PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C07K 14/47, C07H 21/04, C12N 5/06, 15/63, C12P 21/06, C12Q 1/68

A1

(11) International Publication Number: **WO 00/39161**

(43) International Publication Date: 6 July 2000 (06.07.00)

(21) International Application Number: PCT/US99/31328

(22) International Filing Date: 30 December 1999 (30.12.99)

(30) Priority Data:
09/224,669 31 December 1998 (31.12.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/224,669 (CON)
Filed on 31 December 1998 (31.12.98)

(71) Applicant *(for all designated States except US)*: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: BUSFIELD, Samantha, J. [AU/US]; Apartment #1, 15 Trowbridge Street, Cambridge, MA 02138 (US).

(74) Agent: MEIKLEJOHN, Anita, L.; Fish & Richardson, P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, 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).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CLASS II CYTOKINE RECEPTOR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING THEM

(57) Abstract

The invention provides isolated nucleic acid molecules, designated TANGO 241 and TANGO 242. These nucleic acid molecules encode transmembrane proteins that bear substantially sequence similarity to members of the type II cytokine receptor family. The invention also provides antisense nucleic acid molecules, expression vectors containing the nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a nucleic acid molecule of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided. The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes.

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			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

CLASS II CYTOKINE RECEPTOR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING THEM

RELATED APPLICATION INFORMATION

This application is a continuation-in-part of application serial no. 09/224,669, filed December 31, 1998.

Background of the Invention

Cytokine receptors convert an extracellular signal, the binding of a cytokine to its receptor, into an intracellular signal, generally the activation of an enzyme. Cytokine receptors are characteristically transmembrane proteins and have been grouped into several superfamilies based upon the presence of conserved structural motifs.

The members of class II cytokine receptor superfamily, also known as the interferon (IFN) receptor superfamily, are single spanning transmembrane glycoproteins characterized by the presence of one or two homologous extracellular regions of about 200 amino acids, each of which includes two fibronectin III (FNIII) domains. Members of the class II cytokine receptor superfamily include the IFNα/β receptor, the IFNγ receptor, interleukin-10 (IL-10) receptor, and the initiating protein of the coagulation cascade, tissue factor.

Ligands that interact with class II cytokine receptors, e.g., IFN-α, IFN-β, IFN-γ, and IL-10, exert important physiological effects on cells. The interferons have anti-viral activity, antiproliferative activity, and immunomodulatory activity. For example, IFN-α and IFN-β inhibit viral replication, activate natural killer cell lytic functions, and modulate MHC expression. IFN-γ is a potent anti-viral agent and an important immunomodulatory agent. IFN-γ induces expression of class I and class II MHC antigens, nitric oxide synthase, and several cytokines, including interleukin-1. IFN-γ plays a role in inflammatory responses, autoimmune diseases, and activation of macrophages. Moreover, IFN-γ is important in the natural

resolution of bacterial infections. IL-10 can inhibit cytokine production by macrophages and inhibit the accessory functions of macrophages in T-cell function. Specifically, IL-10 can inhibit TNF production triggered by endotoxin (LPS). Thus, IL-10 may reduce the lethality of septic shock.

The identification and characterization of cytokine a receptor permits the identification of both the ligands which bind to the receptor and the intracellular molecules and signal transduction pathways associated with the receptor, and permits the identification or design of modulators of receptor activity, e.g., receptor agonists or antagonists and modulators of signal transduction.

Summary of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding TANGO 241 and TANGO 242, both of which are predicted to be members of the class II cytokine receptor superfamily, also referred to as the interferon receptor superfamily. These receptors and fragments, derivatives, and variants of these receptors are collectively referred to as polypeptides of the invention or proteins of the invention. Nucleic acid molecules encoding polypeptides of the invention are collectively referred to as nucleic acids of the invention.

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, the present invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention features nucleic acid molecules which are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, 24 or the nucleotide sequence of the cDNA of a clone deposited with ATCC as Accession Number 20716 or 20717 (the "cDNA of a clone deposited as ATCC 20716 or 20717"), or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 50 (100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, or 2900) nucleotides of the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, 24 or the cDNA of a clone deposited as ATCC 20716 or 20717, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20 , 23 or the amino acid sequence encoded by the cDNA of a clone deposited as ATCC 20716 or 20717, or a complement thereof.

In preferred embodiments, the nucleic acid molecules have the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, 24 or the nucleotide sequence of the cDNA of a clone deposited as ATCC 20716 or 20717.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, or 23 the fragment including at least 15 (25, 30, 50, 100, 125, 150, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, or 600) contiguous amino acids of any of SEQ ID NOs: 2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, 23 or the polypeptide encoded by the cDNA of a clone deposited as ATCC 20716 or 20717.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, 23 or the amino acid sequence encoded by the cDNA of a clone deposited as ATCC 20716 or 20717, wherein the nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule having a nucleic acid sequence encoding any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof.

Also within the invention are: isolated polypeptides or proteins having an amino acid sequence that is at least about 65%, preferably 75%, 85%, 95%, or 98%

identical to the amino acid sequence of any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, or 23.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 65%, preferably 75%, 85%, or 95% identical the nucleic acid sequence encoding any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20 or 23 and isolated polypeptides or proteins which are encoded by a nucleic acid molecule consisting of the nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule consisting of the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof, or the non-coding strand of the cDNA of a clone deposited as ATCC 20716 or 20717.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, 23 or the amino acid sequence encoded by the cDNA of a clone deposited as ATCC 20716 or 20717, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes under stringent conditions to a nucleic acid molecule consisting of the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, and 13, 19, 21, 22, 24 or the cDNA of a clone deposited as ATCC 20716 or 20717, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 50 (75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, or 1800) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule consisting of the nucleotide sequence of any of SEQ ID NOs:1, 3, 11, 13, 19, 21, 22, and 24 or the cDNA of a clone deposited as ATCC 20716 or 20717, or a complement thereof. In some embodiments, the isolated nucleic acid molecules encode a cytoplasmic, transmembrane, or extracellular domain of a polypeptide of the invention. In other

embodiments, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, e.g., recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment the invention provides isolated host cells containing such a vector. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector encoding a polypeptide of the invention such that the polypeptide of the invention is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, and a functional activity of a polypeptide of the invention refers to an activity exerted by a protein or polypeptide of the invention on a responsive cell as determined in vivo, or in vitro, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein or an indirect activity, such as a cellular signaling activity mediated by interaction of the protein with a second protein. Thus, such activities include, e.g., (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3) the ability to bind to an intracellular target of the naturally-occurring polypeptide. Other activities include: (1) the ability to mediate cytokine signalling, e.g., cytokine signalling in a kinase (e.g., Janus kinase (JAK), src kinase, and/or MAP kinase)/Signal Transducer and Activators of Transcription (STAT) pathway; (2) the ability to modulate cytokine production by an immune cell, e.g., an activated immune cell; (3) the ability to modulate proliferation, differentiation, morphology, and/or function of a cell in which they are expressed, e.g., an immune cell, e.g., an immune cell having IFN, e.g., IFN alpha, IFN beta, IFN gamma, receptors and/or IL-10 receptors, e.g., a T cell, B cell, macrophage, mast cell, and/or natural killer cell; (4) the ability to modulate class I and/or class II MHC expression on a cell,

e.g., a macrophage; (5) the ability to act as a co-stimulator of the growth of hematopoietic cell lineages including, for example, T cells, B cells, and mast cells; (6) the ability modulate an inflammatory response, e.g., an inflammatory response to viral, bacterial, mycoplasma, and/or protozoan infection; and (7) the ability to modulate blood coagulation.

In one embodiment, a polypeptide of the invention has an amino acid sequence sufficiently identical to an identified domain of a polypeptide of the invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 65% identity, preferably 75% identity, more preferably 85%, 95%, or 98% identity are defined herein as sufficiently identical.

In one embodiment, the isolated polypeptide of the invention lacks both a transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

The polypeptides of the present invention, or biologically active portions thereof, can be operably linked to a heterologous amino acid sequence to form fusion proteins. The invention further features antibodies that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies. In addition, the polypeptides of the invention or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides methods for detecting the presence of the activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of

detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small molecule.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of a polypeptide of the invention wherein a wild-type form of the gene encodes a polypeptide having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in

the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence of human TANGO 241 (SEQ ID NO:1) and the predicted amino acid sequence of TANGO 241 (SEQ ID NO:2). The 1722 nucleotide open reading frame of SEQ ID NO:1 extends from nucleotide 58 to 1779 of SEQ ID NO:1 (SEQ ID NO:3). The single underscored region is the signal sequence, and the double underscored region is the transmembrane domain.

Figure 2 depicts a hydropathy plot of human TANGO 241. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line.

Figure 3 depicts the cDNA sequence of human TANGO 242 (SEQ ID NO:11) and predicted amino acid sequence of TANGO 242 (SEQ ID NO:12). The 933 nucleotide open reading frame of SEQ ID NO:11 extends from nucleotide 71 to 1003 of SEQ ID NO:11 (SEQ ID NO:13). The single underscored region is the signal sequence, and the double underscored region is the transmembrane domain.

Figure 4 depicts a hydropathy plot of a human TANGO 242. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line.

Figure 5 depicts an alignment of human IL-10 receptor (SwissProt Accession No. Q13651; SEQ ID NO:19) with human TANGO 241 (SEQ ID NO:2). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the proteins are 21.2% identical. Identical amino acid residues are indicated by a "." between the

aligned amino acids. Similar amino acid residues are indicated by a ":" between the aligned amino acids.

Figure 6 depicts an alignment of human IL-10 receptor (SwissProt Accession No. Q13651; SEQ ID NO:19) with human TANGO 242 (SEQ ID NO:12). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the proteins are 14.9% identical. Identical amino acid residues are indicated by a "." between the aligned amino acids. Similar amino acid residues are indicated by a ":" between the aligned amino acids.

Figures 7A-7B depict a partial cDNA sequence of murine TANGO 241 (SEQ ID NO:19) and the predicted amino acid sequence of murine TANGO 241 (SEQ ID NO:20). The 1060 nucleotide open reading frame of SEQ ID NO:19 extends from nucleotide 1 to 1060 of SEQ ID NO:19 (SEQ ID NO:21) and encodes a 353 amino acid protein.

Figures 8A-8B depict the cDNA sequence of murine TANGO 242 (SEQ ID NO:22) and predicted amino acid sequence of murine TANGO 242 (SEQ ID NO:23). The 924 nucleotide open reading frame of SEQ ID NO:22 extends from nucleotide 107 to 1031 of SEQ ID NO:22 (SEQ ID NO:24) and encodes a 308 amino acid protein.

Figures 9A-9C depict an alignment of a portion of the cDNA sequence of human TANGO 241 (upper line of each pair; SEQ ID NO:25) with the cDNA sequence of murine TANGO 241 (lower line of each pair; SEQ ID NO:22). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 68.6% identical.

Figures 10A-10B depict an alignment of the partial ORF sequence of murine TANGO 241 (upper line of each pair; SEQ ID NO:21) with a portion of the ORF sequence of human TANGO 241 (lower line of each pair; SEQ ID NO:26). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty), the sequences are 78.4% identical.

Figure 11 depicts an alignment of the partial amino acid sequence of murine TANGO 241 (upper line of each pair; SEQ ID NO:20) with a portion of the amino acid sequence of human TANGO 241 (lower line of each pair; SEQ ID NO:27). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 72.5% identical.

Figures 12A-12D depict an alignment of the cDNA sequence of human TANGO 242 (upper line of each pair; SEQ ID NO:11) with the cDNA sequence of murine TANGO 242 (lower line of each pair; SEQ ID NO:22). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 71.1% identical.

Figures 13A-13B depict an alignment of the ORF sequence of human TANGO 242 (upper line of each pair; SEQ ID NO:13) with the ORF sequence of murine TANGO 242 (lower line of each pair; SEQ ID NO:24). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 82% identical.

Figure 14 depicts an alignment of the amino acid sequence of human TANGO 242 (upper line of each pair; SEQ ID NO:12) with the amino acid sequence of murine TANGO 242 (lower line of each pair; SEQ ID NO:23). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman, (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 77.5% identical.

Detailed Description of the Invention

The present invention is based, in part, on the discovery of cDNA molecules encoding TANGO 241 and TANGO 242, transmembrane proteins which are predicted to be members of the class II cytokine receptor superfamily, also referred to as the interferon receptor superfamily.

Human TANGO 241

In one aspect, the present invention is based, in part, on the discovery of a cDNA molecule encoding a TANGO 241, a protein having sequence similarity to members of the class II cytokine receptor family and having both a fibronectin III domain and a box 3 element.

TANGO 241 proteins and nucleic acid molecules comprise a family of molecules having certain conserved structural and functional features. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin. Members of the same family also have common structural domains.

Also included within the scope of the invention are TANGO 241 proteins having a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the N-terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 241 protein contains a signal sequence corresponding to amino acids 1-14 of SEQ ID NO:2 (SEQ ID NO:4). The signal sequence is cleaved during processing of the mature protein.

TANGO 241 proteins can also include an extracellular domain. The human TANGO 241 extracellular domain is located from amino acid 15 to amino acid 226 of SEQ ID NO:2. Within the extracellular domain, TANGO 241 proteins typically also have a fibronectin III domain. As used herein, the term "fibronectin III

domain" refers to a protein domain that includes about 50-105 amino acid residues, more preferably about 60-105 amino acid residues, and most preferably about 65-95 amino acid residues. In addition, a fibronectin III domain includes at least the following consensus sequence: W-Xaa(n1)-P-Xaa(n2)-Y-Xaa(n3)-Y-Xaa(n4)-Y-Xaa(n5)-R-V-Xaa(n6)-A, wherein W is a tryptophan residue, Xaa is any amino acid, n1 is about 1-10 amino acid residues in length, and more preferably about 1-7 amino acid residues in length, n2 is about 1-10 amino acid residues in length, more preferably about 2-10 amino acid residues in length, and most preferably about 3-9 amino acid residues in length, n3 is about 1-5 amino acid residues in length, more preferably about 2-4 amino acid residues in length, and most preferably about 3 amino acid residues in length, n4 is about 20-50 amino acid residues in length, preferably about 25-45 amino acid residues in length, and most preferably about 30-45 amino acid residues in length, n5 is about 1-5 amino acid residues in length, more preferably about 2-4 amino acid residues in length, and most preferably 2 amino acid residues, n6 is about 1-5 amino acid residues in length, R is an arginine residue, V is a valine residue, and A is an alanine residue.

In one embodiment, a TANGO 241 protein includes a fibronectin III domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 38 to 105 of SEQ ID NO:2. In another embodiment, a TANGO 241 protein includes a fibronectin III domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 38 to 105 of SEQ ID NO:2 and includes a fibronectin III consensus sequence as described herein. In yet another embodiment, a TANGO 241 protein includes a fibronectin III domain having an amino acid sequence that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 38 to 105 of SEQ ID NO:2, includes a fibronectin III consensus sequence as described herein and has at least one biological activity as described herein.

In a preferred embodiment, a TANGO 241 protein has the amino acid sequence of SEQ ID NO:2 wherein the fibronectin III domain is located from amino acid 38 to amino acid 105 and the fibronectin consensus sequence is located from amino acid 41 to amino acid 98.

There are also three N-linked glycosylation sites in the extracellular domain of human TANGO 241 protein at positions 80, 87, and 172 of SEQ ID NO:2.

The present invention includes TANGO 241 proteins having a cytoplasmic domain. The human TANGO 241 cytoplasmic domain is located from amino acid 252 to amino acid 574 of SEQ ID NO:2. Within the extracellular domain, TANGO 241 proteins typically also have a Box 3 element. As used herein, the term "Box 3 element" refers to a protein domain that includes about 5-15 amino acid residues, more preferably 8-5, and most preferably about 9-12 amino acid residues. Typically, a Box 3 element includes at least the following consensus sequence (see e.g., Stahl et al. (1995) Science 267:1349-1353): Y-Xaa-Xaa-Q, wherein Y is a tyrosine residue, Xaa is any amino acid, and Q is a glutamine residue. In a preferred embodiment, the Box 3 element is the portion of the TANGO 241 protein which is involved in binding to a STAT, e.g., STAT 1-6 (see e.g., Heim (1996) Eur. J. Clin. Invest. 26:1-12, which describes the various STAT).

In one embodiment, a TANGO 241 protein includes a Box 3 element having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 372 to 381 of SEQ ID NO:2. In another embodiment, a TANGO 241 protein includes a Box 3 element having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 372 to 381 of SEQ ID NO:2 and a Box 3 element consensus sequence as described herein. In yet another embodiment, a TANGO 241 protein includes a Box 3 element having an amino acid sequence that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 372 to 381 of SEQ ID NO:2,

includes a Box 3 element consensus sequence as described herein and has at least one biological activity as described herein.

In a preferred embodiment, a TANGO 241 protein has the amino acid sequence of SEQ ID NO:2 wherein the Box 3 element is located from amino acid 372 to amino acid 381 and the Box 3 element consensus sequence is located from amino acid 377 to amino acid 380.

The present invention also includes TANGO 241 proteins having a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence having at least about 20 to 25 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a preferred embodiment, a transmembrane domain contains at least about 15 to 30 amino acid residues, preferably about 20-25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. Thus, in one embodiment, a TANGO 241 protein contains a transmembrane domain corresponding to amino acids 227-251 of SEQ ID NO:2 (SEQ ID NO:7).

In one embodiment, a TANGO 241 protein of the invention includes a fibronectin III domain or a box 3 element. In another embodiment, a TANGO 241 protein of the invention includes both a fibronectin III domain and a box 3 element. In another embodiment, a TANGO 241 protein of the invention includes a fibronectin III domain, a Box 3 element, and a transmembrane domain.

A cDNA encoding human TANGO 241 was isolated from a human esophagus library. An initial clone was selected based on its sequence similarity to IFNα/β receptor genes. The initial clone was used to isolate a full length TANGO 241 cDNA clone (AthEa20d7). Analysis of the full length TANGO 241 cDNA revealed that TANGO 241 protein has significant sequence similarity to members of the class II cytokine receptor superfamily, e.g., IFNα/β receptor and IL-10 receptor.

The full length human TANGO 241 cDNA (Figure 1; SEQ ID NO:1) is 2856 nucleotides long. The open reading frame of this cDNA, nucleotides 58 to 1779 of

SEQ ID NO:1 (SEQ ID NO:3), encodes a 574 amino acid transmembrane protein (Figure 1; SEQ ID NO:2).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that human TANGO 241 includes a 14 amino acid signal peptide (amino acids 1-14 of SEQ ID NO:2; SEQ ID NO:4) preceding the mature TANGO 241 protein (amino acid 15-574 of SEQ ID NO:2; SEQ ID NO:5). Human TANGO 241 includes an extracellular domain (amino acids 15-226 of SEQ ID NO:2; SEQ ID NO:6); a transmembrane (TM) domain (amino acids 227-251 of SEQ ID NO:2; SEQ ID NO:7); and a cytoplasmic domain (amino acids 252-574 of SEQ ID NO:2; SEQ ID NO:8).

The extracellular domain of human TANGO 241 includes a fibronectin III domain (amino acids 38-105 of SEQ ID NO:2; SEQ ID NO:9). The cytoplasmic domain of human TANGO 241 includes a Box-3-like element (amino acids 372-381 of SEQ ID NO:2; SEQ ID NO:10).

The predicted molecular weight of the entire human TANGO 241 protein without modification and prior to cleavage of the signal sequence is about 62.9 kDa. The predicted molecular weight of the mature human TANGO 241 protein without modification and after cleavage of the signal sequence about 61.5 kDa.

Clone AthEa20d7, encoding human TANGO 241, inserted into pMET7 vector and designated plasmid ApAthEa20d7 was deposited with the American Type Culture Collection (ATCC, 10801 University Boulevard, Manassas, VA 20110-2209) on December 30, 1998 and was assigned Accession Number 20716. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. 112.

Figure 2 depicts a hydropathy plot of human TANGO 241. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The hydrophobic region which corresponds to amino acids 1-14 of SEQ ID NO:2 is the signal sequence of TANGO 241 (SEQ ID NO:4).

The hydrophobic region which corresponds to amino acids 227-251 of SEQ ID NO:2 is the transmembrane domain of TANGO 241 (SEQ ID NO:7).

TANGO 241 protein exhibits considerable sequence similarity to human IL-10 receptor. Figure 5 depicts an alignment of human IL-10 receptor (SwissProt Accession No. Q13651; SEQ ID NO:19) with human TANGO 241 (SEQ ID NO:2). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the proteins are 21.2% identical.

Human TANGO 241 exhibits considerable similarity to interferon alpha/beta receptor 2 (approximately 30% identity and 50% similarity at the amino acid level).

Northern analysis of TANGO 241 expression in human tissues revealed that TANGO 241 is expressed at a high level in the pancreas as an approximately 3.0 kb transcript. No expression was detected in the adrenal medulla, thyroid, adrenal cortex, testis, thymus, small intestine, or stomach.

Genomic mapping of human TANGO 241 revealed that it maps to human chromosomal location 1p36.

Murine TANGO 241

A partial murine TANGO 241 cDNA (Figure 7; SEQ ID NO:19) was identified in a murine esophagus library (clone ftmEa241a5). This cDNA is 1596 nucleotides long and includes a open reading frame (nucleotides 1 to 1060 of SEQ ID NO:19; SEQ ID NO:21) encoding a 354 amino acid protein (Figure 7; SEQ ID NO:20).

Thus, in another aspect, the present invention is based, in part, on the discovery of a cDNA molecule encoding murine TANGO 241. Also included within the scope of the invention are murine TANGO 241 proteins having a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the N-terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least

about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. The signal sequence is cleaved during processing of the mature protein.

Murine TANGO 241 proteins can also include an extracellular domain. Within the extracellular domain, TANGO 241 proteins typically also have a fibronectin III domain. The present invention also includes TANGO 241 proteins having a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence having at least about 20 to 25 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a preferred embodiment, a transmembrane domain contains at least about 15 to 30 amino acid residues, preferably about 20-25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues.

Figures 9A-9C depict an alignment of a portion of the cDNA sequence of human TANGO 241 (upper line of each pair; SEQ ID NO:25) with the cDNA sequence of murine TANGO 241 (lower line of each pair; SEQ ID NO:22). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 68.6% identical.

Figures 10A-10B depict an alignment of the partial ORF sequence of murine TANGO 241 (upper line of each pair; SEQ ID NO:21) with a portion of the ORF sequence of human TANGO 241 (lower line of each pair; SEQ ID NO:26). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 78.4% identical.

Figure 11 depicts an alignment of the partial amino acid sequence of murine TANGO 241 (upper line of each pair; SEQ ID NO:20) with a portion of the amino acid sequence of human TANGO 241 (lower line of each pair; SEQ ID NO:27). In this alignment, created using the Wisconsin™ BestFit software (Smith and

Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 72.5% identical.

A mouse library array was screened to identify tissues and conditions in which murine TANGO 241 is expressed. The library array was screened using PCR to amplify any murine TANGO 241 cDNA present in the individual libraries. The PCR amplification employed a primer for the vector used to create the library and an internal murine TANGO 241 primer. The following libraries were screened: differentiated 3T3 cells;10.5 day mouse fetus; mouse kidney fibrosis model nephrotoxic serum (NTS); LPS-stimulated heart; LPS-stimulated osteoblasts, 1 hour; lung, chronic boyle model, d24/72h, from 4 mice; normal spleen (random primed); 11.5 day mouse; LPS-stimulated lung; lung, day 15, 3 hour Gonzolo inflammation model; LPS-stimulated osteoblasts 24 hour; BL6 lung, day 15, 3 hour Gonzolo inflammation model; LPS-stimulated lung; 12.5 day mouse; LPS-stimulated kidney; LPS-stimulated lymph node; LPS-stimulated osteoblasts, 24 hours; esophagus; choroid plexus; 13.5 day mouse; LPS-stimulated anchored heart; normal thymus; Th2-ova-Tg; brain (random-primed); Balb C liver (bile duct ligation d2); mc/9 mast cells; normal heart; brain polysome (MPB); brain (EAE d10 model); th1-ova-Tg; heart (random primed); long-term bone marrow cells; LPS-stimulated lung, random primed; megakaryocyte; LPS-stimulated spleen; MLTC-1 (leydig cells) (testis); lung (random primed); Th2r; Th2; brain; LPS-stimulated thymus; TM4 (sertoli cells); LPS-stimulated microglial cells; Gonzolo day 15; Th1; Gonzolo day 21, 3 hour; LPS-stimulated brain; LPS-stimulated Alveolar macrophage cell line; mouse lung bleomycin Model d7; and pregnant uterus. Of these libraries, TANGO 241 expression was detected only in the esophagus library.

The predicted molecular weight of the murine TANGO 241 protein encoded by the cDNA described above without modification and prior to cleavage of the signal sequence is about 39.3 kDa.

Human TANGO 242

In another aspect, the present invention is based, in part, on the discovery of a cDNA molecule encoding a human TANGO 242, a protein having sequence

similarity to members of the class II cytokine receptor family and having a fibronectin III domain.

TANGO 242 proteins and nucleic acid molecules comprise a family of molecules having certain conserved structural and functional features. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin. Members of the same family also have common structural domains.

Also included within the scope of the invention are TANGO 242 proteins having a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the N-terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 15 to 30 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a human TANGO 242 protein contains a signal sequence corresponding to amino acids 1-29 of SEQ ID NO:12 (SEQ ID NO:14). The signal sequence is cleaved during processing of the mature protein.

TANGO 242 proteins can also include an extracellular domain. The human TANGO 242 extracellular domain is located from amino acid 30 to amino acid 230 of SEQ ID NO:12. Within the extracellular domain, TANGO 242 proteins typically also have a fibronectin III domain. As used herein, the term "fibronectin III domain" refers to a protein domain that includes about 50-105 amino acid residues, and more preferably about 50-105 amino acid residues, more preferably about 60-105 amino acid residues, and most preferably about 6595 amino acid residues. In

addition, a fibronectin III domain includes at least the following consensus sequence: W-Xaa(n1)-P-Xaa(n2)-Y-Xaa(n3)-Y-Xaa(n4)-Y-Xaa(n5)-R-V-Xaa(n6)-A, wherein W is a tryptophan residue, Xaa is any amino acid, n1 is about 1-10 amino acid residues in length, and more preferably about 1-7 amino acid residues in length, n2 is about 1-10 amino acid residues in length, more preferably about 2-10 amino acid residues in length, and most preferably about 3-9 amino acid residues in length, n3 is about 1-5 amino acid residues in length, more preferably about 2-4 amino acid residues in length, and most preferably about 3 amino acid residues in length, n4 is about 20-50 amino acid residues in length, preferably about 25-45 amino acid residues in length, and most preferably about 30-45 amino acid residues in length, n5 is about 1-5 amino acid residues in length, more preferably about 2-4 amino acid residues in length, and most preferably 2 amino acid residues, n6 is about 1-5 amino acid residues in length, R is an arginine residue, V is a valine residue, and A is an alanine residue.

In one embodiment, a TANGO 242 protein includes a fibronectin III domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 35-125 of SEQ ID NO:12. In another embodiment, a TANGO 242 protein includes a fibronectin III domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 35 to 125 of SEQ ID NO:12 and includes a fibronectin III consensus sequence as described herein. In yet another embodiment, a TANGO 242 protein includes a fibronectin III domain having an amino acid sequence that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 35 to 125 of SEQ ID NO:12, includes a fibronectin III consensus sequence as described herein and has at least one biological activity as described herein.

In a preferred embodiment, a TANGO 242 protein has the amino acid sequence of SEQ ID NO:12 wherein the fibronectin III domain is located from

amino acid 35 to amino acid 125 and the fibronectin consensus sequence is located from amino acid 54 to amino acid 115.

There are also two N-linked glycosylation sites in the extracellular domain of human TANGO 242 protein at positions 40 and 134 of SEQ ID NO:12.

The present invention also includes TANGO 242 proteins having a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence having at least about 20 to 25 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a preferred embodiment, a transmembrane domain contains at least about 15 to 30 amino acid residues, preferably about 20-25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. Thus, in one embodiment, a TANGO 242 protein contains a transmembrane domain corresponding to amino acids 231-255 of SEQ ID NO:12 (SEQ ID NO:17).

In one embodiment, a TANGO 242 protein of the invention includes a fibronectin III domain. In another embodiment, a TANGO 242 protein of the invention includes both a fibronectin III domain and transmembrane domain. In another embodiment, a TANGO 242 protein of the invention includes a fibronectin III domain, a transmembrane domain, and a signal sequence.

A cDNA encoding human TANGO 242 was isolated from a human esophagus library. An initial clone was selected based on its sequence similarity to IFNγ receptor genes. The initial clone was used to isolate a full length TANGO 242 cDNA clone (AthEa89c8). Analysis of the full length TANGO 242 cDNA revealed that TANGO 242 protein has significant sequence similarity to members of the class II cytokine receptor superfamily, e.g., IFNα/β receptor and IL-10 receptor.

The full length human TANGO 242 cDNA (Figure 3; SEQ ID NO:11) is 1832 nucleotides long. The 933 nucleotide open reading frame of this cDNA, nucleotides 71 to 1003 of SEQ ID NO:11 (SEQ ID NO:13), encodes a 311 amino acid transmembrane protein (Figure 3; SEQ ID NO:12).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that human TANGO 242 includes a 29 amino acid signal peptide (amino acids 1-29 of SEQ ID NO:12; SEQ ID NO:14) preceding the mature TANGO 242 protein (amino acids 30-311 of SEQ ID NO:12; SEQ ID NO:15). Human TANGO 242 includes an extracellular domain (amino acids 30-230 of SEQ ID NO:12; SEQ ID NO:16); a transmembrane (TM) domain (amino acids 231-255 of SEQ ID NO:12; SEQ ID NO:17); and a cytoplasmic domain (amino acid 256-311 of SEQ ID NO:12; SEQ ID NO:18).

The extracellular domain of human TANGO 242 includes a fibronectin III domain (amino acids 35-125 of SEQ ID NO:12; SEQ ID NO:19).

The predicted molecular weight of the entire human TANGO 242 protein without modification and prior to cleavage of the signal sequence is about 35.1 kDa. The predicted molecular weight of the mature human TANGO 242 protein without modification and after cleavage of the signal sequence about 31.5 kDa.

Clone AthEa89c8, encoding human TANGO 242, inserted into pMET7 vector and designated plasmid ApAthEa89c8 was deposited with the American Type Culture Collection (ATCC, 10801 University Boulevard, Manassas, VA 20110-2209) on December 30, 1998 and was assigned Accession Number 20717. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. 112.

Figure 4 depicts a hydropathy plot of human TANGO 242. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. As shown in the hydropathy plot, the hydrophobic region at amino acids 1-29 of SEQ ID NO:12 is the signal sequence of TANGO 242, and the hydrophobic regions at amino acids 231-255 of SEQ ID NO:12 is the transmembrane domain of TANGO 242.

TANGO 242 protein exhibits considerable sequence similarity to human IL-10 receptor. Figure 6 depicts an alignment of human IL-10 receptor (SwissProt Accession No. Q13651; SEQ ID NO:19) with human TANGO 242 (SEQ ID

NO:12). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the proteins are 14.9% identical.

Human TANGO 242 exhibits considerable similarity to interferon alpha/beta receptor 2 (approximately 30% identity and 47% similarity at the amino acid level).

Northern analysis of TANGO 242 expression in human tissues revealed that TANGO 241 is expressed at a high level in the brain as an approximately 2.6 kb transcript. No expression was detected in the heart, spleen, lung, liver, skeletal muscle, kidney or testis.

Genomic mapping of human TANGO 242 revealed that it maps to human chromosomal location 3q21.

Murine TANGO 242

A murine TANGO 242 cDNA was identified in a murine esophagus library (clone jtmEa242a01). Figures 8A-8B depict the cDNA 2405 nucleotide sequence of murine TANGO 242 (SEQ ID NO:22) and predicted amino acid sequence of murine TANGO 242 (SEQ ID NO:23). The 1218 nucleotide open reading frame of SEQ ID NO:22 extends from nucleotide 107 to 1031 of SEQ ID NO:22 (SEQ ID NO:24).

Thus, in another aspect, the present invention is based, in part, on the discovery of a cDNA molecule encoding a murine TANGO 242, a protein having sequence similarity to members of the class II cytokine receptor family and having a fibronectin III domain.

Also included within the scope of the invention are murine TANGO 242 proteins having a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the N-terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 15 to 30 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. TANGO 242

proteins can also include an extracellular domain. Within the extracellular domain, TANGO 242 proteins typically also have a fibronectin III domain. As used herein, the term "fibronectin III domain" refers to a protein domain that includes about 50-105 amino acid residues, and more preferably about 50-105 amino acid residues, more preferably about 60-105 amino acid residues, and most preferably about 65-95 amino acid residues. In addition, a fibronectin III domain includes at least the following consensus sequence: W-Xaa(n1)-P-Xaa(n2)-Y-Xaa(n3)-Y-Xaa(n4)-Y-Xaa(n5)-R-V-Xaa(n6)-A, wherein W is a tryptophan residue, Xaa is any amino acid, n1 is about 1-10 amino acid residues in length, and more preferably about 1-7 amino acid residues in length, n2 is about 1-10 amino acid residues in length, more preferably about 2-10 amino acid residues in length, and most preferably about 3-9 amino acid residues in length, n3 is about 1-5 amino acid residues in length, more preferably about 2-4 amino acid residues in length, and most preferably about 3 amino acid residues in length, n4 is about 20-50 amino acid residues in length, preferably about 25-45 amino acid residues in length, and most preferably about 30-45 amino acid residues in length, n5 is about 1-5 amino acid residues in length, more preferably about 2-4 amino acid residues in length, and most preferably 2 amino acid residues, n6 is about 1-5 amino acid residues in length, R is an arginine residue, V is a valine residue, and A is an alanine residue.

In one embodiment, a TANGO 242 protein includes a fibronectin III domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 30-120 of SEQ ID NO:23. In another embodiment, a TANGO 242 protein includes a fibronectin III domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 30 to 120 of SEQ ID NO:23 and includes a fibronectin III consensus sequence as described herein. In yet another embodiment, a TANGO 242 protein includes a fibronectin III domain having an amino acid sequence that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%,

and most preferably at least about 95% identical to amino acids 30 to 120 of SEQ ID NO:23, includes a fibronectin III consensus sequence as described herein and has at least one biological activity as described herein.

The present invention also includes murine TANGO 242 proteins having a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence having at least about 20 to 25 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a preferred embodiment, a transmembrane domain contains at least about 15 to 30 amino acid residues, preferably about 20-25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues.

In one embodiment, a murine TANGO 242 protein of the invention includes a fibronectin III domain. In another embodiment, a TANGO 242 protein of the invention includes both a fibronectin III domain and transmembrane domain. In another embodiment, a TANGO 242 protein of the invention includes a fibronectin III domain, a transmembrane domain, and a signal sequence.

Figures 12A-12D depict an alignment of the cDNA sequence of human TANGO 242 (upper line of each pair; SEQ ID NO:11) with the cDNA sequence of murine TANGO 242 (lower line of each pair; SEQ ID NO:22). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 71.1% identical.

Figures 13A-13B depict an alignment of the ORF sequence of human TANGO 242 (upper line of each pair; SEQ ID NO:13) with the ORF sequence of murine TANGO 242 (lower line of each pair; SEQ ID NO:24). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman, (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 82% identical.

Figure 14 depicts an alignment of the amino acid sequence of human TANGO 242 (upper line of each pair; SEQ ID NO:12) with the amino acid sequence

of murine TANGO 242 (lower line of each pair; SEQ ID NO:23). In this alignment, created using the Wisconsin™ BestFit software (Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, gap opening penalty 10, gap extension penalty 10), the sequences are 77.5% identical.

Murine *in situ* expression analysis revealed that TANGO 242 is expressed in the cerebellum and dentate gyrus of the adult brain. Expression in the cerebellum includes granule cells. However, expression was not observed in the olfactory bulb, a region which, like the dentate gyrus and cerebellum, contains granule cells. No expression was detected in: spinal cord, eye and harderian gland, submandibular gland, white fat, brown fat, stomach, heart, lung, liver, kidney, adrenal gland, colon, small intestine, thymus, lymph node, spleen, pancreas, skeletal muscle, bladder, testes, ovaries, or placenta. Embryonic expression was examined at: E13.5, E14.5, E15.5, E16.5, E18.5, and P1.5. TANGO 242 expression was not detected at any of these stages.

The predicted molecular weight of the entire murine TANGO 242 protein without modification and prior to cleavage of the signal sequence is about 34.4 kDa.

Biological function of TANGO 241 and TANGO 242

Analogous to other class II cytokine receptors, TANGO 241 and TANGO 242 mediate ligand binding though their extracellular domain. Based on sequence homology, ligands of class II cytokine receptors are expected to function as ligands for TANGO 241 and TANGO 242. However, TANGO 241 and TANGO 242 also have their own specific ligands and activities in addition to those reported for other class II cytokine receptors.

Proteins that bind class II cytokine receptors play a role in a large number of cellular processes, e.g., anti-viral and anti-bacterial infection, inflammation, autoimmune disease, vascular injury and disorders associated with osteoclastic bone resorption. Important class II cytokine receptor ligands include IFN-α, IFN-β, IFN-γ, IL-10, and tissue plasmin. Interferons exert a variety of effects on target cells including the induction of enzymes on target cells, e.g., the induction of 2'-5' oligoadenylate synthetase which inhibits viral RNA and DNA replication. Thus,

TANGO 241 and TANGO 242 proteins, nucleic acids, and modulators thereof can be used in the treatment of viral infection (e.g., chronic hepatitis B, chronic hepatitis C, and condyloma acuminata), bacterial infection (e.g., chronic granulomatous disease), inflammatory disorders (e.g., arthritis, multiple sclerosis, and ulcerative colitis), autoimmune disorders, vascular injury (e.g., modulation of would healing, regrowth of vasculature, and regrowth of vasculature into ischemic organs in stroke or coronary bypass patients), and unwanted angiogenesis (e.g., inhibition of angiogenesis that promotes tumor growth).

Interferons have anti-proliferative effects and inhibit normal cell growth. Thus, TANGO 241 and TANGO 242 proteins, nucleic acids, and modulators thereof can be used in the treatment of proliferative disorders (e.g., cancers, hairy cell leukemia, chronic myelogenous leukemia, mycosis fungoides, and Kaposi's sarcoma).

Tissue plasmin is an initiator of blood coagulation. Deregulated expression of tissue plasmin is associated with thrombogenesis in sepsis, cancer, and inflammation. Tissue plasmin also appears to be involved in a variety of non-hemostatic functions including inflammation, brain function, tumor associated angiogenesis. Thus, TANGO 241 and TANGO 242 proteins, nucleic acids, and modulators thereof can be used in the treatment of sepsis.

Soluble forms of TANGO 241 and TANGO 242 (e.g., truncated forms lacking a transmembrane or cytoplasmic domain) can be used to inhibit receptor activity by interfering with the interaction between TANGO 241 or TANGO 242 and their respective ligand(s).

TANGO 241 polypeptides and nucleic acids as well as modulators of the expression or activity of TANGO 241 can be used to treat disorders of the tissue in which it is expressed. Thus, TANGO 241 polypeptides, nucleic acids, or modulators thereof, can be used to treat pancreatic disorders, such as pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational

diabetes), or islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma).

In another example, TANGO 241 polypeptides, nucleic acids, or modulators thereof, can be used to treat esophageal and other digestive system related disorders, e.g., dysphagia (e.g., oropharyngeal dysphagia, esophageal dysphagia), pyrosis, achalasia, diffuse esophageal spasm, nutcracker esophagus, and gastroesophageal reflux disease.

TANGO 242 polypeptides and nucleic acids as well as modulators of the expression or activity of TANGO 242 can be used to treat disorders of the tissue in which it is expressed. Thus, TANGO 242 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain.

Tables 1 and 2 summarize sequence information for TANGO 241 and TANGO 242.

TABLE 1: Summary of TANGO 241 and TANGO 242 Sequence Information.

Gene	cDNA	ORF	Protein	Figure	Accession No.
Human TANGO 241	SEQ ID NO:1	SEQ ID NO:3	SEQ ID NO:2	Fig. 1	20716
Human TANGO 242	SEQ ID NO:11	SEQ ID NO:13	SEQ ID NO:12	Fig. 2	20717
Murine TANGO 241	SEQ ID NO:19	SEQ ID NO:21	SEQ ID NO:20	Fig. 7	
Murine TANGO 242	SEQ ID NO:22	SEQ ID NO:24	SEQ ID NO:23	Fig. 8	

TABLE 2: Summary of the Domains of Human TANGO 241 and Human TANGO 242.

Protein	Signal Sequence	Mature Protein	Extracellular Domain	Transmembrane Domain	Cytoplasmic Domain
Human TANGO 241	aa 1-14 SEQ ID NO:4	aa 15-574 SEQ ID NO:5	aa 15-226 SEQ ID NO:6	aa 227-251 SEQ ID NO:7	aa 252-574 SEQ ID NO:8
Human TANGO 242	aa 1-29 SEQ ID NO:14	aa 30-311 SEQ ID NO:15	aa 30-230 SEQ ID NO:16	aa 231-255 SEQ ID NO:17	aa 256-311 SEQ ID NO:18

Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:1, 3, 11, 13,19, 21, 22, or 24 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al.,

eds., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full-length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologues in other cell types, e.g., from other tissues, as well as homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or of a naturally occurring mutant of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, or 24.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein

molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, e.g., detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

A nucleic acid fragment encoding a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NOs:3 or 13 expressing the encoded portion of the polypeptide protein (e.g., by recombinant expression in vitro) and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of SEQ ID NO:3 or 13.

In addition to the nucleotide sequences of SEQ ID NOs:3 or 13, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (e.g., the human population). Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. As used herein, the phrase allelic variant refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations

that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologues), which have a nucleotide sequence which differs from that of the human protein described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of a cDNA of the invention can be isolated based on their identity to the human nucleic acid molecule disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1290) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, 3, 11 13, 19, 21, 22, 24 or complement thereof. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45 C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65 C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or complement thereof, corresponds to a naturally-

occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologues of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologues of various species (e.g., murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID NO:2, or 12 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:2, or 12.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, or 24 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid

substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein:protein interactions with proteins in a signaling pathway of the polypeptide of the invention; (2) the ability to bind a ligand of the polypeptide of the invention; or (3) the ability to bind to an intracellular target protein of the polypeptide of the invention. In yet another preferred embodiment, the mutant polypeptide can be assayed for the ability to modulate cellular proliferation, cellular migration or chemotaxis, or cellular differentiation.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region

of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier et al. (1987) Nucleic Acids Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) FEBS Lett. 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) Nature 334:585-591)) can be used to catalytically cleave mRNA transcripts to

thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) Science 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (e.g., the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. See generally Helene (1991) Anticancer Drug Des. 6(6):569-84; Helene (1992) Ann. N.Y. Acad. Sci. 660:27-36; and Maher (1992) Bioassays 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al. (1996) Bioorganic & Medicinal Chemistry 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), supra; Perry-O'Keefe et al. (1996) Proc. Natl. Acad. Sci. USA 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996), supra; or as probes or primers for DNA sequence and hybridization (Hyrup (1996), supra; Perry-O'Keefe et al. (1996) Proc. Natl. Acad. Sci. USA 93: 14670-675).

In another embodiment, PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996), supra). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), supra, and Finn et al. (1996) Nucleic Acids Res. 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) Nucleic Acids Res. 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) Nucleic Acids Res. 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) Bioorganic Med. Chem. Lett. 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) Proc. Natl. Acad. Sci. USA 86:6553-6556; Lemaitre et al. (1987) Proc. Natl. Acad. Sci. USA 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol et al. (1988) Bio/Techniques 6:958-976) or intercalating agents (see, e.g., Zon (1988) Pharm. Res. 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating

protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (e.g., the amino acid sequence shown in any of SEQ ID NOs:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, or 23), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of SEQ ID NO:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20 or 23. Other useful proteins are substantially identical (e.g., at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to any of SEQ ID NO:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, or 23 and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for

optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (e.g., overlapping positions) x 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-2268, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) J. Mol. Biol. 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. Id. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) CABIOS 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid

sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., supra) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into

pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction in vivo. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g. promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, e.g., Ausubel et al., supra). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence of a polypeptide of the invention (SEQ ID NO:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20, or 23) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the

signal sequence (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal sequences of the present invention can be used to identify regulatory sequences, e.g., promoters, enhancers, repressors. Since signal sequences are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal sequence on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal sequence can be used as a probe to identify and isolate signal sequences and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, e.g., discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large

gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID NO:2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 20 or 23, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Figures 8-14 are hydrophobicity plots of the proteins of the invention. These plots or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal). An appropriate immunogenic preparation can contain, for example, recombinantly expressed or chemically synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin

molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, e.g., a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')2 fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) Nature 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) Immunol. Today 4:72), the EBV-hybridoma technique (Cole et al. (1985), Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (see generally Current Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science 246:1275-1281; Griffiths et al. (1993) EMBO J. 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) Science 240:1041-1043; Liu et al. (1987) Proc. Natl. Acad. Sci. USA 84:3439-3443; Liu et al. (1987) J. Immunol. 139:3521-3526; Sun et al. (1987) Proc. Natl. Acad. Sci. USA 84:214-218; Nishimura et al. (1987) Canc. Res. 47:999-1005; Wood et al. (1985) Nature 314:446-449; and Shaw et al. (1988) J. Natl. Cancer Inst. 80:1553-1559); Morrison (1985) Science 229:1202-1207; Oi et al. (1986) Bio/Techniques 4:214; U.S. Patent

5,225,539; Jones et al. (1986) Nature 321:552-525; Verhoeyan et al. (1988) Science 239:1534; and Beidler et al. (1988) J. Immunol. 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, Int. Rev. Immunol. 13:65□93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope (Jespers et al. (1994) Bio/technology 12:899-903).

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies

can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Further, an antibody (or fragment thereof) may be conjugated to or administered with a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response. The drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule

capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory

sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (e.g., E. coli) or eukaryotic cells (e.g., insect cells (using baculovirus expression vectors), yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, supra. Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego,

California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident ë prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in E. coli is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in E. coli (Wada et al. (1992) Nucleic Acids Res. 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast S. cerivisae include pYepSec1 (Baldari et al. (1987) EMBO J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al. (1987) Gene 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow and Summers (1989) Virology 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987) EMBO J. 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2,

cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., supra.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) Genes Dev. 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) Adv. Immunol. 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) EMBO J. 8:729-733) and immunoglobulins (Banerji et al. (1983) Cell 33:729-740; Queen and Baltimore (1983) Cell 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) Proc. Natl. Acad. Sci. USA 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) Science 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) Science 249:374-379) and the á-fetoprotein promoter (Campes and Tilghman (1989) Genes Dev. 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or

attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (Reviews - Trends in Genetics, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (e.g., E. coli) or eukaryotic cell (e.g., insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (supra), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced

nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

Alternatively, the expression characteristics of an endogenous TANGO 241 or TANGO 242 gene within a cell line may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or a recombinant cell such that the inserted regulatory element is operatively linked with the endogenous TANGO 241 or TANGO 242 gene. For example, an endogenous TANGO 241 or TANGO 242 gene that is normally transcriptionally silent or is expressed at a very low level may be activated by inserting a regulatory element which is capable of promoting the expression of the gene in the cell line. Alternatively, a promiscuous regulatory element that works across numerous or all cell types may be used. Techniques for inserting regulatory elements by, e.g., homologous recombination, are described in, e.g., Chappel, U.S. Patent No. 5,272,071 and PCT publication No. WO 91/06667, published May 16, 1991.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators

of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic

animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) Cell 51:503 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 gene has homologously recombined with the endogenous gene are selected (see, e.g., Li et al. (1992) Cell 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, 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. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are

described further in Bradley (1991) Current Opinion in Bio/Technology 2:823-829 and in PCT Publication NOS. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lakso et al. (1992) Proc. Natl. Acad. Sci. USA 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of Saccharomyces cerevisiae (O'Gorman et al. (1991) Science 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) Nature 385:810-813 and PCT Publication NOS. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the

active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more addtional active compounds.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the

extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the

compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza

Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) J. Acquired Immune Deficiency Syndromes and Human Retrovirology 14:193).

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (see, e.g., Chen et al. (1994) Proc. Natl. Acad. Sci. USA 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where

the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). For example, polypeptides of the invention can to used to (i) modulate cellular proliferation; (ii) modulate cell migration and chemotaxis; (iii) modulate cellular differentiation; and/or (iv) modulate angiogenesis. The isolated nucleic acid molecules of the invention can be used to express proteins (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (e.g., in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) Anticancer Drug Des. 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) Proc. Natl. Acad. Sci. USA 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al. (1994). J. Med. Chem. 37:2678; Cho et al. (1993) Science 261:1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2061; and Gallop et al. (1994) J. Med. Chem. 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) Bio/Techniques 13:412-421), or on beads (Lam (1991) Nature 354:82-84), chips (Fodor (1993) Nature 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent NOS. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) Proc. Natl. Acad. Sci. USA 89:1865-1869) or phage (Scott and Smith (1990) Science 249:386-390; Devlin (1990) Science 249:404-406; Cwirla et al. (1990)

Proc. Natl. Acad. Sci. USA 87:6378-6382; and Felici (1991) J. Mol. Biol. 222:301-310).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with 125I, 35S, 14C, or 3H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining

the ability of the test compound to modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide protein to bind to or interact with a target molecule.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (e.g., a polypeptide of the invention binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca2+, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test

compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from

the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptidede of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically

significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Bio/Techniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The

mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment. For a review of this technique, see D'Eustachio et al. ((1983) Science 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include in situ hybridization (described in Fan et al. (1990) Proc. Natl. Acad. Sci. USA 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al. (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) Nature 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of

selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 or 11 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, or 13, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence

found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, e.g., fragments derived from noncoding regions having a length of at least 20 or 30 bases.

The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an in situ hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or by organ type.

C. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of

the invention and/or activity of a polypeptide of the invention, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity of a polypeptide of the invention. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, mutations in a gene of the invention can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with aberrant expression or activity of a polypeptide of the invention.

Another aspect of the invention provides methods for expression of a nucleic acid or polypeptide of the invention or activity of a polypeptide of the invention in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent).

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds) on the expression or activity of a polypeptide of the invention in clinical trials. These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent

for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID NO:1, or 11, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')2) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection of genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be

labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention (e.g., a proliferative disorder, e.g., psoriasis or cancer). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits may also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit may comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the

invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit may comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit may also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit may also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit may also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention, e.g., a proliferative disorder, e.g., psoriasis or cancer, or an angiogenic disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a

"test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (e.g., wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level

of a the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) Proc. Natl. Acad. Sci. USA 91:360-364), the latter of which can be particularly useful for detecting point mutations in a gene (see, e.g., Abravaya et al. (1995) Nucleic Acids Res. 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) Bio/Technology 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with

one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) Human Mutation 7:244-255; Kozal et al. (1996) Nature Medicine 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., supra. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) Proc. Natl. Acad. Sci. USA 74:560) or Sanger ((1977) Proc. Natl. Acad. Sci. USA 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) Bio/Techniques 19:448), including sequencing by mass spectrometry (see, e.g., PCT Publication No. WO 94/16101; Cohen et al. (1996) Adv. Chromatogr. 36:127-162; and Griffin et al. (1993) Appl. Biochem. Biotechnol. 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) Science 230:1242). In general, the technique of mismatch cleavage entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton et al. (1988) Proc. Natl. Acad. Sci. USA 85:4397; Saleeba et al. (1992) Methods Enzymol. 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called DNA mismatch repair enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) Carcinogenesis 15:1657-1662). According to an exemplary embodiment, a probe based on a selected sequence, e.g., a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism

(SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) Proc. Natl. Acad. Sci. USA 86:2766; see also Cotton (1993) Mutat. Res. 285:125-144; Hayashi (1992) Genet. Anal. Tech. Appl. 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) Trends Genet. 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) Nature 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a `GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys. Chem. 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) Nature 324:163); Saiki et al. (1989) Proc. Natl. Acad. Sci. USA 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) Nucleic Acids Res. 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner (1993) Tibtech 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) Mol. Cell Probes 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) Proc. Natl. Acad. Sci. USA 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by

altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Linder (1997) Clin. Chem. 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been

identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of a polypeptide of the invention (e.g., the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or

protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the

administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, disorders characterized by abberant expression or activity of the polypeptides of the invention include proliferative disorders, autoimmune disorders, immunomodulatory disorders, and disorders of cell differentiation. In addition, the polypeptides of the invention can be used to promote wound healing and angiogenesis, as well as other uses described herein.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject. For example, an antagonist of an ELVIS protein may be used to treat a proliferative disorder, e.g., psoriasis,

associated with abberant ELVIS expression or activity. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or downregulated and/or in which increased activity is likely to have a beneficial effect, e.g., in wound healing. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high

or upregulated and/or in which decreased activity is likely to have a beneficial effect.

The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Deposit of Clones

Clone AthEa20d7, encoding human TANGO 241, inserted into vecotr pMET7 and designated plasmid ApAthEa20d7 was deposited with the American Type Culture Collection (ATCC, 10801 University Boulevard, Manassas, VA 20110-2209) on December 30, 1998 and was assigned Accession Number 20716.

Clone AthEa89c8, encoding human TANGO 242, inserted into vector pMET7 and designated plasmid ApAthEa89c8 was deposited with the American Type Culture Collection (ATCC, 10801 University Boulevard, Manassas, VA 20110-2209) on December 30, 1998 and was assigned Accession Number 20717.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the polypeptide encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, 12, or the polypeptide encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 12, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2 or 12;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleotide sequence of

SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 65% identical to a nucleic acid consisting of the nucleotide sequence of SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2, or 12.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 12, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 20716 or 20717, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleotide sequence of

SEQ ID NO:1, 3, 11, 13, 19, 21, 22, 24 or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for TANGO 241- or TANGO 242-mediated signal transduction.

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
   1  GTCGACCCACGCGTCCGCCCACGCGTCCGGCCAAGGGAGGGCTCTGTGCCAGCCCCGATG
   1                                                            M

  61  AGGACGCTGCTGACCATCTTGACTGTGGGATCCCTGGCTGCTCACGCCCCTGAGGACCCC
   2  R  T  L  L  T  I  L  T  V  G  S  L  A  A  H  A  P  E  D  P

 121  TCGGATCTGCTCCAGCACGTGAAATTCCAGTCCAGCAACTTTGAAAACATCCTGACGTGG
  22  S  D  L  L  Q  H  V  K  F  Q  S  S  N  F  E  N  I  L  T  W

 181  GACAGCGGGCCAGAGGGCACCCCAGACACGGTCTACAGCATCGAGTATAAGACGTACGGA
  42  D  S  G  P  E  G  T  P  D  T  V  Y  S  I  E  Y  K  T  Y  G

 241  GAGAGGGACTGGGTGGCAAAGAAGGGCTGTCAGCGGATCACCCGGAAGTCCTGCAACCTG
  62  E  R  D  W  V  A  K  K  G  C  Q  R  I  T  R  K  S  C  N  L

 301  ACGGTGGAGACGGGCAACCTCACGGAGCTCTACTATGCCAGGGTCACCGCTGTCAGTGCG
  82  T  V  E  T  G  N  L  T  E  L  Y  Y  A  R  V  T  A  V  S  A

 361  GGAGGCCGGTCAGCCACCAAGATGACTGACAGGTTCAGCTCTCTGCAGCACACTACCCTC
 102  G  G  R  S  A  T  K  M  T  D  R  F  S  S  L  Q  H  T  T  L

 421  AAGCCACCTGATGTGACCTGTATCTCCAAAGTGAGATCGATTCAGATGATTGTTCATCCT
 122  K  P  P  D  V  T  C  I  S  K  V  R  S  I  Q  M  I  V  H  P

 481  ACCCCCACGCCAATCCGTGCAGGCGATGGCCACCGGCTAACCCTGGAAGACATCTTCCAT
 142  T  P  T  P  I  R  A  G  D  G  H  R  L  T  L  E  D  I  F  H

 541  GACCTGTTCTACCACTTAGAGCTCCAGGTCAACCGCACCTACCAAATGCACCTTGGAGGG
 162  D  L  F  Y  H  L  E  L  Q  V  N  R  T  Y  Q  M  H  L  G  G

 601  AAGCAGAGAGAATATGAGTTCTTCGGCCTGACCCCTGACACAGAGTTCCTTGGCACCATC
 182  K  Q  R  E  Y  E  F  F  G  L  T  P  D  T  E  F  L  G  T  I

 661  ATGATTTGCGTTCCCACCTGGGCCAAGGAGAGTGCCCCCTACATGTGCCGAGTGAAGACA
 202  M  I  C  V  P  T  W  A  K  E  S  A  P  Y  M  C  R  V  K  T

 721  CTGCCAGACCGGACATGGACCTACTCCTTCTCCGGAGCCTTCCTGTTCTCCATGGGCTTC
 222  L  P  D  R  T  W  T  Y  S  F  S  G  A  F  L  F  S  M  G  F

 781  CTCGTCGCAGTACTCTGCTACCTGAGCTACAGATATGTCACCAAGCCGCCTGCACCTCCC
 242  L  V  A  V  L  C  Y  L  S  Y  R  Y  V  T  K  P  P  A  P  P

 841  AACTCCCTGAACGTCCAGCGAGTCCTGACTTTCCAGCCGCTGCGCTTCATCCAGGAGCAC
 262  N  S  L  N  V  Q  R  V  L  T  F  Q  P  L  R  F  I  Q  E  H

 901  GTCCTGATCCCTGTCTTTGACCTCAGCGGCCCCAGCAGTCTGGCCCAGCCTGTCCAGTAC
 282  V  L  I  P  V  F  D  L  S  G  P  S  S  L  A  Q  P  V  Q  Y

 961  TCCCAGATCAGGGTGTCTGGACCCAGGGAGCCCGCAGGAGCTCCACAGCGGCATAGCCTG
 302  S  Q  I  R  V  S  G  P  R  E  P  A  G  A  P  Q  R  H  S  L

1021  TCCGAGATCACCTACTTAGGGCAGCCAGACATCTCCATCCTCCAGCCCTCCAACGTGCCA
 322  S  E  I  T  Y  L  G  Q  P  D  I  S  I  L  Q  P  S  N  V  P

1081  CCTCCCCAGATCCTCTCCCCACTGTCCTATGCCCCAAACGCTGCCCCTGAGGTCGGGCCC
 342  P  P  Q  I  L  S  P  L  S  Y  A  P  N  A  A  P  E  V  G  P

1141  CCATCCTATGCACCTCAGGTGACCCCCGAAGCTCAATTCCCATTCTACGCCCCACAGGCC
 362  P  S  Y  A  P  Q  V  T  P  E  A  Q  F  P  F  Y  A  P  Q  A

1201  ATCTCTAAGGTCCAGCCTTCCTCCTATGCCCCTCAAGCCACTCCGGACAGCTGGCCTCCC
 382  I  S  K  V  Q  P  S  S  Y  A  P  Q  A  T  P  D  S  W  P  P

1261  TCCTATGGGGTATGCATGGAAGGTTCTGGCAAAGACTCCCCCACTGGGACACTTTCTAGT
 402  S  Y  G  V  C  M  E  G  S  G  K  D  S  P  T  G  T  L  S  S
```


FIG. 1A

```
1321  TCTAAACACCTTAGGCCTAAAGGTCAGCTTCAGAAAGAGCCACCAGCTGGAAGCTGCATG
 422  P  K  H  L  R  P  K  G  Q  L  Q  K  E  P  P  A  G  S  C  M

1381  TTAGGTGGCCTTTCTCTGCAGGAGGTGACCTCCTTGGCTATGGAGGAATCCCAAGAAGCA
 442  L  G  G  L  S  L  Q  E  V  T  S  L  A  M  E  E  S  Q  E  A

1441  AAATCATTGCACCAGCCCCTGGGGATTTGCACAGACAGAACATCTGACCCAAATGTGCTA
 462  K  S  L  H  Q  P  L  G  I  C  T  D  R  T  S  D  P  N  V  L

1501  CACAGTGGGGAGGAAGGGACACCACAGTACCTAAAGGGCCAGCTCCCCCTCCTCTCCTCA
 482  H  S  G  E  E  G  T  P  Q  Y  L  K  G  Q  L  P  L  L  S  S

1561  GTCCAGATCGAGGGCCACCCCATGTCCCTCCCTTTGCAACCTCCTTCCGGTCCATGTTCC
 502  V  Q  I  E  G  H  P  M  S  L  P  L  Q  P  P  S  G  P  C  S

1621  CCCTCGGACCAAGGTCCAAGTCCCTGGGGCCTGCTGGAGTCCCTTGTGTGTCCCAAGGAT
 522  P  S  D  Q  G  P  S  P  W  G  L  L  E  S  L  V  C  P  K  D

1681  GAAGCCAAGAGCCCAGCCCCTGAGACCTCAGACCTGGAGCAGCCCACAGAACTGGATTCT
 542  E  A  K  S  P  A  P  E  T  S  D  L  E  Q  P  T  E  L  D  S

1741  CTTTTCAGAGGCCTGGCCCTGACTGTGCAGTGGGAGTCCTGAGGGGAATGGGAAAGGCTT
 562  L  F  R  G  L  A  L  T  V  Q  W  E  S  *

1801  GGTGCTTCCTCCCTGTCCCTACCCAGTGTCACATCCTTGGCTGTCAATCCCATGCCTGCC
1861  CATGCCACACACTCTGCGATCTGGCCTCAGACGGGTGCCCTTGAGAGAAGCAGAGGGAGT
1921  GGCATGCAGGGCCCCTGCCATGGGTGCGCTCCTCACCGGAACAAAGCAGCATGATAAGGA
1981  CTGCAGCGGGGGAGCTCTGGGGAGCAGCTTGTGTAGACAAGCGCGTGCTCGCTGAGCCCT
2041  GCAAGGCAGAAATGACAGTGCAAGGAGGAAATGCAGGGAAACTCCCGAGGTCCAGAGCCC
2101  CACCTCCTAACACCATGGATTCAAAGTGCTCAGGGAATTTGCCTCTCCTTGCCCCATTCC
2161  TGGCCAGTTTCACAATCTAGCTCGACAGAGCATGAGGCCCCTGCCTCTTCTGTCATTGTT
2221  CAAAGGTGGGAAGAGAGCCTGGAAAAGAACCAGGCCTGGAAAAGAACCAGAAGGAGGCTG
2281  GGCAGAACCAGAACAACCTGCACTTCTGCCAAGGCCAGGGCCAGCAGGACGGCAGGACTC
2341  TAGGGAGGGGTGTGGCCTGCAGCTCATTCCCAGCCAGGGCAACTGCCTGACGTTGCACGA
2401  TTTCAGCTTCATTCCTCTGATAGAACAAAGCGAAATGCAGGTCCACCAGGGAGGGAGACA
2461  CACAAGCCTTTTCTGCAGGCAGGAGTTTCAGACCCTATCCTGAGAATGGGGTTTGAAAGG
2521  AAGGTGAGGGCTGTGGCCCCTGGACGGGTACAATAACACACTGTACTGATGTCACAACTT
2581  TGCAAGCTCTGCCTTGGGTTCAGCCCATCTGGGCTCAAATTCCAGCCTCACCACTCACAA
2641  GCTGTGTGACTTCAAACAAATGAAATCAGTGCCCAGAACCTCGGTTTCCTCATCTGTAAT
2701  GTGGGGATCATAACACCTACCTCATGGAGTTGTGGTGAAGATGAAATGAAGTCATGTCTT
2761  TAAAGTGCTTAATAGTGCCTGGTACATGGGCAGTGCCCAATAAACGGTAGCTATTTCCAA
2821  AAAAAAAAAAAAAAGGGCGGCCGCTAGACTAGTCTAG
```

FIG. 1B

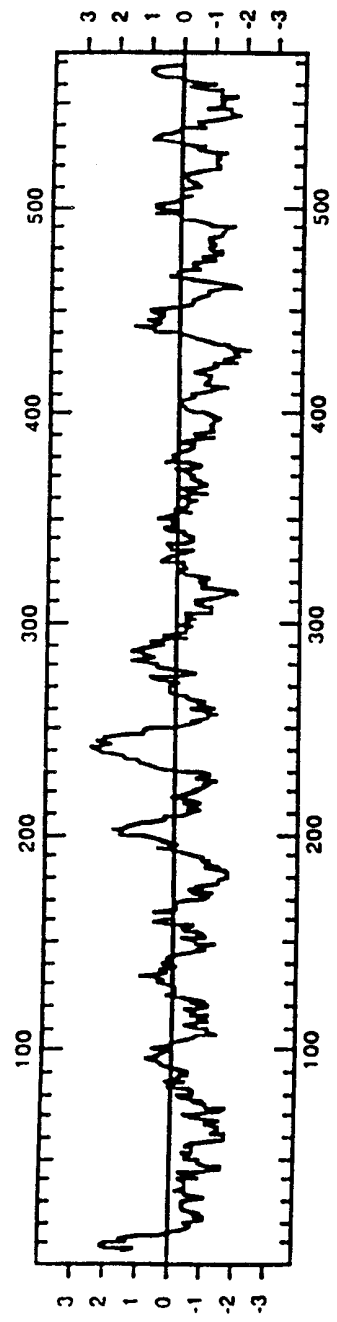


FIG. 2

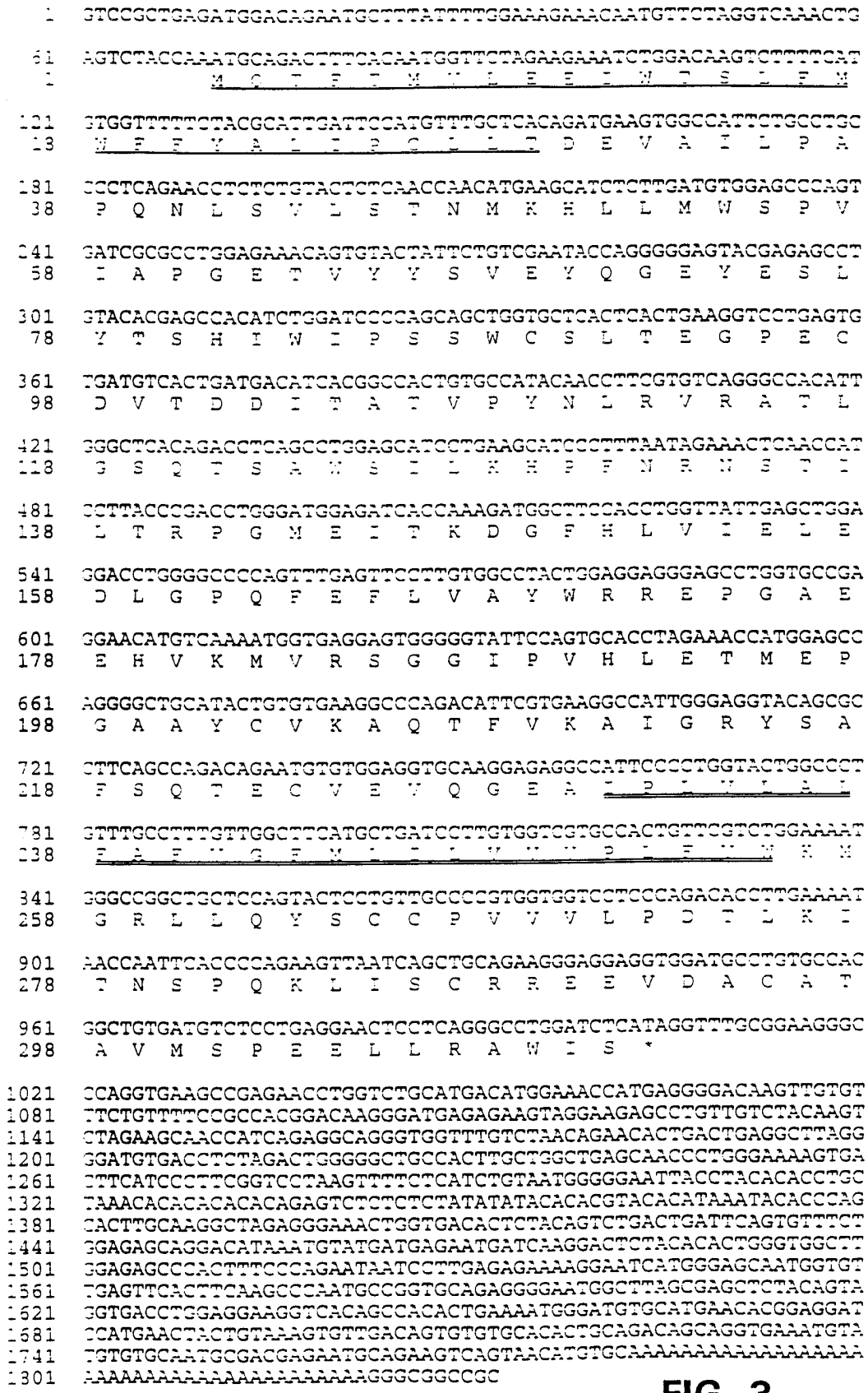


FIG. 3

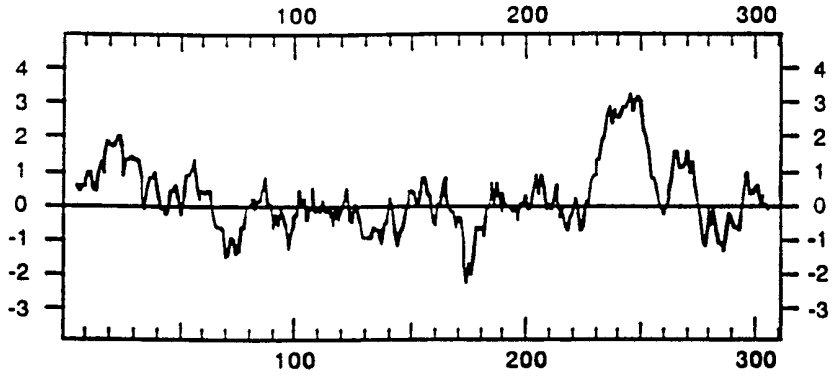


FIG. 4

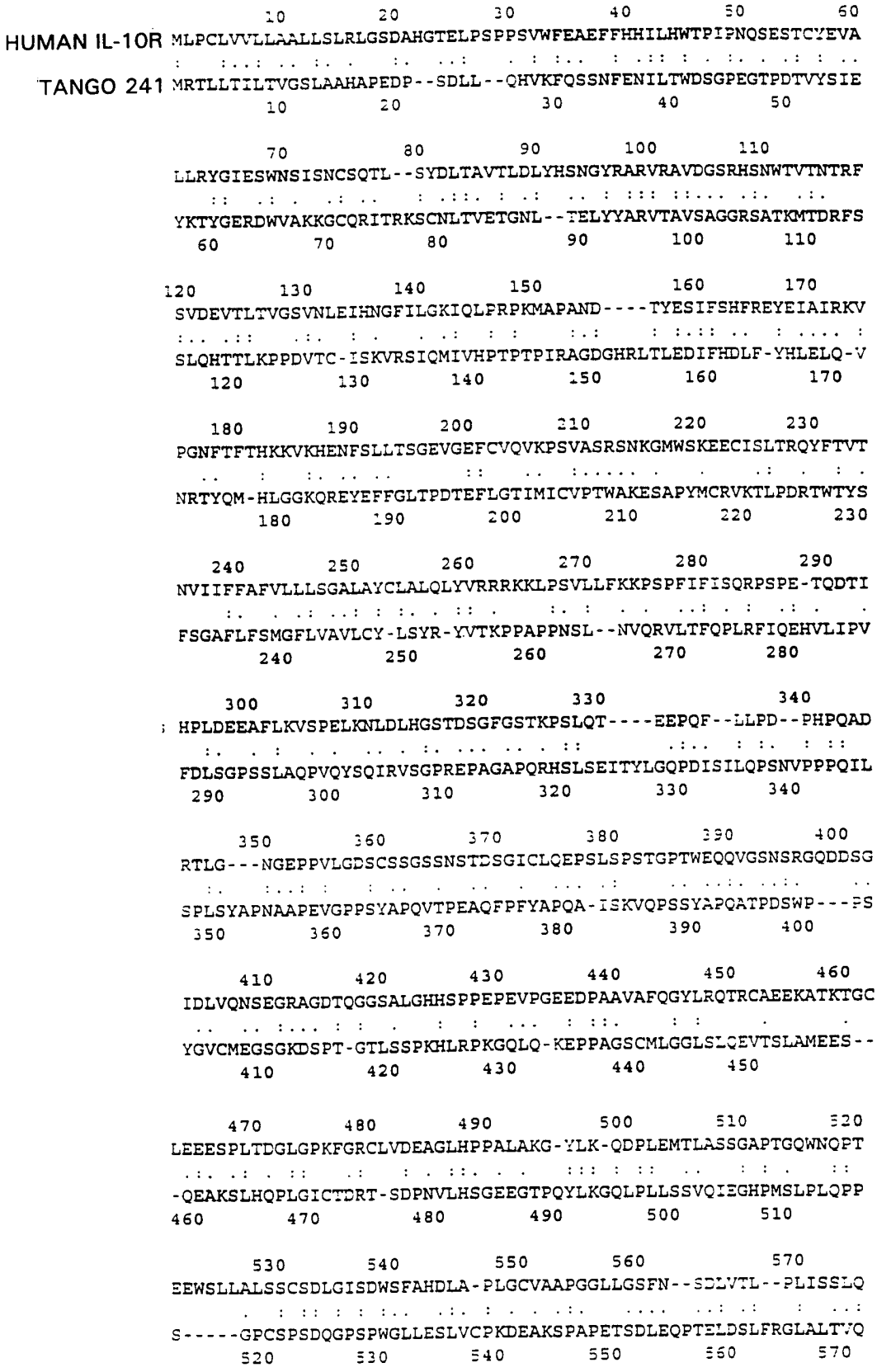


FIG. 5

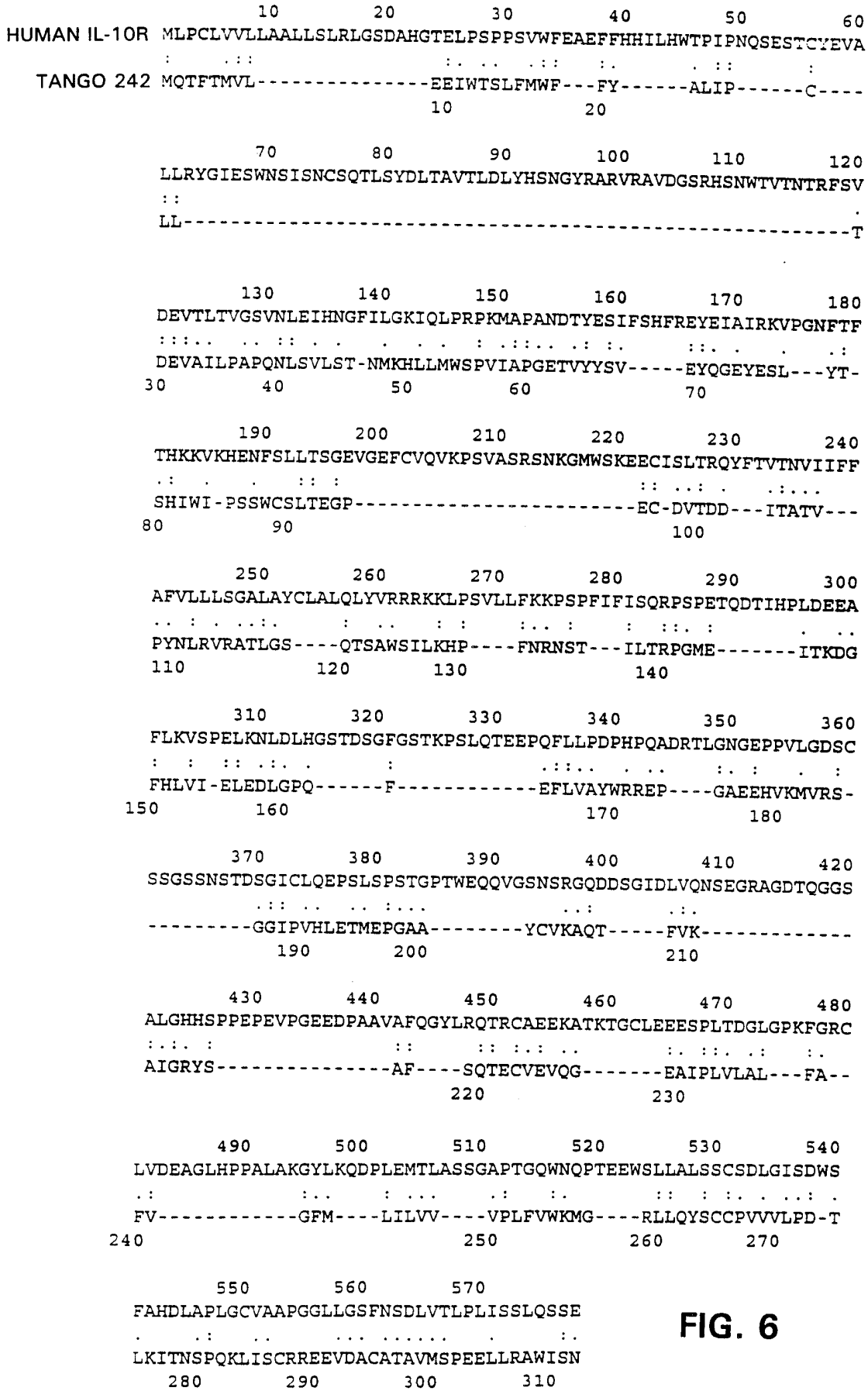


FIG. 6

```
   W   L   A   K   A   G   C   Q   R   I   T   Q   K   F   C   N   L   T   M      19
G TGG CTG GCC AAG GCG GGC TGC CAG CGG ATC ACC CAG AAG TTC TGC AAC CTG ACT ATG      58

 E   T   R   N   H   T   E   F   Y   Y   A   K   V   T   A   V   S   A   G   G    39
GAG ACC CGC AAC CAC ACT GAG TTT TAC TAC GCC AAG GTC ACG GCG GTC AGC GCG GGA GGC   118

 P   P   V   T   K   M   T   D   R   F   S   S   L   Q   H   T   T   I   K   P    59
CCA CCA GTC ACA AAG ATG ACT GAT CGT TTC AGC TCG CTG CAG CAC ACT ACC ATC AAA CCG   178

 P   D   V   T   C   I   P   K   V   R   S   I   Q   M   L   V   H   P   T   L    79
CCT GAT GTG ACC TGT ATC CCC AAA GTG AGG TCC ATT CAG ATG CTG GTC CAC CCC ACA CTC   238

 T   P   V   L   S   E   D   G   H   Q   L   T   L   E   E   I   F   H   D   L    99
ACA CCG GTC CTC TCG GAA GAT GGC CAC CAG CTA ACC CTG GAG GAG ATT TTC CAT GAC CTG   298

 F   Y   R   L   E   L   H   V   N   H   T   Y   Q   M   H   L   E   G   K   Q   119
TTC TAC CGC TTA GAG CTC CAC GTC AAC CAC ACC TAC CAG ATG CAC CTT GAA GGC AAA CAG   358

 R   E   Y   E   F   L   G   L   T   P   D   T   E   F   L   G   S   I   T   I   139
AGA GAA TAC GAG TTC CTT GGC CTG ACT CCC GAC ACA GAG TTC CTC GGC TCC ATC ACA ATT   418

 L   T   P   I   L   S   K   E   S   A   P   Y   V   C   R   V   K   T   L   P   159
TTG ACT CCG ATA TTG TCC AAG GAA AGT GCC CCC TAC GTG TGC CGA GTG AAG ACG CTG CCC   478

 D   R   T   W   A   Y   S   F   S   G   A   V   L   F   S   M   G   F   L   V   179
GAT CGG ACG TGG GCC TAC TCC TTC TCG GGC GCC GTG CTC TTT TCC ATG GGT TTC CTC GTC   538

 G   L   L   C   Y   L   G   Y   K   Y   I   T   K   P   P   V   P   P   N   S   199
GGC TTG CTC TGT TAT CTG GGC TAC AAA TAC ATC ACC AAG CCA CCT GTA CCT CCT AAC TCC   598

 L   N   V   Q   R   V   L   T   F   Q   P   L   R   F   I   Q   E   H   V   L   219
CTG AAC GTC CAA CGT GTC CTG ACC TTT CAA CCC CTA CGC TTC ATC CAA GAA CAC GTA CTG   658

 I   P   V   L   D   L   S   G   P   S   S   L   P   Q   P   I   Q   Y   S   Q   239
ATC CCT GTC TTG GAC CTC AGT GGC CCC AGC AGT CTG CCT CAG CCC ATC CAG TAC TCC CAA   718

 V   V   V   S   G   P   R   E   P   P   G   A   V   W   R   Q   S   L   S   D   259
GTG GTG GTG TCT GGG CCC AGG GAG CCT CCT GGA GCT GTG TGG CGG CAG AGC CTG TCT GAC   778

 L   T   Y   V   G   Q   S   D   V   S   I   L   Q   P   T   N   V   P   A   Q   279
CTC ACC TAC GTA GGG CAG TCA GAT GTC TCC ATC CTG CAA CCT ACC AAC GTG CCA GCT CAG   838

 Q   T   L   S   P   P   S   Y   A   P   K   A   V   P   E   V   Q   P   P   S   299
CAG ACA CTG TCC CCA CCA TCC TAC GCT CCG AAG GCT GTC CCT GAG GTC CAG CCC CCT TCC   898

 Y   A   P   Q   V   A   S   D   A   K   A   L   F   Y   S   P   Q   Q   G   M   319
TAT GCG CCT CAG GTA GCC TCG GAT GCC AAA GCT CTG TTC TAC TCA CCA CAA CAG GGG ATG   958

 K   T   R   P   A   T   Y   D   P   Q   D   I   L   D   S   C   P   A   S   Y   339
AAG ACC AGG CCT GCC ACC TAT GAC CCG CAG GAC ATT CTG GAC AGC TGC CCT GCT TCT TAT  1018

 A   V   C   V   C   A   R   A   S   A   H   V   C   M   *                       354
GCT GTG TGT GTG TGT GCG CGC GCG AGC GCT CAC GTC TGT ATG TGA                      1063
```

FIG. 7A

TGAGAATCTGGTTTTTTTAGAAACCCAGTTATGTTATGTGAATGCGGTTTTGCTTTAATCCTAGGCGTGGGATAAGAG 1142

ACTGCTTCGCAGCAGGTGACTATGATTTGCCTCATGCTCTAGCAAGGGTGTGCTTTTGCCAGCAGCTAGTTACTAT 1221

AAGTGTGTAATGTTTTGGAATTCTGGAAACTTTTCAAGGGGGTATAAATGGCAGAGCCCTGAGCGGGGCTGAGCTCACTC 1300

TGAGGAGGCTGTTGCTCCCCACTGCTGCTGGCTCCTGTTGTGGTGGTTTGTCAAGTGGTTGGGAGCAAAGTGAAAAGAA 1379

GAAATTGTAGATATCCTTACAATGACAAAGACTGAACTTACCCCAAGGAACTCGACACCTTAAATCAGCAGGAAGTAGT 1458

CGAAAAAGGTCAAGGTCCCCTTTCACCTTCTTTCTCTTCTACCTACTGGGGGGTTGGAAGGGAGTAAGGAGGAATAAGG 1537

GTGGTAGATATAAGGACTCAATAAAGTAGCACAAAAAGTAAAAAAAAAAAAAAAAAAAA 1596

FIG. 7B

```
GTCGACCCACGCGTCCGGGATGGACACACAGCCTCCTTCTGGAAGAAAACAGTGTCTCCGGGCAAGCTGAGTTTGCTAG  79

                                M   I   S   Q   G   V   W   T   S   P   I   M      12
AGACCCTCACACCAAACAGACCAGCACA ATG ATT TCC CAG GGA GTC TGG ACA AGT CCG ATC ATG      143

 W   F   F   Y   S   M   T   S   G   F   L   T   D   A   V   S   V   L   P   A    32
TGG TTT TTC TAC AGC ATG ACC TCA GGT TTT CTC ACG GAT GCA GTG TCC GTT TTA CCT GCC   203

 P   Q   N   L   S   V   Q   S   T   N   M   K   H   L   L   M   W   N   P   V    52
CCT CAG AAC CTC TCT GTA CAG TCA ACC AAC ATG AAG CAT CTC TTG ATG TGG AAC CCA GTG   263

 T   Q   P   G   E   T   V   L   Y   C   V   E   Y   Q   G   E   Y   E   S   L    72
ACC CAG CCG GGG GAG ACG GTG CTC TAT TGT GTG GAG TAT CAG GGG GAG TAC GAG AGC CTG   323

 Y   M   S   H   I   W   I   P   S   S   Q   C   S   P   T   K   S   L   E   C    92
TAC ATG AGC CAC ATC TGG ATC CCC AGC AGT CAA TGC TCA CCG ACC AAA AGT CTG GAG TGT   383

 D   V   T   D   D   I   T   A   T   V   P   Y   N   F   R   V   K   A   M   L   112
GAT GTC ACC GAT GAC ATA ACC GCC ACA GTG CCA TAT AAC TTC AGG GTC AAG GCC ATG CTG   443

 G   S   Q   T   S   A   W   S   N   L   E   H   P   F   N   R   N   A   T   V   132
GGC TCA CAG ACT TCA GCC TGG AGC AAC CTG GAG CAC CCC TTT AAC CGA AAT GCA ACT GTC   503

 L   T   P   P   R   M   E   V   T   E   H   G   L   H   L   V   I   E   L   E   152
CTC ACC CCA CCC AGG ATG GAG GTC ACT GAA CAT GGG CTG CAT CTG GTT ATT GAG CTG GAA   563

 D   L   G   P   Q   F   E   F   L   V   V   Y   W   R   R   E   P   G   A   A   172
GAC CTG GGA CCC CAG TTT GAG TTC CTT GTG GTC TAC TGG AGG AGG GAG CCT GGC GCC GCG   623

 E   H   V   K   M   V   R   S   G   D   I   P   V   H   L   E   T   M   E   P   192
GAA CAT GTT AAG ATG GTG AGG AGT GGG GAC ATT CCG GTG CAC CTA GAA ACC ATG GAA CCG   683

 G   A   M   Y   C   V   K   A   Q   A   L   V   K   A   I   G   R   H   S   A   212
GGG GCC ATG TAC TGT GTG AAG GCG CAG GCA CTG GTG AAA GCC ATC GGG AGG CAC AGT GCC   743

 F   S   Q   P   T   C   V   E   M   Q   G   E   S   L   P   L   A   L   A   L   232
TTC AGC CAG CCT ACG TGT GTG GAG ATG CAA GGA GAG TCT CTT CCG CTG GCA CTA GCT CTG   803

 F   A   F   V   G   F   M   L   I   L   V   V   V   L   L   S   V   W   K   M   252
TTT GCG TTT GTT GGC TTC ATG CTG ATT CTC GTG GTT GTA CTA CTC TCC GTC TGG AAG ATG   863

 G   Q   L   L   R   Y   S   C   C   P   A   V   V   L   P   D   T   L   K   I   272
GGC CAG CTG CTC CGG TAT TCT TGC TGC CCC GCC GTT GTC CTC CCA GAC ACC TTG AAA ATA   923

 T   S   S   S   Q   K   L   I   S   C   R   K   E   E   V   D   A   C   A   V   292
ACC AGT TCG TCT CAG AAG CTG ATC AGC TGC AGG AAG GAG GAG GTG GAC GCC TGT GCT GTG   983

 A   V   L   S   S   E   H   L   F   G   V   W   I   S   Q   T   *               309
GCT GTG CTG TCC TCG GAG CAT CTC TTT GGG GTC TGG ATC TCA CAG ACT TGA             1034

GAAGGACCTGATGGATTAACAGCCTGGTCTGCAGGACCCAGAACCCTGGATGGGGTGGGTTGTGTTTCTGTTTTCCTCT  1113

GTGGACAAAAGACAAAGAGCAGTAACAGGAGCCTACTGTGTGCGCGTCTAGCAGTGACCGTCAGAGGCAGAGTGGCGTG  1192
```

FIG. 8A

ACCATCCAAGAGTGACAGAAATAGTCTAGAGAGGTGATCCCCAGACACACAGTTACCATACACTCACTGAATAACAGTG 1271

GAGAAAATGGCTTCATACCTCGGACCCCGCTTTGTCACCTGTAGTGGGGATTAAATGGTTTCTCCCCTCATCTCTGTGT 1350

GTATACAAACTCGCACTTGCAGGTTGGCAAGGAGACTGGTGGCACTGTGCACAACGGTGAGTTCAGATCATGCGGACTG 1429

TGGGGCATACGCGTGTGCTAAGAATGATCAGAGTACTCAGCCCGGGTGCGAAATCTCAGCACTTCGGAGGAGCGGGAGAG 1508

AAAGTCACAGCCTAGACGACGCGGTGAAACCCCGTCTGAAGTAAACACAACCGAACACAACCAAGAGTAATAAACCTCCTTTT 1587

GTCAAATCAAACTCAGACCACTGTTGGTTTGGACTTCCCTCTTTACAAGAGGAAACCCGAGGAGGGACGGAAGAATGGG 1666

AGGGCTGTTTCACTTCAAGCCAAATCCTGTCAGCACGCTCTGCTGAGGGTGACATCGAGAGAGGCCGTAGCCCGTCCTG 1745

GTGACATCGAGAGAGGCCGTAGCCCGTCCATGATGGGACATGCGCGGATGCTGCTGTGGGTTCATGAAATGTCATCAGA 1824

TCCTAGGACCTTGTCCACAGTAAATCACACCAGCTCTCTTCCTAAGGACAATTCACTCAGGGTTTCTCAATCGCTCTCA 1903

GGTACTCATCAGGAACCCCTGACTAGAGGCATAACCATCACTACCTCCAGGCAGCGAGGGAGGGGTGGAGGACAGCTTC 1982

AGTCAGAGGAGGATGAAAGGAGGTGGCTTCATCTCTCTCTGCAGAGTCAGGCTCACCACCACCTCCCCCTCTGGAACCT 2061

GGAGCTGGCTTGTCTCTGCCCCAACTTCCTGGGCTTCACCAGCCAGACAAGGGTAGAGACAGTGTAACGGTTTCAACCA 2140

GGTGCTTCCAGTCCGTCTTCAGCACAGATGCCAACGGCCTCATCTCTACTTTGGGGATTCCAGGAGAGATATGTAGGGG 2219

TGTGGGCAGATATGTAGGGGTGTGGCTGAGGCTCCCTGCCTTTGTCCTGTTCTGTATATTTCTGCCTTGACTCTTTTG 2298

TGAAGCGTGTGTTAAATGTTATATTTCAACTGGTCTGGAAATGTGCTGTTTACAGAATTAAAAATTGCATACTTACATT 2377

AAAAAAAAAAAAAAAAAAGGGCGGCCGC 2405

FIG. 8B

```
201 CCCAGACACGGTCTACAGCATCGAGTATAAGACGTACGGAGAGAGGGACT 250
                                                    |
  1 .............................................GT 2

251 GGGTGGCAAAGAAGGGCTGTCAGCGGATCACCCGGAAGTCCTGCAACCTG 300
    || |||| |||  |||||| |||||||||||| ||||| ||||||||||
  3 GGCTGGCCAAGGCGGGCTGCCAGCGGATCACCCAGAAGTTCTGCAACCTG 52

301 ACGGTGGAGACGGGCAACCTCACGGAGCTCTACTATGCCAGGGTCACCGC 350
    ||  |||||||  |||||| ||| ||| | ||||| |||| |||||| ||
 53 ACTATGGAGACCCGCAACCACACTGAGTTTTACTACGCCAAGGTCACGGC 102

351 TGTCAGTGCGGGAGGCCGGTCAGCCACCAAGATGACTGACAGGTTCAGCT 400
     ||||| ||||||||||   ||| ||| ||||||||||  | |||||||
103 GGTCAGCGCGGGAGGCCCACCAGTCACAAAGATGACTGATCGTTTCAGCT 152

401 CTCTGCAGCACACTACCCTCAAGCCACCTGATGTGACCTGTATCTCCAAA 450
    | ||||||||||||||| |||| || |||||||||||||||||| |||||
153 CGCTGCAGCACACTACCATCAAACCGCCTGATGTGACCTGTATCCCCAAA 202

451 GTGAGATCGATTCAGATGATTGTTCATCCTACCCCCACGCCAATCCGTGC 500
    ||||| || |||||||||| | || || || || | ||| ||  |||   |
203 GTGAGGTCCATTCAGATGCTGGTCCACCCCACACTCACACCGGTCCTCTC 252

501 AGGCGATGGCCACCGGCTAACCCTGGAAGACATCTTCCATGACCTGTTCT 550
     | |||||||||| |||||||||||| || || ||||||||||||||||
253 GGAAGATGGCCACCAGCTAACCCTGGAGGAGATTTTCCATGACCTGTTCT 302

551 ACCACTTAGAGCTCCAGGTCAACCGCACCTACCAAATGCACCTTGGAGGG 600
    ||| |||||||||||| ||||||| ||||||||| |||||||||| |||
303 ACCGCTTAGAGCTCCACGTCAACCACACCTACCAGATGCACCTTGAAGGC 352

601 AAGCAGAGAGAATATGAGTTCTTCGGCCTGACCCCTGACACAGAGTTCCT 650
    || |||||||||| |||||| | |||||||| || ||||||||||||||
353 AAACAGAGAGAATACGAGTTCCTTGGCCTGACTCCCGACACAGAGTTCCT 402

651 TGGCACCATCATGATTTGCGTTCCCACCTGGGCCAAGGAGAGTGCCCCCT 700
     ||| ||||||  ||||    ||| |  | | |||||||  ||||||||||
403 CGGCTCCATCACAATTTTGACTCCGATATTGTCCAAGGAAAGTGCCCCCT 452

701 ACATGTGCCGAGTGAAGACACTGCCAGACCGGACATGGACCTACTCCTTC 750
    || ||||||||||||||| ||||| || ||||| ||| ||||||||||
453 ACGTGTGCCGAGTGAAGACGCTGCCCGATCGGACGTGGGCCTACTCCTTC 502

751 TCCGGAGCCTTCCTGTTCTCCATGGGCTTCCTCGTCGCAGTACTCTGCTA 800
    || || ||| | || || |||||||| ||||||||||  | ||||| ||
503 TCGGGCGCCGTGCTCTTTTCCATGGGTTTCCTCGTCGGCTTGCTCTGTTA 552
```

FIG. 9A

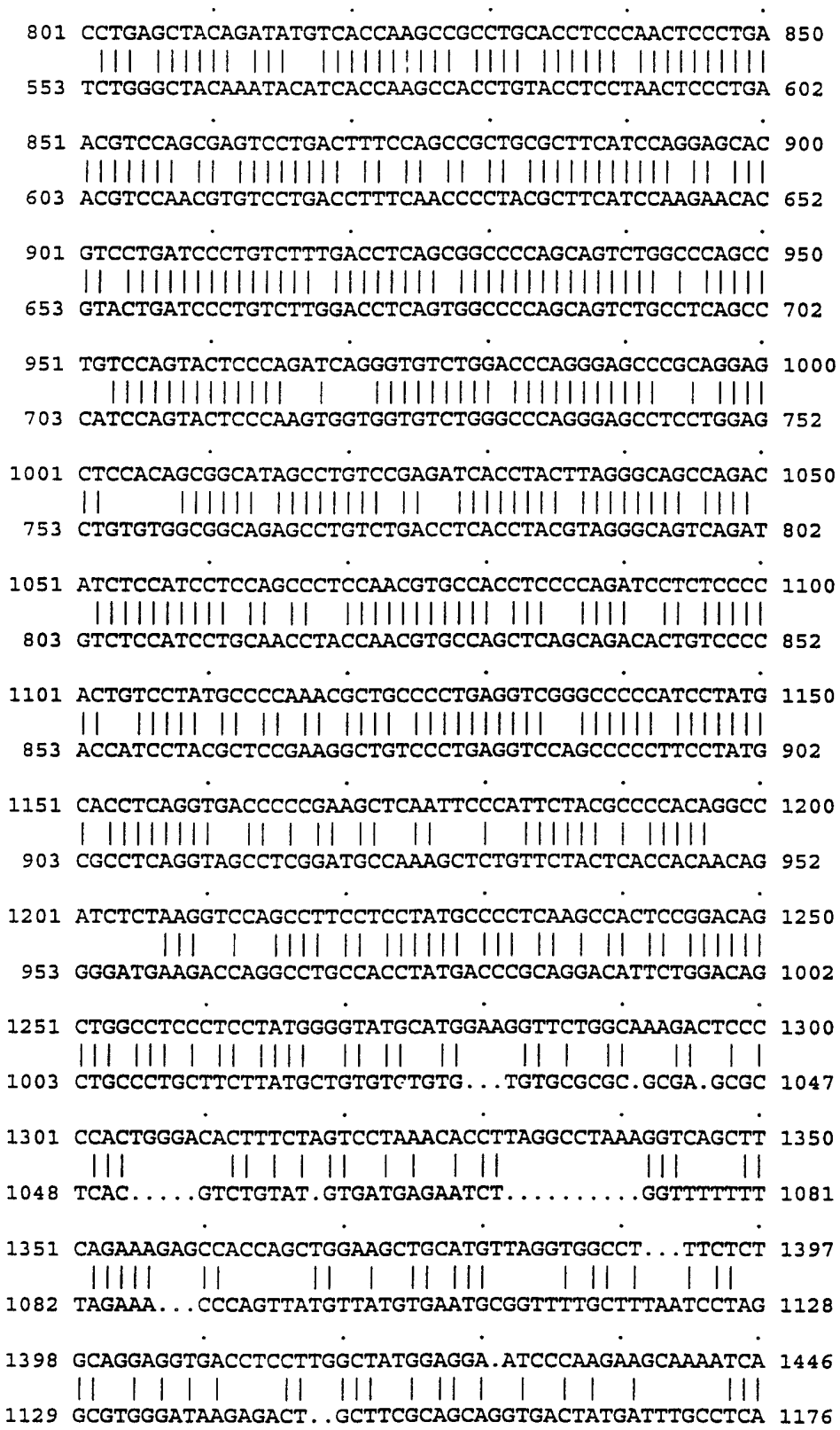

801 CCTGAGCTACAGATATGTCACCAAGCCGCCTGCACCTCCCAACTCCCTGA 850
553 TCTGGGCTACAAATACATCACCAAGCCACCTGTACCTCCTAACTCCCTGA 602
851 ACGTCCAGCGAGTCCTGACTTTCCAGCCGCTGCGCTTCATCCAGGAGCAC 900
603 ACGTCCAACGTGTCCTGACCTTTCAACCCCTACGCTTCATCCAAGAACAC 652
901 GTCCTGATCCCTGTCTTTGACCTCAGCGGCCCCAGCAGTCTGGCCCAGCC 950
653 GTACTGATCCCTGTCTTGGACCTCAGTGGCCCCAGCAGTCTGCCTCAGCC 702
951 TGTCCAGTACTCCCAGATCAGGGTGTCTGGACCCAGGGAGCCCGCAGGAG 1000
703 CATCCAGTACTCCCAAGTGGTGGTGTCTGGGCCCAGGGAGCCTCCTGGAG 752
1001 CTCCACAGCGGCATAGCCTGTCCGAGATCACCTACTTAGGGCAGCCAGAC 1050
753 CTGTGTGGCGGCAGAGCCTGTCTGACCTCACCTACGTAGGGCAGTCAGAT 802
1051 ATCTCCATCCTCCAGCCCTCCAACGTGCCACCTCCCCAGATCCTCTCCCC 1100
803 GTCTCCATCCTGCAACCTACCAACGTGCCAGCTCAGCAGACACTGTCCCC 852
1101 ACTGTCCTATGCCCCAAACGCTGCCCCTGAGGTCGGGCCCCCATCCTATG 1150
853 ACCATCCTACGCTCCGAAGGCTGTCCCTGAGGTCCAGCCCCCTTCCTATG 902
1151 CACCTCAGGTGACCCCCGAAGCTCAATTCCCATTCTACGCCCCACAGGCC 1200
903 CGCCTCAGGTAGCCTCGGATGCCAAAGCTCTGTTCTACTCACCACAACAG 952
1201 ATCTCTAAGGTCCAGCCTTCCTCCTATGCCCCTCAAGCCACTCCGGACAG 1250
953 GGGATGAAGACCAGGCCTGCCACCTATGACCCGCAGGACATTCTGGACAG 1002
1251 CTGGCCTCCCTCCTATGGGGTATGCATGGAAGGTTCTGGCAAAGACTCCC 1300
1003 CTGCCCTGCTTCTTATGCTGTGTCTGTG...TGTGCGCGC.GCGA.GCGC 1047
1301 CCACTGGGACACTTTCTAGTCCTAAACACCTTAGGCCTAAAGGTCAGCTT 1350
1048 TCAC.....GTCTGTAT.GTGATGAGAATCT.....GGTTTTTTT 1081
1351 CAGAAAGAGCCACCAGCTGGAAGCTGCATGTTAGGTGGCCT...TTCTCT 1397
1082 TAGAAA...CCCAGTTATGTTATGTGAATGCGGTTTTGCTTTAATCCTAG 1128
1398 GCAGGAGGTGACCTCCTTGGCTATGGAGGA.ATCCCAAGAAGCAAAATCA 1446
1129 GCGTGGGATAAGAGACT..GCTTCGCAGCAGGTGACTATGATTTGCCTCA 1176

FIG. 9B

```
1447 TTGCACCAGCCCCTG.GGGATTTGCACAGACAGAACATCTGACCCAAATG 1495
1177 .TGCTCTAGCAAGGGTGTGCTTTTGCCAG.CAGCAGCTAGTTACTATAAG 1224

1496 TGCTACACAGTGGGGAGGAAGGGACAC....CACAGTACCTAAAGGGCCA 1541
1225 TG.TGTAATGTTTGGAATTCTGGAAACTTTTCAAGGGGGTATAAATGGCA 1273

1542 GCTCCCCCTCCTCTCCTCAGTCCAGATCGAGGGCCACCCCATGTCCCTCC 1591
1274 GAGCCCTGAGCGGGGCTGAGCTCA.CTCTGAGGAGGCTGTTGCTCCCCAC 1322

1592 CTTTGCAACCTCC..TTCCGGTCCATGTTC...CCCCTCGGACCAAGGTC 1636
1323 TGCTGCTGGCTCCTGTTGTGGTGGTTTGTCAAGTGGTTGGGAGCAAAGTG 1372

1637 CAAGTCCCTGGGGCCTGCTGGAGTCCCTTGTGTGTCCCAAGGATGAA... 1683
1373 AAA....AGAAGAAATTGTAGATATCCTTACAATGACAAAGACTGAACTT 1418

1684 ..GCCAA.GAGCCCAGCCCCTGAGACCTCAGACCTGGAGCAGCCCACAGA 1730
1419 ACCCCAAGGAACTCGACACCTTAAA..TCAG.CAGGAAGTAGTCGAAAAA 1465

1731 ACTGGA.TTCTCTTTTCAGAGGCCTGGCCCTGACTGTGCAGTGGGAGTCC 1779
1466 GGTCAAGGTCCCCTTTCACCTTCTTTCTCTT..CTACCTACTGGGGGGTT 1513

1780 TGAGGGGAATGGGAAAGGCTTGGTGCTTCCTCCCTGTCCCTACCCAGTGT 1829
1514 GGAAGGGAGTAAGGAGGAATAAGGG..TGGTAGATATAAGGACTCAATAA 1561

1830 CACATCCTTGGCTGTCAATCCCATGCCTGCCCATGCCACACACTCTGCGA 1879
1562 AGTAGCACAAAAAGTAAAAAAAAAAAAAAAAAAAA..............  1596
```

FIG. 9C

```
                .         .         .         .         .
  1 .....................................TGGCTGGC 8
                                         ||| ||||
151 ACGGTCTACAGCATCGAGTATAAGACGTACGGAGAGAGGGACTGGGTGGC 200
                .         .         .         .         .
  9 CAAGGCGGGCTGCCAGCGGATCACCCAGAAGTTCTGCAACCTGACTATGG 58
    |||  |||||| |||||||||||||| ||||| ||||||||||||  |||
201 AAAGAAGGGCTGTCAGCGGATCACCCGGAAGTCCTGCAACCTGACGGTGG 250
                .         .         .         .         .
 59 AGACCCGCAACCACACTGAGTTTTACTACGCCAAGGTCACGGCGGTCAGC 108
    ||||  |||||| ||| ||| | ||||| |||| |||||| || |||||
251 AGACGGGCAACCTCACGGAGCTCTACTATGCCAGGGTCACCGCTGTCAGT 300
                .         .         .         .         .
109 GCGGGAGGCCCACCAGTCACAAAGATGACTGATCGTTTCAGCTCGCTGCA 158
    ||||||||||   ||| ||| ||||||||||| | |||||||| |||||
301 GCGGGAGGCCGGTCAGCCACCAAGATGACTGACAGGTTCAGCTCTCTGCA 350
                .         .         .         .         .
159 GCACACTACCATCAAACCGCCTGATGTGACCTGTATCCCCAAAGTGAGGT 208
    |||||||||| |||| || |||||||||||||||||| |||||||||| |
351 GCACACTACCCTCAAGCCACCTGATGTGACCTGTATCTCCAAAGTGAGAT 400
                .         .         .         .         .
209 CCATTCAGATGCTGGTCCACCCCACACTCACACCGGTCCTCTCGGAAGAT 258
    | |||||||||| | || || || || | ||| ||  |||   | |  |||
401 CGATTCAGATGATTGTTCATCCTACCCCCACGCCAATCCGTGCAGGCGAT 450
                .         .         .         .         .
259 GGCCACCAGCTAACCCTGGAGGAGATTTTCCATGACCTGTTCTACCGCTT 308
    ||||||| |||||||||||| || || |||||||||||||||||| |||
451 GGCCACCGGCTAACCCTGGAAGACATCTTCCATGACCTGTTCTACCACTT 500
                .         .         .         .         .
309 AGAGCTCCACGTCAACCACACCTACCAGATGCACCTTGAAGGCAAACAGA 358
    ||||||||| ||||||| ||||||||| |||||||||| ||| || ||||
501 AGAGCTCCAGGTCAACCGCACCTACCAAATGCACCTTGGAGGGAAGCAGA 550
                .         .         .         .         .
359 GAGAATACGAGTTCCTTGGCCTGACTCCCGACACAGAGTTCCTCGGCTCC 408
    ||||||| |||||| | |||||||| || |||||||||||||| ||| ||
551 GAGAATATGAGTTCTTCGGCCTGACCCCTGACACAGAGTTCCTTGGCACC 600
                .         .         .         .         .
409 ATCACAATTTTGACTCCGATATTGTCCAAGGAAAGTGCCCCCTACGTGTG 458
    ||||  ||||    ||| |  | | ||||||| |||||||||||| ||||
601 ATCATGATTTGCGTTCCCACCTGGGCCAAGGAGAGTGCCCCCTACATGTG 650
                .         .         .         .         .
459 CCGAGTGAAGACGCTGCCCGATCCGACGTGGGCCTACTCCTTCTCGGGCG 508
    |||||||||||| ||||| || ||||| ||| |||||||||||| || |
651 CCGAGTGAAGACACTGCCAGACCGGACATGGACCTACTCCTTCTCCGGAG 700
                .         .         .         .         .
509 CCGTGCTCTTTTCCATGGGTTTCCTCGTCGGCTTGCTCTGTTATCTGGGC 558
    || | || || |||||||| ||||||||||  | ||||| || ||| ||
701 CCTTCCTGTTCTCCATGGGCTTCCTCGTCGCAGTACTCTGCTACCTGAGC 750
                .         .         .         .         .
559 TACAAATACATCACCAAGCCACCTGTACCTCCTAACTCCCTGAACGTCCA 608
    |||| |||  |||||||||| |||| |||||| ||||||||||||||||
751 TACAGATATGTCACCAAGCCGCCTGCACCTCCCAACTCCCTGAACGTCCA 800
```

FIG. 10A

```
 609 ACGTGTCCTGACCTTTCAACCCCTACGCTTCATCCAAGAACACGTACTGA 658
     || |||||||| || || || || |||||||||||| || ||||| ||||
 801 GCGAGTCCTGACTTTCCAGCCGCTGCGCTTCATCCAGGAGCACGTCCTGA 850

 659 TCCCTGTCTTGGACCTCAGTGGCCCCAGCAGTCTGCCTCAGCCCATCCAG 708
     |||||||||| |||||||| |||||||||||||||| | |||||  |||||
 851 TCCCTGTCTTTGACCTCAGCGGCCCCAGCAGTCTGGCCCAGCCTGTCCAG 900

 709 TACTCCCAAGTGGTGGTGTCTGGGCCCAGGGAGCCTCCTGGAGCTGTGTG 758
     ||||||||  |   |||||||||| ||||||||||  | ||||||
 901 TACTCCCAGATCAGGGTGTCTGGACCCAGGGAGCCCGCAGGAGCTCCACA 950

 759 GCGGCAGAGCCTGTCTGACCTCACCTACGTAGGGCAGTCAGATGTCTCCA 808
     ||||||  |||||||| ||  |||||||| |||||||| ||||  ||||||
 951 GCGGCATAGCCTGTCCGAGATCACCTACTTAGGGCAGCCAGACATCTCCA 1000

 809 TCCTGCAACCTACCAACGTGCCAGCTCAGCAGACACTGTCCCCACCATCC 858
     ||||  ||  ||  |||||||||||  |||  ||||  ||  |||||||  |||
1001 TCCTCCAGCCCTCCAACGTGCCACCTCCCCAGATCCTCTCCCCACTGTCC 1050

 859 TACGCTCCGAAGGCTGTCCCTGAGGTCCAGCCCCCTTCCTATGCGCCTCA 908
     ||  ||  ||  ||  ||||  ||||||||||  ||||||  ||||||||  |||||
1051 TATGCCCCAAACGCTGCCCCTGAGGTCGGGCCCCCATCCTATGCACCTCA 1100

 909 GGTAGCCTCGGATGCCAAAGCTCTGTTCTACTCACCACAACAGGGGATGA 958
     |||  ||  |  ||  ||  ||   |  ||||||  |  |||||          |
1101 GGTGACCCCCGAAGCTCAATTCCCATTCTACGCCCCACAGGCCATCTCTA 1150

 959 AGACCAGGCCTGCCACCTATGACCCGCAGGACATTCTGGACAGCTGCCCT 1008
     ||  |  ||||  ||  ||||||  |||  ||  |  ||  ||  |||||||||  |||
1151 AGGTCCAGCCTTCCTCCTATGCCCCTCAAGCCACTCCGGACAGCTGGCCT 1200

1009 GCTTCTTATGCTGTGTGTGT.GTGTGCGCGCGCGAGCGCT.CACGTCTGT 1056
     |  ||  ||||  ||  ||  |  |  |  |  |  ||  |  ||  |  |  |||
1201 CCCTCCTATGGGGTATGCATGGAAGGTTCTGGCAAAGACTCCCCCACTGG 1250

1057 ATG..................................................... 1059

1251 GACACTTTCTAGTCCTAAACACCTTAGGCCTAAAGGTCAGCTTCAGAAAG 1300
```

FIG. 10B

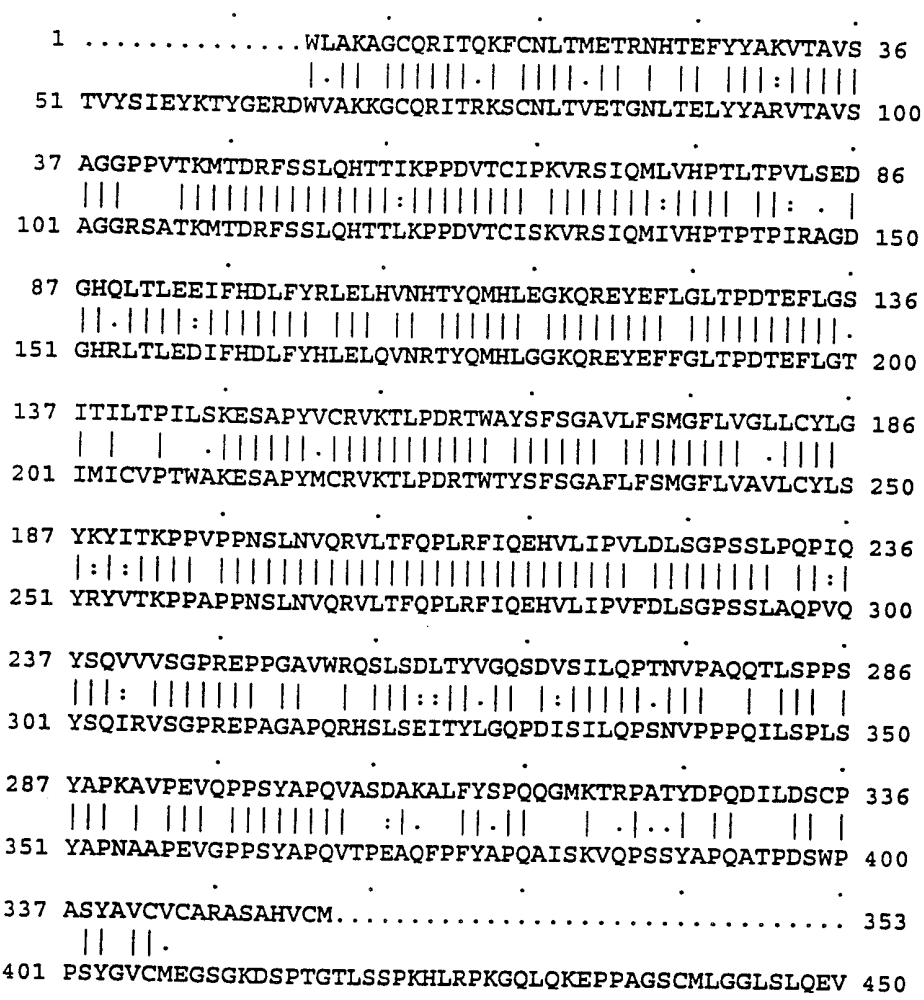

1 .....WLAKAGCQRITQKFCNLTMETRNHTEFYYAKVTAVS 36
51 TVYSIEYKTYGERDWVAKKGCQRITRKSCNLTVETGNLTELYYARVTAVS 100
37 AGGPPVTKMTDRFSSLQHTTIKPPDVTCIPKVRSIQMLVHPTLTPVLSED 86
101 AGGRSATKMTDRFSSLQHTTLKPPDVTCISKVRSIQMIVHPTPTPIRAGD 150
87 GHQLTLEEIFHDLFYRLELHVNHTYQMHLEGKQREYEFLGLTPDTEFLGS 136
151 GHRLTLEDIFHDLFYHLELQVNRTYQMHLGGKQREYEFFGLTPDTEFLGT 200
137 ITILTPILSKESAPYVCRVKTLPDRTWAYSFSGAVLFSMGFLVGLLCYLG 186
201 IMICVPTWAKESAPYMCRVKTLPDRTWTYSFSGAFLFSMGFLVAVLCYLS 250
187 YKYITKPPVPPNSLNVQRVLTFQPLRFIQEHVLIPVLDLSGPSSLPQPIQ 236
251 YRYVTKPPAPPNSLNVQRVLTFQPLRFIQEHVLIPVFDLSGPSSLAQPVQ 300
237 YSQVVVSGPREPPGAVWRQSLSDLTYVGQSDVSILQPTNVPAQQTLSPPS 286
301 YSQIRVSGPREPAGAPQRHSLSEITYLGQPDISILQPSNVPPPQILSPLS 350
287 YAPKAVPEVQPPSYAPQVASDAKALFYSPQQGMKTRPATYDPQDILDSCP 336
351 YAPNAAPEVGPPSYAPQVTPEAQFPFYAPQAISKVQPSSYAPQATPDSWP 400
337 ASYAVCVCARASAHVCM..... 353
401 PSYGVCMEGSGKDSPTGTLSSPKHLRPKGQLQKEPPAGSCMLGGLSLQEV 450

FIG. 11

```
                .         .         .         .         .
  1 .........GTCCGCTGAGATGGACAGAATGCTTTATTTTGGAAAGAAAC 41
              |  | | ||||||||  |  || |  || |||||  ||||
  1 GTCGACCCACGCGTCCGGGATGGACACACAGCCTCCTTCTGGAAGAAAAC 50
                .         .         .         .         .
 42 AATGTTCTAGGTCAAACTGAGTCT..............ACCAAATGCAG 76
    | |||    || ||| ||||||  |               ||||||  |||
 51 AGTGTCTCCGGGCAAGCTGAGTTTGCTAGAGACCCTCACACCAAA..CAG 98
                .         .         .         .         .
 77 ACTTTCACAATGGTTCTAGAAGAAATCTGGACAAGTCTTTTCATGTGGTT 126
    ||   ||||||| ||    | | | ||||||||||||   ||||||||||
 99 ACCAGCACAATGATTTCCCAGGGAGTCTGGACAAGTCCGATCATGTGGTT 148
                .         .         .         .         .
127 TTTCTACGCATTGATTCCATGTTTGCTCACAGATGAAGTGGCCATTCTGC 176
    |||||||    |||   || |||| ||||| |||| |||| || || | |
149 TTTCTACAGCATGACCTCAGGTTTTCTCACGGATGCAGTGTCCGTTTTAC 198
                .         .         .         .         .
177 CTGCCCCTCAGAACCTCTCTGTACTCTCAACCAACATGAAGCATCTCTTG 226
    ||||||||||||||||||||||||  ||||||||||||||||||||||||
199 CTGCCCCTCAGAACCTCTCTGTACAGTCAACCAACATGAAGCATCTCTTG 248
                .         .         .         .         .
227 ATGTGGAGCCCAGTGATCGCGCCTGGAGAAACAGTGTACTATTCTGTCGA 276
    ||||||| ||||||||  |  ||| || || || |||  ||||| ||| ||
249 ATGTGGAACCCAGTGACCCAGCCGGGGGAGACGGTGCTCTATTGTGTGGA 298
                .         .         .         .         .
277 ATACCAGGGGGAGTACGAGAGCCTGTACACGAGCCACATCTGGATCCCCA 326
     || ||||||||||||||||||||||||  ||||||||||||||||||||
299 GTATCAGGGGGAGTACGAGAGCCTGTACATGAGCCACATCTGGATCCCCA 348
                .         .         .         .         .
327 GCAGCTGGTGCTCACTCACTGAAGGTCCTGAGTGTGATGTCACTGATGAC 376
    ||||    |||||||  ||  || |||  ||||||||||||||  ||||||
349 GCAGTCAATGCTCACCGACCAAAAGTCTGGAGTGTGATGTCACCGATGAC 398
                .         .         .         .         .
377 ATCACGGCCACTGTGCCATACAACCTTCGTGTCAGGGCCACATTGGGCTC 426
    || || |||||  ||||||||  |||  |  |||| |||||   |||||||
399 ATAACCGCCACAGTGCCATATAACTTCAGGGTCAAGGCCATGCTGGGCTC 448
                .         .         .         .         .
427 ACAGACCTCAGCCTGGAGCATCCTGAAGCATCCCTTTAATAGAAACTCAA 476
    ||||||  ||||||||||||||  ||||  ||||  ||||||||  ||||  |||
449 ACAGACTTCAGCCTGGAGCAACCTGGAGCACCCCTTTAACCGAAATGCAA 498
                .         .         .         .         .
477 CCATCCTTACCCGACCTGGGATGGAGATCACCAAAGATGGCTTCCACCTG 526
    |  ||||  ||||  |||  ||||||||  ||||  ||  ||||  |  ||  |||
499 CTGTCCTCACCCCACCCAGGATGGAGGTCACTGAACATGGGCTGCATCTG 548
                .         .         .         .         .
527 GTTATTGAGCTGGAGGACCTGGGGCCCCAGTTTGAGTTCCTTGTGGCCTA 576
    ||||||||||||||  ||||||||  ||||||||||||||||||||||  |||
549 GTTATTGAGCTGGAAGACCTGGGACCCCAGTTTGAGTTCCTTGTGGTCTA 598
                .         .         .         .         .
577 CTGGAGGAGGGAGCCTGGTGCCGAGGAACATGTCAAAATGGTGAGGAGTG 626
    ||||||||||||||||||  ||||  |||||||||  ||  ||||||||||||
599 CTGGAGGAGGGAGCCTGGCGCCGCGGAACATGTTAAGATGGTGAGGAGTG 648
```

FIG. 12A

```
627 GGGGTATTCCAGTGCACCTAGAAACCATGGAGCCAGGGGCTGCATACTGT 676
649 GGGACATTCCGGTGCACCTAGAAACCATGGAACCGGGGGCCATGTACTGT 698

677 GTGAAGGCCCAGACATTCGTGAAGGCCATTGGGAGGTACAGCGCCTTCAG 726
699 GTGAAGGCGCAGGCACTGGTGAAAGCCATCGGGAGGCACAGTGCCTTCAG 748

727 CCAGACAGAATGTGTGGAGGTGCAAGGAGAGGCCATTCCCCTGGTACTGG 776
749 CCAGCCTACGTGTGTGGAGATGCAAGGAGAGTCTCTTCCGCTGGCACTAG 798

777 CCCTGTTTGCCTTTGTTGGCTTCATGCTGATCCTTGTGGTCGTGCCACTG 826
799 CTCTGTTTGCGTTTGTTGGCTTCATGCTGATTCTCGTGGTTGTACTACTC 848

827 TTCGTCTGGAAAATGGGCCGGCTGCTCCAGTACTCCTGTTGCCCCGTGGT 876
849 TCCGTCTGGAAGATGGGCCAGCTGCTCCGGTATTCTTGCTGCCCCGCCGT 898

877 GGTCCTCCCAGACACCTTGAAAATAACCAATTCACCCCAGAAGTTAATCA 926
899 TGTCCTCCCAGACACCTTGAAAATAACCAGTTCGTCTCAGAAGCTGATCA 948

927 GCTGCAGAAGGGAGGAGGTGGATGCCTGTGCCACGGCTGTGATGTCTCCT 976
949 GCTGCAGGAAGGAGGAGGTGGACGCCTGTGCTGTGGCTGTGCTGTCCTCG 998

977 GAGGAACTCCTCAGGGCCTGGATCTCATAGGTTTGCGGAAGGGCCCAGGT 1026
999 GAGCATCTCTTTGGGGGTCTGGATCTCACAGACTTGAG..AAGGACCTGAT 1046

1027 GAAGCCGAGAACCTGGTCTGCATGACATGGAAACCATGAGGGGACAAGTT 1076
1047 GGA.TTAACAGCCTGGTCTGCAGGACCCAGAACCCTGGATGGGGTGGGTT 1095

1077 GTGTTTCTGTTTTCCGCCACGGACAAGGGA.TGAGAGAAGTAGGAAGAGC 1125
1096 GTGTTTCTGTTTTCCTCTGTGGACAAAAGACAAAGAGCAGTAACAGGAGC 1145

1126 CTGTTGTCTACAAGTCTAGAAGCAACCATCAGAGGCAGGGTGGTTTGTCT 1175
1146 CTACTGTGTGCGCGTCTAGCAGTGACCGTCAGAGGCAGAGTGGCGTGACC 1195

1176 AACAGAACACTGACTG..AGGCTTAGGGGATGTGACCTCTAGACTGGGGG 1223
1196 ATC.CAAGAGTGACAGAAATAGTCTAGAGAGGTGATCCCCAGACACACAG 1244

1224 CTGCCACTTGCTGGCTGAGCAACCCTGGGAAAAGTGACTTCATCCCTTCG 1273
1245 TTACCATACACTCACTGAATAACAGTGGAGAAAATGGCTTCATACC.TCG 1293
```

FIG. 12B

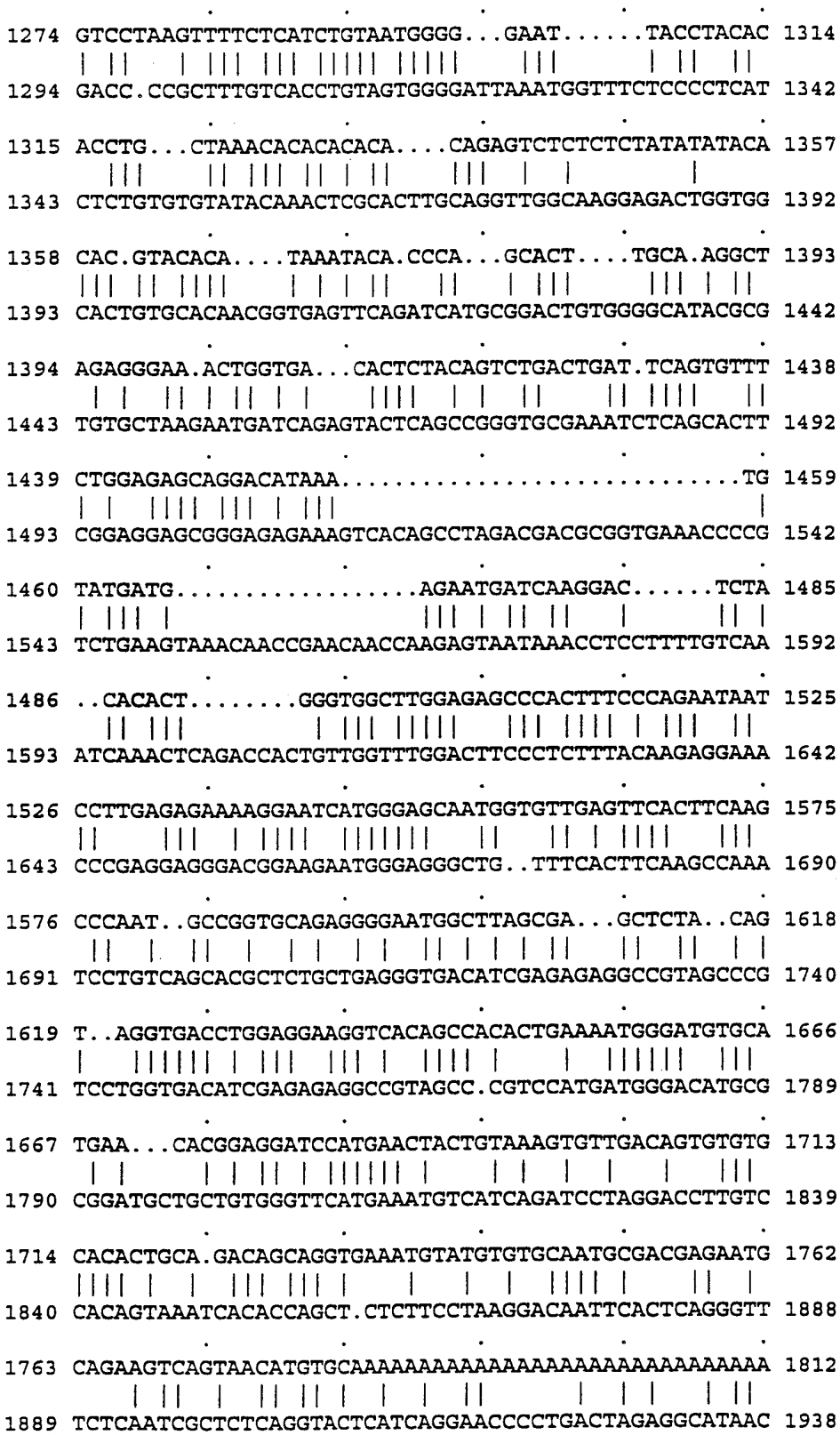

1274 GTCCTAAGTTTTCTCATCTGTAATGGGG...GAAT......TACCTACAC 1314
1294 GACC.CCGCTTTGTCACCTGTAGTGGGGATTAAATGGTTTCTCCCCTCAT 1342
1315 ACCTG...CTAAACACACACACA....CAGAGTCTCTCTCTATATATACA 1357
1343 CTCTGTGTGTATACAAACTCGCACTTGCAGGTTGGCAAGGAGACTGGTGG 1392
1358 CAC.GTACACA....TAAATACA.CCCA...GCACT....TGCA.AGGCT 1393
1393 CACTGTGCACAACGGTGAGTTCAGATCATGCGGACTGTGGGGCATACGCG 1442
1394 AGAGGGAA.ACTGGTGA...CACTCTACAGTCTGACTGAT.TCAGTGTTT 1438
1443 TGTGCTAAGAATGATCAGAGTACTCAGCCGGGTGCGAAATCTCAGCACTT 1492
1439 CTGGAGAGCAGGACATAAA.....TG 1459
1493 CGGAGGAGCGGGAGAGAAAGTCACAGCCTAGACGACGCGGTGAAACCCCG 1542
1460 TATGATG.....AGAATGATCAAGGAC......TCTA 1485
1543 TCTGAAGTAAACAACCGAACAACCAAGAGTAATAAACCTCCTTTTGTCAA 1592
1486 ..CACACT.....GGGTGGCTTGGAGAGCCCACTTTCCCAGAATAAT 1525
1593 ATCAAACTCAGACCACTGTTGGTTTGGACTTCCCTCTTTACAAGAGGAAA 1642
1526 CCTTGAGAGAAAAGGAATCATGGGAGCAATGGTGTTGAGTTCACTTCAAG 1575
1643 CCCGAGGAGGGACGGAAGAATGGGAGGGCTG..TTTCACTTCAAGCCAAA 1690
1576 CCCAAT..GCCGGTGCAGAGGGGAATGGCTTAGCGA...GCTCTA..CAG 1618
1691 TCCTGTCAGCACGCTCTGCTGAGGGTGACATCGAGAGAGGCCGTAGCCCG 1740
1619 T..AGGTGACCTGGAGGAAGGTCACAGCCACACTGAAAATGGGATGTGCA 1666
1741 TCCTGGTGACATCGAGAGAGGCCGTAGCC.CGTCCATGATGGGACATGCG 1789
1667 TGAA...CACGGAGGATCCATGAACTACTGTAAAGTGTTGACAGTGTGTG 1713
1790 CGGATGCTGCTGTGGGTTCATGAAATGTCATCAGATCCTAGGACCTTGTC 1839
1714 CACACTGCA.GACAGCAGGTGAAATGTATGTGTGCAATGCGACGAGAATG 1762
1840 CACAGTAAATCACACCAGCT.CTCTTCCTAAGGACAATTCACTCAGGGTT 1888
1763 CAGAAGTCAGTAACATGTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1812
1889 TCTCAATCGCTCTCAGGTACTCATCAGGAACCCCTGACTAGAGGCATAAC 1938

FIG. 12C

```
                .         .         .         .         .
1813 AAAAAAAAAAGGGCGGCCGC.............................. 1832
       |  |  |      |||  ||
1939 CATCACTACCTCCAGGCAGCGAGGGAGGGGTGGAGGACAGCTTCAGTCAG 1988
```

FIG. 12D

```
  1 ATGCAGACTTTCACAATGGTTCTAGAAGAAATCTGGACAAGTCTTTTCAT 50
                  ||| ||    | | | ||||||||||||   ||||
  1 ..............ATGATTTCCCAGGGAGTCTGGACAAGTCCGATCAT 35

 51 GTGGTTTTTCTACGCATTGATTCCATGTTTGCTCACAGATGAAGTGGCCA 100
    |||||||||||||    |||   || |||| ||||| |||| |||| ||
 36 GTGGTTTTTCTACAGCATGACCTCAGGTTTTCTCACGGATGCAGTGTCCG 85

101 TTCTGCCTGCCCCTCAGAACCTCTCTGTACTCTCAACCAACATGAAGCAT 150
    || | ||||||||||||||||||||||||  ||||||||||||||||||
 86 TTTTACCTGCCCCTCAGAACCTCTCTGTACAGTCAACCAACATGAAGCAT 135

151 CTCTTGATGTGGAGCCCAGTGATCGCGCCTGGAGAAACAGTGTACTATTC 200
    |||||||||||||| ||||||||  |  ||| || || || |||  |||||
136 CTCTTGATGTGGAACCCAGTGACCCAGCCGGGGGAGACGGTGCTCTATTG 185

201 TGTCGAATACCAGGGGGAGTACGAGAGCCTGTACACGAGCCACATCTGGA 250
    ||| || || ||||||||||||||||||||||||  ||||||||||||||
186 TGTGGAGTATCAGGGGGAGTACGAGAGCCTGTACATGAGCCACATCTGGA 235

251 TCCCCAGCAGCTGGTGCTCACTCACTGAAGGTCCTGAGTGTGATGTCACT 300
    ||||||||||    |||||||  ||  || |||  ||||||||||||||
236 TCCCCAGCAGTCAATGCTCACCGACCAAAAGTCTGGAGTGTGATGTCACC 285

301 GATGACATCACGGCCACTGTGCCATACAACCTTCGTGTCAGGGCCACATT 350
    ||||||||  || |||||  ||||||||  |||  |   | ||||  |||||   |
286 GATGACATAACCGCCACAGTGCCATATAACTTCAGGGTCAAGGCCATGCT 335

351 GGGCTCACAGACCTCAGCCTGGAGCATCCTGAAGCATCCCTTTAATAGAA 400
    ||||||||||||  ||||||||||||  ||||  ||||  ||||||||   |||
336 GGGCTCACAGACTTCAGCCTGGAGCAACCTGGAGCACCCCTTTAACCGAA 385

401 ACTCAACCATCCTTACCCGACCTGGGATGGAGATCACCAAAGATGGCTTC 450
    |  ||||  ||||  ||||  |||  ||||||||  ||||  ||  ||||  |
386 ATGCAACTGTCCTCACCCCACCCAGGATGGAGGTCACTGAACATGGGCTG 435

451 CACCTGGTTATTGAGCTGGAGGACCTGGGGCCCCAGTTTGAGTTCCTTGT 500
    ||  ||||||||||||||||  ||||||||  ||||||||||||||||||
436 CATCTGGTTATTGAGCTGGAAGACCTGGGACCCCAGTTTGAGTTCCTTGT 485

501 GGCCTACTGGAGGAGGGAGCCTGGTGCCGAGGAACATGTCAAAATGGTGA 550
    ||  ||||||||||||||||||||  ||||  |||||||||  ||  |||||||
486 GGTCTACTGGAGGAGGGAGCCTGGCGCCGCGGAACATGTTAAGATGGTGA 535

551 GGAGTGGGGGTATTCCAGTGCACCTAGAAACCATGGAGCCAGGGGCTGCA 600
    |||||||||  |||||  ||||||||||||||||||||  ||  |||||
536 GGAGTGGGGACATTCCGGTGCACCTAGAAACCATGGAACCGGGGGCCATG 585

601 TACTGTGTGAAGGCCCAGACATTCGTGAAGGCCATTGGGAGGTACAGCGC 650
    ||||||||||||||  |||  ||  |  |||||  |||||  ||||||  ||||  ||
586 TACTGTGTGAAGGCGCAGGCACTGGTGAAAGCCATCGGGAGGCACAGTGC 635
```

FIG. 13A

```
651 CTTCAGCCAGACAGAATGTGTGGAGGTGCAAGGAGAGGCCATTCCCCTGG 700
    |||||||||| |    ||||||||| |||||||||| |  |||| ||||
636 CTTCAGCCAGCCTACGTGTGTGGAGATGCAAGGAGAGTCTCTTCCGCTGG 685

701 TACTGGCCCTGTTTGCCTTTGTTGGCTTCATGCTGATCCTTGTGGTCGTG 750
    ||| || |||||||| |||||||||||||||||||| || ||||| ||
686 CACTAGCTCTGTTTGCGTTTGTTGGCTTCATGCTGATTCTCGTGGTTGTA 735

751 CCACTGTTCGTCTGGAAAATGGGCCGGCTGCTCCAGTACTCCTGTTGCCC 800
    | ||| | ||||||||| ||||||| |||||||| ||| || || |||||
736 CTACTCTCCGTCTGGAAGATGGGCCAGCTGCTCCGGTATTCTTGCTGCCC 785

801 CGTGGTGGTCCTCCCAGACACCTTGAAAATAACCAATTCACCCCAGAAGT 850
    ||  || |||||||||||||||||||||||||| |||  | ||||||
786 CGCCGTTGTCCTCCCAGACACCTTGAAAATAACCAGTTCGTCTCAGAAGC 835

851 TAATCAGCTGCAGAAGGGAGGAGGTGGATGCCTGTGCCACGGCTGTGATG 900
    | |||||||||||| | |||||||||||| |||||||| |||||||  ||
836 TGATCAGCTGCAGGAAGGAGGAGGTGGACGCCTGTGCTGTGGCTGTGCTG 885

901 TCTCCTGAGGAACTCCTCAGGGCCTGGATCTCA..... 933
    ||  | ||| | ||| |  ||| ||||||||||
886 TCCTCGGAGCATCTCTTTGGGGTCTGGATCTCACAGACT 924
```

FIG. 13B

```
                .         .         .         .         .
  1 MQTFTMVLEEIWTSLFMWFFYALIPCLLTDEVAILPAPQNLSVLSTNMKH 50
        |: : :|||  |||||.:    ||| |.:|||||||||| ||||||
  1 .....MISQGVWTSPIMWFFYSMTSGFLTDAVSVLPAPQNLSVQSTNMKH 45
                .         .         .         .         .
 51 LLMWSPVIAPGETVYYSVEYQGEYESLYTSHIWIPSSWCSLTEGPECDVT 100
    ||||.||  ||||| | |||||||||||| |||||||| || |.  |||||
 46 LLMWNPVTQPGETVLYCVEYQGEYESLYMSHIWIPSSQCSPTKSLECDVT 95
                .         .         .         .         .
101 DDITATVPYNLRVRATLGSQTSAWSILKHPFNRNSTILTRPGMEITKDGF 150
    |||||||||| ||:| ||||||||| |.|||||.|:|| | ||:|. |
 96 DDITATVPYNFRVKAMLGSQTSAWSNLEHPFNRNATVLTPPRMEVTEHGL 145
                .         .         .         .         .
151 HLVIELEDLGPQFEFLVAYWRREPGAEEHVKMVRSGGIPVHLETMEPGAA 200
    |||||||||||||||| |||||||| ||||||||| ||||||||||||
146 HLVIELEDLGPQFEFLVVYWRREPGAAEHVKMVRSGDIPVHLETMEPGAM 195
                .         .         .         .         .
201 YCVKAQTFVKAIGRYSAFSQTECVEVQGEAIPLVLALFAFVGFMLILVVV 250
    ||||||  ||||||:|||||  |||.|||.:|| ||||||||||||||||
196 YCVKAQALVKAIGRHSAFSQPTCVEMQGESLPLALALFAFVGFMLILVVV 245
                .         .         .         .         .
251 PLFVWKMGRLLQYSCCPVVVLPDTLKITNSPQKLISCRREEVDACATAVM 300
     | |||||.||.||||| ||||||||||.| |||||||:||||||| ||:
246 LLSVWKMGQLLRYSCCPAVVLPDTLKITSSSQKLISCRKEEVDACAVAVL 295
                .
301 SPEELLRAWIS.. 311
    | | |   |||
296 SSEHLFGVWISQT 308
```

FIG. 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/31328

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL : 530/350; 536/23.5, 24.31; 435/6, 69.1, 320.1, 361
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350; 536/23.5, 24.31; 435/6, 69.1, 320.1, 361

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

SEQUENCE DATABASE MPSRCH: a-geneseq35, pir 60, swiss-prot37, strembl19, embl-est58, genbank-est111, embl58, genbank111, ngeneseq35

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database embl-est58, WashU-Merck EST Project, (St. Louis, MO, USA), AN AA132964, HILLIER et al, 'zo22b02.s1 Stratagene colon (#937204) Homo sapiens cDNA clone IMAGE:587595 3', mRNA sequence,' November 1996, Nucleotides 2366-2816 of SEQ ID NO:1 are 98.7% identical to nucleotides from this sequence, and so comprise a fragment of at least 300 nucleotides of SEQ ID NO:1.	1, 2 and 5
X	WO 94/13801 A1 (SCHERING CORPORATION) 23 June 1994 (12/06/94), see entire document, especially pages 59-61, SEQ ID NO:2 which is 100% identical to SEQ ID NO:19.	1-10, 12, 16 and 17
A,P	WO 99/07848 A1 (ZYMOGENETICS, INC.) 18 February 1999 (18/02/99), see entire document.	1-10, 12, 16 and 17

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 23 MARCH 2000	Date of mailing of the international search report 26 APR 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer EILEEN B. O'HARA Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/31328

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-10, 12, 16 and 17

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/31328

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

C07K 14/47; C07H 21/04; C12N 5/06, 15/63; C12P 21/06; C12Q 1/68

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-10, 12, 16 and 17, as far as they are drawn to nucleic acid molecules, vectors, host cells containing recombinant nucleic acid molecules, polypeptides and a method of producing polypeptides of human and murine Tango 241, SEQ ID NOS:1-3 and 19-21, respectively.
Group II, claim(s)1-10, 12, 16 and 17, as far as they are drawn to nucleic acid molecules, vectors, host cells containing recombinant nucleic acid molecules, polypeptides and a method of producing polypeptides of human and murine Tango 242, SEQ ID NOS:11-13 and 22-24, respectively.
Group III, claim(s) 11 and 15, as far as they are drawn to antibodies and binding compounds of Tango 241.
Group IV, claim(s) 11 and 15, as far as they are drawn to antibodies and binding compounds of Tango 242.
Group V, claim(s) 13, 14, 19, 20 and 22, as far as they are drawn to a method for detecting a polypeptide or to a method for identifying a compound which binds to or modulates the activity of a Tango 241 polypeptide by binding assays.
Group VI, claim(s) 13, 14, 19, 20 and 22, as far as they are drawn to a method for detecting a polypeptide or to a method for identifying a compound which binds to or modulates the activity of a Tango 242 polypeptide by binding assays.
Group VII, claim(s) 18, as far as it is drawn to a kit comprising a compound which selectively hybridizes to a Tango 241 nucleic acid molecule.
Group VIII, claim(s) 18, as far as it is drawn to a kit comprising a compound which selectively hybridizes to a Tango 242 nucleic acid molecule.
Group IX, claim(s) 21, as far as it is drawn to a method for modulating the activity of a Tango 241 polypeptide.
Group X, claim(s) 21, as far as it is drawn to a method for modulating the activity of a Tango 242 polypeptide.

The inventions listed as Groups I-XII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Pursuant to 37 C.F.R.§ 1.475(d), the ISA/US considers that where multiple products and processes are claimed, the main invention shall consit of the first invention of the category first mentioned in the claims and the first recited invention of each of the other categories related thereto. Accordingly, the main invention (Group I) comprises the first recited product, polynucleotides encoding human and murine Tango 241 proteins, vectors, host cells, a recombinant method of producing Tango 241 protein, and the Tango 241 proteins. Further pursuant to 37 C.F.R.§ 1.475(d), the ISA/US considers that any feature which the subsequently recited products and methods share with the main invention does not constitute a special techincal feature within the meaning of PCT Rule 13.2 and that each of such products and methods accordingly defines a separate invention.