the invention, are prepared for storage by mixing the active molecule having the desired degree of purity with optional pharmaceutically acceptable carriers, excipients or stabilizers (*Remington's Pharmaceutical Sciences* 16th edition, Osol, A. Ed. [1980]), in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride, benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (e.g., Zn-protein complexes); and/or non-ionic surfactants such as TWEEN<sup>TM</sup>, PLURONICS<sup>TM</sup> or polyethylene glycol (PEG).

Compounds identified by the screening assays of the present invention can be formulated in an analogous manner, using standard techniques well known in the art.

Lipofections or liposomes can also be used to deliver the polypeptide, antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment which specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable region sequences of an antibody, peptide molecules can be designed which retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology (see, e.g. Marasco *et al.*, *Proc. Natl. Acad. Sci. USA* 90, 7889-7893 [1993]).

The formulation herein may also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition may comprise a cytotoxic agent, cytokine or growth inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active molecules may also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles and nanocapsules) or in macroemulsions. Such techniques are disclosed in *Remington's Pharmaceutical Sciences* 16th edition, Osol, A. Ed. (1980).

The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations may be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, *e.g.* films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods. When encapsulated antibodies remain in the body for a long time, they may denature or aggregate as a result of exposure to moisture at 37C, resulting in a loss of biological activity and possible changes in immunogenicity. Rational strategies can be devised for stabilization depending on the mechanism involved. For example, if the aggregation mechanism is discovered to be intermolecular S-S bond formation through thio-disulfide interchange, stabilization may be achieved by modifying sulphydryl residues, lyophilizing from acidic solutions, controlling moisture content, using appropriate additives, and developing specific polymer matrix compositions.

#### 11. Methods of Treatment

It is contemplated that the polypeptides, antibodies and other active compounds of the present invention may be used to treat various immune related diseases and conditions, such as T cell mediated diseases, including those characterized by infiltration of inflammatory cells into a tissue, stimulation of T-cell proliferation, inhibition of T-cell proliferation, increased or decreased vascular permeability or the inhibition thereof.

Exemplary conditions or disorders to be treated with the polypeptides, antibodies and other compounds of the invention, include, but are not limited to systemic lupus erythematosis, rheumatoid arthritis, juvenile chronic arthritis, osteoarthritis, spondyloarthropathies, systemic sclerosis (scleroderma), idiopathic inflammatory myopathies (dermatomyositis, polymyositis), Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia (immune pancytopenia,

paroxysmal nocturnal hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic purpura, immune-mediated thrombocytopenia), thyroiditis (Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic thyroiditis), diabetes mellitus, immune-mediated renal disease (glomerulonephritis, tubulointerstitial nephritis), demyelinating diseases of the central and peripheral nervous systems such as multiple sclerosis, idiopathic demyelinating polyneuropathy or Guillain-Barré syndrome, and chronic inflammatory demyelinating polyneuropathy, hepatobiliary diseases such as infectious hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune chronic active hepatitis, primary biliary cirrhosis, granulomatous hepatitis, and sclerosing cholangitis, inflammatory bowel disease (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and Whipple's disease, autoimmune or immune-mediated skin diseases including bullous skin diseases, erythema multiforme and contact dermatitis, psoriasis, allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity and urticaria, immunologic diseases of the lung such as eosinophilic pneumonias, idiopathic pulmonary fibrosis and hypersensitivity pneumonitis, transplantation associated diseases including graft rejection and graft - versus-host-disease.

In systemic lupus erythematosus, the central mediator of disease is the production of auto-reactive antibodies to self proteins/tissues and the subsequent generation of immune-mediated inflammation. Antibodies either directly or indirectly mediate tissue injury. Though T lymphocytes have not been shown to be directly involved in tissue damage, T lymphocytes are required for the development of auto-reactive antibodies. The genesis of the disease is thus T lymphocyte dependent. Multiple organs and systems are affected clinically including kidney, lung, musculoskeletal system, mucocutaneous, eye, central nervous system, cardiovascular system, gastrointestinal tract, bone marrow and blood.

Rheumatoid arthritis (RA) is a chronic systemic autoimmune inflammatory disease that mainly involves the synovial membrane of multiple joints with resultant injury to the articular cartilage. The pathogenesis is T lymphocyte dependent and is associated with the production of rheumatoid factors, auto-antibodies directed against self IgG, with the resultant formation of immune complexes that attain high levels in joint fluid and blood. These complexes in the joint may induce the marked infiltrate of lymphocytes and monocytes into the synovium and subsequent marked synovial changes; the joint space/fluid if infiltrated by similar cells with the addition of numerous neutrophils. Tissues affected are primarily the joints, often in symmetrical pattern. However, extra-articular disease also occurs in two major forms. One form is the development of extra-articular lesions with ongoing progressive joint disease and typical lesions of pulmonary fibrosis, vasculitis, and cutaneous ulcers. The second form of extra-articular disease is the so called Felty's syndrome which occurs late in the RA disease course, sometimes after joint disease has become quiescent, and

involves the presence of neutropenia, thrombocytopenia and splenomegaly. This can be accompanied by vasculitis in multiple organs with formations of infarcts, skin ulcers and gangrene. Patients often also develop rheumatoid nodules in the subcutis tissue overlying affected joints; the nodules late stage have necrotic centers surrounded by a mixed inflammatory cell infiltrate. Other manifestations which can occur in RA include: pericarditis, pleuritis, coronary arteritis, interstitial pneumonitis with pulmonary fibrosis, keratoconjunctivitis sicca, and rheumatoid nodules.

Juvenile chronic arthritis is a chronic idiopathic inflammatory disease which begins often at less than 16 years of age. Its phenotype has some similarities to RA; some patients which are rheumatoid factor positive are classified as juvenile rheumatoid arthritis. The disease is sub-classified into three major categories: pauciarticular, polyarticular, and systemic. The arthritis can be severe and is typically destructive and leads to joint ankylosis and retarded growth. Other manifestations can include chronic anterior uveitis and systemic amyloidosis.

Spondyloarthropathies are a group of disorders with some common clinical features and the common association with the expression of HLA-B27 gene product. The disorders include: ankylosing spondylitis, Reiter's syndrome (reactive arthritis), arthritis associated with inflammatory bowel disease, spondylitis associated with psoriasis, juvenile onset spondyloarthropathy and undifferentiated spondyloarthropathy. Distinguishing features include sacroileitis with or without spondylitis; inflammatory asymmetric arthritis; association with HLA-B27 (a serologically defined allele of the HLA-B locus of class I MHC); ocular inflammation, and absence of autoantibodies associated with other rheumatoid disease. The cell most implicated as key to induction of the disease is the CD8+ T lymphocyte, a cell which targets antigen presented by class I MHC molecules. CD8+ T cells may react against the class I MHC allele HLA-B27 as if it were a foreign peptide expressed by MHC class I molecules. It has been hypothesized that an epitope of HLA-B27 may mimic a bacterial or other microbial antigenic epitope and thus induce a CD8+ T cells response.

Systemic sclerosis (scleroderma) has an unknown etiology. A hallmark of the disease is induration of the skin; likely this is induced by an active inflammatory process. Scleroderma can be localized or systemic; vascular lesions are common and endothelial cell injury in the microvasculature is an early and important event in the development of systemic sclerosis; the vascular injury may be immune mediated. An immunologic basis is implied by the presence of mononuclear cell infiltrates in the cutaneous lesions and the presence of anti-nuclear antibodies in many patients. ICAM-1 is often upregulated on the cell surface of fibroblasts in skin lesions suggesting that T cell interaction with these cells may have a role in the pathogenesis of the disease. Other organs involved include: the gastrointestinal tract: smooth muscle atrophy and fibrosis resulting in abnormal peristalsis/motility; kidney: concentric subendothelial intimal proliferation affecting small arcuate and interlobular arteries with resultant reduced renal cortical blood flow, results in proteinuria, azotemia and hypertension;

skeletal muscle: atrophy, interstitial fibrosis; inflammation; lung: interstitial pneumonitis and interstitial fibrosis; and heart: contraction band necrosis, scarring/fibrosis.

Idiopathic inflammatory myopathies including dermatomyositis, polymyositis and others are disorders of chronic muscle inflammation of unknown etiology resulting in muscle weakness. Muscle injury/inflammation is often symmetric and progressive. Autoantibodies are associated with most forms. These myositis-specific autoantibodies are directed against and inhibit the function of components, proteins and RNA's, involved in protein synthesis.

Sjogren's syndrome is due to immune-mediated inflammation and subsequent functional destruction of the tear glands and salivary glands. The disease can be associated with or accompanied by inflammatory connective tissue diseases. The disease is associated with autoantibody production against Ro and La antigens, both of which are small RNA-protein complexes. Lesions result in keratoconjunctivitis sicca, xerostomia, with other manifestations or associations including biliary cirrhosis, peripheral or sensory neuropathy, and palpable purpura.

Systemic vasculitis are diseases in which the primary lesion is inflammation and subsequent damage to blood vessels which results in ischemia/necrosis/degeneration to tissues supplied by the affected vessels and eventual end-organ dysfunction in some cases. Vasculitides can also occur as a secondary lesion or sequelae to other immune-inflammatory mediated diseases such as rheumatoid arthritis, systemic sclerosis, etc., particularly in diseases also associated with the formation of immune complexes. Diseases in the primary systemic vasculitis group include: systemic necrotizing vasculitis: polyarteritis nodosa, allergic angiitis and granulomatosis, polyangiitis; Wegener's granulomatosis; lymphomatoid granulomatosis; and giant cell arteritis. Miscellaneous vasculitides include: mucocutaneous lymph node syndrome (MLNS or Kawasaki's disease), isolated CNS vasculitis, Behet's disease, thromboangiitis obliterans (Buerger's disease) and cutaneous necrotizing venulitis. The pathogenic mechanism of most of the types of vasculitis listed is believed to be primarily due to the deposition of immunoglobulin complexes in the vessel wall and subsequent induction of an inflammatory response either via ADCC, complement activation, or both.

Sarcoidosis is a condition of unknown etiology which is characterized by the presence of epithelioid granulomas in nearly any tissue in the body; involvement of the lung is most common. The pathogenesis involves the persistence of activated macrophages and lymphoid cells at sites of the disease with subsequent chronic sequelae resultant from the release of locally and systemically active products released by these cell types.

Autoimmune hemolytic anemia including autoimmune hemolytic anemia, immune pancytopenia, and paroxysmal nocturnal hemoglobinuria is a result of production of antibodies that react with antigens expressed on the surface of red blood cells (and in some cases other blood cells including platelets as well) and is a reflection of the removal of those antibody coated cells via complement mediated lysis and/or ADCC/Fc-receptor-mediated mechanisms.

In autoimmune thrombocytopenia including thrombocytopenic purpura, and immune-mediated thrombocytopenia in other clinical settings, platelet destruction/removal occurs as a result of either antibody or complement attaching to platelets and subsequent removal by complement lysis, ADCC or FC-receptor mediated mechanisms.

Thyroiditis including Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, and atrophic thyroiditis, are the result of an autoimmune response against thyroid antigens with production of antibodies that react with proteins present in and often specific for the thyroid gland. Experimental models exist including spontaneous models: rats (BUF and BB rats) and chickens (obese chicken strain); inducible models: immunization of animals with either thyroglobulin, thyroid microsomal antigen (thyroid peroxidase).

Type I diabetes mellitus or insulin-dependent diabetes is the autoimmune destruction of pancreatic islet  $\beta$  cells; this destruction is mediated by auto-antibodies and auto-reactive T cells. Antibodies to insulin or the insulin receptor can also produce the phenotype of insulin-non-responsiveness.

Immune mediated renal diseases, including glomerulonephritis and tubulointerstitial nephritis, are the result of antibody or T lymphocyte mediated injury to renal tissue either directly as a result of the production of autoreactive antibodies or T cells against renal antigens or indirectly as a result of the deposition of antibodies and/or immune complexes in the kidney that are reactive against other, non-renal antigens. Thus other immune-mediated diseases that result in the formation of immune-complexes can also induce immune mediated renal disease as an indirect sequelae. Both direct and indirect immune mechanisms result in inflammatory response that produces/induces lesion development in renal tissues with resultant organ function impairment and in some cases progression to renal failure. Both humoral and cellular immune mechanisms can be involved in the pathogenesis of lesions.

Demyelinating diseases of the central and peripheral nervous systems, including Multiple Sclerosis; idiopathic demyelinating polyneuropathy or Guillain-Barr syndrome; and Chronic Inflammatory Demyelinating Polyneuropathy, are believed to have an autoimmune basis and result in nerve demyelination as a result of damage caused to oligodendrocytes or to myelin directly. In MS there is evidence to suggest that disease induction and progression is dependent on T lymphocytes. Multiple Sclerosis is a demyelinating disease that is T lymphocyte-dependent and has either a relapsing-remitting course or a chronic progressive course. The etiology is unknown; however, viral infections, genetic predisposition, environment, and autoimmunity all contribute. Lesions contain infiltrates of predominantly T lymphocyte mediated, microglial cells and infiltrating macrophages; CD4+T lymphocytes are the predominant cell type at lesions. The mechanism of oligodendrocyte cell death and subsequent demyelination is not known but is likely T lymphocyte driven.

Inflammatory and Fibrotic Lung Disease, including Eosinophilic Pneumonias; Idiopathic Pulmonary Fibrosis, and Hypersensitivity Pneumonitis may involve a disregulated immune-inflammatory response. Inhibition of that response would be of therapeutic benefit.

Autoimmune or Immune-mediated Skin Disease including Bullous Skin Diseases, Erythema Multiforme, and Contact Dermatitis are mediated by auto-antibodies, the genesis of which is T lymphocyte-dependent.

Psoriasis is a T lymphocyte-mediated inflammatory disease. Lesions contain infiltrates of T lymphocytes, macrophages and antigen processing cells, and some neutrophils.

Allergic diseases, including asthma; allergic rhinitis; atopic dermatitis; food hypersensitivity; and urticaria are T lymphocyte dependent. These diseases are predominantly mediated by T lymphocyte induced inflammation, IgE mediated-inflammation or a combination of both.

Transplantation associated diseases, including Graft rejection and Graft-Versus-Host-Disease (GVHD) are T lymphocyte-dependent; inhibition of T lymphocyte function is ameliorative.

Other diseases in which intervention of the immune and/or inflammatory response have benefit are infectious disease including but not limited to viral infection (including but not limited to AIDS, hepatitis A, B, C, D, E and herpes) bacterial infection, fungal infections, and protozoal and parasitic infections (molecules (or derivatives/agonists) which stimulate the MLR can be utilized therapeutically to enhance the immune response to infectious agents), diseases of immunodeficiency (molecules/derivatives/agonists) which stimulate the MLR can be utilized therapeutically to enhance the immune response for conditions of inherited, acquired, infectious induced (as in HIV infection), or iatrogenic (i.e. as from chemotherapy) immunodeficiency), and neoplasia.

It has been demonstrated that some human cancer patients develop an antibody and/or T lymphocyte response to antigens on neoplastic cells. It has also been shown in animal models of neoplasia that enhancement of the immune response can result in rejection or regression of that particular neoplasm. Molecules that enhance the T lymphocyte response in the MLR have utility in vivo in enhancing the immune response against neoplasia. Molecules which enhance the T lymphocyte proliferative response in the MLR (or small molecule agonists or antibodies that affected the same receptor in an agonistic fashion) can be used therapeutically to treat cancer. Molecules that inhibit the lymphocyte response in the MLR also function in vivo during neoplasia to suppress the immune response to a neoplasm; such molecules can either be expressed by the neoplastic cells themselves or their expression can be induced by the neoplasm in other cells. Antagonism of such inhibitory molecules (either with antibody, small molecule antagonists or other means) enhances immune-mediated tumor rejection.

Additionally, inhibition of molecules with proinflammatory properties may have therapeutic benefit in reperfusion injury; stroke; myocardial infarction; atherosclerosis; acute lung injury;

hemorrhagic shock; burn; sepsis/septic shock; acute tubular necrosis; endometriosis; degenerative joint disease and pancreatitis.

The compounds of the present invention, e.g. polypeptides or antibodies, are administered to a mammal, preferably a human, in accord with known methods, such as intravenous administration as a bolus or by continuous infusion over a period of time, by intramuscular, intraperitoneal, intracerebrospinal, subcutaneous, intra-articular, intrasynovial, intrathecal, oral, topical, or inhalation (intranasal, intrapulmonary) routes. Intravenous or inhaled administration of polypeptides and antibodies is preferred.

In immunoadjuvant therapy, other therapeutic regimens, such administration of an anti-cancer agent, may be combined with the administration of the proteins, antibodies or compounds of the instant invention. For example, the patient to be treated with the immunoadjuvant of the invention may also receive an anti-cancer agent (chemotherapeutic agent) or radiation therapy. Preparation and dosing schedules for such chemotherapeutic agents may be used according to manufacturers' instructions or as determined empirically by the skilled practitioner. Preparation and dosing schedules for such chemotherapy are also described in *Chemotherapy Service* Ed., M.C. Perry, Williams & Wilkins, Baltimore, MD (1992). The chemotherapeutic agent may precede, or follow administration of the immunoadjuvant or may be given simultaneously therewith. Additionally, an anti-oestrogen compound such as tamoxifen or an anti-progesterone such as onapristone (see, EP 616812) may be given in dosages known for such molecules.

It may be desirable to also administer antibodies against other immune disease associated or tumor associated antigens, such as antibodies which bind to CD20, CD11a, CD18, ErbB2, EGFR, ErbB3, ErbB4, or vascular endothelial factor (VEGF). Alternatively, or in addition, two or more antibodies binding the same or two or more different antigens disclosed herein may be coadministered to the patient. Sometimes, it may be beneficial to also administer one or more cytokines to the patient. In one embodiment, the polypeptides of the invention are coadministered with a growth inhibitory agent. For example, the growth inhibitory agent may be administered first, followed by a polypeptide of the invention. However, simultaneous administration or administration first is also contemplated. Suitable dosages for the growth inhibitory agent are those presently used and may be lowered due to the combined action (synergy) of the growth inhibitory agent and the polypeptide of the invention.

For the treatment or reduction in the severity of immune related disease, the appropriate dosage of a compound of the invention will depend on the type of disease to be treated, as defined above, the severity and course of the disease, whether the agent is administered for preventive or therapeutic purposes, previous therapy, the patient's clinical history and response to the compound, and the discretion of the attending physician. The compound is suitably administered to the patient at one time or over a series of treatments.

For example, depending on the type and severity of the disease, about 1 ug/kg to 15 mg/kg (e.g. 0.1-20mg/kg) of polypeptide or antibody is an initial candidate dosage for administration to the patient, whether, for example, by one or more separate administrations, or by continuous infusion. A typical daily dosage might range from about 1 ug/kg to 100 mg/kg or more, depending on the factors mentioned above. For repeated administrations over several days or longer, depending on the condition, the treatment is sustained until a desired suppression of disease symptoms occurs. However, other dosage regimens may be useful. The progress of this therapy is easily monitored by conventional techniques and assays.

#### 12. Articles of Manufacture

In another embodiment of the invention, an article of manufacture containing materials useful for the diagnosis or treatment of the disorders described above is provided. The article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers may be formed from a variety of materials such as glass or plastic. The container holds a composition which is effective for diagnosing or treating the condition and may have a sterile access port (for example the container may be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle). The active agent in the composition is usually a polypeptide or an antibody of the invention. The label on, or associated with, the container indicates that the composition is used for diagnosing or treating the condition of choice. The article of manufacture may further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution and dextrose solution. It may further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

#### 13. Diagnosis and Prognosis of Immune Related Disease

Cell surface proteins, such as proteins which are overexpressed in certain immune related diseases, are excellent targets for drug candidates or disease treatment. The same proteins along with secreted proteins encoded by the genes amplified in immune related disease states find additional use in the diagnosis and prognosis of these diseases. For example, antibodies directed against the protein products of genes amplified in multiple sclerosis, rheumatoid arthritis, or another immune related disease, can be used as diagnostics or prognostics.

For example, antibodies, including antibody fragments, can be used to qualitatively or quantitatively detect the expression of proteins encoded by amplified or overexpressed genes ("marker gene products"). The antibody preferably is equipped with a detectable, e.g. fluorescent label, and binding can be monitored by light microscopy, flow cytometry, fluorimetry, or other techniques known in the art. These techniques are particularly suitable, if the overexpressed gene encodes a cell surface protein. Such binding assays are performed essentially as described above.

*In situ* detection of antibody binding to the marker gene products can be performed, for example, by immunofluorescence or immunoelectron microscopy. For this purpose, a histological specimen is removed from the patient, and a labeled antibody is applied to it, preferably by overlaying the antibody on a biological sample. This procedure also allows for determining the distribution of the marker gene product in the tissue examined. It will be apparent for those skilled in the art that a wide variety of histological methods are readily available for *in situ* detection.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

### EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Manassas, VA. Unless otherwise noted, the present invention uses standard procedures of recombinant DNA technology, such as those described hereinabove and in the following textbooks: Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press N.Y., 1989; Ausubel *et al.*, *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y., 1989; Innis *et al.*, *PCR Protocols: A Guide to Methods and Applications*, Academic Press, inc., N.Y., 1990; Harlow *et al.*, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, 1988; Gait, M.J., *Oligonucleotide Synthesis*, IRL Press, Oxford, 1984; R.I. Freshney, *Animal Cell Culture*, 1987; Coligan *et al.*, *Current Protocols in Immunology*, 1991.

### EXAMPLE 1

#### Isolation of cDNA clones Encoding Human PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326

##### I. Isolation of cDNA Clones Encoding Human PRO245

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul *et al.*, *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a

frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington).

A consensus DNA sequence encoding PRO245 was assembled relative to the other identified EST sequences, where the consensus sequence was designated herein as DNA30954, and the polypeptide showed some structural homology to transmembrane protein receptor tyrosine kinase proteins.

Based on the DNA30954 consensus sequence, oligonucleotides were synthesized to identify by PCR a cDNA library that contained the sequence of interest and for use as probes to isolate a clone of the full-length coding sequence for PRO245.

A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-ATCGTTGTGAAGTTAGTGCCCC-3' (SEQ ID NO: 15)

reverse PCR primer 5'-ACCTGCGATATCCAACAGAATTG-3' (SEQ ID NO: 16)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA30954 sequence which had the following nucleotide sequence:

hybridization probe

5'-GGAAGAGGATACAGTCACTCTGGAAAGTATTAGTGGCTCCAGCAGTTCC-3' (SEQ ID NO: 17)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO245 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal liver tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRK5B or pRK5D; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO245 [herein designated as UNQ219 (DNA35638)] and the derived protein sequence for PRO245.

The entire nucleotide sequence of UNQ219 (DNA35638) is shown in Figure 3 (SEQ ID NO: 1). Clone UNQ219 (DNA35638) contains a single open reading frame with an apparent translational

initiation site at nucleotide positions 89-91 [Kozak et al., *supra*] and ending at the stop codon at nucleotide positions 1025-1027 (Fig. 3). The predicted polypeptide precursor is 312 amino acids long (Figure. 4; PRO245; SEQ ID NO: 2). Clone UNQ219 (DNA35638) has been deposited with ATCC on September 17, 1997 and is assigned ATCC Deposit No. 209265.

Analysis of the amino acid sequence of the full-length PRO245 suggests that a portion of it possesses 60% amino acid identity with the human c-myb protein and, therefore, may be a new member of the transmembrane protein receptor tyrosine kinase family.

## II. Isolation of cDNA clones Encoding PRO217

The extracellular domain (ECD) sequences (including the secretion signal sequence, if any) from about 950 known secreted proteins from the Swiss-Prot public database were used to search EST databases. The EST databases included public databases (e.g., Dayhof, GenBank), and proprietary databases (e.g. LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul, SF and Gish (1996), *Methods in Enzymology* 266: 460-80 (1996); <http://blast.wustl.edu/blast/README.html>) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequences. Those comparisons with a Blast score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, WA; (<http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

Consensus DNA sequences encoding EGF-like homologues were assembled (DNA28726, DNA28730 and DNA28760) using phrap. In some cases, the consensus DNA sequence was extended using repeated cycles of blast and phrap to extend the consensus sequence as far as possible using the three sources of EST sequences listed above.

Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. The pair of forward and reverse PCR primers (notated as \*.f and \*.r, respectively) may range from 20 to 30 nucleotides (typically 24), and are designed to give a PCR product of 100-1000 bp in length. The probe sequences (notated as \*.p) are typically 40-55 bp (typically 50) in length. In some cases additional oligonucleotides are synthesized when the consensus sequence is greater than 1-1.5 kbp. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the PCR primers. This library was used to isolate DNA32279, DNA32292 and DNA33094 was fetal kidney, fetal lung and fetal lung, respectively.

RNA for the construction of the cDNA libraries was isolated using standard isolation protocols, e.g., Ausubel *et al.*, *Current Protocols in Molecular Biology*, from tissue or cell line sources or it was purchased from commercial sources (e.g., Clontech). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods (e.g., Ausubel *et al.*) using commercially available reagents (e.g., Invitrogen). The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation in a suitable cloning vector (pRK5B or pRK5D) in the unique XhoI and NotI sites.

A cDNA clone was sequenced in its entirety. The entire nucleotide sequence of EGF-like homologue PRO217 is shown in Figure 5 (SEQ ID NO: 3). The predicted polypeptide is 379 (PRO217; Figure 6; SEQ ID NO: 4) amino acids in length with a molecule weight of approximately 41.52 kDa.

The oligonucleotide sequences used in the above procedure were the following:

28726.p (OLI500) (SEQ ID NO: 18)

GGGTACACCTGCTCCTGCACCGACGGATATTGGCTCTGGAAGGCC

28726.f (OLI 502) (SEQ ID NO: 19)

ACAGATTCCCACCAGTGCAACC

28726.r (OLI 503) (SEQ ID NO: 20)

CACACTCGTTCACATCTTGGC

28730.p (OLI 516) (SEQ ID NO: 21)

AGGGAGCACGGACAGTGTGCAGATGTGGACGTGCTCACTAGCA

28730.f (OLI 517) (SEQ ID NO: 22)

AGAGTGTATCTCTGGCTACGC

28730.r (OLI 518) (SEQ ID NO: 23)

TAAGTCCGGCACATTACAGGTC

28760.p (OLI 617) (SEQ ID NO: 24)

CCCACGATGTATGAATGGTGGACTTGTGTGACTCCTGGTTCTGCATC

28760.f (OLI 618) (SEQ ID NO: 25)

AAAGACGCATCTCGAGTGTCC

28760.r (OLI 619) (SEQ ID NO: 26)

TGCTGATTCACACTGCTCTCCC

### III. Isolation of cDNA clones Encoding Human PRO301

The extracellular domain (ECD) sequences (including the secretion signal sequence, if any) from about 950 known secreted proteins from the Swiss-Prot public database were used to search EST databases. The EST databases included public EST databases (e.g., GenBank), a proprietary EST database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the

computer program BLAST or BLAST2 [Altschul et al., *Methods in Enzymology*, **266**:460-480 (1996); <http://blast.wustl.edu/blast/README.html>] as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequences. Those comparisons resulting in a BLAST score of 70 (or in some cases, 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence encoding DNA35936 was assembled using phrap. In some cases, the consensus DNA sequence was extended using repeated cycles of blast and phrap to extend the consensus sequence as far as possible using the three sources of EST sequences listed above.

Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. Forward and reverse PCR primers (notated as \*.f and \*.r, respectively) may range from 20 to 30 nucleotides (typically about 24), and are designed to give a PCR product of 100-1000 bp in length. The probe sequences (notated as \*.p) are typically 40-55 bp (typically about 50) in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than 1-1.5 kbp. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the in vivo cloning procedure suing the probe oligonucleotide and one of the PCR primers.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO301 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal kidney. The cDNA libraries used to isolated the cDNA clones were constructed by standard methods using commercially available reagents (e.g., Invitrogen, San Diego, CA; Clontech, etc.) The cDNA was primed with oligo dT containing a NotI site, linked with blunt to Sall hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, **253**:1278-1280 (1991)) in the unique XhoI and NotI sites.

A cDNA clone was sequenced in its entirety. The full length nucleotide sequence of native sequence PRO301 is shown in Figure 7 (SEQ ID NO: 5). Clone DNA40628 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 52-54 [Kozak et al., *supra*] (Fig. 7). The predicted polypeptide (PRO301; Figure 8; SEQ ID NO: 6) is 299 amino acids

long with a predicted molecular weight of 32583 daltons and pI of 8.29. Clone DNA40628 has been deposited with ATCC and is assigned ATCC deposit No. 209432.

Based on a BLAST and FastA sequence alignment analysis of the full-length sequence, PRO301 shows amino acid sequence identity to A33 antigen precursor (30%) and coxsackie and adenovirus receptor protein (29%).

The oligonucleotide sequences used in the above procedure were the following:

OLI2162 (35936.f1) (SEQ ID NO: 27)

TCGCGGAGCTGTGTTCTGTTCCC

OLI2163 (35936.p1) (SEQ ID NO: 28)

TGATCGCGATGGGGACAAAGGCGCAAGCTCGAGAGGAAACTGTTGTGCCT

OLI2164 (35936.f2) (SEQ ID NO: 29)

ACACCTGGTTCAAAGATGGG

OLI2165 (35936.r1) (SEQ ID NO: 30)

TAGGAAGAGTTGCTGAAGGCACGG

OLI2166 (35936.f3) (SEQ ID NO: 31)

TTGCCTTACTCAGGTGCTAC

OLI2167 (35936.r2) (SEQ ID NO: 32)

ACTCAGCAGTGGTAGGAAAG

#### IV. Isolation of cDNA Clones Encoding Human PRO266

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

Based on an expression sequence tag oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO266. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5 kbp. In

order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the primer pairs.

A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-GTTGGATCTGGCAACAATAAC-3' (SEQ ID NO: 33)

reverse PCR primer 5'-ATTGTTGTGCAGGCTGAGTTAAG-3' (SEQ ID NO: 34)

Additionally, a synthetic oligonucleotide hybridization probe was constructed which had the following nucleotide sequence:

hybridization probe

5'-GGTGGCTATACATGGATAGCAATTACCTGGACACGCTGTCCCGGG-3' (SEQ ID NO: 35)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO266 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal brain tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO266 [herein designated as UNQ233 (DNA37150)] and the derived protein sequence for PRO266.

The entire nucleotide sequence of UNQ233 (DNA37150) is shown in Figure 9 (SEQ ID NO: 7). Clone UNQ233 (DNA37150) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 1-3 [Kozak et al., *supra*] and ending at the stop codon after nucleotide position 2088. The predicted polypeptide precursor is 696 amino acids long (Figure 10; PRO266: SEQ ID NO: 8). Clone UNQ233 (DNA37150) has been deposited with ATCC and is assigned ATCC deposit no. 209401.

Analysis of the amino acid sequence of the full-length PRO266 polypeptide suggests that portions of it possess significant homology to a SLIT protein, thereby indicating that PRO266 may be a novel leucine rich repeat protein.

## V. Isolation of cDNA Clones Encoding Human PRO335, PRO331 or PRO326

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. Based on the consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO335, PRO331 or PRO326. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the primer pairs.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO335, PRO331 or PRO326 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue (PRO335 and PRO326) and human fetal brain (PRO331). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to Sall hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO335 (Figure 11; SEQ ID NO: 9), PRO331 (Figure 13; SEQ ID NO: 11) or PRO326

(Figure 15; SEQ ID NO: 13) and the derived protein sequence for PRO335 (Figure 12; SEQ ID NO: 10), PRO331 (Figure 14; SEQ ID NO: 12) or PRO326 (Figure 16; SEQ ID NO: 14). The nucleic acid encoding PRO335 was deposited with the ATCC on 2 June 1998 and is assigned ATCC Accession No. 209927; the nucleic acid encoding PRO331 was deposited with the ATCC on 7 November 1997 and is assigned ATCC Accession No. 209439; and the nucleic acid encoding PRO326 was deposited with the ATCC on 21 November 1997 and is assigned ATCC Accession No. 209489.

Analysis of the amino acid sequence of the full-length PRO335, PRO331 or PRO326 polypeptide suggests that portions of it possess significant homology to the LIG-1 protein, thereby indicating that PRO335, PRO331 and PRO326 may be a novel LIG-1-related protein.

## EXAMPLE 2

### Stimulatory Activity in Mixed Lymphocyte Reaction (MLR) Assay (No. 24)

This example shows that the polypeptides of the invention are active as a stimulator of the proliferation of stimulated T-lymphocytes. Compounds which stimulate proliferation of lymphocytes are useful therapeutically where enhancement of an immune response is beneficial. Compounds which inhibit proliferation of lymphocytes are useful therapeutically where suppression of an immune response is beneficial. A therapeutic agent may take the form of antagonists of the polypeptide of the invention, for example, murine-human chimeric, humanized or human antibodies against the polypeptide.

The basic protocol for this assay is described in Current Protocols in Immunology, unit 3.12; edited by J. E. Coligan, A. M. Kruisbeek, D. H. Marglies, E. M. Shevach, W. Strober, National Institutes of Health, Published by John Wiley & Sons, Inc.

More specifically, in one assay variant, peripheral blood mononuclear cells (PBMC) are isolated from mammalian individuals, for example a human volunteer, by leukopheresis (one donor will supply stimulator PBMCs, the other donor will supply responder PBMCs). If desired, the cells are frozen in fetal bovine serum and DMSO after isolation. Frozen cells may be thawed overnight in assay media (37°C, 5% CO<sub>2</sub>) and then washed and resuspended to 3 x 10<sup>6</sup> cells/ml of assay media (RPMI; 10% fetal bovine serum, 1% penicillin/streptomycin, 1% glutamine, 1% HEPES, 1% non-essential amino acids, 1% pyruvate).

The stimulator PBMCs are prepared by irradiating the cells (about 3000 Rads). The assay is prepared by plating in triplicate wells a mixture of:

100µl of test sample diluted to 1% or to 0.1%  
50 µl of irradiated stimulator cells and  
50 µl of responder PBMC cells.

100 microliters of cell culture media or 100 microliter of CD4-IgG is used as the control. The wells are then incubated at 37°C, 5% CO<sub>2</sub> for 4 days. On day 5 and each well is pulsed with tritiated thymidine (1.0 mC/well; Amersham). After 6 hours the cells are washed 3 times and then the uptake of the label is evaluated.

In another variant of this assay, PBMCs are isolated from the spleens of Balb/c mice and C57B6 mice. The cells are teased from freshly harvested spleens in assay media (RPMI; 10% fetal bovine serum, 1% penicillin/streptomycin, 1% glutamine, 1% HEPES, 1% non-essential amino acids, 1% pyruvate) and the PBMCs are isolated by overlaying these cells over Lympholyte M (Organon Teknika), centrifuging at 2000 rpm for 20 minutes, collecting and washing the mononuclear cell layer in assay media and resuspending the cells to 1x 10<sup>7</sup> cells/ml of assay media. The assay is then conducted as described above using a sample having a PRO concentration obtained by diluting a stock solution. The results of this assay for compounds of the invention are shown below. Positive increases over control are considered positive with increases of greater than or equal to 180% being preferred. However, any value greater than control indicates a stimulatory effect for the test protein.

Table 2

<u>PRO</u>	<u>PRO Concentration</u>	<u>Percent Increase Over Control</u>
PRO245	0.1%	189.7
	" 0.1%	193.7
	" 1.0%	212.5
	" 1.0%	300.5
PRO217	0.1%	74.5
	" 1.0%	89.5
	" 0.99 nM	97.0
	" 9.9 nM	122.3
	" 0.25 nM	144.8
	" 2.5 nM	126.9
PRO301	50.0%	109.4
	" 70.0 nM	133.7
	" 700.0 nM	83.6
	" 0.1%	58.7
	1.0%	127.7
PRO301	" 0.1%	181.7
	" 1.0%	187.3
	" 0.1%	127.5
	" 1.0%	108.3

PRO266	0.1%	136.4
"	0.1%	139.2
"	1.0%	189.8
"	1.0%	245.1
PRO335	50.0%	91.0
"	50.0%	103.8
"	0.1%	130.0
"	1.0%	180.2
PRO331	50.0%	155.5
"	0.1%	169.3
"	1.0%	128.1
"	0.1%	129.3
"	1.0%	162.5
PRO326	50.0%	91.0
"	50.0%	103.8
"	0.1%	130.0
"	1.0%	180.2

---

## EXAMPLE 3

Skin Vascular Permeability Assay (No. 64)

This assay shows that certain polypeptides of the invention stimulate an immune response and induce inflammation by inducing mononuclear cell, eosinophil and PMN infiltration at the site of injection of the animal. This skin vascular permeability assay is conducted as follows. Hairless guinea pigs weighing 350 grams or more are anesthetized with ketamine (75-80 mg/Kg) and 5 mg/Kg xylazine intramuscularly (IM). A sample of purified polypeptide of the invention or a conditioned media test sample is injected intradermally onto the backs of the test animals with 100  $\mu$ L per injection site. It is possible to have about 10-30, preferably about 16-24, injection sites per animal. One mL of Evans blue dye (1% in physiologic buffered saline) is injected intracardially. Blemishes at the injection sites are then measured (mm diameter) at 1hr and 6 hr post injection. Animals were sacrificed at 6 hrs after injection. Each skin injection site was biopsied and fixed in formalin. The skins were then prepared for histopathologic evaluation. Each site was evaluated for inflammatory cell infiltration into the skin. Sites with visible inflammatory cell inflammation were scored as

positive. Inflammatory cells can be neutrophilic, eosinophilic, monocytic or lymphocytic. The results of this test for compounds of the invention is shown below.

In the Table below, at least a minimal perivascular infiltrate at the injection site is scored as positive, no infiltrate at the site of injection is scored as negative.

Table 3

<u>PRO</u>	<u>Hours Post Injection</u>	<u>Infiltrate Designation</u>
PRO245	24 hr	positive
PRO217	24 hr	positive
PRO301	24 hr	positive
PRO266	24 hr	positive
PRO335	24 hr	positive
PRO331	24 hr	positive
PRO326	24 hr	positive

#### EXAMPLE 4

##### Human Co-Stimulation Assay

In addition to the activation signal mediated by the T cell receptor, T cell activation requires a costimulatory signal. One costimulatory signal is generated by the interaction of B7 (CD3) with CD28. In this assay, 96 well plates are coated with CD3 with or without CD28 and then human peripheral blood lymphocytes followed by a test protein, are added. Proliferation of the lymphocytes is determined by tritiated thymidine uptake. A positive assay indicates that the test protein provided a stimulatory signal for lymphocyte proliferation.

##### Material:

- 1) Hyclone D-PBS without Calcium, Magnesium
- 2) Nunc 96 well certified plates #4-39454
- 3) Anti-human CD3 Amac 0178 200 µg/ml stock
- 4) Anti-human CD28 Biodesign P42235M
- 5) Media: Gibco RPMI 1640 + 10 % Intergen #1020-90 FBS, 1% Glu, 1% P/S, 50 µg/ml Gentamycin, 10 mM Hepes. Fresh for each assay.
- 6) Tritiated Thymidine
- 7) Frozen human peripheral blood lymphocytes (PBL) collected via a leukaphoresis procedure

Plates are prepared by coating 96 well flat bottom plates with anti-CD3 antibody (Amac) or anti-CD28 antibody (Biodesign) or both in Hyclone D-PBS without calcium and magnesium. Anti-CD3 antibody is used at 10 ng/well (50  $\mu$ l of 200 ng/ml) and anti-CD28 antibody at 25 ng/well (50  $\mu$ l of 0.5  $\mu$ g/ml) in 100  $\mu$ l total volume.

PBLs are isolated from human donors using standard leukophoresis methods. The cell preparations are frozen in 50% fetal bovine serum and 50% DMSO until the assay is conducted. Cells are prepared by thawing and washing PBLs in media, resuspending PBLs in 25 mls of media and incubating at 37°C, 5% CO<sub>2</sub> overnight.

In the assay procedure, the coated plate is washed twice with PBS and the PBLs are washed and resuspended to 1  $\times$  10<sup>6</sup> cells/ml using 100  $\mu$ L /well. 100  $\mu$ l of a test protein or control media are added to the plate making a total volume per well of 200  $\mu$ L. The plate is incubated for 72 hours. The plate is then pulsed for 6 hours with tritiated thymidine (1 mCi/well; Amersham) and the PBLs are harvested from the plates and evaluated for uptake of the tritiated thymidine.

#### EXAMPLE 5

##### In situ Hybridization

*In situ* hybridization is a powerful and versatile technique for the detection and localization of nucleic acid sequences within cell or tissue preparations. It may be useful, for example, to identify sites of gene expression, analyze the tissue distribution of transcription, identify and localize viral infection, follow changes in specific mRNA synthesis and aid in chromosome mapping.

*In situ* hybridization was performed following an optimized version of the protocol by Lu and Gillett, Cell Vision 1: 169-176 (1994), using PCR-generated <sup>33</sup>P-labeled riboprobes. Briefly, formalin-fixed, paraffin-embedded human tissues were sectioned, deparaffinized, deproteinized in proteinase K (20 g/ml) for 15 minutes at 37°C, and further processed for *in situ* hybridization as described by Lu and Gillett, *supra*. A [<sup>33</sup>P] UTP-labeled antisense riboprobe was generated from a PCR product and hybridized at 55°C overnight. The slides were dipped in Kodak NTB2 nuclear track emulsion and exposed for 4 weeks.

##### <sup>33</sup>P-Riboprobe synthesis

6.0  $\mu$ l (125 mCi) of <sup>33</sup>P-UTP (Amersham BF 1002, SA<2000 Ci/mmol) were speed vac dried. To each tube containing dried <sup>33</sup>P-UTP, the following ingredients were added:

2.0  $\mu$ l 5x transcription buffer

1.0  $\mu$ l DTT (100 mM)

2.0  $\mu$ l NTP mix (2.5 mM : 10  $\mu$ l; each of 10 mM GTP, CTP & ATP + 10  $\mu$ l H<sub>2</sub>O)

1.0  $\mu$ l UTP (50  $\mu$ M)

1.0  $\mu$ l Rnasin  
1.0  $\mu$ l DNA template (1 $\mu$ g)  
1.0  $\mu$ l H<sub>2</sub>O

The tubes were incubated at 37°C for one hour. 1.0  $\mu$ L RQ1 DNase were added, followed by incubation at 37°C for 15 minutes. 90  $\mu$ L TE (10 mM Tris pH 7.6/1mM EDTA pH 8.0) were added, and the mixture was pipetted onto DE81 paper. The remaining solution was loaded in a Microcon-50 ultrafiltration unit, and spun using program 10 (6 minutes). The filtration unit was inverted over a second tube and spun using program 2 (3 minutes). After the final recovery spin, 100  $\mu$ L TE were added. 1  $\mu$ L of the final product was pipetted on DE81 paper and counted in 6 ml of Biofluor II.

The probe was run on a TBE/urea gel. 1-3  $\mu$ L of the probe or 5  $\mu$ L of RNA Mrk III were added to 3  $\mu$ L of loading buffer. After heating on a 95C heat block for three minutes, the gel was immediately placed on ice. The wells of gel were flushed, the sample loaded, and run at 180-250 volts for 45 minutes. The gel was wrapped in saran wrap and exposed to XAR film with an intensifying screen in -70C freezer one hour to overnight.

### 33P-Hybridization

*Pretreatment of frozen sections* The slides were removed from the freezer, placed on aluminium trays and thawed at room temperature for 5 minutes. The trays were placed in 55C incubator for five minutes to reduce condensation. The slides were fixed for 10 minutes in 4% paraformaldehyde on ice in the fume hood, and washed in 0.5 x SSC for 5 minutes, at room temperature (25 ml 20 x SSC + 975 ml SQ H<sub>2</sub>O). After deproteination in 0.5  $\mu$ g/ml proteinase K for 10 minutes at 37°C (12.5 $\mu$ L of 10 mg/ml stock in 250 ml prewarmed RNase-free RNase buffer), the sections were washed in 0.5 x SSC for 10 minutes at room temperature. The sections were dehydrated in 70%, 95%, 100% ethanol, 2 minutes each.

*Pretreatment of paraffin-embedded sections* The slides were deparaffinized, placed in SQ H<sub>2</sub>O, and rinsed twice in 2 x SSC at room temperature, for 5 minutes each time. The sections were deproteinated in 20  $\mu$ g/ml proteinase K (500  $\mu$ L of 10 mg/ml in 250 ml RNase-free RNase buffer; 37C, 15 minutes) - human embryo, or 8 x proteinase K (100  $\mu$ L in 250 ml Rnase buffer, 37C, 30 minutes) - formalin tissues. Subsequent rinsing in 0.5 x SSC and dehydration were performed as described above.

*Prehybridization* The slides were laid out in plastic box lined with Box buffer (4 x SSC, 50% formamide) - saturated filter paper. The tissue was covered with 50  $\mu$ L of hybridization buffer (3.75g Dextran Sulfate + 6 ml SQ H<sub>2</sub>O), vortexed and heated in the microwave for 2 minutes with the cap loosened. After cooling on ice, 18.75 ml formamide, 3.75 ml 20 x SSC and 9 ml SQ H<sub>2</sub>O were added, the tissue was vortexed well, and incubated at 42C for 1-4 hours.

*Hybridization* 1.0 x 10<sup>6</sup> cpm probe and 1.0  $\mu$ L tRNA (50 mg/ml stock) per slide were heated at 95C for 3 minutes. The slides were cooled on ice, and 48  $\mu$ L hybridization buffer were added per slide. After vortexing, 50  $\mu$ L <sup>33</sup>P mix were added to 50  $\mu$ L prehybridization on slide. The slides were incubated overnight at 55C.

*Washes* Washing was done 2x10 minutes with 2xSSC, EDTA at room temperature (400 ml 20 x SSC + 16 ml 0.25M EDTA, V<sub>f</sub>=4L), followed by RNaseA treatment at 37C for 30 minutes (500  $\mu$ L of 10 mg/ml in 250 ml Rnase buffer = 20  $\mu$ g/ml), The slides were washed 2x10 minutes with 2 x SSC, EDTA at room temperature. The stringency wash conditions were as follows: 2 hours at 55C, 0.1 x SSC, EDTA (20 ml 20 x SSC + 16 ml EDTA, V<sub>f</sub>=4L).

#### DNA 35638 (1 TM receptor)

Expression was observed in the endothelium lining of a subset of fetal and placental vessels. Endothelial expression was confined to these tissue blocks. Expression was also observed over intermediate trophoblast cells of placenta.

Oligo C-120N: (SEQ ID NO: 36)

GGA TTC TAA TAC GAC TCA CTA TAG GGC TGC GGC GGC TCA GGT CTT CAG TT

Oligo c-120P (SEQ ID NO: 37)

CTA TGA AAT TAA CCC TCA CTA AAG GGA GCA TGG GAT GGG GAG GGA TAC GG

#### DNA 33094 (EGF Homolog)

A highly distinctive expression pattern was observed. In the human embryo expression was observed in outer smooth muscle layer of the GI tract, respiratory cartilage, branching respiratory epithelium, osteoblasts, tendons, gonad, in the optic nerve head and developing dermis. In the adult, expression was observed in the epidermal pegs of the chimp tongue, the basal epithelial / myoepithelial cells of the prostate and urinary bladder. Expression was also found in the alveolar lining cells of the adult lung, mesenchymal cells juxtaposed to erectile tissue in the penis and the cerebral cortex (probably glial cells). In the kidney, expression was only seen in disease, in cells surrounding thyroidized renal tubules.

Oligo D-200V (SEQ ID NO: 38)

CTA TGA AAT TAA CCC TCA CTA AAG GGA ATA GCA GGC CAT CCC AGG ACA

Oligo D-200Z (SEQ ID NO: 39)

CTA TGA AAT TAA CCC TCA CTA AAG GGA TGT CTT CCA TGC CAA CCT TC

#### EXAMPLE 6

##### In situ Hybridization in Cells and Diseased Tissues

The *in situ* hybridization method of Example 5 is used to determine gene expression, analyze the tissue distribution of transcription, and follow changes in specific mRNA synthesis for the genes/DNAs and the proteins of the invention in diseased tissues isolated from human individuals suffering from a specific disease. These results show more specifically where in diseased tissues the genes of the invention are expressed and are more predictive of the particular localization of the therapeutic effect of the inhibitory or stimulatory compounds of the invention (and agonists or antagonists thereof) in a disease. Hybridization is performed according to the method of Example 5 using one or more of the following tissue and cell samples:

- (a) lymphocytes and antigen presenting cells (dendritic cells, langherhans cells, macrophages and monocytes, NK cells);
- (b) lymphoid tissues: normal and reactive lymph node, thymus, Bronchial Associated Lymphoid Tissues, (BALT), Mucosal Associated Lymphoid Tissues (MALT);
- (c) human disease tissues

- Synovium and joint of patients with Arthritis and Degenerative Joint Disease
- Colon from patients with Inflammatory Bowel Disease including Ulcerative Colitis and Crohns' disease
- Skin lesions from Psoriasis and other forms of dermatitis
- Lung tissue including BALT and tissue lymph nodes from Chronic and acute bronchitis, pneumonia, pneumonitis, pleuritis
- Lung tissue including BALT and tissue lymph nodes from Asthma
- nasal and sinus tissue from patients with rhinitis or sinusitis
- Brain and Spinal cord from Multiple Sclerosis. Alzheimer's Disease and Stroke
- Kidney from Nephritis, Glomerulonephritis and Systemic Lupus Erythematosus
- Liver from Infectious and non-infectious Hepatitis
- Tissues from Neoplasms/Cancer.

Expression is observed in one or more cell or tissue samples indicating localization of the therapeutic effect of the compounds of the invention (and agonists or antagonists thereof) in the

disease associated with the cell or tissue sample.

DNA 35638 (PRO245) was found to be expressed in inflamed human tissues (psoriasis, inflammatory bowel disease (IBD), inflamed kidney, inflamed lung, hepatitis (liver block), normal tonsil, adult and chimp (multiblocks). Expression was present in the endothelium/intima of large vessels in the lung afflicted with chronic inflammation, in the superficial dermal vessels of the psoriatic skin, in arterioles in a specimen of chronic sclerosing nephritis, and in capillaries including the perifollicular sinuses of the tonsil. These results indicate that PRO245 is immunostimulatory (enhances T lymphocyte proliferation in the MLR and costimulation) and has proinflammatory properties (induces a neutrophil infiltrate *in vivo*).

#### EXAMPLE 7

##### Use of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 as a hybridization probe.

DNA comprising the coding sequence of full-length or mature PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 (as shown in Figures 4, 6, 8, 10, 12, 14 and 16) is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be identified using standard techniques known in the art.

#### EXAMPLE 8

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in *E. coli*

This example illustrates preparation of an unglycosylated form of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 by recombinant expression in *E. coli*.

The DNA sequence encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., *Gene*, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., *supra*. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 protein can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

PRO245, PRO217, PRO301 and PRO266 were expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding PRO245, PRO217, PRO301 and PRO266 was initially amplified using selected PCR primers. The primers contained restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences were then ligated into an expression vector, which was used to transform an *E. coli* host based on strain 52 (W3110 fuhA(tonA) lon galE rpoHts(htpRts) clpP(lacIq)). Transformants were first grown in LB containing 50 mg/ml carbenicillin at 30C with shaking until an O.D.600 of 3-5 was reached. Cultures were then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71 g sodium citrate.2H<sub>2</sub>O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield

hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v) glucose and 7 mM MgSO<sub>4</sub>) and grown for approximately 20-30 hours at 30C with shaking. Samples were removed to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets were frozen until purification and refolding.

*E. coli* paste from 0.5 to 1 L fermentations (6-10 g pellets) was resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution was stirred overnight at 4C. This step results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution was centrifuged at 40,000 rpm in a Beckman Ultracentrifuge for 30 min. The supernatant was diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. Depending the clarified extract was loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column was washed with additional buffer containing 50 mM imidazole (Calbiochem, Utrol grade), pH 7.4. The protein was eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein were pooled and stored at 4C. Protein concentration was estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

The proteins were refolded by diluting sample slowly into freshly prepared refolding buffer consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes were chosen so that the final protein concentration was between 50 to 100 micrograms/ml. The refolding solution was stirred gently at 4C for 12-36 hours. The refolding reaction was quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution was filtered through a 0.22 micron filter and acetonitrile was added to 2-10% final concentration. The refolded protein was chromatographed on a Poros R1/H reversed phase column using a mobile buffer of 0.1% TFA with elution with a gradient of acetonitrile from 10 to 80%. Aliquots of fractions with A<sub>280</sub> absorbance were analyzed on SDS polyacrylamide gels and fractions containing homogeneous refolded protein were pooled. Generally, the properly refolded species of most proteins are eluted at the lowest concentrations of acetonitrile since those species are the most compact with their hydrophobic interiors shielded from interaction with the reversed phase resin. Aggregated species are usually eluted at higher acetonitrile concentrations. In addition to resolving misfolded forms of proteins from the desired form, the reversed phase step also removes endotoxin from the samples.

Fractions containing the desired folded PRO245, PRO217, PRO301 and PRO266 proteins, respectively, were pooled and the acetonitrile removed using a gentle stream of nitrogen directed at the solution. Proteins were formulated into 20 mM Hepes, pH 6.8 with 0.14 M sodium

chloride and 4% mannitol by dialysis or by gel filtration using G25 Superfine (Pharmacia) resins equilibrated in the formulation buffer and sterile filtered.

#### EXAMPLE 9

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in mammalian cells

This example illustrates preparation of a potentially glycosylated form of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA using ligation methods such as described in Sambrook et al., *supra*. The resulting vector is called pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 ug pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA is mixed with about 1 ug DNA encoding the VA RNA gene [Thimmappaya et al., *Cell*, 31:543 (1982)] and dissolved in 500 uL of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M CaCl<sub>2</sub>. To this mixture is added, dropwise, 500 uL of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM NaPO<sub>4</sub>, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 uCi/ml <sup>35</sup>S-cysteine and 200 uCi/ml <sup>35</sup>S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., *Proc. Natl. Acad. Sci.*, 78:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 ug pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or

PRO326 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5  $\mu$ g/ml bovine insulin and 0.1  $\mu$ g/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be expressed in CHO cells. The pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be transfected into CHO cells using known reagents such as CaPO<sub>4</sub> or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as <sup>35</sup>S-methionine. After determining the presence of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be concentrated and purified by any selected method.

Epitope-tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may also be expressed in host CHO cells. The PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be concentrated and purified by any selected method, such as by Ni<sup>2+</sup>-chelate affinity chromatography.

PRO245, PRO217 and PRO301 were expressed in CHO cells by both a transient and a stable expression procedure.

Stable expression in CHO cells was performed using the following procedure. The proteins were expressed as an IgG construct (immunoadhesin), in which the coding sequences for the soluble forms (e.g. extracellular domains) of the respective proteins were fused to an IgG1 constant region sequence containing the hinge, CH2 and CH2 domains and/or is a poly-His tagged form.

Following PCR amplification, the respective DNAs were subcloned in a CHO expression

vector using standard techniques as described in Ausubel et al., *Current Protocols of Molecular Biology*, Unit 3.16, John Wiley and Sons (1997). CHO expression vectors are constructed to have compatible restriction sites 5' and 3' of the DNA of interest to allow the convenient shuttling of cDNAs. The vector used expression in CHO cells is as described in Lucas *et al.*, *Nucl. Acids Res.* 24: 9 (1774-1779 (1996), and uses the SV40 early promoter/enhancer to drive expression of the cDNA of interest and dihydrofolate reductase (DHFR). DHFR expression permits selection for stable maintenance of the plasmid following transfection.

Twelve micrograms of the desired plasmid DNA were introduced into approximately 10 million CHO cells using commercially available transfection reagents Superfect (Qiagen), Dospers or Fugene (Boehringer Mannheim). The cells were grown and described in Lucas *et al.*, *supra*. Approximately  $3 \times 10^7$  cells are frozen in an ampule for further growth and production as described below.

The ampules containing the plasmid DNA were thawed by placement into water bath and mixed by vortexing. The contents were pipetted into a centrifuge tube containing 10 mLs of media and centrifuged at 1000 rpm for 5 minutes. The supernatant was aspirated and the cells were resuspended in 10 mL of selective media (0.2  $\mu$ m filtered PS20 with 5% 0.2  $\mu$ m diafiltered fetal bovine serum). The cells were then aliquoted into a 100 mL spinner containing 90 mL of selective media. After 1-2 days, the cells were transferred into a 250 mL spinner filled with 150 mL selective growth medium and incubated at 37C. After another 2-3 days, a 250 mL, 500 mL and 2000 mL spinners were seeded with  $3 \times 10^5$  cells/mL. The cell media was exchanged with fresh media by centrifugation and resuspension in production medium. Although any suitable CHO media may be employed, a production medium described in U.S. Patent No. 5,122,469, issued June 16, 1992 was actually used. 3L production spinner is seeded at  $1.2 \times 10^6$  cells/mL. On day 0, the cell number pH were determined. On day 1, the spinner was sampled and sparging with filtered air was commenced. On day 2, the spinner was sampled, the temperature shifted to 33C, and 30 mL of 500 g/L glucose and 0.6 mL of 10% antifoam (e.g., 35% polydimethylsiloxane emulsion, Dow Corning 365 Medical Grade Emulsion). Throughout the production, pH was adjusted as necessary to keep at around 7.2. After 10 days, or until viability dropped below 70%, the cell culture was harvested by centrifugation and filtering through a 0.22  $\mu$ m filter. The filtrate was either stored at 4C or immediately loaded onto columns for purification.

For the poly-His tagged constructs, the proteins were purified using a Ni-NTA column (Qiagen). Before purification, imidazole was added to the conditioned media to a concentration of 5 mM. The conditioned media was pumped onto a 6 ml Ni-NTA column equilibrated in 20 mM Hepes, pH 7.4, buffer containing 0.3 M NaCl and 5 mM imidazole at a flow rate of 4-5 ml/min. at 4C. After loading, the column was washed with additional equilibration buffer and the protein eluted with

equilibration buffer containing 0.25 M imidazole. The highly purified protein was subsequently desalted into a storage buffer containing 10 mM Hepes, 0.14 M NaCl and 4% mannitol, pH 6.8, with a 25 ml G25 Superfine (Pharmacia) column and stored at -80C.

Immunoadhesin (Fc containing) constructs were purified from the conditioned media as follows. The conditioned medium was pumped onto a 5 ml Protein A column (Pharmacia) which had been equilibrated in 20 mM Na phosphate buffer, pH 6.8. After loading, the column was washed extensively with equilibration buffer before elution with 100 mM citric acid, pH 3.5. The eluted protein was immediately neutralized by collecting 1 ml fractions into tubes containing 275  $\mu$ L of 1 M Tris buffer, pH 9. The highly purified protein was subsequently desalted into storage buffer as described above for the poly-His tagged proteins. The homogeneity was assessed by SDS polyacrylamide gels and by N-terminal amino acid sequencing by Edman degradation.

PRO326 was also produced by transient expression in COS cells.

#### EXAMPLE 10

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in Yeast

The following method describes recombinant expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 from the ADH2/GAPDH promoter. DNA encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. For secretion, DNA encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, a native PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 signal peptide or other mammalian signal peptide, or, for example, a yeast alpha-factor or invertase secretory signal/leader sequence, and linker sequences (if needed) for expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may further be

purified using selected column chromatography resins.

#### EXAMPLE 11

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in Baculovirus-infected insect cells.

The sequence coding for PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the sequence encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 or the desired portion of the coding sequence of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 [such as the sequence encoding the extracellular domain of a transmembrane protein or the sequence encoding the mature protein if the protein is extracellular] is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfected the above plasmid and BaculoGold<sup>TM</sup> virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 - 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression are performed as described by O'Reilley et al., Baculovirus expression vectors: A Laboratory Manual, Oxford: Oxford University Press (1994).

Expressed poly-his tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be purified, for example, by Ni<sup>2+</sup>-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM MgCl<sub>2</sub>; 0.1 mM EDTA; 10% glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% glycerol, pH 7.8) and filtered through a 0.45 um filter. A Ni<sup>2+</sup>-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A<sub>280</sub> with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% glycerol, pH

6.0), which elutes nonspecifically bound protein. After reaching  $A_{280}$  baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or Western blot with  $Ni^{2+}$ -NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted  $His_{10}$ -tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

PRO245, PRO301 and PRO266 were expressed in baculovirus infected Sf9 insect cells. While the expression was actually performed in a 0.5-2 L scale, it can be readily scaled up for larger (e.g. 8 L) preparations. The proteins were expressed as an IgG construct (immunoadhesin), in which the protein extracellular region was fused to an IgG1 constant region sequence containing the hinge, CH2 and CH3 domains and/or in poly-His tagged forms.

Following PCR amplification, the respective coding sequences were subcloned into a baculovirus expression vector (pb.PH.IgG for IgG fusions and pb.PH.His.c for poly-His tagged proteins), and the vector and Baculogold baculovirus DNA (Pharmingen) were co-transfected into 105 *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711), using Lipofectin (Gibco BRL). pb.PH.IgG and pb.PH.His are modifications of the commercially available baculovirus expression vector pVL1393 (Pharmingen), with modified polylinker regions to include the His or Fc tag sequences. The cells were grown in Hink's TNM-FH medium supplemented with 10% FBS (Hyclone). Cells were incubated for 5 days at 28C. The supernatant was harvested and subsequently used for the first viral amplification by infecting Sf9 cells in Hink's TNM-FH medium supplemented with 10% FBS at an approximate multiplicity of infection (MOI) of 10. Cells were incubated for 3 days at 28C. The supernatant was harvested and the expression of the constructs in the baculovirus expression vector was determined by batch binding of 1 ml of supernatant to 25 mL of Ni-NTA beads (QIAGEN) for histidine tagged proteins or Protein-A Sepharose CL-4B beads (Pharmacia) for IgG tagged proteins followed by SDS-PAGE analysis comparing to a known concentration of protein standard by Coomassie blue staining.

The first viral amplification supernatant was used to infect a spinner culture (500 ml) of Sf9 cells grown in ESF-921 medium (Expression Systems LLC) at an approximate MOI of 0.1. Cells were incubated for 3 days at 28C. The supernatant was harvested and filtered. Batch binding and SDS-PAGE analysis was repeated, as necessary, until expression of the spinner culture was confirmed.

The conditioned medium from the transfected cells (0.5 to 3 L) was harvested by centrifugation to remove the cells and filtered through 0.22 micron filters. For the poly-His tagged

constructs, the protein construct were purified using a Ni-NTA column (Qiagen). Before purification, imidazole was added to the conditioned media to a concentration of 5 mM. The conditioned media were pumped onto a 6 ml Ni-NTA column equilibrated in 20 mM Hepes, pH 7.4, buffer containing 0.3 M NaCl and 5 mM imidazole at a flow rate of 4-5 ml/min. at 4C. After loading, the column was washed with additional equilibration buffer and the protein eluted with equilibration buffer containing 0.25 M imidazole. The highly purified protein was subsequently desalted into a storage buffer containing 10 mM Hepes, 0.14 M NaCl and 4% mannitol, pH 6.8, with a 25 ml G25 Superfine (Pharmacia) column and stored at -80C.

Immunoadhesin (Fc containing) constructs of proteins were purified from the conditioned media as follows. The conditioned media were pumped onto a 5 ml Protein A column (Pharmacia) which had been equilibrated in 20 mM Na phosphate buffer, pH 6.8. After loading, the column was washed extensively with equilibration buffer before elution with 100 mM citric acid, pH 3.5. The eluted protein was immediately neutralized by collecting 1 ml fractions into tubes containing 275 mL of 1 M Tris buffer, pH 9. The highly purified protein was subsequently desalted into storage buffer as described above for the poly-His tagged proteins. The homogeneity of the proteins was verified by SDS polyacrylamide gel (PEG) electrophoresis and N-terminal amino acid sequencing by Edman degradation.

PRO245, PRO217, PRO301, PRO266, PRO331 and PRO326 were also expressed in baculovirus infected High-5 cells using an analogous procedure.

#### EXAMPLE 12

##### Preparation of Antibodies that Bind PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326

This example illustrates preparation of monoclonal antibodies which can specifically bind PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, *supra*. Immunogens that may be employed include purified PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326, fusion proteins containing PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326, and cells expressing recombinant PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days

later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

#### Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209, USA (ATCC):

<u>Material</u>	<u>ATCC Dep. No.</u>	<u>Deposit Date</u>
DNA40981	209439	7 November 1997
DNA37140	209489	21 November 1997
DNA41388	209927	2 June 1998
DNA35638	209265	17 September 1997
DNA37150	209401	17 October 1997

DNA33094	209256	16 September 1997
DNA32292	209258	16 September 1997
DNA32279	209259	16 September 1997
DNA40628	209432	7 November 1997

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC 122 and the Commissioner's rules pursuant thereto (including 37 CFR 1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

TABLE 4

PRO Comparison Protein      XXXXXXXXXXXXXXXX      (Length = 15 amino acids)  
 Comparison Protein      XXXXXYYYYYYY      (Length = 12 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 15 = 33.3%

---

PRO Comparison Protein      XXXXXXXXXX      (Length = 10 amino acids)  
 Comparison Protein      XXXXXYYYYYYZZYZ (Length = 15 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 10 = 50%

---

PRO-DNA Comparison DNA      NNNNNNNNNNNNNN      (Length = 14 nucleotides)  
 Comparison DNA      NNNNNNNLLLLLLL      (Length = 16 nucleotides)

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

6 divided by 14 = 42.9%

---

PRO-DNA Comparison DNA      NNNNNNNNNNNN      (Length = 12 nucleotides)  
 Comparison DNA      NNNNLLLVV      (Length = 9 nucleotides)

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

4 divided by 12 = 33.3%

TABLE 5

```

/*
* C-C increased from 12 to 15
* Z is average of EQ
* B is average of ND
* match with stop is _M; stop-stop = 0; J (joker) match = 0
*/
#define _M -8 /* value of a match with a stop */

int day[26][26] = {
/* A */ { 2, 0, -2, 0, -4, 1, -1, -1, 0, -1, -2, -1, 0, _M, 1, 0, -2, 1, 1, 0, -6, 0, -3, 0 },
/* B */ { 0, 3, -4, 3, 2, -5, 0, 1, -2, 0, 0, -3, -2, 2, _M, -1, 1, 0, 0, 0, 0, -2, -5, 0, -3, 1 },
/* C */ { -2, -4, 15, -5, -5, -4, -3, -3, -2, 0, -5, -6, -5, -4, _M, -3, -5, -4, 0, -2, 0, -2, -8, 0, 0, -5 },
/* D */ { 0, 3, -5, 4, 3, -6, 1, 1, -2, 0, 0, -4, -3, 2, _M, -1, 2, -1, 0, 0, 0, -2, -7, 0, -4, 2 },
/* E */ { 0, 2, -5, 3, 4, -5, 0, 1, -2, 0, 0, -3, -2, 1, _M, -1, 2, -1, 0, 0, 0, -2, -7, 0, -4, 3 },
/* F */ { -4, -5, -4, -6, -5, -9, -5, -2, 1, 0, -5, 2, 0, -4, _M, -5, -5, -4, -3, -3, 0, -1, 0, 0, 7, 5 },
/* G */ { 1, 0, -3, 1, 0, -5, 5, -2, -3, 0, -2, -4, -3, 0, _M, -1, -1, -3, 1, 0, 0, -1, 7, 0, -5, 0 },
/* H */ { -1, 1, -3, 1, 1, -2, -2, 6, -2, 0, 0, -2, -2, 2, _M, 0, 3, 2, -1, -1, 0, -2, -3, 0, 0, 2 },
/* I */ { -1, -2, -2, -2, 2, 1, -3, -2, 5, 0, -2, 2, 2, -2, _M, -2, -2, -1, 0, 0, 0, 4, -5, 0, -1, -2 },
/* J */ { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, _M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 },
/* K */ { -1, 0, -5, 0, 0, -5, -2, 0, -2, 0, 5, -3, 0, 1, _M, -1, 1, 3, 0, 0, 0, -2, -3, 0, -4, 0 },
/* L */ { -2, -3, -6, -4, -3, 2, -4, -2, 2, 0, -3, 6, 4, -3, _M, -3, -2, -3, -3, -1, 0, 2, -2, 0, -1, -2 },
/* M */ { -1, -2, -5, -3, -2, 0, -3, -2, 2, 0, 0, 4, 6, -2, _M, -2, -1, 0, -2, -1, 0, 2, -4, 0, -2, -1 },
/* N */ { 0, 2, -4, 2, 1, -4, 0, 2, -2, 0, 1, -3, -2, 2, _M, -1, 1, 0, 1, 0, 0, -2, -4, 0, -2, 1 },
/* O */ { _M, _M },
/* P */ { -1, 1, -3, 1, -1, -5, -1, 0, -2, 0, -1, -3, -2, -1, _M, 6, 0, 0, 1, 0, 0, -1, -6, 0, -5, 0 },
/* Q */ { 0, 1, -5, 2, 2, -5, 1, 3, -2, 0, 1, -2, -1, 1, _M, 0, 4, 1, -1, 1, 0, -2, -5, 0, 4, 3 },
/* R */ { -2, 0, -4, -1, 1, -4, -3, 2, -2, 0, 3, -3, 0, 0, _M, 0, 1, 6, 0, -1, 0, -2, 2, 0, -4, 0 },
/* S */ { 1, 0, 0, 0, 0, -3, 1, -1, -1, 0, 0, -3, -2, 1, _M, 1, -1, 0, 2, 1, 0, -1, -2, 0, -3, 0 },
/* T */ { 1, 0, -2, 0, 0, -3, 0, -1, 0, 0, 0, -1, -1, 0, _M, 0, -1, -1, 1, 3, 0, 0, -5, 0, -3, 0 },
/* U */ { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, _M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 },
/* V */ { 0, -2, -2, -2, -2, -1, -1, -2, 4, 0, -2, 2, 2, -2, _M, -1, -2, -2, -1, 0, 0, 4, -6, 0, -2, -2 },
/* W */ { -6, -5, -8, -7, -7, 0, -7, -3, -5, 0, -3, -2, -4, 4, _M, -6, -5, 2, -2, -5, 0, -6, 17, 0, 0, -6 },
/* X */ { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, _M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 },
/* Y */ { -3, -3, 0, -4, -4, -4, -7, -5, -0, -1, 0, -4, -1, -2, -2, _M, -5, -4, -4, -3, -3, 0, -2, 0, 0, 10, -4 },
/* Z */ { 0, 1, -5, 2, 3, -5, 0, 2, -2, 0, 0, -2, -1, 1, _M, 0, 3, 0, 0, 0, 0, -2, -6, 0, -4, 4 }
};

```

Page 1 of day.h

TABLE 5 (cont.)

```

/*
 */
#include <stdio.h>
#include <ctype.h>

#define MAXJMP 16      /* max jumps in a diag */
#define MAXGAP 24      /* don't continue to penalize gaps larger than this */
#define JMPSS 1024      /* max jmps in an path */
#define MX 4           /* save if there's at least MX-1 bases since last jmp */

#define DMAT 3          /* value of matching bases */
#define DMIS 0          /* penalty for mismatched bases */
#define DINS0 8          /* penalty for a gap */
#define DINS1 1          /* penalty per base */
#define PINS0 8          /* penalty for a gap */
#define PINS1 4          /* penalty per residue */

struct jmp {
    short n[MAXJMP];      /* size of jmp (neg for delay) */
    unsigned short x[MAXJMP]; /* base no. of jmp in seq x */
    /* limits seq to 2^16 -1 */
};

struct diag {
    int score;           /* score at last jmp */
    long offset;         /* offset of prev block */
    short ijmp;          /* current jmp index */
    struct jmp *jp;      /* list of jmps */
};

struct path {
    int spc;             /* number of leading spaces */
    short n[JMPSS];     /* size of jmp (gap) */
    int x[JMPSS];        /* loc of jmp (last elem before gap) */
};

char *ofile;           /* output file name */
char *namex[2];        /* seq names: getseqs() */
char *prog;             /* prog name for err msgs */
char *seqx[2];          /* seqs: getseqs() */
int dmax;              /* best diag: nw() */
int dmax0;             /* final diag */
int dna;                /* set if dna: main() */
int endgaps;           /* set if penalizing end gaps */
int gapx, gapy;        /* total gaps in seqs */
int len0, len1;         /* seq lens */
int ngapx, ngapy;       /* total size of gaps */
int smax;               /* max score: nw() */
int *xbm;               /* bitmap for matching */
long offset;             /* current offset in jmp file */
struct diag *dx;        /* holds diagonals */
struct path pp[2];      /* holds path for seqs */

char *calloc(), *malloc(), *index(), *strcpy();
char *getseq(), *g_malloc();

```

Page 1 of nw.h

TABLE 5 (cont.)

```

/* Needleman-Wunsch alignment program
*
* usage: progs file1 file2
* where file1 and file2 are two dna or two protein sequences.
* The sequences can be in upper- or lower-case and may contain ambiguity
* Any lines beginning with ';' or '>' are ignored
* Max file length is 65535 (limited by unsigned short x in the jmp struct)
* A sequence with 1/3 or more of its elements ACGTU is assumed to be DNA
* Output is in the file "align.out"
*
* The program may create a tmp file in /tmp to hold info about traceback.
* Original version developed under BSD 4.3 on a vax 8650
*/
#include "nw.h"
#include "day.h"

static      _dbval[26] = {
    1,14,2,13,0,0,4,11,0,0,12,0,3,15,0,0,0,5,6,8,8,7,9,0,10,0
};

static      _pbval[26] = {
    1, 2,(1<<('D'-'A'))|(1<<('N'-'A')), 4, 8, 16, 32, 64,
    128, 256, 0xFFFFFFFF, 1<<10, 1<<11, 1<<12, 1<<13, 1<<14,
    1<<15, 1<<16, 1<<17, 1<<18, 1<<19, 1<<20, 1<<21, 1<<22,
    1<<23, 1<<24, 1<<25|(1<<('E'-'A'))|(1<<('Q'-'A'))
};

main(ac, av)      main
int               ac;
char              *av[];
{
    prog = av[0];
    if (ac != 3) {
        fprintf(stderr, "usage: %s file1 file2\n", prog);
        fprintf(stderr, "where file1 and file2 are two dna or two protein sequences.\n");
        fprintf(stderr, "The sequences can be in upper- or lower-case\n");
        fprintf(stderr, "Any lines beginning with ';' or '>' are ignored\n");
        fprintf(stderr, "Output is in the file \"align.out\"\n");
        exit(1);
    }
    namex[0] = av[1];
    namex[1] = av[2];
    seqx[0] = getseq(namex[0], &len0);
    seqx[1] = getseq(namex[1], &len1);
    xbm = (dna)?_dbval:_pbval;

    endgaps = 0;                      /* 1 to penalize endgaps */
    ofile = "align.out";               /* output file */

    nw();                            /* fill in the matrix, get the possible jmps */
    readjmps();                      /* get the actual jmps */
    print();                          /* print stats, alignment */

    cleanup(0);                      /* unlink any tmp files */
}

```

Page 1 of nw.c

TABLE 5 (cont.)

```

/* do the alignment, return best score: main()
 * dna: values in Fitch and Smith, PNAS, 80, 1382-1386, 1983
 * pro: PAM 250 values
 * When scores are equal, we prefer mismatches to any gap, prefer
 * a new gap to extending an ongoing gap, and prefer a gap in seqx
 * to a gap in seq y.
 */
nw()    nw
{
    char      *px, *py;          /* seqs and ptrs */
    int       *ndely, *dely;     /* keep track of dely */
    int       ndelx, delx;      /* keep track of delx */
    int       *tmp;             /* for swapping row0, row1 */
    int       mis;              /* score for each type */
    int       ins0, ins1;        /* insertion penalties */
    register  id;              /* diagonal index */
    register  ij;              /* jmp index */
    register  *col0, *coll;     /* score for curr, last row */
    register  xx, yy;          /* index into seqs */

    dx = (struct diag *)g_malloc("to get diags", len0+len1+1, sizeof(struct diag));

    ndely = (int *)g_malloc("to get ndely", len1+1, sizeof(int));
    dely = (int *)g_malloc("to get dely", len1+1, sizeof(int));
    col0 = (int *)g_malloc("to get col0", len1+1, sizeof(int));
    col1 = (int *)g_malloc("to get col1", len1+1, sizeof(int));
    ins0 = (dna)? DINS0 : PINS0;
    ins1 = (dna)? DINS1 : PINS1;

    smax = -10000;
    if(endgaps) {
        for (col0[0] = dely[0] = -ins0, yy = 1; yy <= len1; yy++) {
            col0[yy] = dely[yy] = col0[yy-1] - ins1;
            ndely[yy] = yy;
        }
        col0[0] = 0;           /* Waterman Bull Math Biol 84 */
    }
    else
        for (yy = 1; yy <= len1; yy++)
            dely[yy] = -ins0;

    /* fill in match matrix
    */
    for (px = seqx[0], xx = 1; xx <= len0; px++, xx++) {
        /* initialize first entry in col
        */
        if(endgaps) {
            if(xx == 1)
                col1[0] = delx = -(ins0+ins1);
            else
                col1[0] = delx = col0[0] - ins1;
            ndelx = xx;
        }
        else {
            col1[0] = 0;
            delx = -ins0;
            ndelx = 0;
        }
    }
}

```

Page 2 of nw.c

TABLE 5 (cont.)

...nw

```

for (py = seqx[1], yy = 1; yy <= len1; py++, yy++) {
    mis = col0[yy-1];
    if (dna)
        mis += (xbm[*px-'A']&xbm[*py-'A'])? DMAT : DMIS;
    else
        mis += _day[*px-'A'][*py-'A'];

    /* update penalty for del in x seq;
     * favor new del over ongong del
     * ignore MAXGAP if weighting endgaps
     */
    if (endgaps || ndely[yy] < MAXGAP) {
        if (col0[yy] - ins0 >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else {
            dely[yy] := ins1;
            ndely[yy]++;
        }
    } else {
        if (col0[yy] - (ins0+ins1) >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else
            ndely[yy]++;
    }

    /* update penalty for del in y seq;
     * favor new del over ongong del
     */
    if (endgaps || ndelx < MAXGAP) {
        if (coll[yy-1] - ins0 >= delx) {
            delx = coll[yy-1] - (ins0+ins1);
            ndelx = 1;
        } else {
            delx = ins1;
            ndelx++;
        }
    } else {
        if (coll[yy-1] - (ins0+ins1) >= delx) {
            delx = coll[yy-1] - (ins0+ins1);
            ndelx = 1;
        } else
            ndelx++;
    }

    /* pick the maximum score: we're favoring
     * mis over any del and delx over dely
     */
}

```

Page 3 of nw.c

TABLE 5 (cont.)

```

...nw

id = xx - yy + len1 - 1;
if (mis >= delx && mis >= dely[yy])
    col1[yy] = mis;
else if (delx >= dely[yy]) {
    col1[yy] = delx;
    ij = dx[id].ijmp;
    if (dx[id].jp.n[0] && (!dna || (ndelx >= MAXJMP
    && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
        dx[id].ijmp++;
        if (++ij >= MAXJMP) {
            writejmps(id);
            ij = dx[id].ijmp = 0;
            dx[id].offset = offset;
            offset += sizeof(struct jmp) + sizeof(offset);
        }
    }
    dx[id].jp.n[ij] = nodelx;
    dx[id].jp.x[ij] = xx;
    dx[id].score = delx;
}
else {
    col1[yy] = dely[yy];
    ij = dx[id].ijmp;
    if (dx[id].jp.n[0] && (!dna || (ndely[yy] >= MAXJMP
    && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
        dx[id].ijmp++;
        if (++ij >= MAXJMP) {
            writejmps(id);
            ij = dx[id].ijmp = 0;
            dx[id].offset = offset;
            offset += sizeof(struct jmp) + sizeof(offset);
        }
    }
    dx[id].jp.n[ij] = -ndely[yy];
    dx[id].jp.x[ij] = xx;
    dx[id].score = dely[yy];
}
if (xx == len0 && yy < len1) {
    /* last col
     */
    if (endgaps)
        col1[yy] = ins0+ins1*(len1-yy);
    if (col1[yy] > smax) {
        smax = col1[yy];
        dmax = id;
    }
}
if (endgaps && xx < len0)
    col1[yy-1] = ins0+ins1*(len0-xx);
if (col1[yy-1] > smax) {
    smax = col1[yy-1];
    dmax = id;
}
tmp = col0; col0 = col1; col1 = tmp;
}
(void) free((char *)ndely);
(void) free((char *)dely);
(void) free((char *)col0);
(void) free((char *)col1); }

```

Page 4 of nw.c

TABLE 5 (cont.)

```

/*
*
* print() -- only routine visible outside this module
*
* static:
* getmat() -- trace back best path, count matches: print()
* pr_align() -- print alignment of described in array p[]: print()
* dumpblock() -- dump a block of lines with numbers, stars: pr_align()
* nums() -- put out a number line: dumpblock()
* putline() -- put out a line (name, [num], seq, [num]): dumpblock()
* stars() -- put a line of stars: dumpblock()
* stripname() -- strip any path and prefix from a seqname
*/
#include "nw.h"

#define SPC      3
#define P_LINE   256      /* maximum output line */
#define P_SPC    3         /* space between name or num and seq */

extern _day[26][26];
int    olen;           /* set output line length */
FILE   *fx;            /* output file */

print()  print
{
    int    lx, ly, firstgap, lastgap; /* overlap */

    if ((fx = fopen(ofile, "w")) == 0) {
        fprintf(stderr, "%s: can't write %s\n", prog, ofile);
        cleanup(1);
    }
    fprintf(fx, "<first sequence: %s (length = %d)\n", namex[0], len0);
    fprintf(fx, "<second sequence: %s (length = %d)\n", namex[1], len1);
    olen = 60;
    lx = len0;
    ly = len1;
    firstgap = lastgap = 0;
    if (dmax < len1 - 1) { /* leading gap in x */
        pp[0].spc = firstgap = len1 - dmax - 1;
        ly -= pp[0].spc;
    }
    else if (dmax > len1 - 1) { /* leading gap in y */
        pp[1].spc = firstgap = dmax - (len1 - 1);
        lx -= pp[1].spc;
    }
    if (dmax0 < len0 - 1) { /* trailing gap in x */
        lastgap = len0 - dmax0 - 1;
        lx -= lastgap;
    }
    else if (dmax0 > len0 - 1) { /* trailing gap in y */
        lastgap = dmax0 - (len0 - 1);
        ly -= lastgap;
    }
    getmat(lx, ly, firstgap, lastgap);
    pr_align();
}

```

Page 1 of nwprint.c

TABLE 5 (cont.)

```

/*
 * trace back the best path, count matches
 */
static
getmat(lx, ly, firstgap, lastgap)  getmat
    int      lx, ly;          /* "core" (minus endgaps) */
    int      firstgap, lastgap; /* leading/trailing overlap */
{
    int      nm, i0, i1, siz0, siz1;
    char     outx[32];
    double   pct;
    register int n0, n1;
    register char *p0, *p1;

    /* get total matches, score
     */
    i0 = i1 = siz0 = siz1 = 0;
    p0 = seqx[0] + pp[1].spc;
    p1 = seqx[1] + pp[0].spc;
    n0 = pp[1].spc + 1;
    n1 = pp[0].spc + 1;

    nm = 0;
    while (*p0 && *p1) {
        if (siz0) {
            p1++;
            n1++;
            siz0--;
        }
        else if (siz1) {
            p0++;
            n0++;
            siz1--;
        }
        else {
            if (xbm[*p0-'A']&xbm[*p1-'A'])
                nm++;
            if (n0++ == pp[0].x[i0])
                siz0 = pp[0].n[i0++];
            if (n1++ == pp[1].x[i1])
                siz1 = pp[1].n[i1++];
            p0++;
            p1++;
        }
    }
}

/* pct homology:
 * if penalizing endgaps, base is the shorter seq
 * else, knock off overhangs and take shorter core
 */
if (endgaps)
    lx = (len0 < len1)? len0 : len1;
else
    lx = (lx < ly)? lx : ly;
pct = 100.*(double)nm/(double)lx;
fprintf(fx, "\n");
fprintf(fx, "<%d match%6s in an overlap of %d: %.2f percent similarity\n",
    nm, (nm == 1)? "" : "es", lx, pct);

```

Page 2 of nwprint.c

TABLE 5 (cont.)

```

sprintf(fx, "<gaps in first sequence: %d", gapx);      ...getmat
if (gapx) {
    (void) sprintf(outx, " (%d %s%s)",
        ngapx, (dna)? "base": "residue", (ngapx == 1)? "": "s");
    fprintf(fx, "%s", outx);

    sprintf(fx, ", gaps in second sequence: %d", gapy);
    if (gapy) {
        (void) sprintf(outx, " (%d %s%s",
            ngapy, (dna)? "base": "residue", (ngapy == 1)? "": "s");
        fprintf(fx, "%s", outx);
    }
    if (dna)
        fprintf(fx,
            "\n<score: %d (match = %d, mismatch = %d, gap penalty = %d + %d per base)\n",
            smax, DMAT, DMIS, DINS0, DINS1);
    else
        fprintf(fx,
            "\n<score: %d (Dayhoff PAM 250 matrix, gap penalty = %d + %d per residue)\n",
            smax, PINS0, PINS1);
    if (endgaps)
        fprintf(fx,
            "<endgaps penalized. left endgap: %d %s%s, right endgap: %d %s%s\n",
            firstgap, (dna)? "base": "residue", (firstgap == 1)? "": "s",
            lastgap, (dna)? "base": "residue", (lastgap == 1)? "": "s");
    else
        fprintf(fx, "<endgaps not penalized\n");
}

static      nm;          /* matches in core -- for checking */
static      lmax;        /* lengths of stripped file names */
static      ij[2];        /* jmp index for a path */
static      nc[2];        /* number at start of current line */
static      ni[2];        /* current elem number -- for gapping */
static      siz[2];
static char *ps[2];      /* ptr to current element */
static char *po[2];      /* ptr to next output char slot */
static char out[2][P_LINE];
static char star[P_LINE]; /* set by stars() */

/*
 * print alignment of described in struct path pp[]
 */
static
pr_align() pr_align
{
    int          nn;          /* char count */
    int          more;
    register    i;

    for (i = 0, lmax = 0; i < 2; i++) {
        nn = stripname(namex[i]);
        if (nn > lmax)
            lmax = nn;

        nc[i] = 1;
        ni[i] = 1;
        siz[i] = ij[i] = 0;
        ps[i] = seqx[i];
        po[i] = out[i];
    }
}

```

Page 3 of nwprint.c

TABLE 5 (cont.)

```

for (nn = nm = 0, more = 1; more; ) {           ...pr_align
    for (i = more = 0; i < 2; i++) {
        /*
         * do we have more of this sequence?
         */
        if (!*ps[i])
            continue;

        more++;

        if (pp[i].spc) {           /* leading space */
            *po[i]++ = ' ';
            pp[i].spc--;
        }
        else if (siz[i]) {           /* in a gap */
            *po[i]++ = ' ';
            siz[i]--;
        }
        else {                     /* we're putting a seq element
            /*
             * are we at next gap for this seq?
             */
            if (ni[i] == pp[i].x[ij[i]]) {
                /*
                 * we need to merge all gaps
                 * at this location
                 */
                siz[i] = pp[i].n[ij[i]++];
                while (ni[i] == pp[i].x[ij[i]])
                    siz[i] += pp[i].n[ij[i]++];
            }
            ni[i]++;
        }
    }
    if (++nn == olen || !more && nn) {
        dumpblock();
        for (i = 0; i < 2; i++)
            po[i] = out[i];
        nn = 0;
    }
}

/*
 * dump a block of lines, including numbers, stars: pr_align()
 */
static
dumpblock()      dumpblock
{
    register i;
    for (i = 0; i < 2; i++)
        *po[i]-- = '^0';
}

```

Page 4 of nwprint.c

TABLE 5 (cont.)

```

...dumpblock

(void) putc('\n', fx);
for (i = 0; i < 2; i++) {
    if (*out[i] && (*out[i] != ' ' || *(po[i]) != ' ')) {
        if (i == 0)
            nums(i);
        if (i == 0 && *out[1])
            stars();
        putline(i);
        if (i == 0 && *out[1])
            fprintf(fx, star);
        if (i == 1)
            nums(i);
    }
}
/*
 * put out a number line: dumpblock()
 */
static
nums(ix)  nums
int       ix;      /* index in out[] holding seq line */
{
    char      nline[P_LINE];
    register  i, j;
    register char  *pn, *px, *py;

    for (pn = nline, i = 0; i < lmax+P_SPC; i++, pn++)
        *pn = ' ';
    for (i = nc[ix], py = out[ix]; *py; py++, pn++) {
        if (*py == ' ' || *py == '-')
            *pn = ' ';
        else {
            if (i%10 == 0 || (i == 1 && nc[ix] != 1)) {
                j = (i < 0)? -i : i;
                for (px = pn; j; j /= 10, px--)
                    *px = j%10 + '0';
                if (i < 0)
                    *px = '-';
            }
            else
                *pn = ' ';
        }
        i++;
    }
    *pn = '\0';
    nc[ix] = i;
    for (pn = nline; *pn; pn++)
        (void) putc(*pn, fx);
    (void) putc('\n', fx);
}

/*
 * put out a line (name, [num], seq, [num]): dumpblock()
 */
static
putline(ix) putline
int       ix;
{

```

TABLE 5 (cont.)

```

...putline

int i;
register char *px;

for (px = namex[ix], i = 0; *px && *px != '\'; px++, i++)
  (void) putc(*px, fx);
for (; i < lmax+P_SPC; i++)
  (void) putc(' ', fx);

/* these count from 1:
 * ni[] is current element (from 1)
 * nc[] is number at start of current line
 */
for (px = out[ix]; *px; px++)
  (void) putc(*px&0x7F, fx);
(void) putc('\n', fx);
}

/*
 * put a line of stars (seqs always in out[0], out[1]): dumpblock()
 */
static
stars() stars
{
  int i;
  register char *p0, *p1, cx, *px;

  if (!*out[0] || (*out[0] == ' ' && *(po[0]) == ' ') ||
      !*out[1] || (*out[1] == ' ' && *(po[1]) == ' '))
    return;
  px = star;
  for (i = lmax+P_SPC; i; i--)
    *px++ = ' ';

  for (p0 = out[0], p1 = out[1]; *p0 && *p1; p0++, p1++) {
    if (isalpha(*p0) && isalpha(*p1)) {

      if (xbm[*p0-'A']&xbm[*p1-'A']) {
        cx = '*';
        nm++;
      }
      else if (!dma && _day[*p0-'A'][*p1-'A'] > 0)
        cx = '.';
      else
        cx = ' ';
    }
    else
      cx = ' ';
    *px++ = cx;
  }
  *px++ = '\n';
  *px = '\0';
}
}

```

TABLE 5 (cont.)

```
/*
 * strip path or prefix from pn, return len: pr_align()
 */
static
stripname(pn)      stripname
    char      *pn;      /* file name (may be path) */
{
    register char      *px, *py;

    py = 0;
    for (px = pn; *px; px++)
        if (*px == '/')
            py = px + 1;
    if (py)
        (void) strcpy(pn, py);
    return(strlen(pn));
}
```

TABLE 5 (cont.)

```

/*
 * cleanup() -- cleanup any tmp file
 * getseq() -- read in seq, set dna, len, maxlen
 * g_calloc() -- calloc() with error checkin
 * readjmps() -- get the good jmps, from tmp file if necessary
 * writejmps() -- write a filled array of jmps to a tmp file: nw()
 */
#include "nw.h"
#include <sys/file.h>

char      *jname = "/tmp/homgXXXXXX";           /* tmp file for jmps */
FILE      *fj;

int       cleanup();                         /* cleanup tmp file */
long      lseek();

/*
 * remove any tmp file if we blow
 */
cleanup(i) cleanup
{
    int      i;
{
    if (fj)
        (void) unlink(jname);
    exit(i);
}

/*
 * read, return ptr to seq, set dna, len, maxlen
 * skip lines starting with ';', '<', or '>'
 * seq in upper or lower case
 */
char      *
getseq(file, len) getseq
{
    char      *file;      /* file name */
    int       *len;       /* seq len */
{
    char      line[1024], *pseq;
    register char  *px, *py;
    int       natgc, tlen;
    FILE      *fp;

    if ((fp = fopen(file, "r")) == 0) {
        fprintf(stderr, "%s: can't read %s\n", prog, file);
        exit(1);
    }
    tlen = natgc = 0;
    while (fgets(line, 1024, fp)) {
        if (*line == ';' || *line == '<' || *line == '>')
            continue;
        for (px = line; *px != '\n'; px++)
            if (isupper(*px) || islower(*px))
                tlen++;
    }
    if ((pseq = malloc((unsigned)(tlen+6))) == 0) {
        fprintf(stderr, "%s: malloc() failed to get %d bytes for %s\n", prog, tlen+6, file);
        exit(1);
    }
    pseq[0] = pseq[1] = pseq[2] = pseq[3] = '\0';
}

```

Page 1 of nwsnbr.c

TABLE 5 (cont.)

```

...getseq

py = pseq + 4;
*len = tlen;
rewind(fp);

while (fgets(line, 1024, fp)) {
    if (*line == ';' || *line == '<' || *line == '>')
        continue;
    for (px = line; *px != '\n'; px++) {
        if (isupper(*px))
            *py++ = *px;
        else if (islower(*px))
            *py++ = toupper(*px);
        if (index("ATGCU",*(py-1)))
            natgc++;
    }
}
*py++ = '\0';
*py = '\0';
(void) fclose(fp);
dna = natgc > (tlen/3);
return(pseq+4);
}

char      *
g_calloc(msg, nx, sz) g_calloc
{
    char      *msg;           /* program, calling routine */
    int       nx, sz;         /* number and size of elements */
{
    char      *px, *calloc();

    if ((px = calloc((unsigned)nx, (unsigned)sz)) == 0) {
        if (*msg) {
            fprintf(stderr, "%s: g_calloc() failed %s (n=%d, sz=%d)\n", prog, msg, nx, sz);
            exit(1);
        }
    }
    return(px);
}

/*
 * get final jmps from dx[] or tmp file, set pp[], reset dmax: main()
 */
readjmps() readjmps
{
    int          fd = -1;
    int          siz, i0, i1;
    register    i, j, xx;

    if (fj) {
        (void) fclose(fj);
        if ((fd = open(jname, O_RDONLY, 0)) < 0) {
            fprintf(stderr, "%s: can't open() %s\n", prog, jname);
            cleanup(1);
        }
    }
    for (i = i0 = i1 = 0, dmax0 = dmax, xx = len0; ; i++) {
        while (1) {
            for (j = dx[dmax].ijmp; j >= 0 && dx[dmax].jp.x[j] >= xx; j--)
;
    }
}

```

TABLE 5 (cont.)

```

...readjmps

if (j < 0 && dx[dmax].offset && fj) {
    (void) lseek(fd, dx[dmax].offset, 0);
    (void) read(fd, (char *)&dx[dmax].jp, sizeof(struct jmp));
    (void) read(fd, (char *)&dx[dmax].offset, sizeof(dx[dmax].offset));
    dx[dmax].jmp = MAXJMP-1;
}
else
    break;
}
if (i >= J MPS) {
    fprintf(stderr, "%s: too many gaps in alignment\n", prog);
    cleanup(1);
}
if (j >= 0) {
    siz = dx[dmax].jp.n[j];
    xx = dx[dmax].jp.x[j];
    dmax += siz;
    if (siz < 0) { /* gap in second seq */
        pp[1].n[i1] = -siz;
        xx += siz;
        /* id = xx - yy + len1 - 1
        */
        pp[1].x[i1] = xx - dmax + len1 - 1;
        gapy++;
        ngapy -= siz;
    }
    /* ignore MAXGAP when doing endgaps */
    siz = (-siz < MAXGAP || endgaps)? -siz : MAXGAP;
    i1++;
}
else if (siz > 0) { /* gap in first seq */
    pp[0].n[i0] = siz;
    pp[0].x[i0] = xx;
    gapx++;
    ngapx += siz;
}
/* ignore MAXGAP when doing endgaps */
siz = (siz < MAXGAP || endgaps)? siz : MAXGAP;
i0++;
}
}
else
    break;
}
/* reverse the order of jmps
 */
for (j = 0, i0--; j < i0; j++, i0--) {
    i = pp[0].n[j]; pp[0].n[j] = pp[0].n[i0]; pp[0].n[i0] = i;
    i = pp[0].x[j]; pp[0].x[j] = pp[0].x[i0]; pp[0].x[i0] = i;
}
for (j = 0, i1--; j < i1; j++, i1--) {
    i = pp[1].n[j]; pp[1].n[j] = pp[1].n[i1]; pp[1].n[i1] = i;
    i = pp[1].x[j]; pp[1].x[j] = pp[1].x[i1]; pp[1].x[i1] = i;
}
if (fd >= 0)
    (void) close(fd);
if (fj) {
    (void) unlink(jname);
    fj = 0;
    offset = 0;
}
}

```

Page 3 of nwsbr.c

TABLE 5 (cont.)

```

/*
 * write a filled jmp struct offset of the prev one (if any): nw()
 */
5   writejmps(ix)      writejmps
      int      ix;
{
10  char     *mktemp();
15  if(!fj) {
      if(mktemp(jname)<0) {
          fprintf(stderr, "%s: can't mktemp() %s\n", prog, jname);
          cleanup(1);
      }
      if((fj = fopen(jname, "w")) == 0) {
          fprintf(stderr, "%s: can't write %s\n", prog, jname);
          exit(1);
      }
20  (void) fwrite((char *)&dx[ix].jp, sizeof(struct jmp), 1, fj);
  (void) fwrite((char *)&dx[ix].offset, sizeof(dx[ix].offset), 1, fj);
}

```

TABLE 6

MARRSRHRLLLLLRYLIVVALGYHKAYGFSAPKDQQVVTAVEYQEAILACKTPKKT  
 VSS  
 RLEWKKLGRSVSFVYYQQTLQGDFKNRAEMIDFNIRIKNVTRSDAGKYRCEVSAPSEQG  
 5  
 QNLEEDTVTLEVLVAPAVPSCEVPSSALSGTVVELRCQDKEGNPAPEYTWFKD  
 GIRLLE  
 NPRLGSQSTNNSYTMNTKTGTLQFNTVSKLDTGEYSCEARN  
 SVGYRRC  
 PGKRMQVDDLN  
 ISGIIIAAVVVVALVISV  
 CGLGV  
 CYAQRK  
 GYFSKETSF  
 QKS  
 N  
 SSSKATTM  
 SENVQWLTPV  
 IPALWKA  
 AAGGS  
 RGQEF

## 10 N-glycosylation site at residues:

98-102  
 187-191  
 236-240  
 277-281

## 15 Casein kinase II phosphorylation site at residues:

39-43  
 59-63  
 100-104  
 20  
 149-153  
 205-209  
 284-288

## 20 N-myristoylation site at residues:

182-188  
 239-245  
 255-261  
 257-263  
 305-311

## 30 Amidation site at residues:

226-230

TABLE 7

5 MARRSAFPAAALWLWSILLCLLALRAEAGPPQEEESLYLWIDAHQARVLIGFEEDILIVS  
 EGKMAPFTHDFRKAQQRMPAIPVNIHSMNFTWQAAGQAELYFELSLRSLDKGIMADPT  
 VNVPLLGTVPHKASVVQVGFPCLGKQDGVAAFEVDVIVMNSEGNTILQTPQNAIFFKTC  
 QQAECPGGCRNGGFCNERRICECPDGFHGPCEKALCTPRCMNGGLCVTPGFCICPPGF  
 YGVNCDKANCSTTCFNGGTCFYPGKICPPGLEGEQCEISKCPQPCRNGKCIKGSKCK  
 CSKGYQGDLCSPVCEPGCAHGTCHEPNKCQCQEGWHGRHCNKRYEASLIHALRPAGA  
 QLRQHTPSLKKAEERRDPPESNYIW

10

N-glycosylation site at residues:

88-92

245-249

15

Casein kinase II phosphorylation site at residues:

319-323

Tyrosine kinase phosphorylation site at residues:

370-378

20

N-myristoylation site at residues:

184-190

185-191

189-195

25

315-321

ATP/GTP-binding site motif A (P-loop) at residues:

285-293

30

EGF-like domain cysteine pattern signature at residues:

198-210

230-242

262-274

294-306

35

326-338

TABLE 8

5 MGTKAQVERKLLCLFILAILLCSLALGSVTVHSSEPEVRIPENNPVKLSCAYSGFSSPR  
 VEWKFDQGDTTRLVCYNNKITASYEDRVTFLPTGITFKSVTREDTGYTCMVSEEGGNS  
 YGEVKVKLIVLVPPSKPTVNI  
 10 PSSATIGNRAVLT  
 CSEQDGSPPEYTWFKDGI  
 VMPTNP  
 KSTRAFSNSSYVLNPTTGE  
 LVDPLSASDTGEYSCEAR  
 NGYGT  
 15 PMTSNAVRMEAVERNV  
 GVIVAAVLVTLILLGILV  
 FGIWFAYSRGHFDRTKKGT  
 SSSKKV  
 20 IYSQPSARSEGEFKQTS  
 SFLV

10 N-glycosylation site at residues:  
 185-189

cAMP- and cGMP-dependent protein kinase phosphorylation site at residues:

15 270-274

Casein kinase II phosphorylation site at residues:

34-38

82-86

20 100-108  
 118-122  
 152-156  
 154-158  
 193-197  
 25 203-207  
 287-291

N-myristoylation site at residues:

30 105-111

116-122

158-164

219-225

237-243

256-262

TABLE 9

5 MLLWILLLETSLCFAAGNVTGVCCKEKCSCNEIEGDLHVDCEKKGFTSLQRFTAPTSQ  
 FYHLFLHGNSLTRLFPNEFANFYNASLHMENNGLHEIVPGAFLGLQLVKRLHINNNKI  
 KSFRKQTFLGLDDLEYLQAFNLLRIDPGAFQDLNKLEVILNDNLISTLPANVFQYV  
 PITHLDLRGNRLKTLPYEEVLEQIPGIAEILLEDPWDCTCDLLSLKEWLENIPKNALI  
 GRVVCEAPTRLQGKDLNETTEQDLCPLKNRVDSSLAPPQEEETFAPGPLPTPKTNGQ  
 EDHATPGSAPNGGKIPGNWQIKIRPTAAIATGSSRNKPLANSLPCPGGCSDHIPGSG  
 LKMNCCNNRNVSSLADLKPKLSNVQELFLRDNKIHSIRKSHFVDYKNLILDLGNNNIAT  
 10 VENNTFKNLLDLRWLYMDSNYLDTLSREKFAGLQNLEYLNVEYNNAIQQLILPGTFNAMPK  
 LRILILNNNNLRLSPVDFAGVSLSKLSLHNYYFMYLPVAGVLDQLTSIIQIDLHGNPW  
 ECSCTIVPKQWAERLGSEVLMSDLKCETPVNFFRKDFMLLSNDIECPOLYARISPLT  
 SHSKNSTGLAETGTHSNSYLDTSRVSISVLVPGLLLVFVTSAAFTVVGMLVILRNRKRS  
 KRRDANSSASEINSLQTVCDSSYWHNGPYNADGAHRVYDCGSHSLSD

15

## N-glycosylation site at residues:

18-22  
 253-257  
 363-367  
 20 416-420  
 595-599  
 655-659

20

## cAMP- and cGMP-dependent protein kinase phosphorylation site at residues:

122-126  
 646-650

25

## Casein kinase II phosphorylation site at residues:

30 30-34  
 180-184  
 222-226  
 256-260  
 366-370  
 35 573-577  
 608-612  
 657-661  
 666-670

693-697

N-myristoylation site at residues:

17-23

5 67-73

100-106

302-308

328-334

343-349

10 354-360

465-471

493-499

598-604

603-609

15 Prokaryotic membrane lipoprotein lipid attachment site at

residues:

337-348

TABLE 10

MVDVLLLFSLLCFLFHISRPDLSHNRLSFIKASSMSHLQSLREVKLNNELETIPNLGPV  
 SANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTAFPALQLKYLYLNSNR  
 VTSMEPGYFDNLANTLLVLKLNRRNIRASAPPKMFKLPQLQHLELNRNKKNVGGLTFQG  
 5 LGALKSLKMQRNGVTKLMDGAFWGLSNMEILQLDHNNLTEITKGWLYGLLMLQELHLSQ  
 NAINRISPDAWEFCQKLSELDLTFNHLSRLDDSSFLGLSLLNNTLHIGNNRVSYIADCAF  
 RGLSSLKTLDLKNEISWTIEDMNGAFSGLDKLRRLILQGNRIRSITKKAFTGLDALEH  
 LDLSDNAIMSLQGNAFSQMKKLQQLHLNTSSLLCDCQLKWLQPWVAENNQSFVNASC  
 HPQLLKGRSIFAVSPDGFVCDDFPKPQITVQPETQSAIKGSNLSFICSAASSSDSPMTF  
 10 AWKKDNELLHDAEMENYAHLRAQGGEVMEYTTILRLREVEFASEGKYQCVISNHFGSSY  
 SVKAKLTVNMLPSFTKTPMDLTIRAGAMARLECAAVGHPAPQIAWQKDGGTDFPAARER  
 RMHVMPEDDVFFIVDVKIEDIGVYCTAQNSAGSISANATLTVLETPSLRPLLDRVT  
 KGETAVLQCIAGGSPPPKNWTKDDSPLVTERHFFAAGNQLLIVDSDVSDAGKYTCE  
 MSNTLGTERGNVRLSVIPTPTCDSPQMTAPSLLDDGWATVGVIIAVVCCVVGTSLVWV  
 15 VIIYHTRRNEDCSITNTDETNLPADIPSYLSQGTTLADRQDGYSSESQSHHQFVTSS  
 GAGFFLPQHDSSGTCHIDNSSEADVEAATDLFLCPFLGSTGPMYLKGNVYGSDFETYH  
 TGCPDPRTVLMHYEPSYIKKKECYPCHPSEESCERSFSNISWPSHVRKLLNTSYSH  
 NEGPGMKNLCLNKSSLDFSANPEPASVASSNSFMGTFGKALRRPHLDAYSSFGQPSDCQ  
 PRAFYLKAHSSPDLSGSEEDGKERTDFQEENHICTFKQTLENYRTPNFQSYDLDT

20

## N-glycosylation site at residues:

62-66

96-100

214-218

25

382-386

409-413

455-459

628-632

669-673

30

845-849

927-931

939-943

956-960

35

## Glycosaminoglycan attachment site at residues:

826-830

## Casein kinase II phosphorylation site at residues:

17-21

39-43  
120-124  
203-207  
254-258  
5 264-268  
314-318  
323-327  
347-351  
464-468  
10 548-552  
632-636  
649-653  
671-675  
739-743  
15 783-787  
803-807  
847-851  
943-947  
958-962  
20 1013-1017  
1019-1023  
1021-1025

Tyrosine kinase phosphorylation site at residues:

25 607-615

N-myristoylation site at residues:

30 179-185  
197-203  
320-326  
367-373  
453-459  
528-534  
612-618  
35 623-629  
714-720  
873-879

TABLE 11

5 MLNKMTLHPQQIMIGPRFNRALFDPLLVLALQLLVVAGLVRAQTCPSVCSCSNQFSK  
 VICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIG  
 AFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRL  
 DLGEKLRLSYISEGAFEGLSNLRYLNLCAMCNLREIPNLTLPLIKLDELDLSGNHLSAIRP  
 GSFQGLMHLQKLWMIQSQIQVIERNAFDNLQLSVEINLAHNNLTLLPHDLFTPPLHHLER  
 IHLHHNPWCNCNDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAP  
 10 VIVEPPADLNVTGMAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLFN  
 TNVTQDTGMYTCMVSNVGNTTASATLNVTAATTPFSYFSTVTVETMEPSQDEARTT  
 DNNVGPTPVWDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIG  
 CFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDEITGDTPMESHLPMPAI  
 EHEHLNHYNSYKSPFNHTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI

## 15 N-glycosylation site at residues:

278-282

364-368

390-394

412-416

## 20 415-419

434-438

442-446

488-492

606-610

## 25 cAMP- and cGMP-dependent protein kinase phosphorylation site at residues:

183-187

## 30 Casein kinase II phosphorylation site at residues:

268-272

417-421

465-469

579-583

## 35 620-624

## N-myristoylation site at residues:

40-46

73-79

118-124

191-197

228-234

237-243

5

391-397

422-428

433-439

531-537

TABLE 12

5 MSAPSLRARAAGLGLLLCAVLGRAGRSDSGGRGELGQPSGVAEERPCPTTCRCLGDLLD  
 CSRKRLARLPEPLPSWVARLDLSHNRLSFIKASSMSHLQLSLREVKLNNELETIPNLGP  
 VSANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTAFFPALQLKYLYLNSN  
 RVTSMEPGYFDNLANTLLVLKLNRRNRIASIPPKMFKLPQLQHLELRNRIKNDGLTFQ  
 GLGALKSLKMQRNGVTKLMGAFWGLSNMEILQLDHNNLTEITKGWLYGLLMLQELHLS  
 QNAINRISPDAWEFCQKLSELDLTFNHLRSRLDDSSFLGLSLLNTRHIGNNRVSYIADCA  
 FRGLSSLKTLSDLKNNEISWTIEDMNGAFSGLDKLRRLILQGNRIRSIKKAAFTGLDALE  
 10 HLDLSDNAIMSLQGNAFSQMKKLQQLHLNTSSLCDCQLKWLQPWVAENNQSFVNASC  
 AHPQLLKGRSIFAVSPDGFVCDDFPKPQITVQPETQSAIKGSNLSFICSAASSSDSPMT  
 FAWKKDNELLHDAEMENYAHLRAQGGEVMEYTTILRLREVEFASEGKYQCVISNHFGSS  
 YSVKAKLTVNMLPSFTKTPMDLTIRAGAMARLECAAVGHPPAQIAWQKDGGTDFPAARE  
 RRMHVMPEDDVFFIVDVKIEDIGVYSCTAQNSAGSISANATLTVLETPSFLRPLLRTV  
 15 TKGETAVLQCIAGGSPPPKNWTKDDSPLVTERHFFAAGNQLLIVDSDVSDAGKYTC  
 EMSNTLGTERGNVRLSIVPTPTCDSPQMTAPSLLDDGWATGVGVVIIAVVCCVGTSLVW  
 VVIIYHTRRRNEDCSITNTDETNLPADIPSYLSQGTIADRQDGYSSESQSHHQFVTS  
 SGAGFFLPQHDSSGTCHIDNSSEADVEAATDLFLCPFLGSTGPMYLKGNVYGSDFETY  
 HTGCSPDPRTVLMHDHYEPSYIKKKECYPCHPSEESCERSFSNISWPSHVRKLLNTSYS  
 20 HNEGPGMKNLCLNKSSLDFSANPEPASVASSNSFMGTFGKALRRPHLDAYSSFGQPSDC  
 QPRAFYLKAHSSPDLDGSEEDGKERTDFQEEHICTFKQTLENYRTPNFQSYDLDT

## N-glycosylation site at residues:

25 122-126  
 156-160  
 274-278  
 442-446  
 469-473  
 515-519  
 30 688-692  
 729-733  
 905-909  
 987-991  
 999-1003  
 35 1016-1020

## Glycosaminoglycan attachment site at residues:

886-890

## Casein kinase II phosphorylation site at residues:

99-103  
180-184  
5 263-267  
314-318  
324-328  
374-378  
383-387  
10 407-411  
524-528  
608-612  
692-696  
709-713  
15 731-735  
799-803  
843-847  
863-867  
907-911  
20 1003-1007  
1018-1022  
1073-1077  
1079-1083  
1081-1085  
25

## Tyrosine kinase phosphorylation site at residue:

667-675

## N-myristoylation site at residues:

30 14-20  
36-42  
239-245  
257-263  
380-386  
35 427-433  
513-519  
588-594  
672-678  
683-689

774-780

933-999

Leucine zipper pattern at residues:

5

58-80

65-87

What is claimed:

1. A composition, comprising a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, agonist or fragment thereof and a carrier or excipient, useful for:
  - (a) increasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) stimulating or enhancing an immune response in a mammal in need thereof, or
  - (c) increasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
2. Use of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, agonist or a fragment thereof to prepare a composition useful for:
  - (a) increasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) stimulating or enhancing an immune response in a mammal in need thereof, or
  - (c) increasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
3. A composition, comprising a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, antagonist or a fragment thereof and a carrier or excipient, useful for:
  - (a) decreasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) inhibiting or reducing an immune response in a mammal in need thereof, or
  - (c) decreasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
4. Use of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, antagonist or a fragment thereof to prepare a composition useful for:
  - (a) decreasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) inhibiting or reducing an immune response in a mammal in need thereof, or
  - (c) decreasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
5. A method of treating an immune related disorder, such as a T cell mediated

disorder, in a mammal in need thereof, comprising administering to the mammal an effective amount of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, an agonist antibody thereof, an antagonist antibody thereto, or a fragment thereof.

6. The method of claim 5, wherein the disorder is selected from systemic lupus erythematosis, rheumatoid arthritis, juvenile chronic arthritis, spondyloarthropathies, systemic sclerosis (scleroderma), idiopathic inflammatory myopathies (dermatomyositis, polymyositis), Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia (immune pancytopenia, paroxysmal nocturnal hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic purpura, immune-mediated thrombocytopenia), thyroiditis (Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic thyroiditis), diabetes mellitus, immune-mediated renal disease (glomerulonephritis, tubulointerstitial nephritis), demyelinating diseases of the central and peripheral nervous systems such as multiple sclerosis, idiopathic demyelinating polyneuropathy or Guillain-Barré syndrome, and chronic inflammatory demyelinating polyneuropathy, hepatobiliary diseases such as infectious hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune chronic active hepatitis, primary biliary cirrhosis, granulomatous hepatitis, and sclerosing cholangitis, inflammatory and fibrotic lung diseases such as inflammatory bowel disease (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and Whipple's disease, autoimmune or immune-mediated skin diseases including bullous skin diseases, erythema multiforme and contact dermatitis, psoriasis, allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity and urticaria, immunologic diseases of the lung such as eosinophilic pneumonias, idiopathic pulmonary fibrosis and hypersensitivity pneumonitis, transplantation associated diseases including graft rejection and graft -versus-host-disease.

7. The composition or use of any of the preceding claims, wherein the antibody is a monoclonal antibody.

8. The composition or use of any of the preceding claims, wherein the antibody is an antibody fragment or a single-chain antibody.

9. The composition or use of any of the preceding claims, wherein the antibody has nonhuman complementarity determining region (CDR) residues and human framework region (FR) residues.

10. A method for determining the presence of a PRO245, PRO217, PRO301,

PRO266, PRO335, PRO331 or PRO326 polypeptide, comprising exposing a cell suspected of containing the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide to an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody and determining binding of the antibody to the cell.

11. A method of diagnosing an immune related disease in a mammal, comprising detecting the level of expression of a gene encoding a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (a) in a test sample of tissue cells obtained from the mammal, and (b) in a control sample of known normal tissue cells of the same cell type, wherein a higher expression level in the test sample indicates the presence of immune related disease in the mammal from which the test tissue cells were obtained.

12. A method of diagnosing an immune related disease in a mammal, comprising (a) contacting an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody with a test sample of tissue cells obtained from the mammal, and (b) detecting the formation of a complex between the antibody and the polypeptide in the test sample.

13. An immune related disease diagnostic kit, comprising an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody or fragment thereof and a carrier in suitable packaging.

14. The kit of claim 13, further comprising instructions for using the antibody to detect a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide.

15. An article of manufacture, comprising:  
a container;  
a label on the container; and  
a composition comprising an active agent contained within the container; wherein the composition is effective for stimulating or enhancing an immune response in a mammal, the label on the container indicates that the composition can be used for treating an immune related disease, and the active agent in the composition is an agent inhibiting the expression and/or activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide.

16. The article of manufacture of claim 21 wherein said active agent is an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody.

17. A method for identifying a compound capable of inhibiting the expression or activity of a PRO245 polypeptide, comprising contacting a candidate compound with a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide under conditions and for a time sufficient to allow these two components to interact.

18. The method of claim 17, wherein the candidate compound or the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide is immobilized on a solid support.

19. The method of claim 18, wherein the non-immobilized component carries a detectable label.

FIGURE 1A

PRO                            XXXXXXXXXXXXXXXXXX                    (Length = 15 amino acids)  
Comparison Protein            XXXXXYYYYYYYY                    (Length = 12 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 15 = 33.3%

FIGURE 1B

PRO  
Comparison Protein

XXXXXXXXXX (Length = 10 amino acids)  
XXXXXXYYYYYYZZYZ (Length = 15 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 10 = 50%

FIGURE 1C

PRO-DNA  
Comparison DNA

NNNNNNNNNNNNNN  
NNNNNNL L L L L L L

(Length = 14 nucleotides)  
(Length = 16 nucleotides)

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

6 divided by 14 = 42.9%

FIGURE 1D

PRO-DNA  
Comparison DNA

NNNNNNNNNNNN  
NNNNLLLVV

(Length = 12 nucleotides)  
(Length = 9 nucleotides)

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

4 divided by 12 = 33.3%

**FIGURE 2A**

FIGURE 2B

```

/*
 */
#include <stdio.h>
#include <ctype.h>

#define MAXJMP 16      /* max jumps in a diag */
#define MAXGAP 24      /* don't continue to penalize gaps larger than this */
#define JMPS 1024      /* max jmps in an path */
#define MX 4           /* save if there's at least MX-1 bases since last jmp */

#define DMAT 3          /* value of matching bases */
#define DMIS 0          /* penalty for mismatched bases */
#define DINS0 8          /* penalty for a gap */
#define DINS1 1          /* penalty per base */
#define PINS0 8          /* penalty for a gap */
#define PINS1 4          /* penalty per residue */

struct jmp {
    short n[MAXJMP];      /* size of jmp (neg for delay) */
    unsigned short x[MAXJMP]; /* base no. of jmp in seq x */
};

struct diag {
    int score;           /* score at last jmp */
    long offset;         /* offset of prev block */
    short ijmp;          /* current jmp index */
    struct jmp *jp;      /* list of jmps */
};

struct path {
    int spc;             /* number of leading spaces */
    short n[JMPS]; /* size of jmp (gap) */
    int x[JMPS]; /* loc of jmp (last elem before gap) */
};

char *ofile;           /* output file name */
char *namex[2];        /* seq names: getseqs() */
char *prog;             /* prog name for err msgs */
char *seqx[2];          /* seqs: getseqs() */
int dmax;              /* best diag: nw() */
int dmax0;             /* final diag */
int dna;                /* set if dna: main() */
int endgaps;           /* set if penalizing end gaps */
int gapx, gapy;        /* total gaps in seqs */
int len0, len1;         /* seq lens */
int ngapx, ngapy;       /* total size of gaps */
int smax;               /* max score: nw() */
int *xbm;                /* bitmap for matching */
long offset;             /* current offset in jmp file */
struct diag *dx;        /* holds diagonals */
struct path pp[2];       /* holds path for seqs */

char *calloc(), *malloc(), *index(), *strcpy();
char *getseq(), *g_calloc();

```

FIGURE 2C

```

/* Needleman-Wunsch alignment program
*
* usage: progs file1 file2
* where file1 and file2 are two dna or two protein sequences.
* The sequences can be in upper- or lower-case and may contain ambiguity
* Any lines beginning with ';' or '<' are ignored
* Max file length is 65535 (limited by unsigned short x in the jmp struct)
* A sequence with 1/3 or more of its elements ACGTU is assumed to be DNA
* Output is in the file "align.out"
*
* The program may create a tmp file in /tmp to hold info about traceback.
* Original version developed under BSD 4.3 on a vax 8650
*/
#include "nw.h"
#include "day.h"

static _dbval[26] = {
    1,14,2,13,0,0,4,11,0,0,12,0,3,15,0,0,0,5,6,8,8,7,9,0,10,0
};

static _pbval[26] = {
    1, 2|(1<<('D'-'A'))|(1<<('N'-'A')), 4, 8, 16, 32, 64,
    128, 256, 0xFF, 1<<10, 1<<11, 1<<12, 1<<13, 1<<14,
    1<<15, 1<<16, 1<<17, 1<<18, 1<<19, 1<<20, 1<<21, 1<<22,
    1<<23, 1<<24, 1<<25|(1<<('E'-'A'))|(1<<('Q'-'A'))
};

main(ac, av)      main
int      ac;
char     *av[];
{
    prog = av[0];
    if (ac != 3) {
        fprintf(stderr, "usage: %s file1 file2\n", prog);
        fprintf(stderr, "where file1 and file2 are two dna or two protein sequences.\n");
        fprintf(stderr, "The sequences can be in upper- or lower-case\n");
        fprintf(stderr, "Any lines beginning with ';' or '<' are ignored\n");
        fprintf(stderr, "Output is in the file \"align.out\"\n");
        exit(1);
    }
    namex[0] = av[1];
    namex[1] = av[2];
    seqx[0] = getseq(namex[0], &len0);
    seqx[1] = getseq(namex[1], &len1);
    xbm = (dna)?_dbval:_pbval;

    endgaps = 0;           /* 1 to penalize endgaps */
    ofile = "align.out";   /* output file */

    nw();                  /* fill in the matrix, get the possible jmps */
    readjmps();            /* get the actual jmps */
    print();               /* print stats, alignment */

    cleanup(0);            /* unlink any tmp files */
}

```

FIGURE 2D

```

/* do the alignment, return best score: main()
 * dna: values in Fitch and Smith, PNAS, 80, 1382-1386, 1983
 * pro: PAM 250 values
 * When scores are equal, we prefer mismatches to any gap, prefer
 * a new gap to extending an ongoing gap, and prefer a gap in seqx
 * to a gap in seq y.
 */
nw()    nw
{
    char      *px, *py;          /* seqs and ptrs */
    int       *ndely, *dely;     /* keep track of dely */
    int       ndelx, delx;      /* keep track of delx */
    int       *tmp;             /* for swapping row0, row1 */
    int       mis;              /* score for each type */
    int       ins0, ins1; /* insertion penalties */
    register  id;              /* diagonal index */
    register  ij;              /* jmp index */
    register  *col0, *col1;     /* score for curr, last row */
    register  xx, yy;          /* index into seqs */

    dx = (struct diag *)g_malloc("to get diag", len0+len1+1, sizeof(struct diag));

    ndely = (int *)g_malloc("to get ndely", len1+1, sizeof(int));
    dely = (int *)g_malloc("to get dely", len1+1, sizeof(int));
    col0 = (int *)g_malloc("to get col0", len1+1, sizeof(int));
    col1 = (int *)g_malloc("to get col1", len1+1, sizeof(int));
    ins0 = (dna)? DINS0 : PINS0;
    ins1 = (dna)? DINS1 : PINS1;

    smax = -10000;
    if (endgaps) {
        for (col0[0] = dely[0] = -ins0, yy = 1; yy <= len1; yy++) {
            col0[yy] = dely[yy] = col0[yy-1] - ins1;
            ndely[yy] = yy;
        }
        col0[0] = 0;           /* Waterman Bull Math Biol 84 */
    }
    else
        for (yy = 1; yy <= len1; yy++)
            dely[yy] = -ins0;

    /* fill in match matrix
     */
    for (px = seqx[0], xx = 1; xx <= len0; px++, xx++) {
        /* initialize first entry in col
         */
        if (endgaps) {
            if (xx == 1)
                col1[0] = delx = -(ins0+ins1);
            else
                col1[0] = delx = col0[0] - ins1;
            ndelx = xx;
        }
        else {
            col1[0] = 0;
            delx = -ins0;
            ndelx = 0;
        }
    }
}

```

FIGURE 2E

...nw

```

for (py = seqx[1], yy = 1; yy <= len1; py++, yy++) {
    mis = col0[yy-1];
    if (dna)
        mis += (xbm[*px-'A']&xbm[*py-'A'])? DMAT : DMIS;
    else
        mis += _day[*px-'A'][*py-'A'];

    /* update penalty for del in x seq;
     * favor new del over ongong del
     * ignore MAXGAP if weighting endgaps
     */
    if (endgaps || ndely[yy] < MAXGAP) {
        if (col0[yy] - ins0 >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else {
            dely[yy] = ins1;
            ndely[yy]++;
        }
    } else {
        if (col0[yy] - (ins0+ins1) >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else
            ndely[yy]++;
    }

    /* update penalty for del in y seq;
     * favor new del over ongong del
     */
    if (endgaps || ndelx < MAXGAP) {
        if (col1[yy-1] - ins0 >= delx) {
            delx = col1[yy-1] - (ins0+ins1);
            ndelx = 1;
        } else {
            delx = ins1;
            ndelx++;
        }
    } else {
        if (col1[yy-1] - (ins0+ins1) >= delx) {
            delx = col1[yy-1] - (ins0+ins1);
            ndelx = 1;
        } else
            ndelx++;
    }

    /* pick the maximum score; we're favoring
     * mis over any del and delx over dely
     */
}

```

FIGURE 2F

...nw

```

id = xx - yy + len1 - 1;
if (mis >= delx && mis >= dely[yy])
    col1[yy] = mis;
else if (delx >= dely[yy]) {
    col1[yy] = delx;
    ij = dx[id].ijmp;
    if (dx[id].jp.n[0] && (!dna || (ndelx >= MAXJMP
    && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
        dx[id].ijmp++;
        if (++ij >= MAXJMP) {
            writejmps(id);
            ij = dx[id].ijmp = 0;
            dx[id].offset = offset;
            offset += sizeof(struct jmp) + sizeof(offset);
        }
    }
    dx[id].jp.n[ij] = ndelx;
    dx[id].jp.x[ij] = xx;
    dx[id].score = delx;
}
else {
    col1[yy] = dely[yy];
    ij = dx[id].ijmp;
    if (dx[id].jp.n[0] && (!dna || (ndely[yy] >= MAXJMP
    && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
        dx[id].ijmp++;
        if (++ij >= MAXJMP) {
            writejmps(id);
            ij = dx[id].ijmp = 0;
            dx[id].offset = offset;
            offset += sizeof(struct jmp) + sizeof(offset);
        }
    }
    dx[id].jp.n[ij] = -ndely[yy];
    dx[id].jp.x[ij] = xx;
    dx[id].score = dely[yy];
}
if (xx == len0 && yy < len1) {
    /* last col
     */
    if (endgaps)
        col1[yy] = ins0+ins1*(len1-yy);
    if (col1[yy] > smax) {
        smax = col1[yy];
        dmax = id;
    }
}
if (endgaps && xx < len0)
    col1[yy-1] = ins0+ins1*(len0-xx);
if (col1[yy-1] > smax) {
    smax = col1[yy-1];
    dmax = id;
}
tmp = col0; col0 = col1; col1 = tmp;
}
(void) free((char *)ndely);
(void) free((char *)dely);
(void) free((char *)col0);
(void) free((char *)col1); }

```

FIGURE 2G

```

/*
*
* print() -- only routine visible outside this module
*
* static:
* getmat() -- trace back best path, count matches: print()
* pr_align() -- print alignment of described in array pl[]: print()
* dumpblock() -- dump a block of lines with numbers, stars: pr_align()
* nums() -- put out a number line: dumpblock()
* putline() -- put out a line (name, [num], seq, [num]): dumpblock()
* stars() -- put a line of stars: dumpblock()
* stripname() -- strip any path and prefix from a seqname
*/
#include "nw.h"

#define SPC      3
#define P_LINE   256      /* maximum output line */
#define P_SPC    3        /* space between name or num and seq */

extern  _day[26][26];
int    olen;           /* set output line length */
FILE   *fx;            /* output file */

print()  print
{
    int     lx, ly, firstgap, lastgap;      /* overlap */

    if ((fx = fopen(ofile, "w")) == 0) {
        fprintf(stderr, "%s: can't write %s\n", prog, ofile);
        cleanup(1);
    }
    fprintf(fx, "<first sequence: %s (length = %d)\n", namex[0], len0);
    fprintf(fx, "<second sequence: %s (length = %d)\n", namex[1], len1);
    olen = 60;
    lx = len0;
    ly = len1;
    firstgap = lastgap = 0;
    if (dmax < len1 - 1) { /* leading gap in x */
        pp[0].spc = firstgap = len1 - dmax - 1;
        ly -= pp[0].spc;
    }
    else if (dmax > len1 - 1) { /* leading gap in y */
        pp[1].spc = firstgap = dmax - (len1 - 1);
        lx -= pp[1].spc;
    }
    if (dmax0 < len0 - 1) { /* trailing gap in x */
        lastgap = len0 - dmax0 - 1;
        lx -= lastgap;
    }
    else if (dmax0 > len0 - 1) { /* trailing gap in y */
        lastgap = dmax0 - (len0 - 1);
        ly -= lastgap;
    }
    getmat(lx, ly, firstgap, lastgap);
    pr_align();
}

```

FIGURE 2H

```

/*
 * trace back the best path, count matches
 */
static
getmat(lx, ly, firstgap, lastgap)  getmat
    int      lx, ly;          /* "core" (minus endgaps) */
    int      firstgap, lastgap; /* leading/trailing overlap */
{
    int      nm, i0, i1, siz0, siz1;
    char     outx[32];
    double   pct;
    register int n0, n1;
    register char *p0, *p1;

    /* get total matches, score
     */
    i0 = i1 = siz0 = siz1 = 0;
    p0 = seqx[0] + pp[1].spc;
    p1 = seqx[1] + pp[0].spc;
    n0 = pp[1].spc + 1;
    n1 = pp[0].spc + 1;

    nm = 0;
    while ( *p0 && *p1 ) {
        if (siz0) {
            p1++;
            n1++;
            siz0--;
        }
        else if (siz1) {
            p0++;
            n0++;
            siz1--;
        }
        else {
            if (xbm[*p0-'A']&xbm[*p1-'A'])
                nm++;
            if (n0++ == pp[0].x[i0])
                siz0 = pp[0].n[i0++];
            if (n1++ == pp[1].x[i1])
                siz1 = pp[1].n[i1++];
            p0++;
            p1++;
        }
    }
    /* pct homology:
     * if penalizing endgaps, base is the shorter seq
     * else, knock off overhangs and take shorter core
     */
    if (endgaps)
        lx = (len0 < len1)? len0 : len1;
    else
        lx = (lx < ly)? lx : ly;
    pct = 100.*(double)nm/(double)lx;
    sprintf(fx, "\n");
    sprintf(fx, "<%d match%s in an overlap of %d: %.2f percent similarity\n",
            nm, (nm == 1)? "" : "es", lx, pct);
}

```

FIGURE 2I

```

sprintf(fx, "<gaps in first sequence: %d", gapx);      ...getmat
if (gapx) {
    (void) sprintf(outx, " (%d %s%s)",
                  ngapx, (dna)? "base": "residue", (ngapx == 1)? ":" : "s");
    fprintf(fx, "%s", outx);
}

sprintf(fx, ", gaps in second sequence: %d", gapy);
if (gapy) {
    (void) sprintf(outx, " (%d %s%s)",
                  ngapy, (dna)? "base": "residue", (ngapy == 1)? ":" : "s");
    fprintf(fx, "%s", outx);
}
if (dna)
    fprintf(fx,
            "\n<score: %d (match = %d, mismatch = %d, gap penalty = %d + %d per base)\n",
            smax, DMAT, DMIS, DINS0, DINS1);
else
    fprintf(fx,
            "\n<score: %d (Dayhoff PAM 250 matrix, gap penalty = %d + %d per residue)\n",
            smax, PINS0, PINS1);
if (endgaps)
    fprintf(fx,
            "<endgaps penalized. left endgap: %d %s%s, right endgap: %d %s%s\n",
            firstgap, (dna)? "base": "residue", (firstgap == 1)? ":" : "s",
            lastgap, (dna)? "base": "residue", (lastgap == 1)? ":" : "s");
else
    fprintf(fx, "<endgaps not penalized\n");
}

static          nm;          /* matches in core -- for checking */
static          lmax;        /* lengths of stripped file names */
static          ij[2];        /* jmp index for a path */
static          nc[2];        /* number at start of current line */
static          ni[2];        /* current elem number -- for gapping */
static          siz[2];
static char    *ps[2];        /* ptr to current element */
static char    *po[2];        /* ptr to next output char slot */
static char    out[2][P_LINE]; /* output line */
static char    star[P_LINE]; /* set by stars() */

/*
 * print alignment of described in struct path pp[]
 */
static
pr_align() pr_align
{
    int          nn;          /* char count */
    int          more;
    register    i;
    for (i = 0, lmax = 0; i < 2; i++) {
        nn = stripname(namex[i]);
        if (nn > lmax)
            lmax = nn;

        nc[i] = 1;
        ni[i] = 1;
        siz[i] = ij[i] = 0;
        ps[i] = seqx[i];
        po[i] = out[i];
    }
}

```

FIGURE 2J

```

for (nn = nm = 0, more = 1; more; ) {      ...pr_align
    for (i = more = 0; i < 2; i++) {
        /*
         * do we have more of this sequence?
         */
        if (!*ps[i])
            continue;

        more++;

        if (pp[i].spc) {           /* leading space */
            *po[i]++ = ' ';
            pp[i].spc--;
        }
        else if (siz[i]) {        /* in a gap */
            *po[i]++ = '_';
            siz[i]--;
        }
        else {                   /* we're putting a seq element
            */
            *po[i] = *ps[i];
            if (islower(*ps[i]))
                *ps[i] = toupper(*ps[i]);
            po[i]++;
            ps[i]++;
            /*
             * are we at next gap for this seq?
             */
            if (ni[i] == pp[i].x[ij[i]]) {
                /*
                 * we need to merge all gaps
                 * at this location
                 */
                siz[i] = pp[i].n[ij[i]++];
                while (ni[i] == pp[i].x[ij[i]]) {
                    siz[i] += pp[i].n[ij[i]++];
                }
                ni[i]++;
            }
        }
    }
    if (++nn == olen || !more && nn) {
        dumpblock();
        for (i = 0; i < 2; i++)
            po[i] = out[i];
        nn = 0;
    }
}
/*
 * dump a block of lines, including numbers, stars: pr_align()
 */
static
dumpblock()      dumpblock
{
    register i;
    for (i = 0; i < 2; i++)
        *po[i]-- = '\0';
}

```

FIGURE 2K

```

...dumpblock

(void) putc('\n', fx);
for (i = 0; i < 2; i++) {
    if (*out[i] && (*out[i] != ' ' || *(po[i]) != ' ')) {
        if (i == 0)
            nums(i);
        if (i == 0 && *out[1])
            stars();
        putline(i);
        if (i == 0 && *out[1])
            fprintf(fx, star);
        if (i == 1)
            nums(i);
    }
}

/*
 * put out a number line: dumpblock()
 */
static
nums(ix)  nums
int       ix;      /* index in out[] holding seq line */
{
    char          nline[P_LINE];
    register      i, j;
    register char *pn, *px, *py;

    for (pn = nline, i = 0; i < lmax+P_SPC; i++, pn++)
        *pn = ' ';
    for (i = nc[ix], py = out[ix], *py; py++; pn++) {
        if (*py == ' ' || *py == '-')
            *pn = ' ';
        else {
            if (i%10 == 0 || (i == 1 && nc[ix] != 1)) {
                j = (i < 0)? -i : i;
                for (px = pn; j; j /= 10, px--)
                    *px = j%10 + '0';
                if (i < 0)
                    *px = ' ';
            }
            else
                *pn = ' ';
            i++;
        }
    }
    *pn = '\0';
    nc[ix] = i;
    for (pn = nline; *pn; pn++)
        (void) putc(*pn, fx);
    (void) putc('\n', fx);
}

/*
 * put out a line (name, [num], seq, {num}): dumpblock()
 */
static
putline(ix) putline
int       ix;
{

```

FIGURE 2L

```

...putline

int          i;
register char *px;

for (px = namex[ix], i = 0; *px && *px != ':'; px++, i++)
    (void) putc(*px, fx);
for (; i < lmax+P_SPC; i++)
    (void) putc(' ', fx);

/* these count from 1:
 * ni[] is current element (from 1)
 * nc[] is number at start of current line
 */
for (px = out[ix]; *px; px++)
    (void) putc(*px&0x7F, fx);
(void) putc('\n', fx);
}

/*
 * put a line of stars (seqs always in out[0], out[1]): dumpblock()
 */
static
stars() stars
{
    int          i;
    register char *p0, *p1, cx, *px;

    if (!*out[0] || (*out[0] == ' ' && *(po[0]) == ' ') ||
        !*out[1] || (*out[1] == ' ' && *(po[1]) == ' '))
        return;
    px = star;
    for (i = lmax+P_SPC; i; i--)
        *px++ = ' ';

    for (p0 = out[0], p1 = out[1]; *p0 && *p1; p0++, p1++) {
        if (isalpha(*p0) && isalpha(*p1)) {
            if (xbm[*p0-'A']&xbm[*p1-'A']) {
                cx = '*';
                nm++;
            }
            else if (!dma && _day[*p0-'A'][*p1-'A'] > 0)
                cx = ':';
            else
                cx = ' ';
        }
        else
            cx = ' ';
        *px++ = cx;
    }
    *px++ = '\n';
    *px = '\0';
}

```

FIGURE 2M

```
/*
 * strip path or prefix from pn, return len: pr_align()
 */
static
stripname(pn)      stripname
    char    *pn;      /* file name (may be path) */
{
    register char    *px, *py;

    py = 0;
    for (px = pn; *px; px++)
        if (*px == '/')
            py = px + 1;
    if (py)
        (void) strcpy(pn, py);
    return(strlen(pn));
}
```

FIGURE 2N

```

/*
 * cleanup() -- cleanup any tmp file
 * getseq() -- read in seq, set dna, len, maxlen
 * g_calloc() -- calloc() with error checkin
 * readjmps() -- get the good jmps, from tmp file if necessary
 * writejmps() -- write a filled array of jmps to a tmp file: nw()
 */
#include "nw.h"
#include <sys/file.h>

char      *jname = "/tmp/homgXXXXXX";           /* tmp file for jmps */
FILE      *fj;

int       cleanup();                         /* cleanup tmp file */
long      lseek();

/*
 * remove any tmp file if we blow
 */
cleanup(i) cleanup
{
    int      i;
{
    if (fj)
        (void) unlink(jname);
    exit(i);
}

/*
 * read, return ptr to seq, set dna, len, maxlen
 * skip lines starting with ':', '<', or '>'
 * seq in upper or lower case
 */
char      *
getseq(file, len)      getseq
{
    char      *file;      /* file name */
    int       *len;       /* seq len */
{
    char      line[1024], *pseq;
    register char      *px, *py;
    int       natgc, tlen;
    FILE     *fp;

    if ((fp = fopen(file, "r")) == 0) {
        fprintf(stderr, "%s: can't read %s\n", prog, file);
        exit(1);
    }
    tlen = natgc = 0;
    while (fgets(line, 1024, fp)) {
        if (*line == ':' || *line == '<' || *line == '>')
            continue;
        for (px = line; *px != '\n'; px++)
            if (isupper(*px) || islower(*px))
                tlen++;
    }
    if ((pseq = malloc((unsigned)(tlen+6))) == 0) {
        fprintf(stderr, "%s: malloc() failed to get %d bytes for %s\n", prog, tlen+6, file);
        exit(1);
    }
    pseq[0] = pseq[1] = pseq[2] = pseq[3] = '\0';
}

```

FIGURE 2O

...getseq

```

py = pseq + 4;
*tlen = tlen;
rewind(fp);

while (fgets(line, 1024, fp)) {
    if (*line == ';' || *line == '<' || *line == '>')
        continue;
    for (px = line; *px != '\n'; px++) {
        if (isupper(*px))
            *py++ = *px;
        else if (islower(*px))
            *py++ = toupper(*px);
        if (index("ATGCU", *(py-1)))
            natgc++;
    }
    *py++ = '\0';
    *py = '\0';
    (void) fclose(fp);
    dna = natgc > (tlen/3);
    return(pseq+4);
}

char      *
g_calloc(msg, nx, sz) g_calloc
{
    char      *msg;           /* program, calling routine */
    int       nx, sz;         /* number and size of elements */
{
    char      *px, *calloc();

    if ((px = calloc((unsigned)nx, (unsigned)sz)) == 0) {
        if (*msg) {
            fprintf(stderr, "%s: g_calloc() failed %s (n=%d, sz=%d)\n", prog, msg, nx, sz);
            exit(1);
        }
    }
    return(px);
}

/*
 * get final jmps from dx[] or tmp file, set pp[], reset dmax: main()
 */
readjmps() readjmps
{
    int           fd = -1;
    int           siz, i0, i1;
    register int i, j, xx;

    if (fj) {
        (void) fclose(fj);
        if ((fd = open(jname, O_RDONLY, 0)) < 0) {
            fprintf(stderr, "%s: can't open() %s\n", prog, jname);
            cleanup(1);
        }
    }
    for (i = i0 = i1 = 0, dmax0 = dmax, xx = len0; ; i++) {
        while (1) {
            for (j = dx[dmax].jmp; j >= 0 && dx[dmax].jp.x[j] >= xx; j--)

```

FIGURE 2P

```

...readjmps

if (j < 0 && dx[dmax].offset && fj) {
    (void) lseek(fd, dx[dmax].offset, 0);
    (void) read(fd, (char *)&dx[dmax].jp, sizeof(struct jmp));
    (void) read(fd, (char *)&dx[dmax].offset, sizeof(dx[dmax].offset));
    dx[dmax].ijmp = MAXJMP-1;
}
else
    break;
}
if (i >= JMPS) {
    fprintf(stderr, "%s: too many gaps in alignment\n", prog);
    cleanup(1);
}
if (j >= 0) {
    siz = dx[dmax].jp.n[j];
    xx = dx[dmax].jp.x[j];
    dmax += siz;
    if (siz < 0) { /* gap in second seq */
        pp[1].n[i1] = -siz;
        xx += siz;
        /* id = xx - yy + len1 - 1
        */
        pp[1].x[i1] = xx - dmax + len1 - 1;
        gapy++;
        ngapy -= siz;
    }
    /* ignore MAXGAP when doing endgaps */
    siz = (-siz < MAXGAP || endgaps)? -siz : MAXGAP;
    i1++;
}
else if (siz > 0) { /* gap in first seq */
    pp[0].n[i0] = siz;
    pp[0].x[i0] = xx;
    gapx++;
    ngapx += siz;
}
/* ignore MAXGAP when doing endgaps */
siz = (siz < MAXGAP || endgaps)? siz : MAXGAP;
i0++;
}
else
    break;
}
/* reverse the order of jmps
 */
for (j = 0, i0--; j < i0; j++, i0--) {
    i = pp[0].n[j]; pp[0].n[j] = pp[0].n[i0]; pp[0].n[i0] = i;
    i = pp[0].x[j]; pp[0].x[j] = pp[0].x[i0]; pp[0].x[i0] = i;
}
for (j = 0, i1--; j < i1; j++, i1--) {
    i = pp[1].n[j]; pp[1].n[j] = pp[1].n[i1]; pp[1].n[i1] = i;
    i = pp[1].x[j]; pp[1].x[j] = pp[1].x[i1]; pp[1].x[i1] = i;
}
if (fd >= 0)
    (void) close(fd);
if (fj) {
    (void) unlink(jname);
    fj = 0;
    offset = 0;
}
}

```

FIGURE 2Q

```
/*
 * write a filled jmp struct offset of the prev one (if any): nw()
 */
writejmps(ix)      writejmps
int      ix;
{
    char    *mktemp();

    if (!fj) {
        if (mktemp(jname) < 0) {
            fprintf(stderr, "%s: can't mktemp() %s\n", prog, jname);
            cleanup(1);
        }
        if ((fj = fopen(jname, "w")) == 0) {
            fprintf(stderr, "%s: can't write %s\n", prog, jname);
            exit(1);
        }
    }
    (void) fwrite((char *)&dx[ix].jp, sizeof(struct jmp), 1, fj);
    (void) fwrite((char *)&dx[ix].offset, sizeof(dx[ix].offset), 1, fj);
}
```

Figure 3

CCAGAGTTCAAGGGCCCCGGCCTCCTGCGCTCCTGCCGCCGGACCCCTGACCTCCT  
CAGAGCAGCCGGCTGCCGCCGGAGATGGCGAGGAGGCCACGCCCTCCTCCT  
GCTGCTGCTGCGCTACCTGGTGGCTGCCCTGGCTATCATAAGGCCTATGGGTTTCTGC  
CCAAAAGACCAACAAGTAGTCACAGCAGTAGAGTACCAAGAGGCTATTTAGCCTGCAA  
AACCCAAAAGAAGACTGTTCTCCAGATTAGAGTGGAGAAACTGGTGGAGTGTCTC  
CTTGCTACTATCAACAGACTCTCAAGGTGATTTAAAATCGAGCTGAGATGATAGA  
TTCAATATCCGGATAAAAATGTGACAAGAAGTGTGAGCTGAGTGGAACTGTGGT  
TAGTGGCCCATCTGAGCAAGGCCAAACCTGGAAGAGGATACTGACTCTGGAAAGTATT  
AGTGGCTCCAGCAGTCCATCATGTGAAGTACCCCTTCTGCTCTGAGTGGAACTGTGGT  
AGAGCTACGATGTCAAGACAAAGAAGGGAACTCCAGCTCCTGAATAACACATGGTTAAGGA  
TGGCATCCGTTGCTAGAAAATCCCAGACTTGGCTCCAAAGCACCAACAGCTCATACAC  
AATGAATACAAAAACTGGAACCTGCAATTAAACTGTTCCAAACTGGACACTGGAGA  
ATATTCCGTGAAGCCCGAATTCTGTTGGATATCGCAGGTGCTGGAAACGAATGCA  
AGTAGATGATCTCAACATAAGTGGCATCATAGCAGCCGTAGTAGTTGTGGCCTAGTGAT  
TTCCGTTGTTGGCCTGGTGTATGCTATGCTCAGAGGAAAGGCTACTTTCAAAAGAAAC  
CTCCTTCCAGAAGAGTAATTCTCATCTAAAGCCACGACAATGAGTGAAATGTGAGTG  
GCTCACGCCGTAACTCCAGCACTTGGAAAGGCCGCCGGCGGATCACGAGGTCAAGGA  
GTTCTAGACAGTCTGCCAATATGGTGAACCCCCATCTCTACTAAAATACAAAAATTAG  
CTGGGCATGGTGGCATGTGCCTGCAGTCCAGCTGCTGGAGACAGGAGAATCACTTGA  
ACCCGGGAGGCAGGTTGCAGTGAGCTGAGATCACGCCACTGCAGTCCAGCCTGGTAA  
CAGAGCAAGATTCCATCTAAAAAATAAAATAAAATAAAATAACTGGTTTAC  
TGTAGAATTCTTACAATAATAGCTTGATATTG

Figure 4

MARRSRHRLLLLLLRYLVVALGYHKAYGFSAPKDDQQVVTAVEYQEAILACKTPKKTVSS  
RLEWKKLGRSVSFVYYQQTLQGDFKNRAEMIDFNIRIKNVTRSDAGKYRCEVSAPSEQG  
QNLEEDTVTLEVLVAPAVPSCEVPSSALSGTVVELRCQDKEGNPAPEYTWFKD GIRLLE  
NPRLGSQSTNSSYT MNTKTGTLQFNTVSKLDTGEYSCEARN SVGYRRCPGKRMQVDDLN  
ISGIIIAVVVVALVISVCGLGV CYAQRKGYFSKETSFQKSNSSSKATTMSENVQWLTPV  
IPALWKAAGGSRGQEF

## N-glycosylation site at residues:

98-102  
187-191  
236-240  
277-281

## Casein kinase II phosphorylation site at residues:

39-43  
59-63  
100-104  
149-153  
205-209  
284-288

## N-myristoylation site at residues:

182-188  
239-245  
255-261  
257-263  
305-311

## Amidation site at residues:

226-230

Figure 5

CCAGGCCGGGAGGCGACGCGCCCAGCCGTCTAACGGGAACAGCCCTGGCTGAGGGAGCT  
GCAGCGCAGCAGAGTATCTGACGGGCCAGGTTGCGTAGGTGCGGACGAGGAGTTTCC  
CGGCAGCGAGGAGGTCTGAGCAGCATGGCCCGAGGAGCGCCTCCCTGCCGCCGCGCT  
CTGGCTCTGGAGCATCCTCCTGTGCCGACTGCCGGCGGAGGCCGGGCCGCA  
GGAGGAGAGCCTGTACCTATGGATGCTCACCAAGGAAGAGTACTCATAGGATTGAA  
AGAAAGATATCCTGATTGTTTCAAGAGGGAAAATGGCACCTTACACATGATTTCAGAAA  
AGCGCAACAGAGAATGCCAGCTATTCTGTCAATATCCATTCCATGAATTTCACCTGGCA  
AGCTGCAGGGCAGGCAGAATACTCTATGAATTCTGCTCCCTGCCCTGGATAAAGG  
CATCATGGCAGATCCAACCGTCAATGTCCTCTGCTGGAAACAGTGCCTCACAGGCATC  
AGTTGTTCAAGTTGGTTCCATGTCTGGAAAACAGGATGGGTGGCAGCATTGAAGT  
GGATGTGATTGTTATGAATTCTGAAGGCAACACCATTCTCAAACACCTCAAATGCTAT  
CTTCTTAAAACATGTCAACAAGCTGAGTGCCCAGCGGGTGCCGAAATGGAGGTTTG  
TAATGAAAGACGCATCTGCGAGTGTCTGATGGGTCCACGGACCTCACTGTGAGAAAGC  
CCTTGACCCACGATGTATGAATGGTGGACTTGTGACTCCTGGTTCTGCATCTG  
CCACCTGGATTCTATGGAGTGAACGTGACAAAGCAAACGTCTCAACCACCTGCTTAA  
TGGAGGGACCTGTTCTACCTGGAAAATGTATTGCCCTCAGGACTAGAGGGAGAGCA  
GTGTGAAATCAGCAAATGCCACAACCTGTCGAAATGGAGGTAAATGCATTGGTAAAAG  
CAAATGTAAGTGTCCAAAGGTTACCAAGGGAGACCTCTGTTCAAAGCCTGCTGCGAGC  
TGGCTGTGGTCACATGGAACCTGCCATGAACCCAAACAAATGCCAATGTCAAGAAGGTTG  
GCATGGAAGACACTGCAATAAAGGTACGAAGGCCAGCCTCATACTGCCCTGAGGCCAGC  
AGGCGCCCAGCTCAGGCAGCACACGCCCTCACTTAAAGGCCAGGGAGCGGGATCC  
ACCTGAATCCAATTACATCTGGTAACTCCGACATCTGAAACGTTAAGTTACACCAAG  
TTCATAGCCTTGTAACTTTCATGTGTTGAATGTTCAAATAATGTTCATTACACTTAA  
GAATACTGCCCTGAATTTTATTAGCTTCAATTATAATCACTGAGCTGATATTACTCTC  
CTTTAAGTTTCTAAGTACGTCTGCTAGCATGATGGTATAGATTTCTGTTCACTG  
TTGGGACAGATTTATATTGTCAATTGATCAGGTTAAATTTCACTGTTGAGTTGGC  
AGATATTTCAAATTACAATGCATTGATGGTCTGGGGCAGGGAAACATCAGAAAGG  
TTAAATTGGCAAAATGCGTAAGTCACAAGAATTGGATGGTGCAGTTAATGTTGAAGT  
TACAGCATTTCAGATTTATTGTCACTGAGTATTAGATGTTGTTACATTAAATTG  
TCTTAATTAAACTCTCAATAACATATTTGACCTTACCAATTCCAGAGATTCA  
GTATTAAAAAAAAAAATTACACTGTGGTAGTGGCATTAAACAATATAATATTCTA  
AACACAATGAAATAGGAATATAATGTGAACCTTGCATTGGCTGAAGCAATATAA  
TATATTGTAAACAAACACAGCTTACCTAATAACATTAACTGTTGTTATGTATA  
AAATAAAGGTGCTGTTAGTTGGAAAAA

Figure 6

MARRSAFPAAALWLWSILLCLLALRAEAGPPQEEESLYLWIDAHQARVLIGFEEDILIVS  
EGKMAPFTHDFRKAQQRMPAIPVNIHSMNFTWQAAGQAEYFYEFLSLRSLDKGIMADPT  
VNVPPLLGTVPHKASVVQVGFPCLGKQDGVAAFEVDVIVMNSEGNTILQTPQNAIFFKTC  
QQAECPGGCRNGGFCNERRICECPDGFHGPCEKALCTPRCMNGGLCVTPGFCICPPGF  
YGVNCDKANCSTTCFNNGTCFYPGKICCPPGLEGEQCEISKCPQPCRNGGKIGKSKCK  
CSKGYQGDLCSPVCEPGCAHGTCHEPNKCQCQEGWHGRHCNKRYEASLIHALRPAGA  
QLRQHTPSLKKAEERRDPPESNYIW

N-glycosylation site at residues:

88-92

245-249

Casein kinase II phosphorylation site at residues:

319-323

Tyrosine kinase phosphorylation site at residues:

370-378

N-myristoylation site at residues:

184-190

185-191

189-195

315-321

ATP/GTP-binding site motif A (P-loop) at residues:

285-293

EGF-like domain cysteine pattern signature at residues:

198-210

230-242

262-274

294-306

326-338

Figure 7

GTCTGTTCCCAGGAGTCCTCGCGGCTGTTGTGTCAGTGGCTGATCGCGATGGGGACA  
AAGGCGCAAGTCGAGAGGAAACTGTTGTGCCTCTCATATTGGCGATCCTGTTGTGCTCC  
CTGGCATTGGCAGTGTACAGTCACCTCTGAACCTGAAGTCAGAATTCTGAGAAT  
AATCCTGTGAAGTTGTCTGTGCCTACTCGGGCTTCTCTCCCGTGTGGAGTGGAAAG  
TTTGACCAAGGAGACACCACAGACTCGTTGCTATAATAACAAGATCACAGCTTCTAT  
GAGGACCGGGTGACCTCTTGCAACTGGTATCACCTCAAGTCCGTGACACGGGAAGAC  
ACTGGGACATACACTGTATGGTCTGTGAGGAAGGCGAACAGCAGCTATGGGGAGGTCAAG  
GTCAAGCTCATCGTGCCTGTGCCTCCATCCAAGCCTACAGTTAACATCCCCTCTGCC  
ACCATTGGGAACCGGGCAGTGTGACATGCTCAGAACAAAGATGGTCCCCACCTCTGAA  
TACACCTGGTTCAAAGATGGGATAGTGATGCCTACGAATCCAAAAGCACCCGTGCCTTC  
AGCAACTCTCCTATGTCCTGAATCCCACAACAGGAGAGCTGGTCTTGATCCCCTGTCA  
GCCTCTGATACTGGAGAAATACAGCTGTGAGGCACGGAATGGGTATGGGACACCCATGACT  
TCAAATGCTGTGCGATGGAAGCTGTGGAGCGGAATGTGGGGTCATCGTGGCAGCCGTC  
CTTGTAACCTGATTCTCCTGGGAATCTGGTTTTGGCATCTGGTTGCCTATAGCCGA  
GGCCACTTGACAGAACAAAGAAAGGGACTTCGAGTAAGAAGGTGATTTACAGCCAGCCT  
AGTGCCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCACTCCTGGTGTGAGCCTGGTC  
GCTCACCGCCTATCATCTGCATTGCCTTACTCAGGTGCTACCGGACTCTGGCCCTGAT  
GTCTGTAGTTTCACAGGATGCCTTATTGTCTTCTACACCCCCACAGGGCCCTACTTCT  
TCGGATGTGTTTTAATAATGTCAGCTATGTGCCCATCCTCCTTCATGCCCTCCCTCCC  
TTTCCTACCACTGCTGAGTGGCTGGAACTTGTTAAAGTGTATTCCCCATTCTTTG  
AGGGATCAGGAAGGAATCCTGGGTATGCCATTGACTTCCCTTAAGTAGACAGCAAAA  
TGGCGGGGGTCGCAGGAATCTGCACTCAACTGCCACCTGGCTGGCAGGGATTTGAAT  
AGGTATCTTGAGCTGGTTCTGGCTCTTCCCTGTGACTGACGACCAGGGCAGCTGT  
TCTAGAGCGGAATTAGAGGCTAGAGCGGCTGAAATGGTTGGTGTGACACTGGGG  
TCCTTCCATCTCTGGGCCCCACTCTCTGTCTTCCCATGGGAAGTGCCACTGGGATCC  
CTCTGCCCTGTCCCTGAATACAAGCTGACTGACATTGACTGTGTGAAAGGAAATGGG  
AGCTCTTGTGAGGAGACATAGTAAATTTCAGAGAACTTGAAGCCAAAAGGATTTAA  
ACCGCTGCTCTAAAGAAAAGAAAATGGAGGCTGGGCGCAGTGGCTCACGCCGTAAATCC  
CAGAGGCTGAGGCAGGCGGATCACCTGAGGTGGAGTTGGGATCAGCCTGACCAACAT  
GGAGAAACCCCTACTGAAATACAAAGTTAGCCAGGCATGGTGGTGCATGCCGTAGTCCC  
AGCTGCTCAGGAGCCTGGCAACAAGAGCAAAACTCCAGCTCAAAAAAAAAAAAAAA

Figure 8

MGTKAQVERKLLCLFILAILLCSLALGSVTVHSSEPEVRIPENNPVKLSCAYSGFSSPR  
VEWKFDQGDTTRLVCYNNKITASYEDRVTFLPTGITFKSVTREDTGTYTCMVSEEGGNS  
YGEVKVKLIVLVPPSKPTVNIPSSATIGNRAVLTCSEQDGSPPEYTWFKDGIVMPTNP  
KSTRAFSNSSYVLPNTTGEELVFDPLSASDTGEYSCEARNGYGTPMTSNAVRMEAVERN  
GVIVAAVLVTLILLGILVFGIWFAYSRGHFDRTKKGTSSKKVIYSQPSARSEGEFKQTS  
SFLV

## N-glycosylation site at residues:

185-189

## cAMP- and cGMP-dependent protein kinase phosphorylation site at residues:

270-274

## Casein kinase II phosphorylation site at residues:

34-38

82-86

100-108

118-122

152-156

154-158

193-197

203-207

287-291

## N-myristoylation site at residues:

105-111

116-122

158-164

219-225

237-243

256-262

Figure 9

GGGGGTTAGGGAGGAAGGAATCCACCCCCACCCCCCAAACCCCTTCTCCTTCCCT  
GGCTTCGGACATTGGAGCACTAAATGAACCTGAATTGTGTCTGTGGCGAGCAGGATGGTC  
GCTGTTACTTGTGATGAGATCGGGATGAATTGCTCGCTTAAAATGCTGCTTGGAT  
TCTGTTGCTGGAGACGTCTTTGCGCTGAAACGTTACAGGGGACGTTGCAA  
AGAGAAGATCTGTTCTGCAATGAGATAGAAGGGGACCTACACGTAGACTGTGAAAAAAA  
GGGCTTCACAAGTCTGCAGCGTTCACTGCCCGACTTCCAGTTTACCAATTATTC  
GCATGGCAATTCCCTCACTGACTTTCCCTAATGAGTCGCTAACCTTATAATGCGGT  
TAGTTGACATGGAAAACAATGGCTTGATGAAATCGTCCGGGGCTTCTGGGCT  
GCAGCTGGTAAAAGGCTGCACATCAACAACAAGATCAAGTCTTCGAAAGCAGAC  
TTTCTGGGCTGGACGATCTGGAATATCTCAGGCTGATTTAAATTATTACGAGATAT  
AGACCCGGGGCCTTCCAGGACTTGAACAAGCTGGAGGTGCTCATTTAAATGACAATCT  
CATCAGCACCCCTACCTGCCAACGTGTTCCAGTATGTGCCCATACCCACCTGACCTCCG  
GGGTAACAGGCTGAAAACGCTGCCCTATGAGGAGGTCTGGAGCAAATCCCTGGTATTGC  
GGAGATCTGCTAGAGGATAACCCCTGGGACTGCACCTGTGATCTGCTCCCTGAAAGA  
ATGGCTGGAAAACATTCCAAGAATGCCCTGATCGGCCAGTGGTCTGCGAAGCCCCAC  
CAGACTGCAGGGTAAAGACCTCAATGAAACCACCGAACAGGACTTGTGCTTTGAAAAAA  
CCGAGTGGATTCTAGTCTCCGGCCTGCCAAGAAGAGACCTTGCTCTGGACC  
CCTGCCAACCTTTCAAGACAAATGGCAAGAGGATCATGCCAACACCAGGGCTGCTCC  
AAACGGAGGTACAAAGATCCCAGGCACTGGCAGATCAAATCAGACCCACAGCAGCGAT  
AGCGACGGTAGCTCAGGAACAAACCTTAGCTAACAGTTACCTGCCCTGGGCTG  
CAGCTGCGACCACATCCCAGGGCTGGTTAAAGATGAACCTGCAACAACAGGAACGTGAG  
CAGCTGGCTATTGAAGCCAAGCTCTAACGTGCAAGGAGCTTCTACGAGATAA  
CAAGATCCACAGCATCCGAAAATCGACTTGTGGATTACAAGAACCTCATTCTGTTGGA  
TCTGGCAACAATAACATCGCTACTGTAGAGAACAAACACTTTCAAGAACCTTGGACCT  
CAGGTGGCTACATGGATAGCAATTACCTGGACACGCTGTCCCCGGAGAAATTGCGGG  
GCTGCCAACCTAGAGTACCTGAACGTGGAGTACAACGCTATCCAGCTCATCCTCCGG  
CACTTTCAATGCCATGCCAAACTGAGGATCCTCATTCTAACAAACACCTGCTGAGGTC  
CCTGCCGTGGACGTGGCTGGGTCTCGCTCTCTAAACTCAGCCTGCACAACAATT  
CTTCATGTACCTCCGGTGGCAGGGTGCTGGACCAGTTAACCTCCATCATCCAGATAGA  
CCTCCACGGAAACCCCTGGAGTGCTCCTGCACAATTGTGCCCTTCAAGCAGTGGCAGA  
ACGCTTGGGTTCCGAAGTGCTGATGAGCAGCTAACGTGAGACGCCGGTAACCTCTT  
TAGAAAGGATTCATGCTCCTCTCAATGACGAGATCTGCCCTCAGCTGTACGCTAGGAT  
CTCGCCACGTTAACCTCGCACAGTAAAACAGCACTGGGTTGGCGAGACCGGGACGCA  
CTCCAACCTCCTACCTAGACACCAGCAGGGTGTCCATCTGGTGTGGTCCGGACTGCT  
GCTGGTGTGTCACCTCCGCCCTCACCGTGGTGGCATGCTCGTGTGTTATCCTGAGGAA  
CCGAAAGCGGTCCAAGAGACGAGATGCCAACTCCTCCCGTCCGAGATTAATTCCCTACA

GACAGTCTGTGACTCTTCTACTGGCACAATGGGCCTTACAACGCAGATGGGCCACAG  
AGTGTATGACTGTGGCTCTCACTCGCTCTCAGACTAAGACCCAAACCCAATAGGGGAGG  
GCAGAGGAAGGCATAACATCCTCCCCACCGCAGGCACCCGGGGCTGGAGGGCGTG  
TACCCAAATCCCCGCCATCAGCCTGGATGGCATAAGTAGATAAAACTGTGAGCTC  
GCACAACCGAAAGGGCTGACCCCTTACTTAGCTCCCTCCTGAAACAAAGAGCAGACTG  
TGGAGAGCTGGGAGAGCGCAGCCAGCTCGCTTGCTGAGAGCCCTTGACAGAAAG  
CCCAGCACGACCCCTGCTGGAAGAACTGACAGTGCCCTGCCCTGGCCCCGGGCCTGTG  
GGGTTGGATGCCCGGTTCTATACATATACATATCCACATCTATAGAGAGATAG  
ATATCTATTCTCCCTGTGGATTAGCCCCGTGATGGCTCCCTGTTGGCTACCGCAGGGAT  
GGGCAGTTGCACGAAGGCATGAATGTATTGTAAGTAACCTTGACTTCTGAC

Figure 10

MLLWILLLETSLCFAAGNVTGVCCKICSCNEIEGDLHVDCEKKGFTSLQRFTA  
FYHLFLHGNSLTRLFPNEFANFYNAVSLHMENNGLHEIVPGAFLGLQLVKRLHINNNKI  
KSFRKQTFLGLDDLEYLQADFNLLRDIDPGAFQDLNKLEVILNDNLISTLPANVFQYV  
PITHLDLRLGNRLKTL PYEEVLEQIPGIAEILLEDNPWDCTCDLLSLKEWLENIPKNALI  
GRVVCEAPTRLQGKDLNETTEQDLCPLKNRVDSSLAPPAPAQEETFAPGPLPTPFKTNGQ  
EDHATPGSAPNGGKTPGWNQIKIRPTAAIATGSSRNKPLANSLPCPGGCSDHIPGSG  
LKMNCNNRNVSSLADLKPKLSNVQELFLRDNKIHSIRKSHFVDYKNLILDLGNNNIAT  
VENNTFKNLLDLRWLYMDSNYLDTLSREKFAGLQNLEYLNVEYNQIQLILPGTFNAMPK  
LRILILNNNLLRSLPVDFAGVSLSKLSLHNMYFMYLPVAGVLDQLTSIIQIDLHGNPW  
ECSCTIVPFKQWAERLGSEVLMSDLKCETPVNFFRKDFMLLSNDEICPQLYARISPTLT  
SHSKNSTGLAETGTHSNSYLDTSRVSISVLVPGLL VFVTS AFTVVGMLVFI L RNRKRS  
KRRDANSSASEINSLQTVCDSSYWHNGPYNADGAHRVYDCGSHSLSD

## N-glycosylation site at residues:

18-22

253-257

363-367

416-420

595-599

655-659

## cAMP- and cGMP-dependent protein kinase phosphorylation site at residues:

122-126

646-650

## Casein kinase II phosphorylation site at residues:

30-34

180-184

222-226

256-260

366-370

573-577

608-612

657-661

666-670

693-697

## N-myristoylation site at residues:

17-23

67-73

100-106

302-308

328-334

343-349

354-360

465-471

493-499

598-604

603-609

## Prokaryotic membrane lipoprotein lipid attachment site at residues:

337-348

Figure 11

GTAACGTAGTCAGGCTTTCATTTGGGAAGCCCCCTAACAGAACCGTCATTCTCCA  
AGTTATGGTGGACGTACTCTGTTCTCCCTCTGCTTGCTTTCACATTAGCAGACC  
GGACTTAAGTCACAACAGATTATCTTCATCAAGGCAAGTTCATGAGCCACCTCAAAG  
CCTTCGAGAAGTGAAGACTGAACAACAATGAATTGGAGACCATTCAAATCTGGGACCAGT  
CTCGGCAAATATTACACTTCTCTCCCTGGCTGGAAACAGGATTGTTGAAATACTCCCTGA  
ACATCTGAAAGAGTTCACTCCCTGAAACTTGGACCTTAGCAGCAACAATATTCAGA  
GCTCCAAACTGCATTCAGCCCTACAGCTAAATATCTGTATCTAACAGCAACCGAGT  
CACATCAATGGAACCTGGGTATTTGACAATTGGCAACACACTCCTGTGTTAAAGCT  
GAACAGGAACCGAATCTCAGCTATCCCACCCAAGATGTTAAACTGCCCAACTGCAACA  
TCTCGAATTGAAACCGAAACAAGATTAAAGATGTTAGATGGACTGACATTCCAAGGCCTTGG  
TGCTCTGAAGTCTCTGAAAATGCAAAGAAATGGAGTAACGAAACTTATGGATGGAGCTTT  
TTGGGGGCTGAGCAACATGAAATTGCACTGGCAGCTGGACCATAACACCTAACAGAGATTAC  
CAAAGGCTGGCTTACGGCTTGCTGATGCTGCAGGAACCTCATCTCAGCAAATGCCAT  
CAACAGGATCAGCCCTGATGCCCTGGAGTTCTGCCAGAAGCTCAGTGAGCTGGACCTAAC  
TTTCAATCACTTATCAAGGTTAGATGATTCAAGCTCCTGGCTAACGTTACTAAATAC  
ACTGCACATTGGAAACAACAGAGTCAGCTACATTGCTGATTGTGCCCTCCGGGGCTTTC  
CAGTTAAAGACTTGGATCTGAAGAACATGAAATTCCCTGGACTATTGAAGACATGAA  
TGGTGCTTCTCTGGCTTGACAAACTGAGGCGACTGATACTCCAAGGAAATGGATCCG  
TTCTATTACTAAAAAGCCTTCAGGTTGGATGCTGGAGCATCTAGACCTGAGTGA  
CAACGCAATCATGTCTTACAAGGCAATGCATTTCAAAATGAAGAAACTGCAACAATT  
GCATTTAAATACATCAAGCCTTGTGCGATTGCCAGCTAAATGGCTCCACAGTGGGT  
GGCGAAAACAACCTCAGAGCTTGTAAATGCCAGTTGTGCCAGCTAAAGCTGCTAA  
AGGAAGAACATTTGCTGTTAGCCCAGATGGCTTGTGTGATGATTTCCCAAACC  
CCAGATCACGGTCAGCCAGAAACACAGTCGGCAATAAAAGGTTCAATTGAGTTCT  
CTGCTCAGCTGCCAGCAGCAGTGATTCCCCAATGACTTTGCTTGAAAAAGACAATGA  
ACTACTGCATGATGCTGAAATGGAAAATTATGCAACACCTCCGGGCCAAGGTGGCGAGGT  
GATGGAGTATACCAACATCCTCGGCTGCCAGGGTGAATTGCCAGTGAGGGAAATA  
TCAGTGTGTCATCTCAATCACTTGGTTACCTACTCTGTCAAAGCCAAGCTTACAGT  
AAATATGCTTCCCTCATTCAACAGACCCCCATGGATCTCACCATCCGAGCTGGGCCAT  
GGCACGCTGGAGTGTGCTGCTGTGGGCACCCAGCCCCCAGATAGCCTGGCAGAAGGA  
TGGGGGCACAGACTTCCAGCTGCACGGAGAGACGCATGCATGTGATGCCAGGATGA  
CGTGTCTTATCGTGGATGTGAAGATAGAGGACATTGGGTATACAGCTGCACAGCTCA  
GAACAGTGCAGGAAGTATTCAGCAAATGCAACTCTGACTGTCCTAGAAACACCATCATT  
TTTGCAGGCCACTGTTGGACCGAAGTGAACCAAGGGAGAAACAGCCGCTACAGTGCAT  
TGCTGGAGGAAGCCCTCCCCCTAAACTGAACTGGACCAAAGATGATAGCCCATTGGTGGT  
AACCGAGAGGCACCTTTGCAGCAGGCAATCAGCTCTGATTATTGTGGACTCAGATGT

CAGTGATGCTGGAAATACACATGTGAGATGTCTAACACCCCTGGCACTGAGAGAGGAAA  
CGTGCCTCAGTGTGATCCCCACTCCAACCTGCGACTCCCCTCAGATGACAGCCCCATC  
GTTAGACGATGACGGATGGGCCACTGTGGGTGTCGTGATCATAGCCGTGGTTGCTGTGT  
GGTGGGCACGTCACTCGTGTGGGTGGTCATCATATACCACACAAGGCGGAGGAATGAAGA  
TTGCAGCATTACCAACACAGATGAGACCAACTGCCAGCAGATATCCTAGTTATTTGTC  
ATCTCAGGAAACGTTAGCTGACAGGCAGGATGGGTACGTGTCAGAAAGTGGAAAGCCA  
CCACCAGTTGTCACATCTCAGGTGCTGGATTTTCTTACCAACACATGACAGTAGTGG  
GACCTGCCATATTGACAATAGCAGTGAAGCTGATGTGGAAAGCTGCCACAGATCTGTCCT  
TTGTCGTTTGGATCCACAGGCCATGTATTGAAAGGGAAATGTGTATGGCTCAGA  
TCCTTTGAAACATATCATAACAGGTTGCAGTCCTGACCCAAGAACAGTTTAAATGGACCA  
CTATGAGCCCAGTTACATAAAGAAAAGGAGTGCACCCATGTTCTCATCCTCAGAAGA  
ATCCTGCGAACGGAGTTCACTAATATCGTGGCCTCACATGTGAGGAAGCTACTTAA  
CACTAGTTACTCTACAATGAAGGACCTGGAATGAAAAATCTGTGTCCTAAACAAGTCCTC  
TTTAGATTTTAGTGCAAATCCAGAGCCAGCGTCGGTGCCTCGAGTAATTCTTCATGGG  
TACCTTGAAAGCTCTCAGGAGACCTCACCTAGATGCCATTCAAGCTTGGACAGCC  
ATCAGATTGTCAGCCAAGAGCCTTTATTGAAAGCTCATTCTCCCCAGACTTGGACTC  
TGGGTCAAGAGGAAGATGGAAAGAAAGGACAGATTTCAGGAAGAAAATCACATTGTAC  
CTTTAAACAGACTTTAGAAAACCTACAGGACTCCAAATTTCAGTCTTATGACTTGGACAC  
**ATAGACTGAATGAGACCAAGGAAAAGCTTAACATACTACCTCAAGTGAAC**TTTATT  
AAAGAGAGAAATCTTATGTTTTAAATGGAGTTATGAATTAAAAGGATAAAAATGC  
TTTATTATACAGATGAACAAATTACAAAAAGTTATGAAATTTTATACTGGGAATG  
ATGCTCATATAAGAATACCTTTAAACTATTTTAACTTTGTTTATGCAAAAAGTA  
TCTTACGTAATTAAATGATATAATCATGATTATTTATGTTTATAATGCCAGATT  
TCTTTTATGGAAAATGAGTTACTAAAGCATTAAATAATACCTGCCCTGTACCAATT  
TTAAATAGAAGTTACTTCATTATTTGCACATTATTTAATAAAAATGTGTCATTG  
AA

Figure 12

MVDVLLLFSLCLLFHISRPDLSHNRLSFIKASSMSHLQSLREVKLNNELETIPNLGPV  
 SANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTAFPALQLKYLYLNSNR  
 VTSMEPGYFDNLANTLLVLKLNRRNISAIPPKMFKLPQLQHLELNRNKIKNVDGLTFQG  
 LGALKSLKMQRNGVTKLMGAFWGLSNMEILQLDHNNLTEITKGWLYGLLMLQELHLSQ  
 NAINRISPAWEFCQKLSELDLTFNHLRLDDSSFLGLSLLNTLHIGNNRVSYIADCAF  
 RGLSSLKTLSDLKNNEISWTIEDMNGAFSGLDKLRRRLILQGNRIRSITKKAFTGLDALEH  
 LDLSDNAIMSLQGNAFSQMKKLQQLHLNTSSLCDCQLKWLPOWVAENNQSFVNASC  
 HPQLLKGRSIFAVSPDGFVCDDFPKPQITVQPETQSAIKGSNLSFICSAASSSDSPMTF  
 AWKKDNELLHDAEMENYAHLRAQGGEVMEYTTILRLREVEFASEGKYQCVISNHFGSSY  
 SVKAKLTNVNMLPSFTKTPMDLTIRAGAMARLECAAVGHPAPQIAWQKDGGTDFPAARER  
 RMHVMPEDDVFFIVDVKIEDIGVYSCTAQNSAGSISANATLTVLETPSFLRPLLDRTVT  
 KGETAVLQCIAGGSPPPKNWTKDDSPVVTERHFFAAGNQLLIIVDSDVSDAGKYTCE  
 MSNTLGTERGNVRLSIPPTCDSPQMTAPSLLDDGWATGVVIIAVVCCVVGTSLVVV  
 VIIYHTRRRNEDCSITNTDETNLPADIPSYLSSQGTLADRQDGYSSESGSHHQFVTSS  
 GAGFFLPQHDSSGTCHIDNSSEADVEAATDLFLCPFLGSTGPMYLKGNVYGSDFETYH  
 TGCSPDPRTVLMHDYEPSYIKKKECYPCHPSEESCERSFSNISWPSHVRKLLNTSYSH  
 NEGPGMKNLCLNKSSLDFSANPEPASVASSNSFMGTFGKALRRPHLDAYSSFGQPSDCQ  
 PRAFYLKAHSSPDLDGSEEDGKERTDFQEENHICTFKQTLENYRTPNFQSYDLDT

## N-glycosylation site at residues:

62-66

96-100

214-218

382-386

409-413

455-459

628-632

669-673

845-849

927-931

939-943

956-960

## Glycosaminoglycan attachment site at residues:

826-830

## Casein kinase II phosphorylation site at residues:

17-21  
39-43  
120-124  
203-207  
254-258  
264-268  
314-318  
323-327  
347-351  
464-468  
548-552  
632-636  
649-653  
671-675  
739-743  
783-787  
803-807  
847-851  
943-947  
958-962  
1013-1017  
1019-1023  
1021-1025

## Tyrosine kinase phosphorylation site at residues:

607-615

## N-myristoylation site at residues:

179-185  
197-203  
320-326  
367-373  
453-459  
528-534  
612-618  
623-629  
714-720  
873-879

Figure 13

GGGGAGAGGAATTGACCATGTAAAAGGAGACTTTTTGGTGGTGGCTGTTGGG  
TGCCTTGCAAAATGAAGGATGCAGGACGCAGCTTCTCCTGGAACCGAACGCAATGGAT  
AAACTGATTGTGCAAGAGAGAAGGAAGAACGAAGCTTTCTTGTGAGCCCTGGATCTTA  
ACACAAATGTGTATATGTGCACACAGGGAGCATTCAAGAATGAAATAAACCAAGAGTTAGA  
CCCGCGGGGGTTGGTGTGTTCTGACATAAATAATCTAAAGCAGCTGTTCCCCTCC  
CCACCCCCAAAAAAAAGGATGATTGAAATGAAGAACCGAGGATTACAAAGAAAAAGT  
ATGTTCATTTCTCTATAAAGGAGAAAGTGAGCCAAGGAGATATTTGGAATGAAAG  
TTTGGGCTTTTAGTAAAGTAAAGAACTGGTGTGGTGTGTTCTTCTTTGAA  
TTTCCCACAAGAGGAGAGGAAATAATAACATCTGCAAAGAAATTTCAGAGAAGAAA  
GTTGACCGCGGCAGATTGAGGCATTGATTGGGGAGAGAACCCAGCAGAGCACAGTTGGA  
TTTGTGCCTATGTTGACTAAATTGACGGATAATTGCAAGTGGATTTCTTCATCAACC  
TCCTTTTTAAATTTATTCTTGGTATCAAGATCATGCGTTCTCTGTTCTT  
AACACACTGGATTTCCATCTGGATGTTGCTGATCAGTCTGAAATACAACACTGTTGAAT  
TCCAGAAGGACCAACACCAGATAAATTATGAATGTTGAAACAAGATGACCTTACATCCACA  
GCAGATAATGATAGGTCCTAGGTTAACAGGGCCCTATTGACCCCTGCTGTGGTCT  
GCTGGCTCTCAACTTCTGTGGCTGGTCTGGTGGCTCAGACCTGCCCTCTGT  
GTGCTCTGCAGCAACCAGTTCAAGCAAGGTGATTGTTGCTGGTGGCT  
TCCGGATGGCATCTCCACCAACACACGGCTGCTGAACCTCCATGAGAACCAAATCCAGAT  
CATCAAAGTGAACAGCTTCAAGCACTTGAGGCACTTGAAATCCTACAGTTGAGTAGGAA  
CCATATCAGAACCAATTGAAATTGGGCTTCAATGGTCTGGCAACCTAACACTCTGGA  
ACTCTTGACAATCGTCTTACTACCATCCCAATGGAGCTTTGTATACTTGTCTAAACT  
GAAGGAGCTCTGGTTCGAAACAACCCATTGAAAGCATCCCTCTTATGCTTTAACAG  
AATTCTCTTGCACGGACTAGACTTAGGGAAATTGAAAAGACTTCATACATCTCAGA  
AGGTGCCTTGAAGGCTGTCCAACCTGAGGTATTGAAACCTGCCATGTGCAACCTCG  
GGAAATCCCTAACCTCACACCGCTCATAAAAGTACAGCTGGATCTTCTGGAAATCA  
TTTATCTGCCATCAGGCCCTGGCTTTCCAGGGTTGATGCAACCTCAGTCAGTCAGT  
GATAACAGTCCCAGATTCAAGTGATTGAAACGGAATGCCTTGACACCTCAGTCAGT  
GGAGATCAACCTGGCACACAATAATCTAACATTACTGCCTCATGACCTCTCACTCCCT  
GCATCATCTAGAGCGGATACTTACATCACAAACCTGGAACTGTAACGTGACATACT  
GTGGCTCAGCTGGTGGATAAAAGACATGGCCCCCTGAAACACAGCTTGTGCCCCGGTG  
TAACACTCCCTCAATCTAAAGGGGAGGTACATTGGAGAGCTGACCAGAACATTACTCAC  
ATGCTATGCTCCGGTATTGTGGAGCCCCCTGAGACCTCAATGTCAGTGAAAGGCATGGC  
AGCTGAGCTGAAATGTCGGCCTCCACATCCCTGACATCTGTATCTGGATTACTCCAA  
TGGAACAGTCATGACACATGGGGCGTACAAAGTGCAGGATAGCTGTGCTCAGTGATGGTAC  
GTTAAATTTCACAAATGTAACGTGCAAGATAACAGGCATGTACACATGTATGGTGGTAA  
TTCCGGTGGAAACTACTGCTTCAGCCACCTGAATGTTACTGCAGCAACCAACTACTCC

TTTCTCTTACTTTCAACCGTCACAGTAGAGACTATGGAACCGTCTCAGGATGAGGCACG  
GACCACAGATAACAATGTGGTCCCACTCCAGTGGTCAGTGGAGACCACCAATGTGAC  
CACCTCTCTCACACCAACAGAGCACAAGGTGACAGAGAAAAACCTTCACCATCCCAGTGAC  
TGATATAAACAGTGGGATCCCAGGAATTGATGAGGTATGAAGACTACCAAAATCATCAT  
TGGGTGTTTGTGGCCATCACACTCATGGCTGCAGTGATGCTGGTCATTTCTACAAGAT  
GAGGAAGCAGCACCATCGGCAAAACCATCACGCCCAACAAGGACTGTTGAAATTATTAA  
TGTGGATGATGAGATTACGGGAGACACACCCATGGAAGGCCACTGCCATGCCTGCTAT  
CGAGCATGAGCACCTAAATCACTACTACATAACTCATAAAATCTCCCTCAACCACACAAC  
AGTTAACACAATAAATTCAATACACAGTTCACTGCATGAACCGTTATTGATCCGAATGAA  
CTCTAAAGACAATGTACAAGAGACTCAAATCTAAACATTACAGAGTTACAAAAAAACAA  
ACAATCAAAAAAAAGACAGTTATTAAAAATGACACAAATGACTGGCTAAATCTACTG  
TTTCAAAAAAGTGTCTTACAAAAAAACAAAAAGAAAAGAAATTATTTATTAAAAATT  
CTATTGTGATCTAACAGCAGACAAAAA

Figure 14

MLNKMTLHPQQIMIGPRFNRALFDPLLVLLALQLLVVAGLVRAQTCPSVCSCSNQFSK  
 VICVRKNLREVPGISTNTRLLNLHENQIQIIVKNSFKHLRHEILQLSRNHIRTIEIG  
 AFNGLANLNTLELFDMRLTTIPNGAFVYLSKLKELWLRLNNPIESIPSYAFNRIPSLRRL  
 DLGELKRLSYISEGAFEGLSNLRYLNLMCNLREIPNLTPLIKLDELDLSGNHLSAIRP  
 GSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLER  
 IHLHHNPWNCNCIDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAP  
 VIVEPPADLNVTGMAELKRASTSLSVSWITPNGTVMTHGAYKVRIAVLSDGTLNF  
 TNVTVQDTGMYTCMVSNSVGNTTASATLNTAATTPFSYFSTVTVETMEPSQDEARTT  
 DNNVGPTPVWDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPIGIDEVMKTTKIIIG  
 CFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDEITGDTPMESHPMPAI  
 EHEHLNHYNSYKSPFNHTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI

## N-glycosylation site at residues:

278-282

364-368

390-394

412-416

415-419

434-438

442-446

488-492

606-610

## cAMP- and cGMP-dependent protein kinase phosphorylation site at residues:

183-187

## Casein kinase II phosphorylation site at residues:

268-272

417-421

465-469

579-583

620-624

## N-myristoylation site at residues:

40-46

73-79

118-124

191-197

228-234

237-243

391-397

422-428

433-439

531-537

Figure 15

AGCCGACGCTGCTCAAGCTGCAACTCTGTTGCAGTTGGCAGTTCTTTCGGTTCCCTCC  
TGCTGTTGGGGCATGAAAGGGCTTCGCCGCCGGAGTAAAGAAGGAATTGACCGGGC  
AGCGCGAGGGAGGAGCGCGACCGCGAGGGCGGGCGTGCACCCCTGGCTGGAAGT  
TTGTGCCGGGCCCGAGCGCGCCGGCTGGAGCTCGGGTAGAGACCTAGGCCGCTGG  
ACCGCGATGAGCGCGCCGAGCCTCCGTGCGCGCCGGGTTGGGCTGCTGCTGTGC  
GCGGTGCTGGGCGCGCTGGCCGGTCCGACAGCGCGTGCAGGGAACTCGGGCAGCCC  
TCTGGGGTAGCCGCCAGCGCCATGCCCAACTACCTGCCGCTGCCCTGGGACCTGCTG  
GACTGCAGTCGAAGCGGCTAGCGCGTCTCCCGAGCCACTCCGCTGGGCGCTCGG  
CTGGACTTAAGTCACACAGATTATCTTCATCAAGGCAAGTTCCATGAGCCACCTCAA  
AGCCTTCGAGAAAGTGAACAAACAATGAATTGGAGACCATTCAAATCTGGGACCA  
GTCTCGGCAAATATTACACTTCTCCCTGGCTGGAAACAGGATTGTTGAAATACTCCCT  
GAACATCTGAAAGAGTTCACTCCCTTGAAACTTGGACCTTAGCAGCAACAATATTCA  
GAGCTCCAAACTGCATTCCAGCCCTACAGCTCAAATATCTGTATCTCAACAGCAACC  
GTCACATCAATGGAACCTGGTATTTGACAATTGGCAACACACTCCTGTGTTAAAG  
CTGAACAGGAACCGAATCTCAGCTATCCACCCAAAGATGTTAAACTGCCCAACTGCAA  
CATCTGAATTGAACCGAAACAAGATTAAAATGTAGATGGACTGACATTCCAAGGCCTT  
GGTGTCTGAAGTCTGAAAATGCAAAGAAATGGAGTAACGAAACTTATGGATGGAGCT  
TTTGGGGCTGAGCAACATGAAATTGCACTGGCTGCTGAGGAACTTCATCTCAGCAAATG  
ACCAAAGGCTGGCTTACGGCTTGCTGATGCTGAGGAACTTCATCTCAGCAAATG  
ATCAACAGGATCAGCCCTGATGCCCTGGAGTTCTGCCAGAAGCTCAGTGAGCTGGAC  
ACTTTCAATCACTTATCAAGGTTAGATGATTCAAGCTCCTGGCTAACGTTACTAAAT  
ACACTGCACATTGGAAACAACAGAGTCAGCTACATTGCTGATTGTGCCCTGGGGCTT  
TCCAGTTAAAGACTTGGATCTGAAGAACATGAAATTCTGGACTATTGAAGACATG  
AATGGTGCTTCTCTGGCTTGACAACACTGAGGCGACTGATACTCCAAGGAAATCGGATC  
CGTTCTATTACTAAAAAGCCTTCACTGGTTGGATGCATTGGAGCATCTAGACCTGAGT  
GACAACGCAATCATGTTACAAGGCAATGCATTTCACAAATGAAGAAACTGCAACAA  
TTGCATTAAATACATCAAGCCTTGTGCGATTGCCAGCTAAATGGCTCCACAGTGG  
GTGGCGAAAACAACCTTCAGAGCTTGAAATGCCAGTTGTGCTGCCATCCTCAGTGCTA  
AAAGGAAGAAGCATTGCTGTTAGCCAGATGGCTTGATGATGATTCCAA  
CCCCAGATCACGGTTAGCCAGAAACACAGTCGGCAATAAAGGTTCAATTGAGTT  
ATCTGCTCAGCTGCCAGCAGCAGTGATTCCCAATGACTTTGCTGGAAAAAGACAAT  
GAAACTACTGCATGATGCTGAAATGGAAAATTATGCACACCTCCGGGCCAAGGTGGCGAG  
GTGATGGAGTATACCAACATCCTCGGCTGCGCAGGTGGAATTGCCAGTGAGGGGAAA  
TATCAGTGTGTCATCTCAAATCACTTGGTTCATCCTACTCTGTCAAAGCCAAGCTTACA  
GTAAATATGCTTCCCTCATTCAACAGACCCCCATGGATCTCACCATCCGAGCTGGGCC  
ATGGCACGCTGGAGTGTGCTGCTGTGGGCCACCCAGCCCCCAGATAGCCTGGCAGAAG

GATGGGGCACAGACTCCAGCTGCACGGAGAGACGCATGCATGTGATGCCGAGGAT  
GACGTGTTCTTATCGTGGATGTGAAGATAAGGACATTGGGTATACAGCTGCACAGCT  
CAGAACAGTCAGGAAGTATTCAGCAAATGCAACTCTGACTGTCTAGAAACACCATCA  
TTTTGCGGCCACTGTTGGACCGAACTGTAACCAAGGGAGAAACAGCCGTCTACAGTGC  
ATTGCTGGAGGAAGCCCTCCCCCTAAACTGAACTGGACCAAAAGATGATAGCCCATTGGTG  
GTAACCGAGAGGCACTTTTTGCAAGCAGGCAATCAGCTCTGATTATTGTGGACTCAGAT  
GTCAGTGATGCTGGAAATACACATGTGAGATGTCTAACACCCTGGCACTGAGAGAGGA  
AACGTGCGCCTCAGTGTATCCCCACTCCAACCTGCGACTCCCCCTCAGATGACAGCCCCA  
TCGTTAGACGATGACGGATGGGCCACTGTGGGTGTCGTGATCATAGCGTGGTTGCTGT  
GTGGTGGCACGTCACTCGTGTGGTGGTCATCATATACCACACAAGGGAGGAATGAA  
GATTGCAGCATTACCAACACAGATGAGACCAACTGCCAGCAGATATTCTCTAGTTATTG  
TCATCTCAGGGAACGTTAGCTGACAGGCAGGATGGGTACGTGTCTCAGAAAGTGGAAAGC  
CACCAACAGTTGTACATCTCAGGTGCTGGATTTTCTTACCAACATGACAGTAGT  
GGGACCTGCCATATTGACAATAGCAGTGAAGCTGATGTGGAAGCTGCCACAGATCTGTT  
CTTGTCCGTTTGGATCCACAGGCCCTATGTATTGAAAGGAAATGTGTATGGCTCA  
GATCCTTTGAAACATATCATAACAGGTTGCAGTCCTGACCCAAGAACAGTTAACAG  
CACTATGAGGCCAGTTACATAAAGAAAAGGAGTGCTACCCATGTTCTCATCCTTCAGAA  
GAATCCTGCGAACGGAGCTCAGTAATATATCGTGGCCTCACATGTGAGGAAGCTACTT  
AACACTAGTTACTCTCACAATGAAGGACCTGGAATGAAAATCTGTGTCTAACAAAGTCC  
TCTTTAGATTTAGTGCACATCCAGAGCCAGCGTCGGTGCCTCGAGTAATTCTTCATG  
GGTACCTTGGAAAAGCTCTCAGGAGACCTCACCTAGATGCCTATTCAAGCTTGGACAG  
CCATCAGATTGTCAGCCAAGAGCCTTTATTGAAAGCTCATTCTCCCCAGACTGGAC  
TCTGGGTAGAGGAAGATGGAAAGGAAAGAGCAGATTTCAGGAAGAAAATCACATTGT  
ACCTTAAACAGACTTTAGAAAACACTACAGGACTCCAAATTTCAGTCTTATGACTGGAC  
ACATAGACTGAATGAGACCAAAGGAAAGCTTAACATACTACCTCAAGTGAACTTTATT  
TAAAAGAGAGAGAATCTTATGTTTAAATGGAGTTATGAATTAAAAGGATAAAAAT  
GCTTTATTATACAGATGAACCAAATTACAAAAAGTTATGAAAATTGTTATGCAAAAAAG  
TGATGCTCATATAAGAACCTTTAAACTATTTTAACTTTGTTTATGCAAAAAAG  
TATCTTACGTAAATTAAATGATATAATCATGATTATTTATGTTATGCAATT  
TTTCTTTTATGAAAGTACTAAAGCATTAAATAACCTGCCTGTACCATT  
TTTAAATAGAAGTTACTTCATTATTTGCACATTATTTAATAAAATGTGTCAATT  
TGAAAAAAAAAAAAAAAAAAAAAA

Figure 16

MSAPSLRARAAGLGLLLCAVLGRAGRSDSGGRGELGQPSGVAAERPCPTTCRCLGDLLD  
 CSRKRLARLPEPLPSWVARLDLSHNRLSFIKASSMSHLQSLREVKLNNNELETIPNLGP  
 VSANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTAFPALQLKYLYLNSN  
 RVTSMEPGYFDNLANTLLVLKLNRRNRISSAIPPKMFKLQLQHLELNRNKKIKNVDGLTFQ  
 GLGALKSLKMQRNGVTKLMGAFWGLSNMEIQLQDHNLTEITKGWLYGLLMLOELHLS  
 QNAINRISPDAWEFCQKLSEDLTFNHLRLDDSSFLGLSLLNTHIGNNRVSYIADCA  
 FRGLSSLKTLSDLKNNEISWTIEDMNGAFSGLDKLRRLILQGNRIRSITKKAFTGLDALE  
 HLDLSDNAIMSLQGNAFSQMKKLQQLHLNTSSLCDQCLKWLPQWVAENNQSFVNASC  
 AHPQLLKGRSIFAVSPDGFVCDDFPKPQITVQPETQSAIKGSNLSFICSAASSSDSPMT  
 FAWKKDNELLHDAEMENYAHLRAQGGEVMEYTTILREREFASEGKYQCVISNHFGSS  
 YSVKAKLTNVNMLPSFTKTPMDLTIRAGAMARLECAAVGHPAPQIAWQKDGGTDFPAARE  
 RRMHVMPEDDVFFIVDVKIEDIGVYSCATAQNSAGSISANATLTVLETPSFLRPLLDRTV  
 TKGETAVLQCIAGGSPPPKNWTKDDSPLVVTERHFFAAGNQLLIVDSDVSDAGKYTC  
 EMSNTLGTERGNVRLSIVPTPTCDSPQMTAPSLLDDGWATVGVVIIAVVCCVVGTSLVW  
 VVIIYHTRRRNEDCSITNTDETNLPADIPSYLSSQGTIADRQDGYVSSESGSHHQFVTS  
 SGAGFFLPQHDSSGTCHIDNSSEADVEATDLFLCPFLGSTGPMYLKGNVYGSDFETY  
 HTGCPDPRTVLMHDYEPSYIKKKECYPCHPSEESCERSFSNISWPSHVRKLLNTSYS  
 HNEGPGMKNLCLNKSSLDFSANPEPASVASSNSFMGTFGKALRRHLDAYSSFGQPSDC  
 QPRAFYLKAKSSPDLDGSEEDGKERTDFQEENHICTFKQTLENYRTPNFQSYDLDT

N-glycosylation site at residues:

122-126

156-160

274-278

442-446

469-473

515-519

688-692

729-733

905-909

987-991

999-1003

1016-1020

Glycosaminoglycan attachment site at residues:

886-890

Casein kinase II phosphorylation site at residues:

99-103  
180-184  
263-267  
314-318  
324-328  
374-378  
383-387  
407-411  
524-528  
608-612  
692-696  
709-713  
731-735  
799-803  
843-847  
863-867  
907-911  
1003-1007  
1018-1022  
1073-1077  
1079-1083  
1081-1085

Tyrosine kinase phosphorylation site at residue:

667-675

N-myristoylation site at residues:

14-20  
36-42  
239-245  
257-263  
380-386  
427-433  
513-519  
588-594  
672-678  
683-689

774-780

933-999

Leucine zipper pattern at residues:

58-80

65-87



ggccttggtg tatgctatgc tcagaggaaa ggctactttt caaaagaaaac 900  
 ctccttccag aagagtaatt cttcatctaa agccacgaca atgagtgaaa 950  
 atgtgcagtg gctcacgcct gtaatcccag cacttggaa ggccgcggcg 1000  
 ggcggatcac gaggtcagga gttctagacc agtctggcca atatggtcaa 1050  
 accccatctc tactaaaata caaaaattag ctgggcatgg tggcatgtgc 1100  
 ctgcagttcc agctgcttgg gagacaggag aatcacttga acccgggagg 1150  
 cgaggttgc agtgagctga gatcacgcca ctgcagtcac gcctggtaa 1200  
 cagagcaaga ttccatctca aaaaataaaaa taaataaaata aataaataact 1250  
 ggtttttacc tgtagaattc ttacaataaa tatagcttga tattc 1295

&lt;210&gt; 2

&lt;211&gt; 312

&lt;212&gt; PRT

&lt;213&gt; artificial sequence

&lt;400&gt; 2

Met	Ala	Arg	Arg	Ser	Arg	His	Arg	Leu	Leu	Leu	Leu	Leu	Arg
1								10					15

Tyr	Leu	Val	Val	Ala	Leu	Gly	Tyr	His	Lys	Ala	Tyr	Gly	Phe	Ser
									25					30

Ala	Pro	Lys	Asp	Gln	Gln	Val	Val	Thr	Ala	Val	Glu	Tyr	Gln	Glu
									35					45

Ala	Ile	Leu	Ala	Cys	Lys	Thr	Pro	Lys	Lys	Thr	Val	Ser	Ser	Arg
									50					60

Leu	Glu	Trp	Lys	Lys	Leu	Gly	Arg	Ser	Val	Ser	Phe	Val	Tyr	Tyr
									65					75

Gln	Gln	Thr	Leu	Gln	Gly	Asp	Phe	Lys	Asn	Arg	Ala	Glu	Met	Ile
									80					90

Asp	Phe	Asn	Ile	Arg	Ile	Lys	Asn	Val	Thr	Arg	Ser	Asp	Ala	Gly
									95					105

Lys	Tyr	Arg	Cys	Glu	Val	Ser	Ala	Pro	Ser	Glu	Gln	Gly	Gln	Asn
									110					120

Leu	Glu	Glu	Asp	Thr	Val	Thr	Leu	Glu	Val	Leu	Val	Ala	Pro	Ala
									125					135

Val	Pro	Ser	Cys	Glu	Val	Pro	Ser	Ser	Ala	Leu	Ser	Gly	Thr	Val
									140					150

Val	Glu	Leu	Arg	Cys	Gln	Asp	Lys	Glu	Gly	Asn	Pro	Ala	Pro	Glu
									155					165

Tyr	Thr	Trp	Phe	Lys	Asp	Gly	Ile	Arg	Leu	Leu	Glu	Asn	Pro	Arg
									170					180

Leu	Gly	Ser	Gln	Ser	Thr	Asn	Ser	Ser	Tyr	Thr	Met	Asn	Thr	Lys
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

185	190	195
Thr Gly Thr Leu Gln Phe Asn Thr Val Ser Lys Leu Asp Thr Gly		
200	205	210
Glu Tyr Ser Cys Glu Ala Arg Asn Ser Val Gly Tyr Arg Arg Cys		
215	220	225
Pro Gly Lys Arg Met Gln Val Asp Asp Leu Asn Ile Ser Gly Ile		
230	235	240
Ile Ala Ala Val Val Val Ala Leu Val Ile Ser Val Cys Gly		
245	250	255
Leu Gly Val Cys Tyr Ala Gln Arg Lys Gly Tyr Phe Ser Lys Glu		
260	265	270
Thr Ser Phe Gln Lys Ser Asn Ser Ser Lys Ala Thr Thr Met		
275	280	285
Ser Glu Asn Val Gln Trp Leu Thr Pro Val Ile Pro Ala Leu Trp		
290	295	300
Lys Ala Ala Ala Gly Gly Ser Arg Gly Gln Glu Phe		
305	310	312

&lt;210&gt; 3

&lt;211&gt; 2033

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;400&gt; 3

```

ccaggccggg aggcgacgca cccagccgtc taaacggaa cagccctggc 50
tgagggagct gcagcgcagc agagtatctg acggcgccag gttgcgtagg 100
tgcggcacga ggagtttcc cggcagcggag gaggtcctga gcagcatggc 150
ccggaggagc gcctccctg cccgcgcgt ctggctctgg agcatcctcc 200
tgtgcctgct ggcactgcgg gcggaggccg ggccggcga ggaggagagc 250
ctgtacctat ggatcgatgc tcaccaggca agagtactca taggattga 300
agaagatatac ctgattgttt cagagggaa aatggcacct tttacacatg 350
atttcagaaa agcgcaacag agaatgccag ctattcctgt caatatccat 400
tccatgaatt ttacctggca agctgcaggg caggcagaat acttctatga 450
attcctgtcc ttgcgcctcc tggataaagg catcatggca gatccaaccg 500
tcaatgtccc tctgctggga acagtgcctc acaaggcatc agttgtcaa 550
gttggttcc catgtcttgg aaaacaggat ggggtggcag catttgaagt 600
ggatgtgatt gttatgaatt ctgaaggcaa caccattctc caaacacctc 650
aaaatgctat cttctttaaa acatgtcaac aagctgagtg cccaggcggg 700
tgccgaaatg gaggctttg taatgaaaga cgcacatgcg agtgcctga 750

```

tgggttccac ggacccact gtgagaaaagc cctttgtacc ccacgatgta 800  
 tgaatggtgg actttgtgtg actcctggtt tctgcacatcg cccacctgga 850  
 ttctatggag tgaactgtga caaagcaaac tgctcaacca cctgctttaa 900  
 tggagggacc tgtttctacc ctggaaaatg tatttgcctt ccaggactag 950  
 agggagagca gtgtgaaatc agcaaattgcc cacaaccctg tcgaaatgga 1000  
 ggtaaatgca ttggtaaaag caaatgtaaag tgttccaaag gttaccagg 1050  
 agacctctgt tcaaagcctg tctgcagcc tggctgtggt gcacatggaa 1100  
 cctgccatga acccaacaaa tgccaatgtc aagaaggttg gcatggaaga 1150  
 cactgcaata aaaggtaacga agccagcctc atacatgccc tgaggccagc 1200  
 aggccccccag ctcaggcagc acacgccttc actaaaaaag gccgaggagc 1250  
 ggcgggatcc acctgaatcc aattacatct ggtgaactcc gacatctgaa 1300  
 acgttttaag ttacaccaag ttcatagcct ttgttaacct ttcatgtgtt 1350  
 gaatgttcaa ataatgttca ttacacttaa gaatactggc ctgaatttt 1400  
 ttagcttcat tataaatcac tgagctgata tttactcttc cttttaaagg 1450  
 ttcttaagtac gtctgttagca tgatggata gatttcttg tttcagtgtc 1500  
 ttgggacaga ttttatatta tgtcaattga tcaggttaaa attttcagtg 1550  
 ttagttggc agatattttc aaaattacaa tgcatatttg gtgtctgggg 1600  
 gcaggggaac atcagaaagg ttaaattggg caaaaatgcg taagtcacaa 1650  
 gaatttggat ggtgcagtta atgttgaagt tacagcattt cagattttat 1700  
 tgcagatat ttagatgttt gttacatttt taaaaattgc tcttaatttt 1750  
 taaactctca atacaatata tttgacctt accattttc cagagattca 1800  
 gtattaaaaa aaaaaaaaaatt acactgtggt agtggcattt aaacaatata 1850  
 atatattctta aacacaatga aataggaaat ataatgtatg aacttttgc 1900  
 atggcttga agcaatataa tatattgtaa aaaaaacaca gctcttacct 1950  
 aataaacatt ttatactgtt tttatgtata aaataaaggt gctgctttag 2000  
 tttttggaa aaaaaaaaaa aaaaaaaaaa aaa 2033

&lt;210&gt; 4

&lt;211&gt; 379

&lt;212&gt; PRT

&lt;213&gt; artificial sequence

&lt;400&gt; 4

Met	Ala	Arg	Arg
Ser	Ala	Phe	Pro
1	5	10	15



Ala His Gly Thr Cys His Glu Pro Asn Lys Cys Gln Cys Gln Glu  
 320 325 330  
 Gly Trp His Gly Arg His Cys Asn Lys Arg Tyr Glu Ala Ser Leu  
 335 340 345  
 Ile His Ala Leu Arg Pro Ala Gly Ala Gln Leu Arg Gln His Thr  
 350 355 360  
 Pro Ser Leu Lys Lys Ala Glu Glu Arg Arg Asp Pro Pro Glu Ser  
 365 370 375  
 Asn Tyr Ile Trp  
 379

<210> 5  
 <211> 1857  
 <212> DNA  
 <213> artificial sequence

<400> 5  
 gtcgttccc aggagtcctt cggcggtgtgt tgcgtcagtg gcctgatcgc 50  
 gatggggaca aaggcgcaag tcgagagggaa actgttgtgc ctcttcata 100  
 tggcgatcct gttgtgctcc ctggcattgg gcagtgttac agtgcactct 150  
 tctgaacctg aagtcagaat tcctgagaat aatcctgtga agttgtcctg 200  
 tgcctactcg ggctttctt ctccccgtgt ggagtggaaag tttgaccaag 250  
 gagacaccac cagactcggt tgctataata acaagatcac agtttcctat 300  
 gaggaccggg tgaccttctt gccaactgggt atcaccttca agtccgtgac 350  
 acgggaagac actgggacat acacttgtat ggtctctgag gaaggcggca 400  
 acagctatgg ggaggtcaag gtcaagctca tcgtgtttgt gcctccatcc 450  
 aagcctacag ttaacatccc ctcccttgcc accattggga accggggcagt 500  
 gctgacatgc tcagaacaag atggttcccc accttctgaa tacacctgg 550  
 tcaaagatgg gatagtgatg cctacgaatc ccaaaagcac ccgtgccttc 600  
 agcaactctt cctatgtcct gaatcccaca acaggagagc tggtctttga 650  
 tccccgtca gcctctgata ctggagaata cagctgtgag gcacggaatg 700  
 ggtatggac acccatgact tcaaattgtc tgcgcatgga agctgtggag 750  
 cggaatgtgg gggtcatcgt ggcagccgtc cttgttaaccc tgattctcct 800  
 gggaaatcttg gttttggca tctgggtttgc ctatagccga ggcactttg 850  
 acagaacaaa gaaaggact tcgagtaaga aggtgattta cagccagcct 900  
 agtccccgaa gtgaaggaga attcaaacag acctcgatcat tcctgggtgtg 950  
 agcctggtcg gctcaccgccc tatcatctgc atttgcctta ctcaggtgct 1000

accggactct ggccctgat gtctgttagtt tcacaggatg ccttatttgt 1050  
 cttctacacc ccacagggcc ccctacttct tcggatgtgt tttaataat 1100  
 gtcagctatg tgccccatcc tccttcatgc cctccctccc tttcctacca 1150  
 ctgctgagtg gcctggaact tgtttaaagt gtttattccc catttcttg 1200  
 agggatcagg aaggaatcct ggttatgcca ttgacttccc ttctaaatgt 1250  
 acagcaaaaa tggcgggggt cgccaggaatc tgcactcaac tgcccacctg 1300  
 gctggcaggg atctttgaat aggtatcttgc agcttgggttc tgggctctt 1350  
 ccttgcgtac tgacgaccag gcccagctgt tctagagcgg gaatttagagg 1400  
 ctagagcggc tgaaatgggt gtttgggtat gacactgggg tccttccatc 1450  
 tctggggccc actctcttct gtcttccat gggaaatgtcc actgggatcc 1500  
 ctctgccctg tcctcctgaa tacaagctga ctgacattga ctgtgtctgt 1550  
 gaaaaatggg agctctgtt gtggagagca tagtaaattt tcagagaact 1600  
 tgaagccaaa aggattaaa accgctgctc taaagaaaag aaaactggag 1650  
 gctggcgca gtggctcactg cctgtaatcc cagaggctga ggcaggcgga 1700  
 tcacctgagg tcgggagttc gggatcagcc tgaccaacat ggagaaaccc 1750  
 tactggaaat acaaagttag ccaggcatgg tggtgcatgc ctgttagtccc 1800  
 agctgctcag gagcctggca acaagagcaa aactccagct caaaaaaaaaa 1850  
 aaaaaaaaaa 1857

&lt;210&gt; 6

&lt;211&gt; 299

&lt;212&gt; PRT

&lt;213&gt; artificial sequence

&lt;400&gt; 6

Met	Gly	Thr	Lys	Ala	Gln	Val	Glu	Arg	Lys	Leu	Leu	Cys	Leu	Phe
1														15

Ile	Leu	Ala	Ile	Leu	Leu	Cys	Ser	Leu	Ala	Leu	Gly	Ser	Val	Thr
														30
20														

Val	His	Ser	Ser	Glu	Pro	Glu	Val	Arg	Ile	Pro	Glu	Asn	Asn	Pro
														45
35														

Val	Lys	Leu	Ser	Cys	Ala	Tyr	Ser	Gly	Phe	Ser	Ser	Pro	Arg	Val
														60
50														

Glu	Trp	Lys	Phe	Asp	Gln	Gly	Asp	Thr	Thr	Arg	Leu	Val	Cys	Tyr
														75
65														

Asn	Asn	Lys	Ile	Thr	Ala	Ser	Tyr	Glu	Asp	Arg	Val	Thr	Phe	Leu
														90
80														

Pro	Thr	Gly	Ile	Thr	Phe	Lys	Ser	Val	Thr	Arg	Glu	Asp	Thr	Gly
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

95	100	105
Thr Tyr Thr Cys Met Val Ser Glu Glu Gly Gly Asn Ser Tyr Gly		
110	115	120
Glu Val Lys Val Lys Leu Ile Val Leu Val Pro Pro Ser Lys Pro		
125	130	135
Thr Val Asn Ile Pro Ser Ser Ala Thr Ile Gly Asn Arg Ala Val		
140	145	150
Leu Thr Cys Ser Glu Gln Asp Gly Ser Pro Pro Ser Glu Tyr Thr		
155	160	165
Trp Phe Lys Asp Gly Ile Val Met Pro Thr Asn Pro Lys Ser Thr		
170	175	180
Arg Ala Phe Ser Asn Ser Ser Tyr Val Leu Asn Pro Thr Thr Gly		
185	190	195
Glu Leu Val Phe Asp Pro Leu Ser Ala Ser Asp Thr Gly Glu Tyr		
200	205	210
Ser Cys Glu Ala Arg Asn Gly Tyr Gly Thr Pro Met Thr Ser Asn		
215	220	225
Ala Val Arg Met Glu Ala Val Glu Arg Asn Val Gly Val Ile Val		
230	235	240
Ala Ala Val Leu Val Thr Leu Ile Leu Leu Gly Ile Leu Val Phe		
245	250	255
Gly Ile Trp Phe Ala Tyr Ser Arg Gly His Phe Asp Arg Thr Lys		
260	265	270
Lys Gly Thr Ser Ser Lys Lys Val Ile Tyr Ser Gln Pro Ser Ala		
275	280	285
Arg Ser Glu Gly Glu Phe Lys Gln Thr Ser Ser Phe Leu Val		
290	295	299

<210> 7  
 <211> 2755  
 <212> DNA  
 <213> artificial sequence

<400> 7  
 gggggttagg gaggaaggaa tccaccccca cccccccaaa ccctttctt 50  
 ctcccttcct ggcttcggac attggagcac taaatgaact tgaattgtgt 100  
 ctgtggcgag caggatggtc gctgttactt tgtgatgaga tcggggatga 150  
 attgctcgct ttaaaaatgc tgctttggat tctgttgctg gagacgtctc 200  
 tttgtttgc cgctggaaac gttacagggg acgtttgcaa agagaagatc 250  
 tttcctgca atgagataga aggggaccta cacgtagact gtaaaaaaaaa 300  
 gggcttcaca agtctgcagc gttcactgc cccgacttcc cagtttacc 350

atttatttct gcatggcaat tccctcactc gactttccc taatgagttc 400  
gctaactttt ataatgcggt tagttgcac atggaaaaca atggcttgca 450  
tgaaatcggtt ccggggcctt ttctgggct gcagctggtg aaaaggctgc 500  
acatcaacaa caacaagatc aagtctttc gaaagcagac ttttctgggg 550  
ctggacgatc tggaatatct ccaggctgat ttttaatttat tacgagatat 600  
agacccgggg gccttccagg acttgaacaa gctggaggtg ctcattttaa 650  
atgacaatct catcagcacc ctacctgcca acgtttcca gtatgtgcc 700  
atcacccacc tcgacctccg ggtaacagg ctgaaaacgc tgccctatga 750  
ggaggtcttg gagcaaattcc ctggatttgc ggagatcctg ctagaggata 800  
acccttggga ctgcacctgt gatctgctct ccctgaaaga atggctggaa 850  
aacattccca agaatgccct gatcggccga gtggctgctg aagccccac 900  
cagactgcag ggttaagacc tcaatgaaac caccgaacag gacttgtgtc 950  
ctttgaaaaa ccgagtgat tctagtctcc cggcgcccccc tgcccaagaa 1000  
gagacctttg ctctggacc cctgccaact ccttcaaga caaatggca 1050  
agaggatcat gccacaccag ggtctgctcc aaacggaggt acaaagatcc 1100  
caggcaactg gcagatcaaa atcagaccca cagcagcgt agcgacgggt 1150  
agctccagga acaaaccctt agctaacagt ttaccctgccc ctggggctg 1200  
cagctgcac cacatccag ggtcggttt aaagatgaac tgcaacaaca 1250  
ggaacgtgag cagcttgct gatttgaagc ccaagctctc taacgtgcag 1300  
gagctttcc tacgagataa caagatccac agcatccaa aatcgactt 1350  
tgtggattac aagaacctca ttctgttggta tctggcaac aataacatcg 1400  
ctactgtaga gaacaacact ttcaagaacc ttttggacct caggtggcta 1450  
tacatggata gcaattaccc ggacacgctg tcccgggaga aattcgccgg 1500  
gctgaaaaac ctagagtacc tgaacgtgga gtacaacgct atccagctca 1550  
tcctcccggtt cacttcaat gccatgccc aactgaggat cctcattctc 1600  
aacaacaacc tgctgaggc cctgcctgtg gacgtttcg ctggggctc 1650  
caggggtgct ggaccaggta acctccatca tccagataga cctccacgga 1750  
aaccctggg agtgctcctg cacaattgtg ccttcaagc agtggggcaga 1800  
acgcttgggt tccgaagtgc tgatgagcga cctcaagtgt gagacgcccgg 1850  
tgaacttctt tagaaaggat ttcatgctcc tctccaatga cgagatctgc 1900

cctcagctgt acgctaggat ctcgcccacg ttaacttcgc acagtaaaaa 1950  
 cagcactggg ttggcggaga cggggacgca ctccaaactcc tacctagaca 2000  
 ccagcagggt gtccatctcg gtgttggtcc cgggactgct gctgggttt 2050  
 gtcacctccg ctttaccgt ggtgggcatg ctcgtgtta tcctgagaa 2100  
 ccgaaagcgg tccaagagac gagatgccaa ctcctccgca tccgagatta 2150  
 attccctaca gacagtctgt gactcttct actggcacaa tgggcattac 2200  
 aacgcagatg gggcccacag agtgtatgac tgtggctctc actcgctctc 2250  
 agactaagac cccaaacccca ataggggagg gcagagggaa ggcgatacat 2300  
 cttcccccac cgccaggcacc ccgggggctg gaggggcgtg tacccaaatc 2350  
 cccgcgcat cagcctggat gggcataagt agataaataa ctgtgagctc 2400  
 gcacaaccga aaggcctga ccccttactt agctccctcc ttgaaacaaa 2450  
 gagcagactg tggagagctg ggagagcga gccagctcg tcttgctga 2500  
 gagcccttt tgacagaaag cccagcacga ccctgctgga agaactgaca 2550  
 gtgccttcgc ctcggcccc ggggcctgtg ggttggatg ccgcggttct 2600  
 atacatatat acatatatcc acatctatat agagagatag atatctattt 2650  
 ttccccctgtg gattagcccc gtgatggctc cctgtggct acgcagggat 2700  
 gggcagttgc acgaaggcat gaatgtattt taaataagta actttgactt 2750  
 ctgac 2755

<210> 8  
 <211> 696  
 <212> PRT  
 <213> artificial sequence

<400> 8  
 Met Leu Leu Trp Ile Leu Leu Leu Glu Thr Ser Leu Cys Phe Ala  
 1 5 10 15  
 Ala Gly Asn Val Thr Gly Asp Val Cys Lys Glu Lys Ile Cys Ser  
 20 25 30  
 Cys Asn Glu Ile Glu Gly Asp Leu His Val Asp Cys Glu Lys Lys  
 35 40 45  
 Gly Phe Thr Ser Leu Gln Arg Phe Thr Ala Pro Thr Ser Gln Phe  
 50 55 60  
 Tyr His Leu Phe Leu His Gly Asn Ser Leu Thr Arg Leu Phe Pro  
 65 70 75  
 Asn Glu Phe Ala Asn Phe Tyr Asn Ala Val Ser Leu His Met Glu  
 80 85 90

Asn Asn Gly Leu His Glu Ile Val Pro Gly Ala Phe Leu Gly Leu  
 95 100 105  
 Gln Leu Val Lys Arg Leu His Ile Asn Asn Asn Lys Ile Lys Ser  
 110 115 120  
 Phe Arg Lys Gln Thr Phe Leu Gly Leu Asp Asp Leu Glu Tyr Leu  
 125 130 135  
 Gln Ala Asp Phe Asn Leu Leu Arg Asp Ile Asp Pro Gly Ala Phe  
 140 145 150  
 Gln Asp Leu Asn Lys Leu Glu Val Leu Ile Leu Asn Asp Asn Leu  
 155 160 165  
 Ile Ser Thr Leu Pro Ala Asn Val Phe Gln Tyr Val Pro Ile Thr  
 170 175 180  
 His Leu Asp Leu Arg Gly Asn Arg Leu Lys Thr Leu Pro Tyr Glu  
 185 190 195  
 Glu Val Leu Glu Gln Ile Pro Gly Ile Ala Glu Ile Leu Leu Glu  
 200 205 210  
 Asp Asn Pro Trp Asp Cys Thr Cys Asp Leu Leu Ser Leu Lys Glu  
 215 220 225  
 Trp Leu Glu Asn Ile Pro Lys Asn Ala Leu Ile Gly Arg Val Val  
 230 235 240  
 Cys Glu Ala Pro Thr Arg Leu Gln Gly Lys Asp Leu Asn Glu Thr  
 245 250 255  
 Thr Glu Gln Asp Leu Cys Pro Leu Lys Asn Arg Val Asp Ser Ser  
 260 265 270  
 Leu Pro Ala Pro Pro Ala Gln Glu Glu Thr Phe Ala Pro Gly Pro  
 275 280 285  
 Leu Pro Thr Pro Phe Lys Thr Asn Gly Gln Glu Asp His Ala Thr  
 290 295 300  
 Pro Gly Ser Ala Pro Asn Gly Gly Thr Lys Ile Pro Gly Asn Trp  
 305 310 315  
 Gln Ile Lys Ile Arg Pro Thr Ala Ala Ile Ala Thr Gly Ser Ser  
 320 325 330  
 Arg Asn Lys Pro Leu Ala Asn Ser Leu Pro Cys Pro Gly Gly Cys  
 335 340 345  
 Ser Cys Asp His Ile Pro Gly Ser Gly Leu Lys Met Asn Cys Asn  
 350 355 360  
 Asn Arg Asn Val Ser Ser Leu Ala Asp Leu Lys Pro Lys Leu Ser  
 365 370 375  
 Asn Val Gln Glu Leu Phe Leu Arg Asp Asn Lys Ile His Ser Ile  
 380 385 390  
 Arg Lys Ser His Phe Val Asp Tyr Lys Asn Leu Ile Leu Leu Asp

395	400	405
Leu Gly Asn Asn Asn Ile Ala Thr Val Glu Asn Asn Thr Phe Lys		
410	415	420
Asn Leu Leu Asp Leu Arg Trp Leu Tyr Met Asp Ser Asn Tyr Leu		
425	430	435
Asp Thr Leu Ser Arg Glu Lys Phe Ala Gly Leu Gln Asn Leu Glu		
440	445	450
Tyr Leu Asn Val Glu Tyr Asn Ala Ile Gln Leu Ile Leu Pro Gly		
455	460	465
Thr Phe Asn Ala Met Pro Lys Leu Arg Ile Leu Ile Leu Asn Asn		
470	475	480
Asn Leu Leu Arg Ser Leu Pro Val Asp Val Phe Ala Gly Val Ser		
485	490	495
Leu Ser Lys Leu Ser Leu His Asn Asn Tyr Phe Met Tyr Leu Pro		
500	505	510
Val Ala Gly Val Leu Asp Gln Leu Thr Ser Ile Ile Gln Ile Asp		
515	520	525
Leu His Gly Asn Pro Trp Glu Cys Ser Cys Thr Ile Val Pro Phe		
530	535	540
Lys Gln Trp Ala Glu Arg Leu Gly Ser Glu Val Leu Met Ser Asp		
545	550	555
Leu Lys Cys Glu Thr Pro Val Asn Phe Phe Arg Lys Asp Phe Met		
560	565	570
Leu Leu Ser Asn Asp Glu Ile Cys Pro Gln Leu Tyr Ala Arg Ile		
575	580	585
Ser Pro Thr Leu Thr Ser His Ser Lys Asn Ser Thr Gly Leu Ala		
590	595	600
Glu Thr Gly Thr His Ser Asn Ser Tyr Leu Asp Thr Ser Arg Val		
605	610	615
Ser Ile Ser Val Leu Val Pro Gly Leu Leu Val Phe Val Thr		
620	625	630
Ser Ala Phe Thr Val Val Gly Met Leu Val Phe Ile Leu Arg Asn		
635	640	645
Arg Lys Arg Ser Lys Arg Arg Asp Ala Asn Ser Ser Ala Ser Glu		
650	655	660
Ile Asn Ser Leu Gln Thr Val Cys Asp Ser Ser Tyr Trp His Asn		
665	670	675
Gly Pro Tyr Asn Ala Asp Gly Ala His Arg Val Tyr Asp Cys Gly		
680	685	690
Ser His Ser Leu Ser Asp		
695	696	

<210> 9  
<211> 3659  
<212> DNA  
<213> artificial sequence

<400> 9  
gtaactgaag tcaggcttt catttggaa gccccctcaa cagaattcgg 50  
tcattctcca agttatggtg gacgtacttc tgggtttctc cctctgcttg 100  
cttttcaca ttagcagacc ggacttaagt cacaacagat tatctttcat 150  
caaggcaagt tccatgagcc accttcaaag ctttcgagaa gtgaaaactga 200  
acaacaatga attggagacc attccaaatc tgggaccagt ctcggcaaat 250  
attacacttc tctccttggc tggaaacagg attgttgaaa tactccctga 300  
acatctgaaa gagttcagt ccctgaaac tttggacctt agcagcaaca 350  
atatttcaga gctccaaact gcatttccag ccctacagct caaatatctg 400  
tatctcaaca gcaaccgagt cacatcaatg gaacctgggt attttgacaa 450  
tttggccaac acactccttg tggtaaagct gaacaggaac cgaatctcag 500  
ctatcccacc caagatgtt aaactgcccc aactgcaaca tctcgaattt 550  
aaccgaaaca agattaaaaa tggatgtt cttgacattcc aaggcattgg 600  
tgctctgaag tctctgaaaa tgcaaagaaa tggagtaacg aaacttatgg 650  
atggagctt ttggggctg agcaacatgg aaatttgca gctggaccat 700  
aacaaccaa cagagattac caaaggctgg cttaacggct tgctgatgct 750  
gcaggaactt catctcagcc aaaatgccat caacaggatc agccctgatg 800  
cctggagtt ctgccagaag ctcaatggc tggacctaacttcaatcac 850  
ttatcaaggt tagatgattc aagcttcctt ggcctaagct tactaaatac 900  
ctgcacattt ggaacaacag agtcagctac attgctgatt gtgccttcgg 950  
ggggctttcc agttaaaga ctggatccgt tctattacta aaaaagcctt 1000  
cgactgatac tccaaggaaa tcggatccgt tctattacta aaaaagcctt 1100  
cactggttt gatgcattgg agcatctaga cctgagtgac aacgcaatca 1150  
tgtctttaca aggcaatgca ttttccaaaa tgaagaaaact gcaacaattt 1200  
catttaaata catcaagcct tttgtgcgt tgccagctaa aatggctccc 1250  
acagtgggtg gcgaaaaaca actttcagag ctggtaaat gccagttgtg 1300  
cccatcctca gctgctaaaa ggaagaagca ttttgctgt tagcccagat 1350

ggctttgtgt gtgatgattt tcccaaacc cagatcacgg ttcagccaga 1400  
aacacagtcg gcaataaaag gttccaattt gagttcatac tgctcagctg 1450  
ccagcagcag tgattcccc aatgactttt cttggaaaaa agacaatgaa 1500  
ctactgcatg atgctgaaat ggaaaattat gcacacctcc gggcccaagg 1550  
tggcgaggtg atggaggtata ccaccatcct tcggctgcgc gaggtggaat 1600  
ttgccagtga ggggaaatat cagtgtgtca tctccaatca ctttggttca 1650  
tcctactctg tcaaagccaa gcttacagta aatatgcttc cctcattcac 1700  
caagacccccc atggatctca ccatccgagc tggggccatg gcacgcttgg 1750  
agtgtgctgc tgtggggcac ccagcccccc agatagcctg gcagaaggat 1800  
gggggcacag acttcccagc tgcacgggag agacgcatgc atgtgatgcc 1850  
cgaggatgac gtgttctta tcgtggatgt gaagatagag gacattgggg 1900  
tatacagctg cacagctcag aacagtgcag gaagtatttc agcaaatgca 1950  
actctgactg tcctagaaac accatcattt ttgcggccac tggggaccg 2000  
aactgttaacc aagggagaaa cagccgtcct acagtgcatt gctggaggaa 2050  
gccctcccccc taaaactgaac tggaccaaag atgatagccc attgggtgta 2100  
accgagagggc actttttgc agcaggcaat cagcttctga ttattgtgga 2150  
ctcagatgtc agtgatgctg gaaaaatacac atgtgagatg tctaacaccc 2200  
ttggcactga gagaggaaac gtgcgcctca gtgtgatccc cactccaacc 2250  
tgcgactccc ctcagatgac agccccatcg ttagacgatg acggatgggc 2300  
cactgtgggt gtcgtgatca tagccgtggt ttgctgtgtg gtgggcacgt 2350  
cactcgtgtg ggtggcatac atataccaca caaggcggag gaatgaagat 2400  
tgcagcatta ccaacacaga tgagaccaac ttgcacgcag atattcctag 2450  
ttatttgtca tctcagggaa cgtagtgcata caggcaggat gggtagtgc 2500  
cttcagaaag tggaaagccac caccagtttgc tcaatcattc aggtgctgga 2550  
ttttcttac cacaacatga cagtagtggg acctgcata ttgacaatag 2600  
cagtgaagct gatgtggaa ctgcccacaga tctgttcctt tggccgtttt 2650  
tgggatccac agggccatag tatttgaagg gaaatgtgta tggctcagat 2700  
cctttgaaa catatcatac aggttgcagt cctgacccaa gaacagttt 2750  
aatggaccac atgagccag ttacataaaag aaaaaggagt gctacccatg 2800  
ttctcatcct tcagaagaat cctgcgaacg gagcttcagt aatatacgt 2850  
ggccttcaca tgtgaggaag ctacttaaca ctagttactc tcacaatgaa 2900

ggacctggaa tgaaaaatct gtgtctaaac aagtcccttt tagattttag 2950  
 tgcaaatcca gagccagcgt cgggtgcctc gagtaattct ttcatggta 3000  
 cctttggaaa agctctcagg agacctcacc tagatgccta ttcaagcttt 3050  
 ggacagccat cagattgtca gccaaagagcc ttttatttga aagctcattc 3100  
 ttccccagac ttggactctg ggtcagagga agatgggaaa gaaaggacag 3150  
 attttcagga agaaaaatcac atttgtacct taaacagact ttagaaaaact 3200  
 acaggactcc aaattttcag tcttatgact tggacacata gactgaatga 3250  
 gaccaaagga aaagcttaac atactaccc aagtgaactt ttattnaaa 3300  
 gagagagaat cttatgttt ttaaatggag ttatgaattt taaaaggata 3350  
 aaaatgcttt atttatacag atgaacccaa attacaaaaa gttatgaaaa 3400  
 ttttatact gggaaatgatg ctcataaag aataccttt taaactat 3450  
 tttaactttt ttttatgcaa aaaagtatct tacgtaaattt aatgatataa 3500  
 atcatgatta ttttatgtat ttttataatg ccagatttct ttttatggaa 3550  
 aatgagttac taaagcattt taaataatac ctgccttgta ccattttta 3600  
 aatagaagtt acttcattat atttgcaca ttatattaa taaaatgtgt 3650  
 caatttgaa 3659

&lt;210&gt; 10

&lt;211&gt; 1059

&lt;212&gt; PRT

&lt;213&gt; artificial sequence

&lt;400&gt; 10

Met	Val	Asp	Val	Leu	Leu	Leu	Phe	Ser	Leu	Cys	Leu	Leu	Phe	His
1				5					10				15	

Ile	Ser	Arg	Pro	Asp	Leu	Ser	His	Asn	Arg	Leu	Ser	Phe	Ile	Lys
				20				25					30	

Ala	Ser	Ser	Met	Ser	His	Leu	Gln	Ser	Leu	Arg	Glu	Val	Lys	Leu
				35				40				45		

Asn	Asn	Asn	Glu	Leu	Glu	Thr	Ile	Pro	Asn	Leu	Gly	Pro	Val	Ser
				50					55			60		

Ala	Asn	Ile	Thr	Leu	Leu	Ser	Leu	Ala	Gly	Asn	Arg	Ile	Val	Glu
				65				70				75		

Ile	Leu	Pro	Glu	His	Leu	Lys	Glu	Phe	Gln	Ser	Leu	Glu	Thr	Leu
				80				85				90		

Asp	Leu	Ser	Ser	Asn	Asn	Ile	Ser	Glu	Leu	Gln	Thr	Ala	Phe	Pro
				95				100				105		

Ala	Leu	Gln	Leu	Lys	Tyr	Leu	Tyr	Leu	Asn	Ser	Asn	Arg	Val	Thr
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

110	115	120
Ser Met Glu Pro Gly Tyr Phe Asp Asn Leu Ala Asn Thr Leu Leu		
125	130	135
Val Leu Lys Leu Asn Arg Asn Arg Ile Ser Ala Ile Pro Pro Lys		
140	145	150
Met Phe Lys Leu Pro Gln Leu Gln His Leu Glu Leu Asn Arg Asn		
155	160	165
Lys Ile Lys Asn Val Asp Gly Leu Thr Phe Gln Gly Leu Gly Ala		
170	175	180
Leu Lys Ser Leu Lys Met Gln Arg Asn Gly Val Thr Lys Leu Met		
185	190	195
Asp Gly Ala Phe Trp Gly Leu Ser Asn Met Glu Ile Leu Gln Leu		
200	205	210
Asp His Asn Asn Leu Thr Glu Ile Thr Lys Gly Trp Leu Tyr Gly		
215	220	225
Leu Leu Met Leu Gln Glu Leu His Leu Ser Gln Asn Ala Ile Asn		
230	235	240
Arg Ile Ser Pro Asp Ala Trp Glu Phe Cys Gln Lys Leu Ser Glu		
245	250	255
Leu Asp Leu Thr Phe Asn His Leu Ser Arg Leu Asp Asp Ser Ser		
260	265	270
Phe Leu Gly Leu Ser Leu Leu Asn Thr Leu His Ile Gly Asn Asn		
275	280	285
Arg Val Ser Tyr Ile Ala Asp Cys Ala Phe Arg Gly Leu Ser Ser		
290	295	300
Leu Lys Thr Leu Asp Leu Lys Asn Asn Glu Ile Ser Trp Thr Ile		
305	310	315
Glu Asp Met Asn Gly Ala Phe Ser Gly Leu Asp Lys Leu Arg Arg		
320	325	330
Leu Ile Leu Gln Gly Asn Arg Ile Arg Ser Ile Thr Lys Lys Ala		
335	340	345
Phe Thr Gly Leu Asp Ala Leu Glu His Leu Asp Leu Ser Asp Asn		
350	355	360
Ala Ile Met Ser Leu Gln Gly Asn Ala Phe Ser Gln Met Lys Lys		
365	370	375
Leu Gln Gln Leu His Leu Asn Thr Ser Ser Leu Leu Cys Asp Cys		
380	385	390
Gln Leu Lys Trp Leu Pro Gln Trp Val Ala Glu Asn Asn Phe Gln		
395	400	405
Ser Phe Val Asn Ala Ser Cys Ala His Pro Gln Leu Leu Lys Gly		
410	415	420

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

Arg Leu Ser Val Ile Pro Thr Pro Thr Cys Asp Ser Pro Gln Met  
 725 730 735  
 Thr Ala Pro Ser Leu Asp Asp Asp Gly Trp Ala Thr Val Gly Val  
 740 745 750  
 Val Ile Ile Ala Val Val Cys Cys Val Val Gly Thr Ser Leu Val  
 755 760 765  
 Trp Val Val Ile Ile Tyr His Thr Arg Arg Arg Asn Glu Asp Cys  
 770 775 780  
 Ser Ile Thr Asn Thr Asp Glu Thr Asn Leu Pro Ala Asp Ile Pro  
 785 790 795  
 Ser Tyr Leu Ser Ser Gln Gly Thr Leu Ala Asp Arg Gln Asp Gly  
 800 805 810  
 Tyr Val Ser Ser Glu Ser Gly Ser His His Gln Phe Val Thr Ser  
 815 820 825  
 Ser Gly Ala Gly Phe Phe Leu Pro Gln His Asp Ser Ser Gly Thr  
 830 835 840  
 Cys His Ile Asp Asn Ser Ser Glu Ala Asp Val Glu Ala Ala Thr  
 845 850 855  
 Asp Leu Phe Leu Cys Pro Phe Leu Gly Ser Thr Gly Pro Met Tyr  
 860 865 870  
 Leu Lys Gly Asn Val Tyr Gly Ser Asp Pro Phe Glu Thr Tyr His  
 875 880 885  
 Thr Gly Cys Ser Pro Asp Pro Arg Thr Val Leu Met Asp His Tyr  
 890 895 900  
 Glu Pro Ser Tyr Ile Lys Lys Lys Glu Cys Tyr Pro Cys Ser His  
 905 910 915  
 Pro Ser Glu Glu Ser Cys Glu Arg Ser Phe Ser Asn Ile Ser Trp  
 920 925 930  
 Pro Ser His Val Arg Lys Leu Leu Asn Thr Ser Tyr Ser His Asn  
 935 940 945  
 Glu Gly Pro Gly Met Lys Asn Leu Cys Leu Asn Lys Ser Ser Leu  
 950 955 960  
 Asp Phe Ser Ala Asn Pro Glu Pro Ala Ser Val Ala Ser Ser Asn  
 965 970 975  
 Ser Phe Met Gly Thr Phe Gly Lys Ala Leu Arg Arg Pro His Leu  
 980 985 990  
 Asp Ala Tyr Ser Ser Phe Gly Gln Pro Ser Asp Cys Gln Pro Arg  
 995 1000 1005  
 Ala Phe Tyr Leu Lys Ala His Ser Ser Pro Asp Leu Asp Ser Gly  
 1010 1015 1020  
 Ser Glu Glu Asp Gly Lys Glu Arg Thr Asp Phe Gln Glu Glu Asn

1025	1030	1035
------	------	------

His Ile Cys Thr Phe Lys Gln Thr Leu Glu Asn Tyr Arg Thr Pro  
 1040 1045 1050  
 Asn Phe Gln Ser Tyr Asp Leu Asp Thr  
 1055 1059

<210> 11  
 <211> 2906  
 <212> DNA  
 <213> artificial sequence

<400> 11  
 ggggagagga attgaccatg taaaaggaga cttttttttt tgggggtgg 50  
 ggctgttggg tgccttgcaa aatgaagga tgcaggacgc agctttctcc 100  
 tggAACCGAA cgcaatggat aaactgattt tgcaagagag aaggaagaac 150  
 gaagctttttt cttgtgagcc ctggatctta acacaaatgt gtatatgtc 200  
 acacagggag cattcaagaa tgaataaac cagagttaga cccgcggggg 250  
 ttgggtgtgtt ctgacataaa taaataatct taaaggcagct gttccctcc 300  
 ccaccccaa aaaaaaggat gattggaaat gaagaaccga ggattcacaa 350  
 agaaaaaaagt atgttcattt ttctctataa aggagaaaagt gagccaagga 400  
 gatatttttgaatgaaaatg tttgggctt ttttagtaaa gtaaaagaact 450  
 ggtgtgggtgg tgtttcctt tcttttgaa tttcccacaa gaggagagga 500  
 aattaataat acatctgcaa agaaatttca gagaagaaaa gttgaccgcg 550  
 gcagattgag gcattgattt ggggagagaa accagcagag cacagttgga 600  
 ttgtgccta tggtgactaa aattgacgga taattgcagt tggatttttc 650  
 ttcatcaacc tcctttttt taaattttta ttccctttgg tatcaagatc 700  
 atgcgttttc tcttggctt aaccacctgg atttccatct ggatgttgct 750  
 gtgatcagtc taaaatacaa ctgtttgaat tccagaagga ccaacaccag 800  
 ataaattatg aatgttgaac aagatgacct tacatccaca gcagataatg 850  
 ataggtccta ggttaacag ggcccttattt gacccctgc ttgtggtgct 900  
 gctggctctt caacttcttg tggggctgg tctggcgg gctcagacct 950  
 gcccttctgt gtgctcctgc agcaaccagt tcagcaaggt gatttgggtt 1000  
 cgaaaaacc tgcgtgaggt tccggatggc atctccacca acacacggct 1050  
 gctgaacctc catgagaacc aaatccagat catcaaagtg aacagcttca 1100  
 agcacttgag gcacttggaa atcctacagt tgagtaggaa ccatatcaga 1150  
 accattgaaa ttggggctt caatggctg gcgaacctca acactctgga 1200

actcttgac aatcgctta ctaccatccc gaatggagct tttgtatact 1250  
tgtctaaact gaaggagctc tgggtgcgaa acaacccat taaaagcatac 1300  
ccttcattatg ctttaacag aattcattct ttgcggccgac tagacttagg 1350  
gaaattgaaa agactttcat acatctcaga aggtgcctt gaaggctgt 1400  
ccaactttagt gtatttgaac cttgccatgt gcaaccttcg ggaaatccct 1450  
aacctcacac cgctcataaa actagatgag ctggatctt ctggaaatca 1500  
tttatctgcc atcaggcctg gctttcca gggtttgatg caccttcaaa 1550  
aactgtggat gatacagtcc cagattcaag tgattgaacg gaatgcctt 1600  
gacaaccttc agtcactagt ggagatcaac ctggcacaca ataatcta 1650  
attactgcct catgacctct tcactccctt gcatcatcta gagcggatac 1700  
atttacatca caacccttgg aactgttaact gtgacatact gtggctcagc 1750  
tgggtggataa aagacatggc cccctcgaac acagcttggt gtgcccggtg 1800  
taacactcct cccaatctaa aggggaggtt cattggagag ctcgaccaga 1850  
attacttcac atgctatgct ccgggtattt tggagcccc tgcagacctc 1900  
aatgtcactg aaggcatggc agctgagctg aaatgtcggg cctccacatc 1950  
cctgacatct gtatcttggta ttactccaaa tggAACAGTC atgacacatg 2000  
gggcgtacaa agtgcggata gctgtgctca gtgatggtaac gttaaatttc 2050  
acaaatgtaa ctgtgcaaga tacaggcatg tacacatgta tggtagttaa 2100  
ttccgttggg aatactactg cttcagccac cctgaatgtt actgcagcaa 2150  
ccactactcc tttcttttac tttcaaccg tcacagttaga gactatggaa 2200  
ccgtctcagg atgaggcactg gaccacagat aacaatgtgg gtcccactcc 2250  
agtggtcgac tgggagacca ccaatgtgac cacctctctc acaccacaga 2300  
gcacaagggtc gacagagaaa acttcacca tcccagtgac tgatataaac 2350  
agtgggatcc caggaattga tgaggtcatg aagactacca aaatcatcat 2400  
tgggtgtttt gtggccatca cactcatggc tgcagtgtatg ctggtcattt 2450  
tctacaagat gaggaagcag caccatcggtt aaaaccatca cgccccaaaca 2500  
aggactgttggaaattattaa tgtggatgtt gagattacgg gagacacacc 2550  
catggaaagc cacctgccc tgcctgctat cgagcatgag cacctaaatc 2600  
actataactc atacaaatct cccttcaacc acacaacaac agttaacaca 2650  
ataaattcaa tacacagtcc agtgcattgaa ccgttattga tccgaatgaa 2700

ctctaaagac aatgtacaag agactcaaat ctaaaacatt tacagagttt 2750

caaaaaacaa acaatcaaaa aaaaagacag tttttaaaaa atgacacaaa 2800

tgactggcct aaatctactg tttcaaaaaa gtgtcttac aaaaaaaca 2850

aaaagaaaaag aaatttattt attaaaaattt ctattgtgat ctaaagcaga 2900

caaaaa 2906

<210> 12

<211> 640

<212> PRT

<213> artificial sequence

<400> 12

Met Leu Asn Lys Met Thr Leu His Pro Gln Gln Ile Met Ile Gly  
1 5 10 15

Pro Arg Phe Asn Arg Ala Leu Phe Asp Pro Leu Leu Val Val Leu  
20 25 30

Leu Ala Leu Gln Leu Leu Val Val Ala Gly Leu Val Arg Ala Gln  
35 40 45

Thr Cys Pro Ser Val Cys Ser Cys Ser Asn Gln Phe Ser Lys Val  
50 55 60

Ile Cys Val Arg Lys Asn Leu Arg Glu Val Pro Asp Gly Ile Ser  
65 70 75

Thr Asn Thr Arg Leu Leu Asn Leu His Glu Asn Gln Ile Gln Ile  
80 85 90

Ile Lys Val Asn Ser Phe Lys His Leu Arg His Leu Glu Ile Leu  
95 100 105

Gln Leu Ser Arg Asn His Ile Arg Thr Ile Glu Ile Gly Ala Phe  
110 115 120

Asn Gly Leu Ala Asn Leu Asn Thr Leu Glu Leu Phe Asp Asn Arg  
125 130 135

Leu Thr Thr Ile Pro Asn Gly Ala Phe Val Tyr Leu Ser Lys Leu  
140 145 150

Lys Glu Leu Trp Leu Arg Asn Asn Pro Ile Glu Ser Ile Pro Ser  
155 160 165

Tyr Ala Phe Asn Arg Ile Pro Ser Leu Arg Arg Leu Asp Leu Gly  
170 175 180

Glu Leu Lys Arg Leu Ser Tyr Ile Ser Glu Gly Ala Phe Glu Gly  
185 190 195

Leu Ser Asn Leu Arg Tyr Leu Asn Leu Ala Met Cys Asn Leu Arg  
200 205 210

Glu Ile Pro Asn Leu Thr Pro Leu Ile Lys Leu Asp Glu Leu Asp  
215 220 225

Leu Ser Gly Asn His Leu Ser Ala Ile Arg Pro Gly Ser Phe Gln  
 230 235 240  
 Gly Leu Met His Leu Gln Lys Leu Trp Met Ile Gln Ser Gln Ile  
 245 250 255  
 Gln Val Ile Glu Arg Asn Ala Phe Asp Asn Leu Gln Ser Leu Val  
 260 265 270  
 Glu Ile Asn Leu Ala His Asn Asn Leu Thr Leu Leu Pro His Asp  
 275 280 285  
 Leu Phe Thr Pro Leu His His Leu Glu Arg Ile His Leu His His  
 290 295 300  
 Asn Pro Trp Asn Cys Asn Cys Asp Ile Leu Trp Leu Ser Trp Trp  
 305 310 315  
 Ile Lys Asp Met Ala Pro Ser Asn Thr Ala Cys Cys Ala Arg Cys  
 320 325 330  
 Asn Thr Pro Pro Asn Leu Lys Gly Arg Tyr Ile Gly Glu Leu Asp  
 335 340 345  
 Gln Asn Tyr Phe Thr Cys Tyr Ala Pro Val Ile Val Glu Pro Pro  
 350 355 360  
 Ala Asp Leu Asn Val Thr Glu Gly Met Ala Ala Glu Leu Lys Cys  
 365 370 375  
 Arg Ala Ser Thr Ser Leu Thr Ser Val Ser Trp Ile Thr Pro Asn  
 380 385 390  
 Gly Thr Val Met Thr His Gly Ala Tyr Lys Val Arg Ile Ala Val  
 395 400 405  
 Leu Ser Asp Gly Thr Leu Asn Phe Thr Asn Val Thr Val Gln Asp  
 410 415 420  
 Thr Gly Met Tyr Thr Cys Met Val Ser Asn Ser Val Gly Asn Thr  
 425 430 435  
 Thr Ala Ser Ala Thr Leu Asn Val Thr Ala Ala Thr Thr Pro  
 440 445 450  
 Phe Ser Tyr Phe Ser Thr Val Thr Val Glu Thr Met Glu Pro Ser  
 455 460 465  
 Gln Asp Glu Ala Arg Thr Thr Asp Asn Asn Val Gly Pro Thr Pro  
 470 475 480  
 Val Val Asp Trp Glu Thr Thr Asn Val Thr Thr Ser Leu Thr Pro  
 485 490 495  
 Gln Ser Thr Arg Ser Thr Glu Lys Thr Phe Thr Ile Pro Val Thr  
 500 505 510  
 Asp Ile Asn Ser Gly Ile Pro Gly Ile Asp Glu Val Met Lys Thr  
 515 520 525  
 Thr Lys Ile Ile Ile Gly Cys Phe Val Ala Ile Thr Leu Met Ala

530	535	540
Ala Val Met Leu Val Ile Phe Tyr Lys Met Arg Lys Gln His His		
545	550	555
Arg Gln Asn His His Ala Pro Thr Arg Thr Val Glu Ile Ile Asn		
560	565	570
Val Asp Asp Glu Ile Thr Gly Asp Thr Pro Met Glu Ser His Leu		
575	580	585
Pro Met Pro Ala Ile Glu His Glu His Leu Asn His Tyr Asn Ser		
590	595	600
Tyr Lys Ser Pro Phe Asn His Thr Thr Val Asn Thr Ile Asn		
605	610	615
Ser Ile His Ser Ser Val His Glu Pro Leu Leu Ile Arg Met Asn		
620	625	630
Ser Lys Asp Asn Val Gln Glu Thr Gln Ile		
635	640	

<210> 13  
 <211> 4051  
 <212> DNA  
 <213> artificial sequence

<400> 13  
 agccgacgct gctcaagctg caactctgtt gcagttggca gttctttcg 50  
 gttccctcc tgctgtttgg gggcatgaaa gggcttcgcc gccgggagta 100  
 aaagaaggaa ttgaccgggc agcgcgaggg aggagcgcgc acgcgaccgc 150  
 gagggcgggc gtgcaccctc ggctggaagt ttgtccggg ccccgagcgc 200  
 gcgcggcgtg ggagcttcgg gtagagacct aggccgctgg accgcgatga 250  
 gcgcgcccag cctccgtgcg cgccgcgcgg ggttggggct gctgctgtgc 300  
 gcggtgctgg ggccgcgtgg ccggtccgac agcggcggtc gcgggaaact 350  
 cgggcagccc tctgggttag ccgcccagcg cccatgcccc actacctgcc 400  
 gctgcctcgg ggacctgctg gactgcagtc gtaagcggct agcgcgtctt 450  
 cccgagccac tcccgtcctg ggtcgctcgg ctggacttaa gtcacaacag 500  
 attatctttc atcaaggcaa gttccatgag ccaccttcaa agccttcgag 550  
 aagtgaaact gaacaacaat gaattggaga ccattccaaa tctgggacca 600  
 gtctcggcaa atattacact tctctccttg gctggaaaca ggattgttg 650  
 aatactccct gaacatctga aagagttca gtcccttgaa actttggacc 700  
 ttagcagcaa caatattca gagctccaaa ctgcatttcc agccctacag 750  
 ctcaaatac tgcatactcaa cagcaaccga gtcacatcaa tggaacctgg 800

gtattttgac aatttggcca acacactcct tgtgttaaag ctgaacagga 850  
accgaatctc agctatccca cccaagatgt ttaaactgcc ccaactgcaa 900  
catctcgaat tgaaccgaaa caagattaaa aatgttagatg gactgacatt 950  
ccaaggcctt ggtgctctga agtctctgaa aatgcaaaga aatggagtaa 1000  
cgaaacttat ggtggagct ttttgggggc tgagcaacat ggaaattttg 1050  
cagctggacc ataacaacct aacagagatt accaaaggct ggcttacgg 1100  
cttgctgatg ctgcaggaac ttcatctcag ccaaaatgcc atcaacagga 1150  
tcagccctga tgcctggag ttctgccaga agctcagtga gctggaccta 1200  
actttcaatc acttatcaag gtttagatgat tcaagcttcc ttggcctaag 1250  
cttactaaat acactgcaca ttgggaacaa cagagtcagc tacattgctg 1300  
attgtgcctt ccgggggctt tccagttaa agactttgga tctgaagaac 1350  
aatgaaattt cctggactat tgaagacatg aatggtgctt tctctggct 1400  
tgacaaactg aggcgactga tactccaagg aaatcgatc cgttcttatta 1450  
ctaaaaaaagc cttcactggt ttggatgcat tggagcatct agacctgagt 1500  
acaacgcaat catgtcttta caaggcaatg cattttcaca aatgaagaaa 1550  
ctgcaacaat tgcatttaaa tacatcaagc cttttgcg attgccagct 1600  
aaaatggctc ccacagtggg tggcgaaaaa caactttcag agctttgtaa 1650  
atgccagttg tgcccatcct cagctgctaa aaggaagaag catttttgct 1700  
gttagcccaag atggctttgt gtgtgatgat tttcccaaac cccagatcac 1750  
ggtcagcca gaaacacagt cggcaataaa aggttccaat ttgagttca 1800  
tctgctcagc tgccagcagc agtgattccc caatgacttt tgcttgaaa 1850  
aaagacaatg aactactgca tcatgctgaa atggaaaatt atgcacaccc 1900  
ccgggccccaa ggtggcgagg tcatggagta taccaccatc ctccggctgc 1950  
gcgaggtgga atttgcaggat gaggggaaat atcagtggt catctccat 2000  
cactttggtt catcctactc tgtcaaagcc aagcttacag taaatatgct 2050  
tccctcattc accaagaccc ccatggatct caccatccga gctggggcca 2100  
tggcacgctt ggagtgtgct gctgtggggc acccagcccc ccagatagcc 2150  
tggcagaagg atggggcac agacttccca gctgcacggg agagacgcat 2200  
gcatgtgatg cccgaggatg acgtgttctt tatacgatgt gtgaagatag 2250  
aggacattgg ggtatacagc tgacagctc agaacagtgc aggaagtatt 2300  
tcagcaaatg caactctgac tgccttagaa acaccatcat ttttgcggcc 2350

actgttggac cgaactgtaa ccaaggaga aacagccgtc ctacagtgc 2400  
ttgctggagg aagccctccc cctaaactga actggaccaa agatgatagc 2450  
ccattggtgg taaccgagag gcacttttt gcagcaggca atcagcttct 2500  
gattattgtg gactcagatg tcagtgtgc tggaaatac acatgtgaga 2550  
tgtctaacac ccttggcact gagagaggaa acgtgcgcct cagtgtgatc 2600  
cccactccaa cctgcgactc ccctcagatg acagccccat cgtagacga 2650  
tgacggatgg gccactgtgg gtgtcgtgat catagccgtg gtttgctgtg 2700  
tggtgggcac gtcactcgtg tgggtggtca tcatatacca cacaaggcgg 2750  
aggaatgaag attgcagcat taccaacaca gatgagacca acttgccagc 2800  
agatattcct agttatttgt catctcaggg aacgttagct gacaggcagg 2850  
atgggtacgt gtcttcagaa agtggaaagcc accaccagtt tgtcacatct 2900  
tcaggtgctg gatTTTcTTT accacaacat gacagtagtg ggacctgcca 2950  
tattgacaat agcagtgaag ctgatgtgga agctgccaca gatctgttcc 3000  
ttgtccgttt ttgggatcca caggccctat gtatttgaag ggaaatgtgt 3050  
atggctcaga tcctttgaa acatatcata caggttgcag tcctgaccca 3100  
agaacagttt taatggacca ctatgagccc agttacataa agaaaaagga 3150  
gtgctaccca ttttctcatc cttcagaaga atcctgcgaa cggagcttca 3200  
gtaatatatc gtggccttca catgtgagga agctacttaa cactagttac 3250  
tctcacaatg aaggacctgg aatgaaaaat ctgtgtctaa acaagtcctc 3300  
tttagatttt agtgcaaatac cagagccagc gtcgggtgcc tcgagtaatt 3350  
cttcatggg tacctttgga aaagctctca ggagacctca cctagatgcc 3400  
tattcaagct ttggacagcc atcagattgt cagccaagag ctttttattt 3450  
gaaagctcat tttccccag acttggactc tgggtcagag gaagatggga 3500  
aagaaaggac agatttcag gaagaaaatc acattgtac cttaaacag 3550  
actttagaaaa actacaggac tccaaatttt cagtcttatg acttggacac 3600  
atagactgaa tgagaccaaa ggaaaagctt aacatactac ctcaagtgaa 3650  
cttttattta aaagagagag aatcttatgt ttttaaatg gagttatgaa 3700  
ttttaaaagg ataaaaatgc ttatattata cagatgaacc aaaattacaa 3750  
aaagttatga aaattttat actgggaatg atgctcatat aagaataacct 3800  
ttttaaacta ttttttaact ttgtttatg caaaaaagta tcttacgtaa 3850

attaatgata taaatcatga ttatTTATG tATTTTATA atGCCAGATT 3900  
 tCTTTTATG gAAAATGAGT tactAAAGCA tTTAAATAA tacCTGCCTT 3950  
 gtaccATTT tAAATAGAA gttactTCAT tatATTTGC acATTATATT 4000  
 taATAAAATG tgtcaATTG aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4050  
 a 4051

<210> 14  
 <211> 1119  
 <212> PRT  
 <213> artificial sequence

<400> 14  
 Met Ser Ala Pro Ser Leu Arg Ala Arg Ala Ala Gly Leu Gly Leu  
 1 5 10 15  
 Leu Leu Cys Ala Val Leu Gly Arg Ala Gly Arg Ser Asp Ser Gly  
 20 25 30  
 Gly Arg Gly Glu Leu Gly Gln Pro Ser Gly Val Ala Ala Glu Arg  
 35 40 45  
 Pro Cys Pro Thr Thr Cys Arg Cys Leu Gly Asp Leu Leu Asp Cys  
 50 55 60  
 Ser Arg Lys Arg Leu Ala Arg Leu Pro Glu Pro Leu Pro Ser Trp  
 65 70 75  
 Val Ala Arg Leu Asp Leu Ser His Asn Arg Leu Ser Phe Ile Lys  
 80 85 90  
 Ala Ser Ser Met Ser His Leu Gln Ser Leu Arg Glu Val Lys Leu  
 95 100 105  
 Asn Asn Asn Glu Leu Glu Thr Ile Pro Asn Leu Gly Pro Val Ser  
 110 115 120  
 Ala Asn Ile Thr Leu Leu Ser Leu Ala Gly Asn Arg Ile Val Glu  
 125 130 135  
 Ile Leu Pro Glu His Leu Lys Glu Phe Gln Ser Leu Glu Thr Leu  
 140 145 150  
 Asp Leu Ser Ser Asn Asn Ile Ser Glu Leu Gln Thr Ala Phe Pro  
 155 160 165  
 Ala Leu Gln Leu Lys Tyr Leu Tyr Leu Asn Ser Asn Arg Val Thr  
 170 175 180  
 Ser Met Glu Pro Gly Tyr Phe Asp Asn Leu Ala Asn Thr Leu Leu  
 185 190 195  
 Val Leu Lys Leu Asn Arg Asn Arg Ile Ser Ala Ile Pro Pro Lys  
 200 205 210  
 Met Phe Lys Leu Pro Gln Leu Gln His Leu Glu Leu Asn Arg Asn  
 215 220 225

Lys Ile Lys Asn Val Asp Gly Leu Thr Phe Gln Gly Leu Gly Ala  
 230 235 240  
 Leu Lys Ser Leu Lys Met Gln Arg Asn Gly Val Thr Lys Leu Met  
 245 250 255  
 Asp Gly Ala Phe Trp Gly Leu Ser Asn Met Glu Ile Leu Gln Leu  
 260 265 270  
 Asp His Asn Asn Leu Thr Glu Ile Thr Lys Gly Trp Leu Tyr Gly  
 275 280 285  
 Leu Leu Met Leu Gln Glu Leu His Leu Ser Gln Asn Ala Ile Asn  
 290 295 300  
 Arg Ile Ser Pro Asp Ala Trp Glu Phe Cys Gln Lys Leu Ser Glu  
 305 310 315  
 Leu Asp Leu Thr Phe Asn His Leu Ser Arg Leu Asp Asp Ser Ser  
 320 325 330  
 Phe Leu Gly Leu Ser Leu Leu Asn Thr Leu His Ile Gly Asn Asn  
 335 340 345  
 Arg Val Ser Tyr Ile Ala Asp Cys Ala Phe Arg Gly Leu Ser Ser  
 350 355 360  
 Leu Lys Thr Leu Asp Leu Lys Asn Asn Glu Ile Ser Trp Thr Ile  
 365 370 375  
 Glu Asp Met Asn Gly Ala Phe Ser Gly Leu Asp Lys Leu Arg Arg  
 380 385 390  
 Leu Ile Leu Gln Gly Asn Arg Ile Arg Ser Ile Thr Lys Lys Ala  
 395 400 405  
 Phe Thr Gly Leu Asp Ala Leu Glu His Leu Asp Leu Ser Asp Asn  
 410 415 420  
 Ala Ile Met Ser Leu Gln Gly Asn Ala Phe Ser Gln Met Lys Lys  
 425 430 435  
 Leu Gln Gln Leu His Leu Asn Thr Ser Ser Leu Leu Cys Asp Cys  
 440 445 450  
 Gln Leu Lys Trp Leu Pro Gln Trp Val Ala Glu Asn Asn Phe Gln  
 455 460 465  
 Ser Phe Val Asn Ala Ser Cys Ala His Pro Gln Leu Leu Lys Gly  
 470 475 480  
 Arg Ser Ile Phe Ala Val Ser Pro Asp Gly Phe Val Cys Asp Asp  
 485 490 495  
 Phe Pro Lys Pro Gln Ile Thr Val Gln Pro Glu Thr Gln Ser Ala  
 500 505 510  
 Ile Lys Gly Ser Asn Leu Ser Phe Ile Cys Ser Ala Ala Ser Ser  
 515 520 525  
 Ser Asp Ser Pro Met Thr Phe Ala Trp Lys Lys Asp Asn Glu Leu

530	535	540
Leu His Asp Ala Glu Met Glu Asn Tyr Ala His Leu Arg Ala Gln		
545	550	555
Gly Gly Glu Val Met Glu Tyr Thr Thr Ile Leu Arg Leu Arg Glu		
560	565	570
Val Glu Phe Ala Ser Glu Gly Lys Tyr Gln Cys Val Ile Ser Asn		
575	580	585
His Phe Gly Ser Ser Tyr Ser Val Lys Ala Lys Leu Thr Val Asn		
590	595	600
Met Leu Pro Ser Phe Thr Lys Thr Pro Met Asp Leu Thr Ile Arg		
605	610	615
Ala Gly Ala Met Ala Arg Leu Glu Cys Ala Ala Val Gly His Pro		
620	625	630
Ala Pro Gln Ile Ala Trp Gln Lys Asp Gly Gly Thr Asp Phe Pro		
635	640	645
Ala Ala Arg Glu Arg Arg Met His Val Met Pro Glu Asp Asp Val		
650	655	660
Phe Phe Ile Val Asp Val Lys Ile Glu Asp Ile Gly Val Tyr Ser		
665	670	675
Cys Thr Ala Gln Asn Ser Ala Gly Ser Ile Ser Ala Asn Ala Thr		
680	685	690
Leu Thr Val Leu Glu Thr Pro Ser Phe Leu Arg Pro Leu Leu Asp		
695	700	705
Arg Thr Val Thr Lys Gly Glu Thr Ala Val Leu Gln Cys Ile Ala		
710	715	720
Gly Gly Ser Pro Pro Pro Lys Leu Asn Trp Thr Lys Asp Asp Ser		
725	730	735
Pro Leu Val Val Thr Glu Arg His Phe Phe Ala Ala Gly Asn Gln		
740	745	750
Leu Leu Ile Ile Val Asp Ser Asp Val Ser Asp Ala Gly Lys Tyr		
755	760	765
Thr Cys Glu Met Ser Asn Thr Leu Gly Thr Glu Arg Gly Asn Val		
770	775	780
Arg Leu Ser Val Ile Pro Thr Pro Thr Cys Asp Ser Pro Gln Met		
785	790	795
Thr Ala Pro Ser Leu Asp Asp Gly Trp Ala Thr Val Gly Val		
800	805	810
Val Ile Ile Ala Val Val Cys Cys Val Val Gly Thr Ser Leu Val		
815	820	825
Trp Val Val Ile Ile Tyr His Thr Arg Arg Arg Asn Glu Asp Cys		
830	835	840

Ser Ile Thr Asn Thr Asp Glu Thr Asn Leu Pro Ala Asp Ile Pro  
 845 850 855  
 Ser Tyr Leu Ser Ser Gln Gly Thr Leu Ala Asp Arg Gln Asp Gly  
 860 865 870  
 Tyr Val Ser Ser Glu Ser Gly Ser His His Gln Phe Val Thr Ser  
 875 880 885  
 Ser Gly Ala Gly Phe Phe Leu Pro Gln His Asp Ser Ser Gly Thr  
 890 895 900  
 Cys His Ile Asp Asn Ser Ser Glu Ala Asp Val Glu Ala Ala Thr  
 905 910 915  
 Asp Leu Phe Leu Cys Pro Phe Leu Gly Ser Thr Gly Pro Met Tyr  
 920 925 930  
 Leu Lys Gly Asn Val Tyr Gly Ser Asp Pro Phe Glu Thr Tyr His  
 935 940 945  
 Thr Gly Cys Ser Pro Asp Pro Arg Thr Val Leu Met Asp His Tyr  
 950 955 960  
 Glu Pro Ser Tyr Ile Lys Lys Lys Glu Cys Tyr Pro Cys Ser His  
 965 970 975  
 Pro Ser Glu Glu Ser Cys Glu Arg Ser Phe Ser Asn Ile Ser Trp  
 980 985 990  
 Pro Ser His Val Arg Lys Leu Leu Asn Thr Ser Tyr Ser His Asn  
 995 1000 1005  
 Glu Gly Pro Gly Met Lys Asn Leu Cys Leu Asn Lys Ser Ser Leu  
 1010 1015 1020  
 Asp Phe Ser Ala Asn Pro Glu Pro Ala Ser Val Ala Ser Ser Asn  
 1025 1030 1035  
 Ser Phe Met Gly Thr Phe Gly Lys Ala Leu Arg Arg Pro His Leu  
 1040 1045 1050  
 Asp Ala Tyr Ser Ser Phe Gly Gln Pro Ser Asp Cys Gln Pro Arg  
 1055 1060 1065  
 Ala Phe Tyr Leu Lys Ala His Ser Ser Pro Asp Leu Asp Ser Gly  
 1070 1075 1080  
 Ser Glu Glu Asp Gly Lys Glu Arg Thr Asp Phe Gln Glu Glu Asn  
 1085 1090 1095  
 His Ile Cys Thr Phe Lys Gln Thr Leu Glu Asn Tyr Arg Thr Pro  
 1100 1105 1110  
 Asn Phe Gln Ser Tyr Asp Leu Asp Thr  
 1115 1119  
 <210> 15  
 <211> 22  
 <212> DNA

<213> artificial sequence

<400> 15  
atcggttgtga agtttagtgcc cc 22

<210> 16  
<211> 23  
<212> DNA  
<213> artificial sequence

<400> 16  
acctgcgata tccaaacagaa ttg 23

<210> 17  
<211> 48  
<212> DNA  
<213> artificial sequence

<400> 17  
ggaagaggat acagtcactc tggaaagtatt agtggctcca gcagttcc 48

<210> 18  
<211> 46  
<212> DNA  
<213> artificial sequence

<400> 18  
gggtacacacct gtcctgcac cgacggatat tggcttctgg aaggcc 46

<210> 19  
<211> 22  
<212> DNA  
<213> artificial sequence

<400> 19  
acagattcccc accagtgcaa cc 22

<210> 20  
<211> 21  
<212> DNA  
<213> artificial sequence

<400> 20  
cacactcggtt cacatcttgg c 21

<210> 21  
<211> 45  
<212> DNA  
<213> artificial sequence

<400> 21  
agggagcacg gacagtgtgc agatgtggac gagtgctcac tagca 45

<210> 22  
<211> 21  
<212> DNA  
<213> artificial sequence

<400> 22  
agagtgtatc tctggctacg c 21

<210> 23  
<211> 22  
<212> DNA  
<213> artificial sequence

<400> 23  
taagtccggc acattacagg tc 22

<210> 24  
<211> 49  
<212> DNA  
<213> artificial sequence

<400> 24  
cccacgatgt atgaatggtg gactttgtgt gactcctggt ttctgcac 49

<210> 25  
<211> 22  
<212> DNA  
<213> artificial sequence

<400> 25  
aaagacgcat ctgcgagtgt cc 22

<210> 26  
<211> 23  
<212> DNA  
<213> artificial sequence

<400> 26  
tgctgatttc acactgctct ccc 23

<210> 27  
<211> 24  
<212> DNA  
<213> artificial sequence

<400> 27  
tcgcggagct gtgttctgtt tccc 24

<210> 28  
<211> 50  
<212> DNA  
<213> artificial sequence

<400> 28  
tgatcgcat ggggacaaaag gcgcaagctc gagaggaaac tgttgtgcct 50

<210> 29  
<211> 20  
<212> DNA  
<213> artificial sequence

<400> 29  
acacctgggtt caaagatggg 20

<210> 30  
<211> 24  
<212> DNA  
<213> artificial sequence

<400> 30  
taggaagagt tgctgaaggc acgg 24

<210> 31  
<211> 20  
<212> DNA  
<213> artificial sequence

<400> 31  
ttgccttact caggtgctac 20

<210> 32  
<211> 20  
<212> DNA  
<213> artificial sequence

<400> 32  
actcagcagt ggttaggaaag 20

<210> 33  
<211> 22  
<212> DNA  
<213> artificial sequence

<400> 33  
gttggatctg ggcaacaata ac 22

<210> 34  
<211> 24  
<212> DNA  
<213> artificial sequence

<400> 34  
attgttgtgc aggctgagtt taag 24

<210> 35  
<211> 45  
<212> DNA  
<213> artificial sequence

<400> 35  
ggtggctata catggatagc aattacctgg acacgctgtc ccggg 45

<210> 36  
<211> 50  
<212> DNA  
<213> artificial sequence

<400> 36  
ggattctaat acgactcact atagggctgc ggcggctcag gtcttcagtt 50

<210> 37  
<211> 50  
<212> DNA  
<213> artificial sequence

<400> 37  
ctatgaaatt aaccctcact aaaggagca tggatgggg agggatacgg 50

<210> 38  
<211> 48

<212> DNA  
<213> artificial sequence

<400> 38  
ctatgaaatt aaccctcaact aaagggaata gcaggccatc ccaggaca 48

<210> 39  
<211> 47

<212> DNA  
<213> artificial sequence

<400> 39  
ctatgaaatt aaccctcaact aaaggatgt cttccatgcc aaccttc 47

1

1