



Office de la Propriété  
Intellectuelle  
du Canada

Un organisme  
d'Industrie Canada

Canadian  
Intellectual Property  
Office

An agency of  
Industry Canada

CA 2427701 A1 2002/05/10

(21) **2 427 701**

(12) **DEMANDE DE BREVET CANADIEN  
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2001/10/29

(87) Date publication PCT/PCT Publication Date: 2002/05/10

(85) Entrée phase nationale/National Entry: 2003/04/29

(86) N° demande PCT/PCT Application No.: US 2001/050938

(87) N° publication PCT/PCT Publication No.: 2002/036633

(30) Priorité/Priority: 2000/10/30 (60/244,285) US

(51) Cl.Int.<sup>7</sup>/Int.Cl.<sup>7</sup> C12N 15/12, C07K 14/72, C07K 14/705

(71) Demandeur/Applicant:  
LEXICON GENETICS INCORPORATED, US

(72) Inventeurs/Inventors:  
HU, YI, US;  
ZAMBROWICZ, BRIAN, US;  
NEPOMNICHY, BORIS, US

(74) Agent: OGILVY RENAULT

(54) Titre : NOUVELLES PROTEINES HUMAINES 7<sup>TM</sup> ET POLYNUCLEOTIDES CODANT POUR CES PROTEINES

(54) Title: HUMAN 7<sup>TM</sup> PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abrégé/Abstract:

The nucleotide and amino acid sequences of human G protein coupled receptors are described.



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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
10 May 2002 (10.05.2002)

PCT

(10) International Publication Number  
**WO 02/036633 A3**

(51) International Patent Classification<sup>7</sup>: **C12N 15/12**,  
C07K 14/705, 14/72

(21) International Application Number: PCT/US01/50938

(22) International Filing Date: 29 October 2001 (29.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/244,285 30 October 2000 (30.10.2000) US

(71) Applicant: **LEXICON GENETICS INCORPORATED**  
[US/US]; 4000 Research Forest Drive, The Woodlands, TX  
77381 (US).

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

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

**Published:**

— with international search report

(88) Date of publication of the international search report:  
27 March 2003

(72) Inventors: **HU, Yi**; 333 Holly Creek Ct. #203, The Woodlands, TX 77381 (US). **ZAMBROWICZ, Brian**; 18 Firethorne Place, The Woodlands, TX 77382 (US). **NEPOMNICHY, Boris**; 905 Cypress Station Dr. #1515, Houston, TX 77909 (US).

(74) Agents: **ISHIMOTO, Lance, K.** et al.; Lexicon Genetics Incorporated, 8800 Technology Forest Place, The Woodlands, TX 77381 (US).

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

(54) Title: HUMAN 7<sup>TM</sup> PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: The nucleotide and amino acid sequences of human G protein coupled receptors are described.



**WO 02/036633 A3**

NOVEL HUMAN 7<sup>TM</sup> PROTEINS AND  
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S.  
Provisional Application Number 60/244,285, which was filed on  
October 30, 2000, which is herein incorporated by reference in  
5 its entirety.

1. INTRODUCTION

The present invention relates to the discovery,  
identification and characterization of novel human  
10 polynucleotides that encode membrane associated proteins and  
receptors. The invention encompasses the described  
polynucleotides, host cell expression systems, the encoded  
proteins, fusion proteins, polypeptides and peptides, antibodies  
to the encoded proteins and peptides, and genetically engineered  
15 animals that lack the disclosed genes, or overexpress the  
disclosed genes, or antagonists and agonists of the proteins,  
and other compounds that modulate the expression or activity of  
the proteins encoded by the disclosed genes that can be used for  
diagnosis, drug screening, clinical trial monitoring, the  
20 treatment of physiological or behavioral disorders, and/or  
cosmetic or nutraceutical applications.

2. BACKGROUND OF THE INVENTION

Membrane receptor proteins can serve as integral components  
25 of cellular mechanisms for sensing their environment, and  
maintaining cellular homeostasis and function. Accordingly,  
membrane receptor proteins are often involved in transduction  
pathways that control cell physiology, chemical communication,  
and gene expression. A particularly relevant class of membrane  
30 receptors are those typically characterized by the presence of 7  
conserved transmembrane domains that are interconnected by

nonconserved hydrophilic loops. Such, "7TM receptors" include a superfamily of receptors known as G-protein coupled receptors (GPCRs). GPCRs are typically involved in transduction pathways involving G-proteins or PPG proteins. As such, the GPCR family  
5 includes many receptors that are known to serve as drug targets for therapeutic agents.

### 3. SUMMARY OF THE INVENTION

The present invention relates to the discovery,  
10 identification, and characterization of nucleotides that encode novel GPCRs, and the corresponding novel GPCR (NGPCR) amino acid sequences. The NGPCRs described for the first time herein are transmembrane proteins that span the cellular membrane and are involved in signal transduction after ligand binding. The  
15 described NGPCRs have structural motifs found in the 7TM receptor family. Expression of the described NGPCRs can be detected in a variety of human cells.

The novel human GPCR sequences described herein encode proteins of 1,210, 733, 1,138, and 662 amino acids in length  
20 (see respectively SEQ ID NOS: 2, 4, 6, and 8). The described NGPCRs have multiple transmembrane regions (of about 20-30 amino acids) characteristic of 7TM proteins, as well as several predicted cytoplasmic domains.

Additionally contemplated are "knockout" ES cells that have  
25 been engineered using conventional methods (see, for example, PCT Patent Application No. PCT/US98/03243, filed February 20, 1998, herein incorporated by reference). A knockout ES cell line has been produced in a murine homolog of the disclosed sequences. Accordingly, an additional aspect of the present  
30 invention includes knockout cells and animals having genetically engineered mutations in the sequences encoding the presently described NGPCRs.

The invention encompasses the nucleotides presented in the Sequence Listing, expression vectors that have been engineered to incorporate one or more of the nucleotides presented in the Sequence Listing, host cells expressing such nucleotides, and  
5 the expression products of such nucleotides, and: (a) nucleotides that encode mammalian homologs of the described NGPCRs, including the specifically described human NGPCRs, and the human NGPCR gene products; (b) nucleotides that encode one or more portions of the NGPCRs that correspond to functional  
10 domains, and the polypeptide products specified by such nucleotide sequences, including, but not limited to, the novel regions of the described extracellular domain(s) (ECD), one or more transmembrane domain(s) (TM) first disclosed herein, and the cytoplasmic domain(s) (CD); (c) isolated nucleotides that  
15 encode mutants, engineered or naturally occurring, of the described NGPCRs, in which all or a part of at least one of the domains is deleted or altered, and the polypeptide products specified by such nucleotide sequences, including, but not limited to, soluble receptors in which all or a portion of a TM  
20 is deleted (in the case of the described 7TMs a soluble product can be generated by engineering a protein to include only the region upstream from the first TM such that all downstream TMs are deleted), and nonfunctional receptors in which all or a portion of one or more of the CD(s) is deleted; (d) nucleotides  
25 that encode fusion proteins containing all or a portion of the coding region from a NGPCR, or one of its domains (e.g., an extracellular domain) fused to another peptide or polypeptide; and (e) therapeutic or diagnostic derivatives of the described polynucleotides, such as oligonucleotides, antisense  
30 polynucleotides, ribozymes, dsRNA, or gene therapy constructs, comprising one or more of the sequences first disclosed in the Sequence Listing.

The invention also encompasses agonists and antagonists of the NGPCRs, including small molecules, large molecules, mutant NGPCR proteins, or portions thereof that compete with the native NGPCR, and antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NGPCRs (e.g., antisense and ribozyme molecules, and gene or regulatory sequence replacement constructs) or to enhance the expression of the described NGPCR gene (e.g., expression constructs that place the described sequence under the control of a strong promoter system), and transgenic animals that express a NGPCR transgene or "knock-outs" that do not express a functional NGPCR. Knock-out mice can be produced in several ways, one of which involves the use of mouse embryonic stem cell ("ES cell") lines that contain gene trap mutations in a murine homolog of at least one of the described NGPCRs. When the unique NGPCR sequences described in SEQ ID NOS:1-9 are "knocked-out" they provide a method of identifying phenotypic expression of the particular gene, as well as a method of assigning function to previously unknown genes. In addition, animals in which the unique NGPCR sequences described in SEQ ID NOS:1-9 are "knocked-out" provide a unique source in which to elicit antibodies to homologous and orthologous proteins, which would have been previously viewed by the immune system as "self" and therefore would have failed to elicit significant antibody responses. To these ends, gene trapped knockout ES cells have been generated in murine homologs of the described NGPCRs.

Additionally, the unique NGPCR sequences described in SEQ ID NOS:1-9 are useful for the identification of protein coding sequences and mapping unique genes to one or more particular chromosome (the gene encoding the described NGPCRs is apparently X-linked, see GENBANK accession no. AL161778). These sequences identify actual, biologically relevant, exon splice junctions, as opposed to those that might have been predicted

bioinformatically from genomic sequence alone. The sequences of the present invention are also useful as additional DNA markers for restriction fragment length polymorphism (RFLP) analysis, and in forensic biology.

5 Further, the present invention also relates to methods of using the described NGPCR nucleotide sequences and/or NGPCR gene products for the identification of compounds that modulate, *i.e.*, act as agonists or antagonists of, NGPCR gene expression and/or NGPCR gene product activity. Such compounds can be used  
10 as therapeutic agents for the treatment of various symptomatic representations of biological disorders or imbalances.

#### 4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

The Sequence Listing provides the sequences of the  
15 described NGPCR ORFs and the amino acid sequences encoded thereby. SEQ ID NO:9 describes a NGPCR ORF and flanking regions.

#### 5. DETAILED DESCRIPTION OF THE INVENTION

20 The human NGPCRs described for the first time herein are novel receptor proteins that are expressed in human testis, small intestine, and uterus cells. The described NGPCR sequences were obtained using human genomic sequences in conjunction with cDNAs generated from mRNAs from human testis,  
25 small intestine, and uterus (Edge Biosystems, Gaithersburg, MD, and Clontech, Palo Alto, CA). The described NGPCRs are transmembrane proteins of the 7<sup>TM</sup> family of receptors. As with other GPCRs, signal transduction is triggered when a ligand binds to the receptor. Interfering with the binding of the  
30 natural ligand, or neutralizing or removing the ligand, or interference with its binding to a NGPCR will effect NGPCR mediated signal transduction. Because of their biological significance, 7<sup>TM</sup>, and particularly GPCR, proteins have been

subjected to intense scientific and commercial scrutiny (see, for example, U.S. Patent Nos. 5,942,416 and 5,891,720, both of which are herein incorporated by reference in their entirety, for applications, uses, and assays involving NGPCRs). In  
5 addition to 7<sup>TM</sup> proteins, the presently described NGPCRs share significant homology with GPCRs of the human epididymis 6 (HE6), secretin, and latrotoxin receptor families among others.

The invention encompasses the use of the described NGPCR nucleotides, NGPCR proteins and peptides, as well as antibodies,  
10 preferably humanized monoclonal antibodies, or binding fragments, domains, or fusion proteins thereof, to the NGPCRs (which can, for example, act as NGPCR agonists or antagonists), antagonists that inhibit receptor activity or expression, or agonists that activate receptor activity or increase its  
15 expression, in the diagnosis and treatment of disease.

In particular, the invention described in the subsections below encompasses NGPCR polypeptides or peptides corresponding to functional domains of NGPCR (e.g., ECD, TM or CD), mutated, truncated or deleted NGPCRs (e.g., NGPCRs missing one or more  
20 functional domains or portions thereof, such as,  $\Delta$ ECD,  $\Delta$ TM, and/or  $\Delta$ CD), NGPCR fusion proteins (e.g., a NGPCR or a functional domain of a NGPCR, such as the ECD, fused to an unrelated protein or peptide such as an immunoglobulin constant region, i.e., IgFc), nucleotide sequences encoding such  
25 products, and host cell expression systems that can produce such NGPCR products.

The invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of the NGPCRs, as well as compounds or nucleotide  
30 constructs that inhibit expression of a NGPCR gene (transcription factor inhibitors, antisense and ribozyme molecules, or gene or regulatory sequence replacement

constructs), or promote expression of NGPCR (e.g., expression constructs in which NGPCR coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, etc.). The invention also relates to host  
5 cells and animals genetically engineered to express the human NGPCRs (or mutants thereof) or to inhibit or "knock-out" expression of the animal's endogenous NGPCR genes.

The NGPCR proteins or peptides, NGPCR fusion proteins, NGPCR nucleotide sequences, antibodies, antagonists and agonists  
10 can be useful for the detection of mutant NGPCRs or inappropriately expressed NGPCRs for the diagnosis of disease. The NGPCR proteins or peptides, NGPCR fusion proteins, NGPCR nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and  
15 animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of perturbing the normal function of a NGPCR in the body. The use of engineered host cells and/or animals may offer an advantage in  
20 that such systems allow not only for the identification of compounds that bind to an ECD of a NGPCR, but can also identify compounds that affect the signal transduced by an activated NGPCR.

Finally, the NGPCR protein products (especially soluble  
25 derivatives such as peptides corresponding to a NGPCR ECD, or truncated polypeptides lacking one or more TM domains) and fusion protein products (especially NGPCR-Ig fusion proteins, i.e., fusions of a NGPCR, or a domain of a NGPCR, e.g., ECD,  $\Delta$ TM to an IgFc), antibodies and anti-idiotypic antibodies (including  
30 Fab fragments), antagonists or agonists (including compounds that modulate signal transduction, which may act on downstream targets in a NGPCR-mediated signal transduction pathway) can be

used for therapy of such diseases. For example, the administration of an effective amount of soluble NGPCR ECD,  $\Delta$ TM, or an ECD-IgFc fusion protein or an anti-idiotypic antibody (or its Fab) that mimics the NGPCR ECD would "mop up" or  
5 "neutralize" the endogenous NGPCR ligand, and prevent or reduce binding and receptor activation. Nucleotide constructs encoding such NGPCR products can be used to genetically engineer host cells to express such products *in vivo*; these genetically engineered cells function as "bioreactors" in the body,  
10 delivering a continuous supply of a NGPCR, a NGPCR peptide, soluble ECD or  $\Delta$ TM or a NGPCR fusion protein, that will "mop up" or neutralize a NGPCR ligand. Nucleotide constructs encoding functional NGPCRs, mutant NGPCRs, as well as antisense and ribozyme molecules can be used in "gene therapy" approaches for  
15 the modulation of NGPCR expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

Various aspects of the invention are described in greater detail in the subsections below.

20

### 5.1 THE NGPCR POLYNUCLEOTIDES

The cDNA sequences and deduced amino acid sequences of the described human NGPCRs are presented in the Sequence Listing. The described NGPCRs can be expressed in a variety of human  
25 tissues, as described above, and are apparently encoded on the human X chromosome.

Several polymorphism were identified during the sequencing of the NGPCRs, including a T/C polymorphism at nucleotide position 3601 of SEQ ID NO:1 (both of which result in a thr  
30 being present at the corresponding amino acid (aa) position 1201 of SEQ ID NO:2), and a T/C polymorphism at nucleotide position 2173 of SEQ ID NO:3 (both of which result in a thr being present at the corresponding aa position 725 of SEQ ID NO:4).

The NGPCRs of the present invention include the human DNA sequences presented in the Sequence Listing (and vectors comprising the same), and additionally contemplates any nucleotide sequence encoding a contiguous and functional NGPCR open reading frame (ORF) that hybridizes to a complement of the DNA sequences presented in the Sequence Listing under highly stringent conditions, e.g., hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York, at p. 2.10.3) and encodes a functionally equivalent gene product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of DNA sequences that encode and express an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, e.g., washing in 0.2x SSC/0.1% SDS at 42°C (Ausubel et al., 1989, supra), yet that still encode a functionally equivalent NGPCR gene product. Functional equivalents of a NGPCR include naturally occurring NGPCRs present in other species, and mutant NGPCRs, whether naturally occurring or engineered (by site directed mutagenesis, gene shuffling, directed evolution as described in, for example, U.S. Patent Nos. 5,837,458 and 5,723,323, both of which are herein incorporated by reference in their entirety). The invention also includes degenerate nucleic acid variants of the disclosed NGPCR polynucleotide sequences.

Additionally contemplated are polynucleotides encoding NGPCR ORFs, or their functional equivalents, encoded by polynucleotide sequences that are about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the polynucleotide sequences described in the Sequence Listing (as measured by BLAST sequence comparison analysis using, for

example, the GCG sequence analysis package described herein using standard default parameters).

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore  
5 the complements of, the described NGPCR nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described herein. In instances wherein the nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such molecules are about 16 to about 100 bases long, about 20 to  
10 about 80 bases long, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein, incorporating a contiguous region of nucleotide sequence first disclosed in the present Sequence Listing, and can be used in conjunction with the polymerase chain reaction (PCR) to screen  
15 libraries, isolate clones, and prepare cloning and sequencing templates, *etc.*

Alternatively, such NGPCR oligonucleotides can be used as hybridization probes for screening libraries, and assessing gene expression patterns (particularly using a microarray or high-  
20 throughput "chip" format). Additionally, a series of the described NGPCR oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NGPCR sequences. An oligonucleotide or polynucleotide sequence first disclosed in at least a portion of one or more of  
25 the sequences of SEQ ID NOS:1-9 can be used as a hybridization probe in conjunction with a solid support matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or polycrystalline substrates, *etc.*). Of particular note are spatially addressable  
30 arrays (*i.e.*, gene chips, microtiter plates, *etc.*) of oligonucleotides and polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the

biopolymers present on the spatially addressable array comprises an oligonucleotide or polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS:1-9, or an amino acid sequence encoded thereby. Methods for attaching  
5 biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding studies thereon are disclosed in, *inter alia*, U.S. Patent Nos. 5,700,637, 5,556,752, 5,744,305, 4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186, and 4,689,405, the disclosures of which are herein  
10 incorporated by reference in their entirety.

Addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-9 can be used to identify and characterize the temporal and tissue specific expression of a gene. These addressable arrays incorporate oligonucleotide sequences of  
15 sufficient length to confer the required specificity, yet be within the limitations of the production technology. The length of these probes is within a range of between about 8 to about 2000 nucleotides. Preferably the probes consist of 60 nucleotides and more preferably 25 nucleotides from the  
20 sequences first disclosed in SEQ ID NOS:1-9.

For example, a series of the described oligonucleotide sequences, or the complements thereof, can be used in chip format to represent all or a portion of the described NGPCRs. The oligonucleotides, typically between about 16 to about 40 (or  
25 any whole number within the stated range) nucleotides in length, can partially overlap each other and/or the NGPCR sequence may be represented using oligonucleotides that do not overlap. Accordingly, the described polynucleotide sequences shall typically comprise at least about two or three distinct  
30 oligonucleotide sequences of at least about 18 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences can begin at any nucleotide present within a sequence in the Sequence Listing and

proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense (3'-to-5') orientation.

Microarray-based analysis allows the discovery of broad patterns of genetic activity, providing new understanding of gene functions and generating novel and unexpected insight into transcriptional processes and biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-9 provides detailed information about transcriptional changes involved in a specific pathway, potentially leading to the identification of novel components or gene functions that manifest themselves as novel phenotypes.

Probes consisting of sequences first disclosed in SEQ ID NOS:1-9 can also be used in the identification, selection and validation of novel molecular targets for drug discovery. The use of these unique sequences permits the direct confirmation of drug targets and recognition of drug dependent changes in gene expression that are modulated through pathways distinct from the intended target of the drug. These unique sequences therefore also have utility in defining and monitoring both drug action and toxicity.

As an example of utility, the sequences first disclosed in SEQ ID NOS:1-9 can be utilized in microarrays or other assay formats, to screen collections of genetic material from patients who have a particular medical condition. These investigations can also be carried out using the sequences first disclosed in SEQ ID NOS:1-9 *in silico* and by comparing previously collected genetic databases and the disclosed sequences using computer software known to those in the art.

Thus the sequences first disclosed in SEQ ID NOS:1-9 can be used to identify mutations associated with a particular disease, and also in diagnostic or prognostic assays.

Although the presently described sequences have been specifically described using nucleotide sequence, it should be

appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present  
5 within a given region of the sequence in conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in SEQ ID NOS:1-9. Alternatively, a restriction map specifying the relative positions of restriction  
10 endonuclease digestion sites, or various palindromic or other specific oligonucleotide sequences can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (e.g., the University of Wisconsin GCG sequence analysis package, SEQUENCHER 3.0, Gene Codes Corp., Ann Arbor, MI, etc.), can  
15 optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more additional sequence(s) or one or more restriction sites present in the disclosed sequence.

20 For oligonucleotides probes, highly stringent conditions may refer, e.g., to washing in 6x SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). The described oligonucleotides may encode or act as NGPCR antisense  
25 molecules, useful, for example, in NGPCR gene regulation (and/or as antisense primers in amplification reactions of NGPCR gene nucleic acid sequences). With respect to NGPCR gene regulation, such techniques can be used to regulate biological functions. Further, such sequences may be used as part of ribozyme and/or  
30 triple helix sequences, also useful for NGPCR gene regulation.

Additionally, the antisense oligonucleotides may comprise at least one modified base moiety that is selected from the

group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) 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.

The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier *et al.*, 1987, Nucl. Acids Res. 15:6625-

6641). The oligonucleotide is a 2'-O-methylribonucleotide (Inoue *et al.*, 1987, *Nucl. Acids Res.* 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987, *FEBS Lett.* 215:327-330).

Alternatively, double stranded RNA can be used to disrupt the  
5 expression and function of a targeted NGPCR.

Oligonucleotides of the invention may be synthesized by standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples,

10 phosphorothioate oligonucleotides may be synthesized (Stein *et al.*, 1988, *Nucl. Acids Res.* 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:7448-7451), etc.

15 Low stringency conditions are well-known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory*  
20 *Manual*, Cold Spring Harbor Press, N.Y. (and periodic updates thereof); and Ausubel *et al.*, 1989, *supra* (and periodic updates thereof).

Alternatively, suitably labeled NGPCR nucleotide probes may be used to screen a human genomic library using appropriately  
25 stringent conditions, or by PCR. The identification and characterization of human genomic clones is helpful for identifying polymorphisms (including, but not limited to, nucleotide repeats, microsatellite alleles, single nucleotide polymorphisms, or coding single nucleotide polymorphisms),  
30 determining the genomic structure of a given locus/allele, and designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the human gene

can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (e.g., splice acceptor and/or donor sites), etc., that can be used in diagnostics and pharmacogenomics.

5       For example, the present sequences can be used in restriction fragment length polymorphism (RFLP) analysis to identify specific individuals. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield  
10 unique bands for identification (as generally described in U.S. Patent No. 5,272,057, incorporated herein by reference). In addition, 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  
15 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). Actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction  
20 enzyme generated fragments.

      Further, NGPCR homologs may be isolated from nucleic acids of the organism of interest by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of amino acid sequences within the NGPCR gene products disclosed  
25 herein. The template for the reaction may be total RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from, for example, human or non-human cell lines or tissue(s) known to express, or suspected of expressing, a NGPCR gene allele.

30       The PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NGPCR gene. The PCR fragment may then be used to

isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to isolate  
5 genomic clones via the screening of a genomic library.

PCR technology may also be utilized to isolate full length NGPCR cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source (*i.e.*, one known to express, or suspected of  
10 expressing, a NGPCR gene). A reverse transcription (RT) reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal  
15 transferase reaction, the hybrid may be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies that may be used, see *e.g.*, Sambrook *et al.*,  
20 1989, *supra*.

A cDNA of a mutant NGPCR gene can be isolated, for example, by using PCR. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known to express, or suspected of  
25 expressing, a mutant NGPCR allele, in an individual putatively carrying a mutant NGPCR allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized using an oligonucleotide that hybridizes specifically to the 5' end of the normal gene. Using these two  
30 primers, the product is then amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well-known to those of skill in the

art. By comparing the DNA sequence of the mutant NGPCR allele to that of the normal NGPCR allele, the mutation(s) responsible for the loss or alteration of function of the mutant NGPCR gene product can be ascertained.

5 Alternatively, a genomic library can be constructed using DNA obtained from an individual suspected of carrying, or known to carry, a mutant NGPCR allele, or a cDNA library can be constructed using RNA from a tissue known to express, or suspected of expressing, a mutant NGPCR allele. A normal NGPCR  
10 gene, or any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NGPCR allele in such libraries. Clones containing the mutant NGPCR gene sequences can then be purified and subjected to sequence analysis according to methods well-known to those of skill in  
15 the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known to express, or suspected of expressing, a mutant NGPCR allele in an individual suspected of carrying, or known to  
20 carry, such a mutant allele. In this manner, gene products made by the putatively mutant tissue may be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal NGPCR gene product, as described, below, in Section 5.3 (for screening techniques, see,  
25 for example, Harlow and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, incorporated herein in its entirety by reference).

Additionally, screening can be accomplished by screening with labeled NGPCR fusion proteins, such as, for example,  
30 alkaline phosphatase-NGPCR or NGPCR-alkaline phosphatase fusion proteins. In cases where a NGPCR mutation results in an expressed gene product with altered function (e.g., as a result of a missense or a frameshift mutation), a polyclonal set of

antibodies to NGPCR are likely to cross-react with the mutant NGPCR gene product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well-known to those of skill in the art.

The invention also encompasses nucleotide sequences that encode mutant NGPCRs, peptide fragments of the NGPCRs, truncated NGPCRs, and NGPCR fusion proteins. These include, but are not limited to, nucleotide sequences encoding mutant NGPCRs described below, polypeptides or peptides corresponding to one or more ECD, TM and/or CD domains of the NGPCR or portions of these domains, truncated NGPCRs in which one or two of the domains is deleted, e.g., a soluble NGPCR lacking the TM or both the TM and CD regions, or a truncated, nonfunctional NGPCR lacking all or a portion of the CD region. Nucleotides encoding fusion proteins may include, but are not limited to, full length NGPCR sequences, truncated NGPCRs, or nucleotides encoding peptide fragments of NGPCR fused to an unrelated protein or peptide, such as, for example, a transmembrane sequence, which anchors the NGPCR ECD to the cell, an IgFc domain, which increases the stability and half life of the resulting fusion protein (e.g., NGPCR-Ig) in the bloodstream, or an enzyme, fluorescent protein, or luminescent protein that can be used as a marker.

The invention also encompasses: (a) DNA vectors that contain any of the foregoing NGPCR coding sequences and/or their complements (*i.e.*, antisense); (b) DNA expression vectors that contain any of the foregoing NGPCR coding sequences operatively associated with at least a first regulatory element that directs the expression of the coding sequences (for example, baculovirus vectors as described in U.S. Patent No. 5,869,336 herein incorporated by reference); (c) genetically engineered host

cells that contain any of the foregoing NGPCR coding sequences operatively associated with at least a first regulatory element that directs the expression of the coding sequences in the host cell; and (d) genetically engineered host cells that express an  
5 endogenous NGPCR sequence under the control of an exogenously introduced regulatory element (*i.e.*, gene activation). As used herein, regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and  
10 regulate expression. Such regulatory elements include, but are not limited to, the cytomegalovirus hCMV immediate early gene, regulatable, viral (particularly retroviral LTR promoters) the early or late promoters of SV40 or adenovirus, the *lac* system, the *trp* system, the *tet* system, the *TAC* system, the *TRC* system,  
15 the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast  $\alpha$ -mating factors.

An additional application of the described NGPCR  
20 polynucleotide sequences is their use in the molecular mutagenesis/evolution of proteins that are at least partially encoded by the described novel sequences using, for example, polynucleotide shuffling or related methodologies. Such approaches are described in U.S. Patent Nos. 5,830,721,  
25 5,723,323, and 5,837,458, which are herein incorporated by reference in their entirety.

Additionally contemplated uses for the described sequences include the engineering of constitutively "on" variants for use in cell assays and genetically engineered animals, using the  
30 methods and applications described in U.S. Provisional Patent Application Ser Nos. 60/110,906, 60/106,300, 60/094,879, and

60/121,851, each of which are herein incorporated by reference in their entirety.

NGPCR gene products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, worms, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, birds, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees, may be used to generate NGPCR transgenic animals.

Any technique known in the art may be used to introduce a NGPCR transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Patent No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci., USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell 57:717-723); *etc.* For a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. 115:171-229, which is incorporated by reference herein in its entirety.

The present invention provides for transgenic animals that carry the NGPCR transgene in all their cells, as well as animals that carry the transgene in some, but not all their cells, *i.e.*, mosaic animals or somatic cell transgenic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89:6232-6236. The regulatory sequences required for such a cell-type specific activation will depend upon the particular

cell type of interest, and will be apparent to those of skill in the art.

When it is desired that a NGPCR transgene be integrated into the chromosomal site of the endogenous NGPCR gene, gene  
5 targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous NGPCR gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the  
10 nucleotide sequence of the endogenous NGPCR gene (*i.e.*, "knockout" animals).

The transgene can also be selectively introduced into a particular cell type, thus inactivating the endogenous NGPCR gene in only that cell type, by following, for example, the  
15 teaching of Gu *et al.*, 1994, *Science*, 265:103-106. The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression  
20 of the recombinant NGPCR gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the  
25 transgenic animals may also be assessed using techniques that include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of NGPCR gene-expressing tissue may also be evaluated immunocytochemically using antibodies  
30 specific for the NGPCR transgene product.

## 5.2 NGPCR PROTEINS AND POLYPEPTIDES

NGPCR proteins, polypeptides and peptide fragments, mutated, truncated or deleted forms of the NGPCRs and/or NGPCR fusion proteins can be prepared for a variety of uses, including, but not limited to, use as protein therapeutics, the generation of antibodies, as reagents in diagnostic assays, the identification of other cellular gene products related to a NGPCR, as reagents in assays for screening for compounds that can be used as pharmaceutical reagents useful in the therapeutic treatment of mental, biological, or medical disorders (*i.e.*, kidney disorders, digestive disorders, infertility, improper blood pressure, body weight disorders, *etc.*) and disease. The described NGPCRs share structural similarity with compounds including, but not limited to, latrotoxin receptors and latrophilins. Given the similarity information and expression data, the described NGPCRs can be targeted (by drugs, oligos, antibodies, *etc.*) in order to treat disease, or to therapeutically augment the efficacy of therapeutic agents.

The Sequence Listing discloses the amino acid sequences encoded by the described NGPCR nucleotide sequences. The NGPCRs have initiator methionines in DNA sequence contexts consistent with translation initiation sites, followed by hydrophobic signal sequences typical of membrane associated proteins or secreted proteins. The sequence data presented herein indicate that alternatively spliced forms of the NGPCRs exist.

The NGPCR amino acid sequences of the invention include the amino acid sequences presented in the Sequence Listing, as well as analogues and derivatives thereof. Further, corresponding NGPCR homologues from other species are encompassed by the invention. In fact, any NGPCR protein encoded by the NGPCR nucleotide sequences described above are within the scope of the invention, as are any novel polynucleotide sequences encoding

all or any novel portion of an amino acid sequence presented in the Sequence Listing. The degenerate nature of the genetic code is well-known, and, accordingly, each amino acid presented in the Sequence Listing is generically representative of the well-known nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As such, as contemplated herein, the amino acid sequences presented in the Sequence Listing, when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, Darnell et al., eds., Scientific American Books, New York, NY, herein incorporated by reference), are generically representative of all the various permutations and combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are functionally equivalent to the NGPCRs encoded by the described nucleotide sequences as judged by any of a number of criteria, including, but not limited to, the ability to bind a ligand of a NGPCR, the ability to effect an identical or complementary signal transduction pathway, a change in cellular metabolism (e.g., ion flux, tyrosine phosphorylation, etc.), or a change in phenotype when the NGPCR equivalent is present in an appropriate cell type (such as the amelioration, prevention or delay of a biochemical, biophysical, or overt phenotype). Such functionally equivalent NGPCR proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequences encoded by the NGPCR nucleotide sequences described above, but that result in a silent change, thus producing a functionally equivalent gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine,

leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include  
5 arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

While random mutations can be made to a NGPCR DNA (using random mutagenesis techniques well-known to those skilled in the art), and the resulting mutant NGPCRs tested for activity, site-  
10 directed mutations of the NGPCR coding sequence can be engineered (using site-directed mutagenesis techniques well-known to those skilled in the art) to generate mutant NGPCRs with increased function (e.g., higher binding affinity for the target ligand and/or greater signaling capacity) or decreased  
15 function (e.g., weaker binding affinity for the target ligand and/or decreased signal transduction capacity). One starting point for such analysis is by aligning the disclosed human sequences with corresponding gene/protein sequences from other mammals, for example, in order to identify amino acid sequence  
20 motifs that are conserved between different species. Non-conservative changes can be engineered at variable positions to alter function, signal transduction capability, or both. Alternatively, where alteration of function is desired, deletion or non-conservative alterations of the conserved regions (*i.e.*,  
25 identical amino acids) can be engineered, for example, deletion or non-conservative alterations (substitutions or insertions) of the various conserved transmembrane domains.

Other mutations in the NGPCR coding sequence can be made to generate NGPCRs that are better suited for expression, scale up,  
30 *etc.*, in the host cells chosen. For example, cysteine residues can be deleted or substituted with another amino acid in order to eliminate disulfide bridges, and N-linked glycosylation sites

can be altered or eliminated to achieve, for example, expression of a homogeneous product that is more easily recovered and purified from yeast hosts, which are known to hyperglycosylate N-linked sites. To this end, a variety of amino acid

5 substitutions at one or both of the first or third amino acid positions of any one or more of the glycosylation recognition sequences that occur in the ECD (N-X-S or N-X-T), and/or an amino acid deletion at the second position of any one or more such recognition sequences in the ECD, will prevent  
10 glycosylation of a NGPCR at the modified tripeptide sequence (see, e.g., Miyajima *et al.*, 1986, EMBO J. 5:1193-1197).

Peptides corresponding to one or more domains of the NGPCR (e.g., ECD, TM, CD, etc.), truncated or deleted NGPCRs (e.g., NGPCR in which a ECD, TM and/or CD is deleted), as well as  
15 fusion proteins in which a full length NGPCR, a NGPCR peptide, or truncated NGPCR is fused to an unrelated protein, are also within the scope of the invention, and can be designed on the basis of the presently disclosed NGPCR nucleotide and NGPCR amino acid sequences. Such fusion proteins include, but are not  
20 limited to: IgFc fusions, which stabilize the NGPCR protein or peptide and prolong half-life *in vivo*; fusions to any amino acid sequence that allows the fusion protein to be anchored to the cell membrane, allowing an ECD to be exhibited on the cell surface; or fusions to an enzyme, fluorescent protein, or  
25 luminescent protein that provides a marker function.

While the NGPCR polypeptides and peptides can be chemically synthesized (e.g., see Creighton, 1983, *Proteins: Structures and Molecular Principles*, W.H. Freeman & Co., N.Y.), large polypeptides derived from a NGPCR and full length NGPCRs can be  
30 advantageously produced by recombinant DNA technology using techniques well-known in the art for expressing nucleic acid containing NGPCR gene sequences and/or coding sequences. Such

methods can be used to construct expression vectors containing one or more of the presently described NGPCR nucleotide sequences and appropriate transcriptional and translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (see, for example, the techniques described in Sambrook *et al.*, 1989, *supra*, and Ausubel *et al.*, 1989, *supra*). Alternatively, RNA corresponding to all or a portion of a transcript encoded by a NGPCR nucleotide sequence may be chemically synthesized using, for example, synthesizers (see, for example, the techniques described in "Oligonucleotide Synthesis", 1984, Gait, M.J., ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety).

A variety of host-expression vector systems may be utilized to express the NGPCR nucleotide sequences of the invention. Where the NGPCR peptide or polypeptide is a soluble derivative (e.g., NGPCR peptides corresponding to an ECD; truncated or deleted NGPCR in which a TM and/or CD are deleted), the peptide or polypeptide can be recovered from the culture, *i.e.*, from the host cell in cases where the NGPCR peptide or polypeptide is not secreted, and from the culture media in cases where the NGPCR peptide or polypeptide is secreted by the host cell. However, such expression systems also encompass engineered host cells that express a NGPCR, or functional equivalent, *in situ*, *i.e.*, anchored in the cell membrane. Purification or enrichment of NGPCR from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well-known to those skilled in the art. However, such engineered host cells themselves may be used in situations where it is important not only to retain the structural and functional characteristics of the NGPCR, but to assess biological activity, e.g., in drug screening assays.

The expression systems that may be used for purposes of the invention include, but are not limited to, microorganisms such as bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA  
5 expression vectors containing NGPCR nucleotide sequences; yeast (e.g., *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NGPCR nucleotide sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing NGPCR nucleotide sequences; plant  
10 cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing NGPCR nucleotide sequences; or mammalian cell systems (e.g., COS, CHO, BHK, 293, 3T3) harboring  
15 recombinant expression constructs containing NGPCR nucleotide sequences and promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter or the vaccinia virus 7.5K promoter).

20 In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NGPCR gene product being expressed. For example, when a large quantity of such a protein is to be produced, such as in the generation of pharmaceutical compositions of NGPCR protein or  
25 for raising antibodies to a NGPCR protein, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which a NGPCR coding  
30 sequence may be ligated individually into the vector in frame with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye and Inouye, 1985, Nucleic Acids

Res. 13:3101-3109; Van Heeke and Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors may also be used to express NGPCR proteins, polypeptides or peptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the NGPCR protein, polypeptide or peptide can be released from the GST moiety.

In an exemplary insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express NGPCR nucleotide sequences. The virus grows in *Spodoptera frugiperda* cells. A NGPCR gene coding sequence may be cloned individually into a non-essential region (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of NGPCR coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (e.g., see Smith et al., 1983, J. Virol. 46:584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a NGPCR nucleotide sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric sequence may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination.

Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a NGPCR gene product in infected hosts (e.g., see Logan and Shenk, 1984, Proc. Natl.

5 Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted NGPCR nucleotide sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire NGPCR gene or cDNA, including its own initiation codon and adjacent  
10 sequences, is inserted into an appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a NGPCR coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, may be provided.  
15 Furthermore, the initiation codon should be in phase with the reading frame of the desired NGPCR coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of  
20 expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bitter et al., 1987, Methods in Enzymol. 153:516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies  
25 and processes the expression product, in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-  
30 translational processing and modification of proteins and gene or expression products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing

of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells  
5 include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, and WI38 cell lines.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the NGPCR sequences described above  
10 may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a  
15 selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the  
20 plasmid into their chromosomes and grow to form foci, which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express a NGPCR gene product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds  
25 that affect the endogenous activity of a NGPCR gene product.

A number of selection systems can be used, including, but not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, 1977, *Cell* 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, 1962, *Proc.*  
30 *Natl. Acad. Sci. USA* 48:2026), and adenine phosphoribosyltransferase (Lowy *et al.*, 1980, *Cell* 22:817) genes, which can be employed in tk<sup>-</sup>, hgprt<sup>-</sup> or aprt<sup>-</sup> cells,

respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler *et al.*, 1980, Natl. Acad. Sci. USA 77:3567; O'Hare *et al.*, 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan and Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, 1981, J. Mol. Biol. 150:1); and hygro, which confers resistance to hygromycin (Santerre *et al.*, 1984, Gene 30:147).

Alternatively, any fusion protein can be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, one such system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht *et al.*, 1991, Proc. Natl. Acad. Sci. USA 88:8972-8976). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni<sup>2+</sup>-nitriloacetic acid-agarose columns, and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are fusion proteins that direct a NGPCR to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation of NGPCRs to antibody molecules, or their Fab fragments, could be used to target cells bearing a particular epitope. Attaching an appropriate signal sequence to a NGPCR would also transport the NGPCR to a desired location within the cell. Alternatively, targeting of NGPCRs or their nucleic acid sequences might be achieved using liposomes or lipid complex based delivery

systems. Such technologies are described in "Liposomes: A Practical Approach", New, R.R.C., ed., Oxford University Press, New York, and in U.S. Patent Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures, which  
5 are herein incorporated by reference in their entirety.

Additionally embodied are novel protein constructs engineered in such a way that they facilitate transport of a NGPCR to a target site or desired organ, where the NGPCR crosses the cell membrane and/or the nucleus and exerts its functional activity. This  
10 goal may be achieved by coupling of a NGPCR to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain (see generally U.S. Patent Application Ser. Nos. 60/111,701 and 60/056,713, both of which are herein incorporated by reference, for examples of such  
15 transducing sequences), to facilitate passage across cellular membranes, and/or can optionally be engineered to include one or more nuclear localization signal(s).

### 5.3 ANTIBODIES TO NGPCR PROTEINS

20 Antibodies that specifically recognize one or more epitopes of a NGPCR, or epitopes of conserved variants of a NGPCR, or peptide fragments of a NGPCR, are also encompassed by the invention. Such antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized  
25 or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')<sub>2</sub> fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

The antibodies of the invention may be used, for example,  
30 in the detection of NGPCRs in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal amounts of one or more NGPCR. Such antibodies may also be utilized in

conjunction with, for example, compound screening schemes, as described below, for the evaluation of the effect of test compounds on expression and/or activity of a NGPCR expression or gene product. Additionally, such antibodies can be used in  
5 conjunction with gene therapy to, for example, evaluate the normal and/or engineered NGPCR-expressing cells prior to their introduction into a patient. Such antibodies may additionally be used as a method for the inhibition of abnormal NGPCR activity. Thus, such antibodies may be utilized as part of a  
10 variety of therapeutic regimens, such as, for example, in kidney disorder, digestive disorder, infertility, improper blood pressure, and/or body weight disorder treatment methods.

For the production of antibodies, various host animals may be immunized by injection with a NGPCR, one or more NGPCR  
15 peptide (e.g., one corresponding to a functional domain of the receptor, such as an ECD, TM or CD), truncated NGPCR polypeptide(s) (NGPCR in which one or more domains, e.g., a TM or CD, has been deleted), functional equivalents of a NGPCR, or mutants of a NGPCR. Such host animals may include, but are not  
20 limited to, rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including, but not limited to, Freund's adjuvant (complete and incomplete), mineral salts such as aluminum hydroxide or aluminum phosphate, chitosan, surface  
25 active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Alternatively, the immune response could be enhanced by combination and/or coupling with molecules  
30 such as keyhole limpet hemocyanin, tetanus toxoid, diphtheria toxoid, ovalbumin, cholera toxin, or fragments thereof.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any  
5 technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique (Kohler and Milstein, 1975, Nature 256:495-497; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor *et al.*, 1983, Immunology  
10 Today 4:72; Cole *et al.*, 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030), and the EBV-hybridoma technique (Cole *et al.*, 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class, including IgG, IgM, IgE, IgA, IgD, and any subclass  
15 thereof. The hybridoma producing the mAbs of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of  
20 "chimeric antibodies" (Morrison *et al.*, 1984, Proc. Natl. Acad. Sci., 81:6851-6855; Neuberger *et al.*, 1984, Nature, 312:604-608; Takeda *et al.*, 1985, Nature, 314:452-454), by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule  
25 of appropriate biological activity, can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. Such technologies are described in U.S. Patent  
30 Nos. 6,075,181 and 5,877,397 and their respective disclosures, which are herein incorporated by reference in their entirety. Also encompassed by the present invention is the use of fully

humanized monoclonal antibodies, as described in U.S. Patent No. 6,150,584 and respective disclosures, which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of  
5 single chain antibodies (U.S. Patent No. 4,946,778; Bird, 1988, Science 242:423-426; Huston *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward *et al.*, 1989, Nature 341:544-546) can be adapted to produce single chain antibodies against NGPCR peptides, polypeptides and/or proteins. Single chain antibodies  
10 are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments  
15 include, but are not limited to: F(ab')<sub>2</sub> fragments, which can be produced by pepsin digestion of an antibody molecule; and Fab fragments, which can be generated by reducing the disulfide bridges of F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, Science,  
20 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to a NGPCR can; in turn, be utilized to generate anti-idiotypic antibodies that "mimic" a given NGPCR, using techniques well-known to those skilled in the art (see, e.g.,  
25 Greenspan and Bona, 1993, FASEB J 7:437-444; and Nissinoff, 1991, J. Immunol. 147:2429-2438). For example antibodies that bind to a NGPCR ECD and competitively inhibit the binding of a ligand of NGPCR can be used to generate anti-idiotypes that "mimic" a NGPCR ECD and, therefore, bind and neutralize a  
30 ligand. Such neutralizing anti-idiotypes, or Fab fragments of such anti-idiotypes, can be used in therapeutic regimens involving the NGPCR signaling pathway.

Additionally given the high degree of relatedness of mammalian NGPCRs, the presently described knock-out mice (having never seen a particular NGPCR, and thus never been tolerized to the NGPCR), have a unique utility, as they can be advantageously  
5 applied to the generation of antibodies against the disclosed mammalian NGPCRs (*i.e.*, NGPCRs will be immunogenic in NGPCR knock-out animals).

#### 5.4 DIAGNOSIS OF ABNORMALITIES RELATED TO A NGPCR

10 A variety of methods can be employed for the diagnostic and prognostic evaluation of disorders related to NGPCR function, and for the identification of subjects having a predisposition to such disorders.

Such methods can, for example, utilize reagents such as the  
15 NGPCR nucleotide sequences described in Section 5.1, and/or NGPCR antibodies, as described, in Section 5.3. Specifically, such reagents may be used, for example, for: (1) the detection of the presence of NGPCR gene mutations, or the detection of either over- or under-expression of NGPCR mRNA relative to a  
20 given (*i.e.*, normal) phenotype; (2) the detection of either an over- or an under-abundance of NGPCR gene product relative to a given (*i.e.*, normal) phenotype; and (3) the detection of perturbations or abnormalities in the signal transduction pathway mediated by NGPCRs.

25 The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific NGPCR nucleotide sequence and/or NGPCR antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting medical  
30 disorders or abnormalities, such as, for example, kidney disorder, digestive disorder, infertility, improper blood pressure, and/or body weight disorder abnormalities.

For the detection of NGPCR mutations, any nucleated cell can be used as a starting source for genomic nucleic acid. For the detection of NGPCR gene expression or NGPCR gene products, any cell type or tissue in which a NGPCR gene is expressed can  
5 be utilized.

Nucleic acid-based detection techniques and peptide detection techniques are described in greater detail below.

#### 5.4.1 DETECTION OF NGPCR GENES AND TRANSCRIPTS

10 Mutations within a NGPCR gene or nucleotide sequence can be detected by utilizing a number of techniques. Nucleic acids from any nucleated cell can be used as the starting point for such assay techniques, and may be isolated according to standard nucleic acid preparation procedures that are well-known to those  
15 of skill in the art.

DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving NGPCR gene structure, including point mutations, insertions, deletions and chromosomal rearrangements. Such assays may include, but are  
20 not limited to, Southern analyses, single stranded conformational polymorphism analyses (SSCP), restriction fragment length polymorphisms (RFLP, as generally described in U.S. Patent No. 5,272,057, incorporated herein by reference), coding single nucleotide polymorphisms (cSNP), and PCR analyses.

25 Such diagnostic methods for the detection of NGPCR gene-specific mutations can involve, for example, contacting and incubating nucleic acids, including recombinant DNA molecules, cloned genes or degenerate variants thereof, obtained from a sample, e.g., derived from a patient sample or other appropriate  
30 cellular source, with one or more labeled nucleic acid reagents, including recombinant DNA molecules, cloned genes or degenerate variants thereof, as described in Section 5.1, under conditions favorable for the specific annealing of these reagents to their

complementary sequences within a given NGPCR gene or sequence. Preferably, the lengths of these nucleic acid reagents are at least 15 to 30 nucleotides. After incubation, all non-annealed nucleic acids are removed from the nucleic acid:NGPCR molecule  
5 hybrid. The presence of nucleic acids that have hybridized, if any such molecules exist, is then detected. In conjunction with such a detection scheme, the nucleic acid from the cell type or tissue of interest can be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on  
10 a microtiter plate or polystyrene beads. In such a case, after incubation, non-annealed, labeled nucleic acid reagents of the type described in Section 5.1 are easily removed. Detection of the remaining, annealed, labeled NGPCR nucleic acid reagents is accomplished using standard techniques well-known to those in  
15 the art. The NGPCR sequences to which the nucleic acid reagents have annealed can be compared to the annealing pattern expected from a normal NGPCR sequence in order to determine whether a NGPCR mutation is present.

Alternative diagnostic methods for the detection of NGPCR  
20 gene specific nucleic acid molecules, in patient samples or other appropriate cell sources, may involve their amplification, e.g., by PCR (the experimental embodiment set forth in U.S. Patent Nos. 4,683,195; 4,683,202 and 4,800,159, which are incorporated herein by reference in their entirety), followed by  
25 the detection of the amplified molecules using techniques well-known to those of skill in the art. The resulting amplified sequences can be compared to those that would be expected if the nucleic acid being amplified contained only normal copies of a NGPCR gene in order to determine whether a NGPCR gene mutation  
30 exists.

Additionally, well-known genotyping techniques can be performed to identify individuals carrying NGPCR gene mutations. Such techniques include, for example, the use of restriction

fragment length polymorphisms (RFLPs), which involve sequence variations in one of the recognition sites for the specific restriction enzyme used.

Additionally, improved methods for analyzing DNA polymorphisms that can be utilized for the identification of NGPCR gene mutations have been described, which capitalize on the presence of variable numbers of short, tandemly repeated DNA sequences between restriction enzyme sites. For example, Weber (U.S. Patent No. 5,075,217, which is incorporated herein by reference in its entirety) describes a DNA marker based on length polymorphisms in blocks of (dC-dA)<sub>n</sub>-(dG-dT)<sub>n</sub> short tandem repeats. The average separation of (dC-dA)<sub>n</sub>-(dG-dT)<sub>n</sub> blocks is estimated to be 30,000-60,000 bp. Markers that are so closely spaced exhibit a high frequency co-inheritance, and are extremely useful in the identification of genetic mutations, such as, for example, mutations within a given NGPCR gene, and the diagnosis of diseases and disorders related to NGPCR mutations.

Also, Caskey et al. (U.S. Patent No. 5,364,759, which is incorporated herein by reference in its entirety) describe a DNA profiling assay for detecting short tri- and tetra-nucleotide repeat sequences. The process includes extracting the DNA of interest, such as the NGPCR gene, amplifying the extracted DNA, and labeling the repeat sequences to form a genotypic map of the individual's DNA.

The level of NGPCR gene expression can also be assayed by detecting and measuring NGPCR transcription. For example, RNA from a cell type or tissue known to express, or suspected of expressing, a NGPCR gene can be isolated and tested utilizing hybridization or PCR techniques such as those described herein. The isolated cells can be derived from cell culture or from a patient. The analysis of cells taken from culture may be a necessary step in the assessment of cells to be used as part of

a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of a NGPCR gene. Such analyses may reveal both quantitative and qualitative aspects of the expression pattern of a NGPCR gene, including activation or  
5 inactivation of NGPCR gene expression.

In one embodiment of such a detection scheme, cDNAs are synthesized from the RNAs of interest (e.g., by reverse transcription of the RNA molecule into cDNA). A sequence within the cDNA is then used as the template for a nucleic acid  
10 amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (e.g., primers) in the reverse transcription and nucleic acid amplification steps of this method are chosen from among the NGPCR nucleic acid reagents described in Section  
15 5.1. The preferred lengths of such nucleic acid reagents are at least 9-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that  
20 the product may be visualized by standard ethidium bromide staining or any other suitable nucleic acid staining method, or by sequencing.

Additionally, it is possible to perform such NGPCR gene expression assays "*in situ*", i.e., directly upon tissue sections  
25 (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents, such as those described above, may be used as probes and/or primers for such *in situ* procedures (see, for example, Nuovo, 1992, "PCR *In Situ*  
30 Hybridization: Protocols And Applications", Raven Press, NY).

Alternatively, if a sufficient quantity of the appropriate cells can be obtained, standard Northern analysis can be performed to determine the level of NGPCR mRNA expression.

Additionally, NGPCR oligonucleotide or polynucleotide  
5 sequences can be used as hybridization probes in conjunction with a solid support matrix/substrate (e.g., resins, beads, membranes, plastics, polymers, metal or metallized substrates, gene chips, and crystalline or polycrystalline substrates, etc.).

10

#### 5.4.2 DETECTION OF NGPCR GENE PRODUCTS

Antibodies directed against wild-type or mutant NGPCR gene products, or conserved variants or peptide fragments thereof, which are discussed above, may also be used as diagnostics and  
15 prognostics, as described herein. Such diagnostic methods may be used to detect abnormalities in the level of NGPCR gene expression, or abnormalities in the structure and/or temporal, tissue, cellular, or subcellular location of a NGPCR, and may be performed *in vivo* or *in vitro*, such as, for example, on biopsy  
20 tissue.

For example, antibodies directed to epitopes of a NGPCR ECD can be used *in vivo* to detect the pattern and level of expression of the particular NGPCR in the body. Such antibodies can be labeled, e.g., with a radio-opaque or other appropriate  
25 compound, and injected into a subject in order to visualize binding to NGPCRs expressed in the body, using methods such as X-rays, CAT-scans, or MRI. Labeled antibody fragments, e.g., the Fab or single chain antibody comprising the smallest portion of the antigen binding region, are preferred for this purpose,  
30 in order to promote crossing the blood-brain barrier and permit labeling of NGPCRs expressed in the brain.

Additionally, any NGPCR fusion protein or NGPCR conjugated protein, whose presence can be detected, can be administered. For example, NGPCR fusion or conjugated proteins labeled with a radio-opaque or other appropriate compound can be administered  
5 and visualized *in vivo*, as discussed above for labeled antibodies. Further, such NGPCR fusion proteins as alkaline phosphatase-NGPCR or NGPCR-alkaline phosphatase fusion proteins can be utilized for *in vitro* diagnostic procedures.

Alternatively, immunoassays or fusion protein detection  
10 assays, as described above, can be utilized on biopsy and autopsy samples *in vitro* to permit assessment of the expression pattern of a NGPCR. Such assays are not confined to the use of antibodies that define a NGPCR ECD, but can include the use of antibodies directed to epitopes of any of the domains of a  
15 NGPCR, e.g., the ECD, the TM and/or CD. The use of each or all of these labeled antibodies will yield useful information regarding translation and intracellular transport of a NGPCR to the cell surface, and can identify defects in processing.

The tissue or cell type to be analyzed will generally  
20 include those that are known to express, or suspected of expressing, a NGPCR gene. The protein isolation methods employed herein may, for example, be such as those previously described (Harlow and Lane, 1988, *supra*). The isolated cells can be derived from cell culture or from a patient. The  
25 analysis of cells taken from culture may be a necessary step in the assessment of cells that could be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of a NGPCR gene.

For example, antibodies, or fragments of antibodies, such  
30 as those described above in Section 5.3 as useful in the present invention, may be used to quantitatively or qualitatively detect the presence of NGPCR gene products, or conserved variants or

peptide fragments thereof. This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below, this Section), coupled with light microscopic, flow cytometric, or fluorimetric  
5 detection. Such techniques are especially preferred if such NGPCR gene products are expressed on the cell surface.

The antibodies (or fragments thereof), or NGPCR fusion or conjugated proteins, useful in the present invention may, additionally, be employed histologically, as in  
10 immunofluorescence, immunoelectron microscopy or non-immuno assays, for *in situ* detection of NGPCR gene products or conserved variants or peptide fragments thereof, or for NGPCR binding (in the case of labeled NGPCR ligand fusion proteins).

*In situ* detection may be accomplished by removing a  
15 histological specimen from a patient, and applying thereto a labeled antibody or fusion protein of the present invention. The antibody (or fragment) or fusion protein is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is  
20 possible to determine not only the presence of a NGPCR gene product, or conserved variants or peptide fragments, or NGPCR binding, but also NGPCR distribution in the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological  
25 methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Immunoassays and non-immunoassays for NGPCR gene products, or conserved variants or peptide fragments thereof, will typically comprise incubating a sample, such as a biological  
30 fluid, a tissue extract, freshly harvested cells, or lysates of cells that have been incubated in cell culture, in the presence of a detectably labeled antibody capable of identifying NGPCR

gene products, or conserved variants or peptide fragments thereof, and detecting the bound antibody by any of a number of techniques well-known in the art.

The biological sample may be brought in contact with and  
5 immobilized onto a solid phase support or carrier, such as nitrocellulose or any other solid support that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled NGPCR antibody or NGPCR  
10 ligand fusion protein. The solid phase support may then be washed with the buffer a second time to remove unbound antibody or fusion protein. The amount of bound label remaining on the solid support may then be detected by conventional means.

By "solid phase support or carrier" is intended any support  
15 capable of binding an antigen or an antibody. Well-known supports or carriers include, but are not limited to, glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amyloses, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either  
20 soluble to some extent or insoluble for the purposes of the present invention. The support material can have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or  
25 cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding an antibody or antigen,  
30 or will be able to ascertain the same by use of routine experimentation.

The binding activity of a given lot of NGPCR antibody or NGPCR ligand fusion protein may be determined according to well-

known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

With respect to antibodies, one of the ways in which a  
5 NGPCR antibody can be detectably labeled is by linking it to an enzyme for use in an enzyme immunoassay (EIA) (Voller, 1978, Diagnostic Horizons 2:1-7, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller *et al.*, 1978, J. Clin. Pathol. 31:507-520; Butler, 1981, Meth. Enzymol. 73:482-523;  
10 Maggio, E. (ed.), 1980, Enzyme Immunoassay, CRC Press, Boca Raton, FL,; and Ishikawa *et al.*, (eds.), 1981, Enzyme Immunoassay, Kigaku Shoin, Tokyo). The enzyme that is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to  
15 produce a chemical moiety that can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes that can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol  
20 dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The  
25 detection can be accomplished by colorimetric methods, which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

30 Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect a

NGPCR through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference  
5 herein). The radioactive isotope can be detected by such means as the use of a gamma counter, a scintillation counter, or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody  
10 is exposed to light of the proper wavelength, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

15 The antibody can also be detectably labeled using fluorescence emitting metals such as  $^{152}\text{Eu}$ , or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentaacetic acid (DTPA) or  
20 ethylenediaminetetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a  
25 chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label  
30 the antibodies (or fragments thereof) of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems, in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a

bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin (green fluorescent protein and mutants thereof; as described in U. S. Patent Nos. 5,491,084, 5,625,048, 5,777,079, 5,795,737, 5,804,387, 5,874,304, 5,968,750, 5,976,796, 6,020,192, 6,027,881, 6,054,321, 6,096,865, 6,146,826, 6,172,188 and 6,265,548, each of which is hereby incorporated by reference).

10                    5.5        SCREENING ASSAYS FOR COMPOUNDS THAT  
                                         MODULATE NGPCR EXPRESSION OR ACTIVITY

The following assays are designed to identify: compounds that interact with (e.g., bind to) NGPCRs (including, but not limited to, an ECD or CD of a NGPCR); intracellular proteins that interact with a NGPCR (including, but not limited to, the TM and CD of NGPCR); compounds that interfere with the interaction of NGPCR with transmembrane or intracellular proteins involved in NGPCR-mediated signal transduction; and compounds that modulate the activity of a NGPCR gene (i.e., modulate the level of NGPCR gene expression) or modulate the level of NGPCR. Assays may additionally be utilized that identify compounds that bind to NGPCR gene regulatory sequences (e.g., promoter sequences), and that may modulate NGPCR gene expression (see e.g., Platt, 1994, J. Biol. Chem. 269:28558-28562, which is incorporated herein by reference in its entirety).

The compounds that can be screened in accordance with the invention include, but are not limited to, peptides, antibodies and fragments thereof, and other organic compounds (e.g., peptidomimetics) that bind to an ECD of a NGPCR and either mimic the activity triggered by the natural ligand (i.e., agonists) or inhibit the activity triggered by the natural ligand (i.e.,

antagonists); as well as peptides, antibodies or fragments thereof, and other organic compounds that mimic the ECD of the NGPCR (or a portion thereof) and bind to and "neutralize" the natural ligand.

5        Such compounds may include, but are not limited to, peptides, such as, for example, soluble peptides, including, but not limited to, members of random peptide libraries (see, e.g., Lam *et al.*, 1991, *Nature* 354:82-84; Houghten *et al.*, 1991, *Nature* 354:84-86), and combinatorial chemistry-derived molecular  
10       libraries made of D- and/or L- configuration amino acids, phosphopeptides (including, but not limited to, members of random or partially degenerate, directed phosphopeptide libraries; see, e.g., Songyang *et al.*, 1993, *Cell* 72:767-778), antibodies (including, but not limited to, polyclonal,  
15       monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')<sub>2</sub> and Fab expression library fragments, and epitope-binding fragments thereof), and small organic or inorganic molecules.

Other compounds that can be screened in accordance with the  
20       invention include, but are not limited to, small organic molecules that are able to cross the blood-brain barrier, gain entry into an appropriate cell (e.g., in the cerebellum, the hypothalamus, etc.), and affect the expression of a NGPCR gene or some other gene involved in the NGPCR signal transduction  
25       pathway (e.g., by interacting with the regulatory region or transcription factors involved in gene expression); or such compounds that affect the activity of a NGPCR (e.g., by inhibiting or enhancing the enzymatic activity of a CD) or the activity of some other intracellular factor involved in the  
30       NGPCR signal transduction pathway.

Computer modeling and searching technologies permit identification of compounds, or the improvement of already

identified compounds, that can modulate NGPCR expression or activity. Having identified such a compound or composition, the active sites or regions are identified. Such active sites might typically be ligand binding sites. The active site can be  
5 identified using methods known in the art including, for example, from the amino acid sequences of peptides, from the nucleotide sequences of nucleic acids, or from study of complexes of the relevant compound or composition with its natural ligand. In the latter case, chemical or X-ray  
10 crystallographic methods can be used to find the active site by finding where on the factor the complexed ligand is found.

Next, the three dimensional geometric structure of the active site is determined. This can be done by known methods, including X-ray crystallography, which can determine a complete  
15 molecular structure. On the other hand, solid or liquid phase NMR can be used to determine certain intra-molecular distances. Any other experimental method of structure determination can be used to obtain partial or complete geometric structures. The geometric structures may be measured with a complexed ligand,  
20 natural or artificial, which may increase the accuracy of the active site structure determined.

If an incomplete or insufficiently accurate structure is determined, the methods of computer based numerical modeling can be used to complete the structure or improve its accuracy. Any  
25 recognized modeling method may be used, including parameterized models specific to particular biopolymers such as proteins or nucleic acids, molecular dynamics models based on computing molecular motions, statistical mechanics models based on thermal ensembles, or combined models. For most types of models,  
30 standard molecular force fields, representing the forces between constituent atoms and groups, are necessary, and can be selected from force fields known in physical chemistry. The incomplete or less accurate experimental structures can serve as

constraints on the complete and more accurate structures computed by such modeling methods.

Finally, having determined the structure of the active site, either experimentally, by modeling, or by a combination  
5 thereof, candidate modulating compounds can be identified by searching databases containing compounds along with information on their molecular structure. Such a search seeks compounds having structures that match the determined active site structure and that interact with the groups defining the active  
10 site. Such a search can be manual, but is preferably computer assisted. The compounds found from such a search are potential NGPCR modulating compounds.

Alternatively, these methods can be used to identify improved modulating compounds from an already known modulating  
15 compound or ligand. The composition of the known compound can be modified, and the structural effects of modification can be determined using the experimental and computer modeling methods described above applied to the new composition. The altered structure is then compared to the active site structure of the  
20 compound to determine if an improved fit or interaction results. In this manner systematic variations in composition, such as by varying side groups, can be quickly evaluated to obtain modified modulating compounds or ligands of improved specificity or activity.

25 Further experimental and computer modeling methods useful to identify modulating compounds based upon identification of the active sites of a NGPCR, and related transduction and transcription factors, will be apparent to those of skill in the art.

30 Examples of molecular modeling systems are the CHARMM and QUANTA programs (Polygen Corporation, Waltham, MA). CHARMM performs the energy minimization and molecular dynamics functions, while QUANTA performs the construction, graphic

modeling and analysis of molecular structure. QUANTA allows interactive construction, modification, visualization, and analysis of the behavior of molecules with each other.

A number of articles review computer modeling of drugs  
5 interactive with specific proteins, such as: Rotivinen *et al.*,  
1988, *Acta Pharmaceutical Fennica* 97:159-166; Ripka, *New*  
*Scientist* 54-57 (June 16, 1988); McKinaly and Rossmann, 1989,  
*Annu. Rev. Pharmacol. Toxiciol.* 29:111-122; Perry and Davies,  
OSAR: Quantitative Structure-Activity Relationships in Drug  
10 Design, pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis and Dean,  
1989 *Proc. R. Soc. Lond.* 236:125-140 and 141-162; and, with  
respect to a model receptor for nucleic acid components, Askew  
*et al.*, 1989, *J. Am. Chem. Soc.* 111:1082-1090. Other computer  
programs that screen and graphically depict chemicals are  
15 available from companies such as BioDesign, Inc. (Pasadena,  
CA.), Allelix, Inc. (Mississauga, Ontario, Canada), and  
Hypercube, Inc. (Cambridge, Ontario). Although these are  
primarily designed for application to drugs specific to  
particular proteins, they can be adapted to design of drugs  
20 specific to NGPCRs, or regions of NGPCR DNA or RNA, once that  
region is identified.

Although described above with reference to design and  
generation of compounds that could alter binding, one could also  
screen libraries of known compounds, including natural products  
25 or synthetic chemicals, and biologically active materials,  
including proteins, for compounds that are inhibitors or  
activators.

Cell-based systems can also be used to identify compounds  
that bind NGPCRs, as well as assess the altered activity  
30 associated with such binding in living cells. One tool of  
particular interest for such assays is green fluorescent  
protein, which is described, *inter alia*, in U.S. Patent No.

5,625,048, herein incorporated by reference. Cells that may be used in such cellular assays include, but are not limited to, leukocytes, or cell lines derived from leukocytes, lymphocytes, stem cells, including embryonic stem cells, and the like. In addition, expression host cells (e.g., B95 cells, COS cells, CHO cells, OMK cells, fibroblasts, Sf9 cells) genetically engineered to express a functional NGPCR of interest and to respond to activation by the test, or natural, ligand, as measured by a chemical or phenotypic change, or induction of another host cell gene, can be used as an end point in the assay.

#### 5.5.1 *IN VITRO* SCREENING ASSAYS FOR COMPOUNDS THAT BIND TO NGPCRs

*In vitro* systems can be designed to identify compounds capable of interacting with (e.g., binding to) a NGPCR (including, but not limited to, a ECD or CD of a NGPCR). Compounds identified may be useful, for example, in modulating the activity of wild-type and/or mutant NGPCR gene products, or in elaborating the biological function of the NGPCR; may be utilized in screens for identifying compounds that disrupt normal NGPCR interactions; or may in themselves disrupt such interactions.

The principle of the assays used to identify compounds that bind to a NGPCR involves preparing a reaction mixture of a NGPCR and a test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. The NGPCR species used can vary depending upon the goal of the screening assay. For example, where agonists of the natural ligand are sought, the full length NGPCR, or a soluble truncated NGPCR, e.g., in which the TM and/or CD is deleted from the molecule, a peptide corresponding to an ECD, or a fusion protein containing one or more NGPCR ECD fused to a protein or

polypeptide that affords advantages in the assay system (e.g., labeling, isolation of the resulting complex, etc.), can be utilized. Where compounds that interact with the cytoplasmic domain are sought to be identified, peptides corresponding to a  
5 NGPCR CD or fusion proteins containing a NGPCR CD can be used.

The screening assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the NGPCR protein, polypeptide, peptide, or fusion protein, or the test substance, onto a solid phase and detecting  
10 NGPCR/test compound complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, the NGPCR reactant may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly. Examples of some of the technologies  
15 available to immobilize the molecules are discussed in Cass, ed., "Immobilized Biomolecules In Analysis: A Practical Approach", Oxford University Press, NY.

In practice, microtiter plates may conveniently be utilized as the solid phase. The anchored component may be immobilized  
20 by non-covalent or covalent attachments. Non-covalent attachment may be accomplished by simply coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein to be immobilized may be used  
25 to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components  
30 are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be

accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, e.g., using a labeled antibody specific for the previously non-immobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

10 Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected, e.g., using an immobilized antibody specific for a NGPCR protein, polypeptide, peptide, or fusion protein, or the test compound, to anchor any complexes  
15 formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

Alternatively, cell-based assays can be used to identify compounds that interact with a NGPCR. To this end, cell lines  
20 that express a NGPCR, or cell lines (e.g., COS cells, CHO cells, fibroblasts, etc.) that have been genetically engineered to express a NGPCR (e.g., by transfection or transduction of NGPCR DNA) can be used. Interaction of the test compound with, for example, a ECD of a NGPCR expressed by the host cell can be  
25 determined by comparison or competition with native ligand.

#### 5.5.2. ASSAYS FOR INTRACELLULAR PROTEINS THAT INTERACT WITH NGPCRS

30 Any method suitable for detecting protein-protein interactions may be employed for identifying transmembrane proteins or intracellular proteins that interact with a NGPCR. Among the traditional methods that may be employed are

co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns of cell lysates, or proteins obtained from cell lysates, and a NGPCR to identify proteins in the lysate that interact with the NGPCR. For these  
5 assays, the NGPCR component used can be a full length NGPCR, a soluble derivative lacking the membrane-anchoring region (e.g., a truncated NGPCR in which a TM is deleted resulting in a truncated molecule containing a ECD fused to a CD), a peptide corresponding to a CD, or a fusion protein containing a CD of a  
10 NGPCR. Once isolated, such an intracellular protein can be identified and can, in turn, be used in conjunction with standard techniques to identify proteins with which it interacts. For example, at least a portion of the amino acid sequence of an intracellular protein that interacts with a NGPCR  
15 can be ascertained using techniques well-known to those of skill in the art, such as via the Edman degradation technique (see, e.g., Creighton, 1983, "Proteins: Structures and Molecular Principles", W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained may be used as a guide for the generation  
20 of oligonucleotide mixtures that can be used to screen for nucleotide sequences encoding such intracellular proteins. Screening can be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known  
25 (see, e.g., Ausubel, *supra*, and Innis *et al.*, eds., PCR Protocols: A Guide to Methods and Applications, 1990, Academic Press, Inc., New York).

Additionally, methods may be employed that result in the simultaneous identification of genes that encode transmembrane  
30 or intracellular proteins that interact with a NGPCR. These methods include, for example, probing expression libraries, in a manner similar to the well-known technique of antibody probing

of  $\lambda$ gt11 libraries, using labeled NGPCR protein, or an NGPCR polypeptide, peptide or fusion protein, e.g., an NGPCR polypeptide or NGPCR domain fused to a marker (e.g., an enzyme, fluor, luminescent protein, or dye), or an Ig-Fc domain.

5 One method that detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration only, and not by way of limitation. One version of this system utilizes yeast cells (Chien *et al.*, 1991, Proc. Natl. Acad. Sci. USA, 88:9578-9582), while another uses mammalian cells (Luo *et*  
10 *al.*, 1997, Biotechniques, 22:350-352). Both the yeast and mammalian two-hybrid systems are commercially available from Clontech (Palo Alto, CA), and are further described in U.S. Patent Nos. 5,283,173; 5,468,614, and 5,667,973, which are herein incorporated by reference in their entirety.

15 Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one plasmid consists of nucleotides encoding the DNA-binding domain of a transcription activator protein fused to a NGPCR nucleotide sequence encoding a NGPCR protein, polypeptide, peptide, or fusion protein, and  
20 the other plasmid consists of nucleotides encoding an activation domain of a transcription activator protein fused to a cDNA encoding an unknown protein to be tested for interaction with a NGPCR, which has been recombined into this plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the  
25 cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* or a mammalian cell (such as Saos-2, CHO, CV1, Jurkat or HeLa) that contains a reporter gene (e.g., HBS, lacZ, CAT, or a gene encoding an essential amino acid) whose regulatory region contains the binding site of the  
30 transcription activator. Either hybrid protein alone cannot activate transcription of the reporter gene: the DNA-binding domain hybrid cannot because it does not provide activation

function; and the activation domain hybrid cannot because it cannot localize to the binding site of the activator.

Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system, or related methodologies, may be used to screen activation domain libraries for proteins that interact with the "bait" gene product. By way of example, and not by way of limitation, a NGPCR may be used as the bait gene product. Total genomic or cDNA sequences are fused to DNA encoding an activation domain. This library and a plasmid encoding a hybrid of a bait NGPCR gene product fused to the DNA-binding domain are co-transformed into a reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, a bait NGPCR gene sequence, such as an open reading frame of a NGPCR (or a domain of a NGPCR), can be cloned into a vector such that it is translationally fused to DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the library plasmids.

A cDNA library of the cell line from which proteins that interact with bait NGPCR gene product are to be detected can be made using methods routinely practiced in the art. According to one particular system, for example, the cDNA fragments can be inserted into a vector such that they are translationally fused to the transcriptional activation domain of GAL4. This library can be co-transformed along with the bait NGPCR gene-GAL4 fusion plasmid into a yeast strain that: a) cannot grow without added histidine; and b) contains a HIS3 gene driven by a promoter that contains GAL4 activation sequence. A cDNA encoded protein,

fused to a GAL4 transcriptional activation domain, which interacts with bait MARK3 gene product will reconstitute an active GAL4 protein, and thereby drive expression of the HIS3 gene. Colonies that express HIS3 can be detected by growth on petri dishes containing semi-solid agar based media lacking histidine. The cDNA can then be purified from these strains, and used to produce and isolate the bait NGPCR gene-interacting protein using techniques routinely practiced in the art.

10                   5.5.3. ASSAYS FOR COMPOUNDS THAT INTERFERE  
                              WITH NGPCR/INTRACELLULAR OR NGPCR/  
                              TRANSMEMBRANE MACROMOLECULE INTERACTION

          The macromolecules that interact with a NGPCR are referred to, for purposes of this discussion, as "binding partners."  
15 These binding partners are likely to be involved in a NGPCR signal transduction pathway. Therefore, it is desirable to identify compounds that interfere with or disrupt the interaction of such binding partners with NGPCRs, which compounds may be useful in regulating the activity of a NGPCR  
20 and controlling disorders associated with NGPCR activity. For example, given their expression pattern, the described NGPCRs are contemplated to be particularly useful in methods for identifying compounds useful in the therapeutic treatment of high or low blood pressure (and associated symptoms), kidney  
25 disorders, weight control disorders, metabolic disorders, and cancer.

          The basic principle of the assay systems used to identify compounds that interfere with the interaction between a NGPCR and its binding partner or partners involves preparing a  
30 reaction mixture containing a NGPCR protein, polypeptide, peptide, or fusion protein, as described in Sections 5.5.1 and 5.5.2 above, and the binding partner under conditions and for a time sufficient to allow the two components to interact and

bind, thus forming a complex. In order to test a compound for inhibitory activity, the reaction mixture is prepared in the presence and absence of the test compound. The test compound may be initially included in the reaction mixture, or may be  
5 added at a time subsequent to the addition of the NGPCR moiety and its binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the NGPCR moiety and the binding partner is then detected. The formation of a complex in  
10 the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the NGPCR and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and a normal NGPCR protein  
15 may also be compared to complex formation within reaction mixtures containing the test compound and a mutant NGPCR. This comparison may be important in those cases wherein it is desirable to identify compounds that specifically disrupt interactions of mutant, or mutated, NGPCRs, but not normal  
20 NGPCRs.

The assay for compounds that interfere with the interaction of a NGPCR and its binding partner(s) can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the NGPCR moiety or the binding partner  
25 onto a solid phase and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested.  
30 For example, test compounds that interfere with the interaction by competition can be identified by conducting the reaction in the presence of the test substance, *i.e.*, by adding the test substance to the reaction mixture prior to, or simultaneously

with, a NGPCR moiety and interactive binding partner.  
Alternatively, test compounds that disrupt preformed complexes,  
e.g., compounds with higher binding constants that displace one  
of the components from the complex, can be tested by adding the  
5 test compound to the reaction mixture after complexes have been  
formed. The various formats are described briefly below.

In a heterogeneous assay system, either a NGPCR moiety or  
an interactive binding partner is anchored onto a solid surface,  
while the non-anchored species is labeled, either directly or  
10 indirectly. In practice, microtiter plates are conveniently  
utilized. The anchored species may be immobilized by non-  
covalent or covalent attachments. Non-covalent attachment may  
be accomplished simply by coating the solid surface with a  
solution of a NGPCR gene product or binding partner and drying.  
15 Alternatively, an immobilized antibody specific for the species  
to be anchored may be used to anchor the species to the solid  
surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the partner of the  
immobilized species is exposed to the coated surface with or  
20 without the test compound. After the reaction is complete,  
unreacted components are removed (e.g., by washing) and any  
complexes formed will remain immobilized on the solid surface.  
The detection of complexes anchored on the solid surface can be  
accomplished in a number of ways. Where the non-immobilized  
25 species is pre-labeled, the detection of label immobilized on  
the surface indicates that complexes were formed. Where the  
non-immobilized species is not pre-labeled, an indirect label  
can be used to detect complexes anchored on the surface, e.g.,  
using a labeled antibody specific for the initially non-  
30 immobilized species (the antibody, in turn, may be directly  
labeled or indirectly labeled with a labeled anti-Ig antibody).  
Depending upon the order of addition of reaction components,

test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the  
5 reaction products separated from unreacted components, and complexes detected, e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the  
10 order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

In an alternate embodiment of the invention, a homogeneous assay can be used. In this approach, a preformed complex of a  
15 NGPCR moiety and an interactive binding partner is prepared in which either the NGPCR or its binding partners is labeled, but the signal generated by the label is quenched due to formation of the complex (see, e.g., U.S. Patent No. 4,109,496, which utilizes this approach for immunoassays). The addition of a  
20 test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt NGPCR/intracellular binding partner interactions can be identified.

25 In a particular embodiment, a NGPCR fusion protein can be prepared for immobilization. For example, a NGPCR or a peptide fragment, e.g., corresponding to a CD, can be fused to a glutathione-S-transferase (GST) gene using a fusion vector, such as pGEX-5X-1, in such a manner that its binding activity is  
30 maintained in the resulting fusion protein. The interactive binding partner can be purified and used to raise a monoclonal antibody, using methods routinely practiced in the art and

described above, in Section 5.3. This antibody can be labeled with a radioactive isotope,  $^{125}\text{I}$  for example, by methods routinely practiced in the art. In a heterogeneous assay, e.g., the GST-NGPCR fusion protein can be anchored to glutathione-  
5 agarose beads. The interactive binding partner can then be added in the presence or absence of the test compound in a manner that allows interaction and binding to occur. At the end of the reaction period, unbound material can be washed away, and the labeled monoclonal antibody can be added to the system and  
10 allowed to bind to the complexed components. The interaction between a NGPCR gene product and the interactive binding partner can be detected by measuring the amount of radioactivity that remains associated with the glutathione-agarose beads. A successful inhibition of the interaction by the test compound  
15 will result in a decrease in measured radioactivity.

Alternatively, the GST-NGPCR fusion protein and the interactive binding partner can be mixed together in liquid in the absence of the solid glutathione-agarose beads. The test compound can be added either during or after the species are  
20 allowed to interact. This mixture can then be added to the glutathione-agarose beads, and unbound material is washed away. Again the extent of inhibition of the NGPCR/binding partner interaction can be detected by adding a labeled antibody, and measuring the radioactivity associated with the beads.

25 In another embodiment of the invention, these same techniques can be employed using peptide fragments that correspond to the binding domains of a NGPCR and/or the interactive or binding partner (in cases where the binding partner is a protein), in place of one or both of the full  
30 length proteins. Any number of methods routinely practiced in the art can be used to identify and isolate the binding sites. These methods include, but are not limited to, mutagenesis of the gene encoding one of the proteins and screening for

disruption of binding in a co-immunoprecipitation assay. Compensatory mutation(s) in the sequence encoding the second species in the complex can then be selected. Sequence analysis of the genes encoding the respective proteins will reveal the mutations that correspond to the region of the protein involved in interactive binding. Alternatively, one protein can be anchored to a solid surface using methods described above, and allowed to interact with and bind to its labeled binding partner, which has been treated with a proteolytic enzyme, such as trypsin. After washing, a relatively short, labeled peptide comprising the binding domain may remain associated with the solid material, which can be isolated and identified by amino acid sequencing. Also, once the gene coding for the intracellular binding partner is obtained, short gene segments can be engineered to express peptide fragments of the protein, which can then be tested for binding activity and purified or synthesized.

For example, and not by way of limitation, a NGPCR gene product can be anchored to a solid material, as described above, by making a GST-NGPCR fusion protein and allowing it to bind to glutathione-agarose beads. The interactive binding partner can be labeled with a radioactive isotope, such as  $^{35}\text{S}$ , and cleaved with a proteolytic enzyme, such as trypsin. Cleavage products can then be added to the anchored GST-NGPCR fusion protein and allowed to bind. After washing away unbound peptides, labeled bound material, representing the intracellular binding partner binding domain, can be eluted, purified, and analyzed for amino acid sequence by well-known methods. Peptides so identified can be produced synthetically or fused to appropriate facilitative proteins using recombinant DNA technology.

## 5.6 PHARMACEUTICAL COMPOSITIONS

Compounds identified via assays such as those described herein may be useful, for example, in elaborating the biological function of a NGPCR gene product. Such compounds can be  
5 administered to a patient at therapeutically effective doses to treat any of a variety of physiological or mental disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in any delay in onset, or any amelioration, impediment, prevention, or alteration of any  
10 biological or overt symptom.

### 5.6.1 EFFECTIVE DOSE

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell  
15 cultures or experimental animals, e.g., for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Compounds  
20 that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used in certain embodiments, care should be taken to design a delivery system that targets such compounds to the site of affected tissue, in order to minimize potential damage to uninfected cells and,  
25 thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with  
30 little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the methods of the invention, the therapeutically effective dose can be

estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the  $IC_{50}$  (*i.e.*, the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

When the therapeutic treatment of disease is contemplated, the appropriate dosage may also be determined using animal studies to determine the maximal tolerable dose, or MTD, of a bioactive agent per kilogram weight of the test subject. In general, at least one animal species tested is mammalian. Those skilled in the art regularly extrapolate doses for efficacy and avoiding toxicity to other species, including humans. Before human studies of efficacy are undertaken, Phase I clinical studies in normal subjects help establish safe doses.

Additionally, the bioactive agent may be complexed with a variety of well established compounds or structures that, for instance, enhance the stability of the bioactive agent or otherwise enhance its pharmacological properties (*e.g.*, increase *in vivo* half-life, reduce toxicity, *etc.*).

The therapeutic agents will be administered by any number of methods known to those of ordinary skill in the art including, but not limited to, inhalation, subcutaneous (sub-q), intravenous (I.V.), intraperitoneal (I.P.), intramuscular (I.M.) or intrathecal injection, or topically applied (transderm, ointments, creams, salves, eye drops, and the like).

#### 5.6.2 FORMULATIONS AND USE

Pharmaceutical compositions for use in accordance with the present invention may be formulated in a conventional manner

using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral, intracranial, topical, intrathecal, or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well-known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection, or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by

intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for  
5 example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device that may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The  
10 pack or dispenser device may be accompanied by instructions for administration.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as single illustrations of individual aspects of the invention, and  
15 functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein, will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the  
20 scope of the appended claims. All referenced publications, patents, and patent applications are herein incorporated by reference.

## SEQUENCE LISTING

&lt;110&gt; LEXICON GENETICS INCORPORATED

<120> Novel Human 7<sup>TM</sup> Proteins and Polynucleotides Encoding the Same

&lt;130&gt; LEX-0262-PCT

&lt;150&gt; US 60/244,285

&lt;151&gt; 2000-10-30

&lt;160&gt; 9

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 3633

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 1

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| atgacatcta  | gtaataccca  | acctctgctt  | atgacttcct  | ggaacatacc  | cacagctgaa  | 60   |
| ggttctcagt  | ttccaatttc  | caccactatt  | aatgtaccta  | catccaatga  | gatggaaaca  | 120  |
| gagactctac  | accttggtcc  | tgggcctttg  | tcaacattca  | cagcctctca  | gactgggtcta | 180  |
| gtatctaaag  | atgtcatggc  | aatgtcatca  | attcctatgt  | caggaattct  | tcctaaccat  | 240  |
| gggctttctg  | agaacccttc  | attatcaaca  | tctttaagag  | ctatcacttc  | cacattggct  | 300  |
| gacgttaagc  | acacatttga  | gaaaatgacc  | acatctgtaa  | ctcctgggac  | cacactccca  | 360  |
| tcaattcttt  | ctggtgccac  | ttcaggatct  | gtaatttcaa  | agtcacccat  | tctgacatgg  | 420  |
| ctcttatcta  | gtctcccttc  | tggctccctt  | ccggcaactg  | tatctaatac  | ccctcatgtt  | 480  |
| atgacttcct  | ctacagtaga  | ggtgtcaaaa  | tcaacatttc  | tgacatctga  | catgatatac  | 540  |
| gcgcacccat  | tcactaactt  | gacaacacta  | ccctctgcta  | ctatgagcac  | catactcacc  | 600  |
| cgaaccattc  | ctacacctac  | actgggtggg  | atcactactg  | gcttcccaac  | ttctctccct  | 660  |
| atgtctataa  | atgtcacaga  | tgacattgtg  | tacattttcca | cacaccctga  | ggcatcctcc  | 720  |
| agaaccacaa  | taactgccaa  | ccccaggact  | gtgtctcatc  | cttcatcctt  | cagcagaaag  | 780  |
| actatgtcac  | cttctacaac  | tgaccacact  | ctatctgttg  | gtgccatgcc  | tctgcctagc  | 840  |
| tctacaataa  | catcttcatg  | gaacagaatt  | ccaactgcat  | catcaccttc  | tactttaatt  | 900  |
| attcctaagc  | ccacactgga  | ctcccttcta  | aataataatga | ctactacatc  | cactgttcct  | 960  |
| ggagcctcat  | ttccactcat  | atccactggg  | gtgacataatc | ctttttacagc | aactgtgtct  | 1020 |
| tcaccaatat  | cgctcctttt  | tgaacaactt  | tggctggact  | ccacaccttc  | ctttctatct  | 1080 |
| acggaagcat  | cgacttcgcc  | tactgccacc  | aagtccacag  | tttccttcta  | caatgttgaa  | 1140 |
| atgagcttct  | ctgtctttgt  | tgaagagcca  | aggatcccta  | ttaccagtgt  | tataaatgaa  | 1200 |
| tttacggaaa  | attcggtgaa  | ttctatatatt | cagaacagtg  | aattttctct  | tgctactctg  | 1260 |
| gaaacccaaa  | ttaaaagcag  | ggacatttca  | gaggaagaga  | tgggtcatgga | tcgagctatt  | 1320 |
| ttggaacaga  | gagaaggaca  | agaaatggct  | acaatttcct  | atgtaccata  | cagttgtgtt  | 1380 |
| tgtcagggtca | tcataaaaagc | cagctcttcc  | ttagcatcct  | ctgaattgat  | gagaaaaatc  | 1440 |
| aaaagtaaaa  | tacatggcaa  | cttcacacat  | ggaaacttca  | cacaagatca  | attgacgtta  | 1500 |
| ttagtaaaact | gtgaacacgt  | tgcagtgaag  | aaactagagc  | ctggaaattg  | caaagctgat  | 1560 |
| gaaacagcct  | ctaaatacaa  | agggacctat  | aagtggctat  | taaccaaccc  | tacggagaca  | 1620 |
| gcccacacca  | gatgcataaa  | aatgagggat  | ggaaatgcca  | caagattctg  | ttcaatcagc  | 1680 |
| atcaacacgg  | gcaaattctca | gtgggaaaag  | ccaaagttta  | aacaatgcaa  | attgcttcaa  | 1740 |
| gaacttcctg  | acaagattgt  | ggatcttgct  | aatattacca  | taagtgatga  | gaatcctgag  | 1800 |
| gatgttgtag  | agcatatttt  | aaatttgata  | aatgaatccc  | cagccctggg  | ttaaagaagag | 1860 |
| acaaagatta  | ttgttttcta  | aatatcagat  | atttcacaat  | gtgatgagat  | aagtatgaac  | 1920 |
| ctaactcatg  | ttatgttaca  | aataatcaac  | gttgtttttg  | aaaagcaaaa  | caattccgcc  | 1980 |
| tctgatctgc  | atgaaataag  | caatgaaatt  | ctgaggataa  | ttgagcgtcc  | tggtcacaag  | 2040 |

```

atggagtttt ctgggcagat agcaaactctg gcggtggccg ggctggcttt ggctgtgctg 2100
cggggggacc acacgtttga tggcatggct ttcagcatte actcctatga agaaggcca 2160
gaccctgaca ttttcctagg caatgtccct gtgggagga ttttggcttc catatatttg 2220
cctaaatcac tgacggagag aattcctctt agcaacttac aaccgatctt gtttaatttc 2280
tttggccaaa cttcactctt taagaccaa aatgtcacta aagcattaac cacatatgtt 2340
gtgagtgccg gcatttcaga tatgttcatt caaaacttag ctgaccaggt gggtatcact 2400
ctgcagcata ttggaggaaa ccagaattat ggtcaagttc actgtgcctt ttgggatttt 2460
gagaataata atgggctggg tggatggaat tcgtcaggct gtaaagtaaa ggaaacaaat 2520
gtaaattaca caatctgtca gtgtgaccac ctcaccatt ttggagtctt aatggattta 2580
tccaggtcta cagtggattc agtgaatgaa cagatattag cgcttataac atacaccgga 2640
tgtggaatct cctccatttt cctgggagtt gcagtgggtg catacatagc ttttcacaaa 2700
cttcgaaaag attatcctgc caaaattctg atcaacctgt gcacagcact actgatgcta 2760
aacctgggat ttttgatcaa ttcttgggtt tcatcatttc agaaagtggg agtttgtatc 2820
acagctgcag tggcacttca ttacttcctg cttgtttctt ttacttggat gggcctggag 2880
gcagtccaca tgtatttggc tctagtcaaa gtcttcaaca tatacattcc aaattatata 2940
cttaaatttt gtctagtgtg ttggggaatc cgggctatca tgggtggcaat cacagtcagt 3000
gtgaaaaaag atctgtatgg aactctgagc ccaacaactc cgttttgttg gattaaagat 3060
gattctatct tttacatctc agtgggtggc tatttttgcc tcatatttct catgaatctc 3120
tccatgttct gcactgttct tgttcaactg aattctgtga aatcccaaat ccagaagact 3180
cggcggaaga tgatcctgca tgacctcaa ggcacaatga gcctgacatt cttacttggc 3240
ctcacctggg ggtttgcatt ttttgcttgg ggacccatga ggaacttttt cttgtatttg 3300
tttgccattt ttaacacttt gcaaggattc ttcatttttg tgtttcactg tgtgatgaag 3360
gagagtgtgc gggagcagtg gcagatacac ctctgctgtg ggtggttgcg attggataac 3420
tcttctgatg ggagcagccg gtgtcagata aaggttggat ataaacagga gggactaaag 3480
aaaatctttg agcacaaact gttgacgcca tctctcaagt caactgcaac tagctccact 3540
ttcaaactct taggctctgc acaaggcaca ccttcagaaa taagctttcc aatgatgac 3600
yttgacaaag atccttactg ttcctctcct tga 3633

```

&lt;210&gt; 2

&lt;211&gt; 1210

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```

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

```

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Met | Ile | Ser | Ala | His | Pro | Phe | Thr | Asn | Leu | Thr | Thr | Leu | Pro | Ser | 180 | 185 | 190 |
| Ala | Thr | Met | Ser | Thr | Ile | Leu | Thr | Arg | Thr | Ile | Pro | Thr | Pro | Thr | Leu | 195 | 200 | 205 |
| Gly | Gly | Ile | Thr | Thr | Gly | Phe | Pro | Thr | Ser | Leu | Pro | Met | Ser | Ile | Asn | 210 | 215 | 220 |
| Val | Thr | Asp | Asp | Ile | Val | Tyr | Ile | Ser | Thr | His | Pro | Glu | Ala | Ser | Ser | 225 | 230 | 235 |
| Arg | Thr | Thr | Ile | Thr | Ala | Asn | Pro | Arg | Thr | Val | Ser | His | Pro | Ser | Ser | 245 | 250 | 255 |
| Phe | Ser | Arg | Lys | Thr | Met | Ser | Pro | Ser | Thr | Thr | Asp | His | Thr | Leu | Ser | 260 | 265 | 270 |
| Val | Gly | Ala | Met | Pro | Leu | Pro | Ser | Ser | Thr | Ile | Thr | Ser | Ser | Trp | Asn | 275 | 280 | 285 |
| Arg | Ile | Pro | Thr | Ala | Ser | Ser | Pro | Ser | Thr | Leu | Ile | Ile | Pro | Lys | Pro | 290 | 295 | 300 |
| Thr | Leu | Asp | Ser | Leu | Leu | Asn | Ile | Met | Thr | Thr | Thr | Ser | Thr | Val | Pro | 305 | 310 | 315 |
| Gly | Ala | Ser | Phe | Pro | Leu | Ile | Ser | Thr | Gly | Val | Thr | Tyr | Pro | Phe | Thr | 325 | 330 | 335 |
| Ala | Thr | Val | Ser | Ser | Pro | Ile | Ser | Ser | Phe | Phe | Glu | Thr | Thr | Trp | Leu | 340 | 345 | 350 |
| Asp | Ser | Thr | Pro | Ser | Phe | Leu | Ser | Thr | Glu | Ala | Ser | Thr | Ser | Pro | Thr | 355 | 360 | 365 |
| Ala | Thr | Lys | Ser | Thr | Val | Ser | Phe | Tyr | Asn | Val | Glu | Met | Ser | Phe | Ser | 370 | 375 | 380 |
| Val | Phe | Val | Glu | Glu | Pro | Arg | Ile | Pro | Ile | Thr | Ser | Val | Ile | Asn | Glu | 385 | 390 | 395 |
| Phe | Thr | Glu | Asn | Ser | Leu | Asn | Ser | Ile | Phe | Gln | Asn | Ser | Glu | Phe | Ser | 405 | 410 | 415 |
| Leu | Ala | Thr | Leu | Glu | Thr | Gln | Ile | Lys | Ser | Arg | Asp | Ile | Ser | Glu | Glu | 420 | 425 | 430 |
| Glu | Met | Val | Met | Asp | Arg | Ala | Ile | Leu | Glu | Gln | Arg | Glu | Gly | Gln | Glu | 435 | 440 | 445 |
| Met | Ala | Thr | Ile | Ser | Tyr | Val | Pro | Tyr | Ser | Cys | Val | Cys | Gln | Val | Ile | 450 | 455 | 460 |
| Ile | Lys | Ala | Ser | Ser | Ser | Leu | Ala | Ser | Ser | Glu | Leu | Met | Arg | Lys | Ile | 465 | 470 | 475 |
| Lys | Ser | Lys | Ile | His | Gly | Asn | Phe | Thr | His | Gly | Asn | Phe | Thr | Gln | Asp | 485 | 490 | 495 |
| Gln | Leu | Thr | Leu | Leu | Val | Asn | Cys | Glu | His | Val | Ala | Val | Lys | Lys | Leu | 500 | 505 | 510 |
| Glu | Pro | Gly | Asn | Cys | Lys | Ala | Asp | Glu | Thr | Ala | Ser | Lys | Tyr | Lys | Gly | 515 | 520 | 525 |
| Thr | Tyr | Lys | Trp | Leu | Leu | Thr | Asn | Pro | Thr | Glu | Thr | Ala | Gln | Thr | Arg | 530 | 535 | 540 |
| Cys | Ile | Lys | Asn | Glu | Asp | Gly | Asn | Ala | Thr | Arg | Phe | Cys | Ser | Ile | Ser | 545 | 550 | 555 |
| Ile | Asn | Thr | Gly | Lys | Ser | Gln | Trp | Glu | Lys | Pro | Lys | Phe | Lys | Gln | Cys | 565 | 570 | 575 |
| Lys | Leu | Leu | Gln | Glu | Leu | Pro | Asp | Lys | Ile | Val | Asp | Leu | Ala | Asn | Ile | 580 | 585 | 590 |
| Thr | Ile | Ser | Asp | Glu | Asn | Pro | Glu | Asp | Val | Ala | Glu | His | Ile | Leu | Asn | 595 | 600 | 605 |
| Leu | Ile | Asn | Glu | Ser | Pro | Ala | Leu | Gly | Lys | Glu | Glu | Thr | Lys | Ile | Ile | 610 | 615 | 620 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |      |      |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| Val | Ser | Lys | Ile | Ser | Asp | Ile | Ser | Gln | Cys | Asp | Glu | Ile | Ser | Met | Asn | 625  | 630  | 635  | 640  |
| Leu | Thr | His | Val | Met | Leu | Gln | Ile | Ile | Asn | Val | Val | Leu | Glu | Lys | Gln | 645  | 650  | 655  |      |
| Asn | Asn | Ser | Ala | Ser | Asp | Leu | His | Glu | Ile | Ser | Asn | Glu | Ile | Leu | Arg | 660  | 665  | 670  |      |
| Ile | Ile | Glu | Arg | Pro | Gly | His | Lys | Met | Glu | Phe | Ser | Gly | Gln | Ile | Ala | 675  | 680  | 685  |      |
| Asn | Leu | Ala | Val | Ala | Gly | Leu | Ala | Leu | Ala | Val | Leu | Arg | Gly | Asp | His | 690  | 695  | 700  |      |
| Thr | Phe | Asp | Gly | Met | Ala | Phe | Ser | Ile | His | Ser | Tyr | Glu | Glu | Gly | Pro | 705  | 710  | 715  | 720  |
| Asp | Pro | Asp | Ile | Phe | Leu | Gly | Asn | Val | Pro | Val | Gly | Gly | Ile | Leu | Ala | 725  | 730  | 735  |      |
| Ser | Ile | Tyr | Leu | Pro | Lys | Ser | Leu | Thr | Glu | Arg | Ile | Pro | Leu | Ser | Asn | 740  | 745  | 750  |      |
| Leu | Gln | Pro | Ile | Leu | Phe | Asn | Phe | Phe | Gly | Gln | Thr | Ser | Leu | Phe | Lys | 755  | 760  | 765  |      |
| Thr | Lys | Asn | Val | Thr | Lys | Ala | Leu | Thr | Thr | Tyr | Val | Val | Ser | Ala | Ser | 770  | 775  | 780  |      |
| Ile | Ser | Asp | Met | Phe | Ile | Gln | Asn | Leu | Ala | Asp | Pro | Val | Val | Ile | Thr | 785  | 790  | 795  | 800  |
| Leu | Gln | His | Ile | Gly | Gly | Asn | Gln | Asn | Tyr | Gly | Gln | Val | His | Cys | Ala | 805  | 810  | 815  |      |
| Phe | Trp | Asp | Phe | Glu | Asn | Asn | Asn | Gly | Leu | Gly | Gly | Trp | Asn | Ser | Ser | 820  | 825  | 830  |      |
| Gly | Cys | Lys | Val | Lys | Glu | Thr | Asn | Val | Asn | Tyr | Thr | Ile | Cys | Gln | Cys | 835  | 840  | 845  |      |
| Asp | His | Leu | Thr | His | Phe | Gly | Val | Leu | Met | Asp | Leu | Ser | Arg | Ser | Thr | 850  | 855  | 860  |      |
| Val | Asp | Ser | Val | Asn | Glu | Gln | Ile | Leu | Ala | Leu | Ile | Thr | Tyr | Thr | Gly | 865  | 870  | 875  | 880  |
| Cys | Gly | Ile | Ser | Ser | Ile | Phe | Leu | Gly | Val | Ala | Val | Val | Thr | Tyr | Ile | 885  | 890  | 895  |      |
| Ala | Phe | His | Lys | Leu | Arg | Lys | Asp | Tyr | Pro | Ala | Lys | Ile | Leu | Ile | Asn | 900  | 905  | 910  |      |
| Leu | Cys | Thr | Ala | Leu | Leu | Met | Leu | Asn | Leu | Val | Phe | Leu | Ile | Asn | Ser | 915  | 920  | 925  |      |
| Trp | Leu | Ser | Ser | Phe | Gln | Lys | Val | Gly | Val | Cys | Ile | Thr | Ala | Ala | Val | 930  | 935  | 940  |      |
| Ala | Leu | His | Tyr | Phe | Leu | Leu | Val | Ser | Phe | Thr | Trp | Met | Gly | Leu | Glu | 945  | 950  | 955  | 960  |
| Ala | Val | His | Met | Tyr | Leu | Ala | Leu | Val | Lys | Val | Phe | Asn | Ile | Tyr | Ile | 965  | 970  | 975  |      |
| Pro | Asn | Tyr | Ile | Leu | Lys | Phe | Cys | Leu | Val | Gly | Trp | Gly | Ile | Pro | Ala | 980  | 985  | 990  |      |
| Ile | Met | Val | Ala | Ile | Thr | Val | Ser | Val | Lys | Lys | Asp | Leu | Tyr | Gly | Thr | 995  | 1000 | 1005 |      |
| Leu | Ser | Pro | Thr | Thr | Pro | Phe | Cys | Trp | Ile | Lys | Asp | Asp | Ser | Ile | Phe | 1010 | 1015 | 1020 |      |
| Tyr | Ile | Ser | Val | Val | Ala | Tyr | Phe | Cys | Leu | Ile | Phe | Leu | Met | Asn | Leu | 1025 | 1030 | 1035 | 1040 |
| Ser | Met | Phe | Cys | Thr | Val | Leu | Val | Gln | Leu | Asn | Ser | Val | Lys | Ser | Gln | 1045 | 1050 | 1055 |      |
| Ile | Gln | Lys | Thr | Arg | Arg | Lys | Met | Ile | Leu | His | Asp | Leu | Lys | Gly | Thr | 1060 | 1065 | 1070 |      |

Met Ser Leu Thr Phe Leu Leu Gly Leu Thr Trp Gly Phe Ala Phe Phe  
 1075 1080 1085  
 Ala Trp Gly Pro Met Arg Asn Phe Phe Leu Tyr Leu Phe Ala Ile Phe  
 1090 1095 1100  
 Asn Thr Leu Gln Gly Phe Phe Ile Phe Val Phe His Cys Val Met Lys  
 1105 1110 1115 1120  
 Glu Ser Val Arg Glu Gln Trp Gln Ile His Leu Cys Cys Gly Trp Leu  
 1125 1130 1135  
 Arg Leu Asp Asn Ser Ser Asp Gly Ser Ser Arg Cys Gln Ile Lys Val  
 1140 1145 1150  
 Gly Tyr Lys Gln Glu Gly Leu Lys Lys Ile Phe Glu His Lys Leu Leu  
 1155 1160 1165  
 Thr Pro Ser Leu Lys Ser Thr Ala Thr Ser Ser Thr Phe Lys Ser Leu  
 1170 1175 1180  
 Gly Ser Ala Gln Gly Thr Pro Ser Glu Ile Ser Phe Pro Asn Asp Asp  
 1185 1190 1195 1200  
 Phe Asp Lys Asp Pro Tyr Cys Ser Ser Pro  
 1205 1210

&lt;210&gt; 3

&lt;211&gt; 2205

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 3

|             |            |             |            |             |            |      |
|-------------|------------|-------------|------------|-------------|------------|------|
| atgagaaaaa  | tcaaaagtaa | aatacatggc  | aacttcacac | atggaaactt  | cacacaagat | 60   |
| caattgacgt  | tattagtaaa | ctgtgaacac  | gttgcagtga | aaaaactaga  | gcctggaaat | 120  |
| tgcaaagctg  | atgaaacagc | ctctaaatac  | aaagggacct | ataagtggct  | attaaccaac | 180  |
| cctacggaga  | cagcccaaac | cagatgcata  | aaaaatgagg | atggaaatgc  | cacaagattc | 240  |
| tgttcaatca  | gcatcaacac | gggcaaattc  | cagtgggaaa | agccaaagtt  | taaacaatgc | 300  |
| aaattgcttc  | agaactttcc | tgacaagatt  | gtggatcttg | ctaataattac | cataagtgat | 360  |
| gagaatcctg  | aggatgttgc | agagcatatt  | ttaaatttga | taaatgaatc  | cccagccctg | 420  |
| ggtaaagaag  | agacaaagat | tattgtttct  | aaaatatcag | atatttcaca  | atgtgatgag | 480  |
| ataagtatga  | acctaactca | tgttatgtta  | caaataatca | acgttggttt  | ggaaaagcaa | 540  |
| aacaattccg  | cctctgatct | gcatgaaata  | agcaatgaaa | ttctgaggat  | aattgagcgt | 600  |
| cctggtcaca  | agatggagtt | ttctgggcag  | atagcaaatc | tggcgggtggc | cgggctggct | 660  |
| ttggctgtgc  | tgcgggggga | ccacacgttt  | gatggcatgg | ctttcagcat  | tcactcctat | 720  |
| gaagaaggcc  | cagaccctga | catttttccta | ggcaatgtcc | ctgtggggagg | gatttttggt | 780  |
| tccatatatt  | tgcctaaatc | actgacggag  | agaattcctc | ttagcaactt  | acaaccgatc | 840  |
| ttgtttaatt  | tctttggcca | aacttcactc  | tttaagacca | aaaatgtcac  | taaagcatta | 900  |
| accacatatg  | ttgtgagtgc | cagcatttca  | gatatgttca | ttcaaaactt  | agctgaccca | 960  |
| gtggttatca  | ctctgcagca | tattggagga  | aaccagaatt | atggtcaagt  | tcactgtgcc | 1020 |
| ttttgggatt  | ttgagaataa | taatgggctg  | ggtggatgga | attcgtcagg  | ctgtaaagta | 1080 |
| aaggaaacaa  | atgtaaatta | cacaatctgt  | cagtgtgacc | acctcaccca  | ttttggagtc | 1140 |
| ttaatggatt  | tatccaggtc | tacagtggat  | tcagtgaatg | aacagatatt  | agcgcttata | 1200 |
| acatacaccg  | gatgtggaat | ctcctccatt  | ttcctgggag | ttgcagtggg  | gacatacata | 1260 |
| gctttttcaca | aacttcgaaa | agattatcct  | gccaaaattc | tgatcaacct  | gtgcacagca | 1320 |
| ctactgatgc  | taaacctggt | atttttgatc  | aattcttggg | tgatcatcatt | tcagaaagtg | 1380 |
| ggagtttgta  | tcacagctgc | agtggcactt  | cattacttcc | tgcttgtttc  | ttttacttgg | 1440 |
| atgggcctgg  | aggcagtcca | catgtatttg  | gctctagtca | aagtcttcaa  | catatacatt | 1500 |
| ccaaattata  | tccttaaatt | ttgtctagtt  | ggttggggaa | tcccggctat  | catggtggca | 1560 |
| atcacagtca  | gtgtgaaaaa | agatctgtat  | ggaactctga | gcccaacaac  | tccgttttgt | 1620 |
| tggattaaag  | atgattctat | cttttacatc  | tcagtgggtg | cttatttttg  | cctcatattt | 1680 |
| ctcatgaatc  | tctccatggt | ctgcactggt  | cttgttcaac | tgaattctgt  | gaaatcccaa | 1740 |
| atccagaaga  | ctcggcggaa | gatgatcctg  | catgacctca | aaggcacaat  | gagcctgaca | 1800 |
| ttcttacttg  | gcctcacctg | ggggtttgca  | ttttttgctt | ggggacccat  | gaggaacttt | 1860 |

WO 02/36633

PCT/US01/50938

```

ttcttgtatt tgtttgccat ttttaacact ttgcaaggat tcttcatttt tgtgtttcac 1920
tgtgtgatga aggagagtgt gcgggagcag tggcagatac acctctgctg tgggtggttg 1980
cgattggata actcttctga tgggagcagc cgggtgtcaga taaagggttg atataaacag 2040
gagggactaa agaaaatctt tgagcacaaa ctgttgacgc catctctcaa gtcaactgca 2100
actagctcca ctttcaaatac tttaggctct gcacaaggca caccttcaga aataagcttt 2160
ccaaatgatg acyttgacaa agatccttac tgttcctctc cttga 2205

```

&lt;210&gt; 4

&lt;211&gt; 733

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

```

Met Arg Lys Ile Lys Ser Lys Ile His Gly Asn Phe Thr His Gly Asn
 1          5          10          15
Phe Thr Gln Asp Gln Leu Thr Leu Leu Val Asn Cys Glu His Val Ala
          20          25          30
Val Lys Lys Leu Glu Pro Gly Asn Cys Lys Ala Asp Glu Thr Ala Ser
          35          40          45
Lys Tyr Lys Gly Thr Tyr Lys Trp Leu Leu Thr Asn Pro Thr Glu Thr
          50          55          60
Ala Gln Thr Arg Cys Ile Lys Asn Glu Asp Gly Asn Ala Thr Arg Cys
65          70          75          80
Ser Ile Ser Ile Asn Thr Gly Lys Ser Gln Trp Glu Lys Pro Lys Phe
          85          90          95
Lys Gln Cys Lys Leu Leu Gln Glu Leu Pro Asp Lys Ile Val Asp Leu
          100          105          110
Ala Asn Ile Thr Ile Ser Asp Glu Asn Pro Glu Asp Val Ala Glu His
          115          120          125
Ile Leu Asn Leu Ile Asn Glu Ser Pro Ala Leu Gly Lys Glu Glu Thr
          130          135          140
Lys Ile Ile Val Ser Lys Ile Ser Asp Ile Ser Gln Cys Asp Glu Ile
145          150          155          160
Ser Met Asn Leu Thr His Val Met Leu Gln Ile Ile Asn Val Val Leu
          165          170          175
Glu Lys Gln Asn Asn Ser Ala Ser Asp Leu His Glu Ile Ser Asn Glu
          180          185          190
Ile Leu Arg Ile Ile Glu Arg Pro Gly His Lys Met Glu Phe Ser Gly
          195          200          205
Gln Ile Ala Asn Leu Ala Val Ala Gly Leu Ala Leu Ala Val Leu Arg
210          215          220
Gly Asp His Thr Phe Asp Gly Met Ala Phe Ser Ile His Ser Tyr Glu
225          230          235          240
Glu Gly Pro Asp Pro Asp Ile Phe Leu Gly Asn Val Pro Val Gly Gly
          245          250          255
Ile Leu Ala Ser Ile Tyr Leu Pro Lys Ser Leu Thr Glu Arg Ile Pro
260          265          270
Leu Ser Asn Leu Gln Pro Ile Leu Phe Asn Phe Phe Gly Gln Thr Ser
          275          280          285
Leu Phe Lys Thr Lys Asn Val Thr Lys Ala Leu Thr Thr Tyr Val Val
290          295          300
Ser Ala Ser Ile Ser Asp Met Phe Ile Gln Asn Leu Ala Asp Pro Val
305          310          315          320
Val Ile Thr Leu Gln His Ile Gly Gly Asn Gln Asn Tyr Gly Gln Val
          325          330          335
His Cys Ala Phe Trp Asp Phe Glu Asn Asn Asn Gly Leu Gly Gly Trp

```

WO 02/36633

PCT/US01/50938

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |     |  |
| Asn | Ser | Ser | Gly | Cys | Lys | Val | Lys | Glu | Thr | Asn | Val | Asn | Tyr | Thr | Ile |     |  |
|     |     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |  |
| Cys | Gln | Cys | Asp | His | Leu | Thr | His | Phe | Gly | Val | Leu | Met | Asp | Leu | Ser |     |  |
|     |     |     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |  |
| Arg | Ser | Thr | Val | Asp | Ser | Val | Asn | Glu | Gln | Ile | Leu | Ala | Leu | Ile | Thr |     |  |
| 385 |     |     |     |     |     |     | 390 |     |     |     |     |     | 395 |     |     | 400 |  |
| Tyr | Thr | Gly | Cys | Gly | Ile | Ser | Ser | Ile | Phe | Leu | Gly | Val | Ala | Val | Val |     |  |
|     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     |     | 415 |     |  |
| Thr | Tyr | Ile | Ala | Phe | His | Lys | Leu | Arg | Lys | Asp | Tyr | Pro | Ala | Lys | Ile |     |  |
|     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |     |  |
| Leu | Ile | Asn | Leu | Cys | Thr | Ala | Leu | Leu | Met | Leu | Asn | Leu | Val | Phe | Leu |     |  |
|     |     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |  |
| Ile | Asn | Ser | Trp | Leu | Ser | Ser | Phe | Gln | Lys | Val | Gly | Val | Cys | Ile | Thr |     |  |
|     |     |     | 450 |     |     |     | 455 |     |     |     |     |     | 460 |     |     |     |  |
| Ala | Ala | Val | Ala | Leu | His | Tyr | Phe | Leu | Leu | Val | Ser | Phe | Thr | Trp | Met |     |  |
| 465 |     |     |     |     |     | 470 |     |     |     |     |     |     | 475 |     |     | 480 |  |
| Gly | Leu | Glu | Ala | Val | His | Met | Tyr | Leu | Ala | Leu | Val | Lys | Val | Phe | Asn |     |  |
|     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     |     | 495 |     |  |
| Ile | Tyr | Ile | Pro | Asn | Tyr | Ile | Leu | Lys | Phe | Cys | Leu | Val | Gly | Trp | Gly |     |  |
|     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |     |  |
| Ile | Pro | Ala | Ile | Met | Val | Ala | Ile | Thr | Val | Ser | Val | Lys | Lys | Asp | Leu |     |  |
|     |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |  |
| Tyr | Gly | Thr | Leu | Ser | Pro | Thr | Thr | Pro | Phe | Cys | Trp | Ile | Lys | Asp | Asp |     |  |
|     |     |     | 530 |     |     |     | 535 |     |     |     |     |     | 540 |     |     |     |  |
| Ser | Ile | Phe | Tyr | Ile | Ser | Val | Val | Ala | Tyr | Phe | Cys | Leu | Ile | Phe | Leu |     |  |
| 545 |     |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |  |
| Met | Asn | Leu | Ser | Met | Phe | Cys | Thr | Val | Leu | Val | Gln | Leu | Asn | Ser | Val |     |  |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     |     | 575 |     |  |
| Lys | Ser | Gln | Ile | Gln | Lys | Thr | Arg | Arg | Lys | Met | Ile | Leu | His | Asp | Leu |     |  |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     |     | 590 |     |     |  |
| Lys | Gly | Thr | Met | Ser | Leu | Thr | Phe | Leu | Leu | Gly | Leu | Thr | Trp | Gly | Phe |     |  |
|     |     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |  |
| Ala | Phe | Phe | Ala | Trp | Gly | Pro | Met | Arg | Asn | Phe | Phe | Leu | Tyr | Leu | Phe |     |  |
|     |     |     | 610 |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |  |
| Ala | Ile | Phe | Asn | Thr | Leu | Gln | Gly | Phe | Phe | Ile | Phe | Val | Phe | His | Cys |     |  |
| 625 |     |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     | 640 |     |  |
| Val | Met | Lys | Glu | Ser | Val | Arg | Glu | Gln | Trp | Gln | Ile | His | Leu | Cys | Cys |     |  |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     |     | 655 |     |  |
| Gly | Trp | Leu | Arg | Leu | Asp | Asn | Ser | Ser | Asp | Gly | Ser | Ser | Arg | Cys | Gln |     |  |
|     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |     |  |
| Ile | Lys | Val | Gly | Tyr | Lys | Gln | Glu | Gly | Leu | Lys | Lys | Ile | Phe | Glu | His |     |  |
|     |     |     | 675 |     |     |     | 680 |     |     |     |     |     | 685 |     |     |     |  |
| Lys | Leu | Leu | Thr | Pro | Ser | Leu | Lys | Ser | Thr | Ala | Thr | Ser | Ser | Thr | Phe |     |  |
|     |     |     | 690 |     |     |     | 695 |     |     |     |     |     | 700 |     |     |     |  |
| Lys | Ser | Leu | Gly | Ser | Ala | Gln | Gly | Thr | Pro | Ser | Glu | Ile | Ser | Phe | Pro |     |  |
| 705 |     |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     | 720 |     |  |
| Asn | Asp | Asp | Phe | Asp | Lys | Asp | Pro | Tyr | Cys | Ser | Ser | Pro |     |     |     |     |  |
|     |     |     |     | 725 |     |     |     |     | 730 |     |     |     |     |     |     |     |  |

&lt;210&gt; 5

&lt;211&gt; 3416

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| atgacatcta  | gtaataccca  | acctctgctt  | atgacttcct  | ggaacatacc  | cacagctgaa  | 60   |
| ggttctcagt  | ttccaatttc  | caccactatt  | aatgtaccta  | catccaatga  | gatggaaaca  | 120  |
| gagactctac  | accttggttc  | tgggcctttg  | tcaacattca  | cagcctctca  | gactgggtcta | 180  |
| gtatctaaag  | atgtcatggc  | aatgtcatca  | attcctatgt  | caggaattct  | tcctaaccat  | 240  |
| gggctttctg  | agaacccttc  | attatcaaca  | tctttaagag  | ctatcacttc  | cacattggct  | 300  |
| gacgttaagc  | acacatttga  | gaaaatgacc  | acatctgtaa  | ctcctgggac  | cacactccca  | 360  |
| tcaattcttt  | ctggtgccac  | ttcaggatct  | gtaatttcaa  | agtcacccat  | tctgacatgg  | 420  |
| ctcttatcta  | gtctcccttc  | tggctccctt  | ccggcaactg  | tatctaatac  | ccctcatggt  | 480  |
| atgacttcct  | ctacagtaga  | ggtgtcaaaa  | tcaacatttc  | tgacatctga  | catgatatac  | 540  |
| gcgcacccat  | tcactaactt  | gacaacacta  | ccctctgcta  | ctatgagcac  | catactcacc  | 600  |
| cgaaccattc  | ctacacctac  | actgggtggg  | atcactactg  | gcttcccaac  | ttctctccct  | 660  |
| atgtctataa  | atgtcacaga  | tgacattgtg  | tacatttcca  | cacaccctga  | ggcatcctcc  | 720  |
| agaaccacaa  | taactgcca   | ccccaggact  | gtgtctcctc  | cttcctcctt  | cagcagaaaag | 780  |
| actatgtcac  | cttctacaac  | tgaccacact  | ctatctgttg  | gtgccatgcc  | tctgcctagc  | 840  |
| tctacaataa  | catcttcatg  | gaacagaatt  | ccaactgcat  | catcaccctc  | tactttaatt  | 900  |
| attcctaagc  | ccacactgga  | ctcccttcta  | aatataatga  | ctactacatc  | caactgttct  | 960  |
| ggagcctcat  | ttccactcat  | atccactggg  | gtgacataat  | ctttttacagc | aactgtgtct  | 1020 |
| tcaccaatat  | cgctcctttt  | tgaacaactt  | tggctggact  | ccacaccttc  | ctttctatct  | 1080 |
| acggaagcat  | cgacttcgcc  | tactgccacc  | aagtccacag  | tttccttcta  | caatgttgaa  | 1140 |
| atgagcttct  | ctgtctttgt  | tgaagagcca  | aggatcccta  | ttaccagtgt  | tataaatgaa  | 1200 |
| tttacggaaa  | attcggtgaa  | ttctatatatt | cagaacagtg  | aattttctct  | tgctactctg  | 1260 |
| gaaacccaaa  | ttaaaagcag  | ggacatttca  | gaggaagaga  | tgggtcatgga | tcgagctatt  | 1320 |
| ttggaacaga  | gagaaggaca  | agaaatggct  | acaatttcct  | atgtaccata  | cagttgtgtt  | 1380 |
| tgtcagggtca | tcataaaaagc | cagctcttcc  | ttagcatcct  | ctgaattgat  | gagaaaaatc  | 1440 |
| aaaagtaaaa  | tacatggcaa  | cttcacacat  | ggaaacttca  | cacaagatca  | attgacgtta  | 1500 |
| ttagtaaaact | gtgaacacgt  | tgcagtgaat  | aaactagagc  | ctggaaattg  | caaagctgat  | 1560 |
| gaaacagcct  | ctaaatacaa  | agggacctat  | aagtggctat  | taaccaaccc  | tacggagaca  | 1620 |
| gcccacacca  | gatgcataaa  | aatgaggat   | ggaaatgcca  | caagattctg  | ttcaatcagc  | 1680 |
| atcaacacgg  | gcaaattcta  | gtgggaaaag  | ccaaagttta  | aacaatgcaa  | attgcttcaa  | 1740 |
| gaacttcctg  | acaagattgt  | ggatcttgct  | aatattacca  | taagtgatga  | gaatcctgag  | 1800 |
| gatgttgtag  | agcatatttt  | aaatttgata  | aatgaatccc  | cagccctggg  | taaagaagag  | 1860 |
| acaagatta   | ttgttttcta  | aatatcagat  | atttcacaa   | gtgatgagat  | aagtatgaac  | 1920 |
| ctaactcatg  | ttatgttaca  | aataatcaac  | gttgtttttg  | aaaagcaaaa  | caattccgcc  | 1980 |
| tctgatctgc  | atgaaataag  | caatgaaatt  | ctgaggataa  | ttgagcgtcc  | tggtcacaag  | 2040 |
| atggagtttt  | ctgggcagat  | agcaaatctg  | gcgggtggccg | ggctggcttt  | ggctgtgctg  | 2100 |
| cggggggacc  | acacgtttga  | tggcatggct  | ttcagcattc  | actcctatga  | agaaggccca  | 2160 |
| gaccttgaca  | ttttcctagg  | caatgtccct  | gtgggaggga  | ttttggcttc  | catataattg  | 2220 |
| cctaaatcac  | tgacggagag  | aattcctctt  | agcaacttac  | aaccgatctt  | gtttaatttc  | 2280 |
| tttggccaaa  | cttcactctt  | taagaccaaa  | aatgtcacta  | aagcattaac  | cacatatggt  | 2340 |
| gtgagtgcc   | gcatttccaga | tatgttcatt  | caaaacttag  | ctgacccagt  | ggttatcact  | 2400 |
| ctgcagcata  | ttggaggaaa  | ccagaattat  | ggtcaagttc  | actgtgcctt  | ttgggatttt  | 2460 |
| gagaataata  | atgggctggg  | tggatggaat  | tcgtcaggct  | gtaaagtaaa  | ggaaacaaat  | 2520 |
| gtaaattaca  | caatctgtca  | gtgtgaccac  | ctcaccattt  | ttggagtctt  | aatggattta  | 2580 |
| tccagggtcta | cagtggattc  | agtgaatgaa  | cagatattag  | cgcttataac  | atacaccgga  | 2640 |
| tgtggaatct  | cctccatttt  | cctgggagtt  | gcagtgggtg  | catacatagc  | ttttcacaaa  | 2700 |
| cttcgaaaag  | attatcctgc  | caaaattctg  | atcaacctgt  | gcacagcact  | actgatgcta  | 2760 |
| aacctggtat  | ttttgatcaa  | ttcttggttg  | tcattcatctt | agaaagtggg  | agtttgtatc  | 2820 |
| acagctgcag  | tggcacttca  | ttacttcctg  | cttgtttctt  | ttacttggat  | gggcctggag  | 2880 |
| gcagtccaca  | tgtatttggc  | tctagtcaaa  | gtcttcaaca  | tatacattcc  | aaattatata  | 2940 |
| cttaaatttt  | gtctagtgtg  | ttgggggaatc | ccggctatca  | tgggtggcaat | cacagtcagt  | 3000 |
| gtgaaaaaag  | atctgtatgg  | aactctgagc  | ccaacaactc  | cgttttgttg  | gattaaagat  | 3060 |
| gattctatct  | tttacatctc  | agtgggtggc  | tatttttgcc  | tcataattct  | catgaatctc  | 3120 |
| tccatgttct  | gcactgttct  | tgttcaactg  | aattctgtga  | aatcccaaat  | ccagaagact  | 3180 |
| cggcggaaga  | tgatcctgca  | tgacctcaaa  | ggcacaatga  | gcctgacatt  | cttacttggc  | 3240 |
| ctcacctggg  | ggtttgcatt  | ttttgcttgg  | ggacccatga  | ggaacttttt  | cttgtatttg  | 3300 |
| tttgccattt  | ttaacacttt  | gcaaggtaac  | tgggtgctttt | ttgccttttc  | tgtggccagc  | 3360 |

WO 02/36633

PCT/US01/50938

tacacatgca gcaaagcttt tggttgctttg gaaaataatc acctgttgga aacatt

3416

&lt;210&gt; 6

&lt;211&gt; 1138

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 6

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Thr | Ser | Ser | Asn | Thr | Gln | Pro | Leu | Leu | Met | Thr | Ser | Trp | Asn | Ile |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     |     | 15  |     |
| Pro | Thr | Ala | Glu | Gly | Ser | Gln | Phe | Pro | Ile | Ser | Thr | Thr | Ile | Asn | Val |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Pro | Thr | Ser | Asn | Glu | Met | Glu | Thr | Glu | Thr | Leu | His | Leu | Val | Pro | Gly |
|     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |
| Pro | Leu | Ser | Thr | Phe | Thr | Ala | Ser | Gln | Thr | Gly | Leu | Val | Ser | Lys | Asp |
|     | 50  |     |     |     |     | 55  |     |     |     | 60  |     |     |     |     |     |
| Val | Met | Ala | Met | Ser | Ser | Ile | Pro | Met | Ser | Gly | Ile | Leu | Pro | Asn | His |
| 65  |     |     |     |     |     | 70  |     |     |     | 75  |     |     |     |     | 80  |
| Gly | Leu | Ser | Glu | Asn | Pro | Ser | Leu | Ser | Thr | Ser | Leu | Arg | Ala | Ile | Thr |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Ser | Thr | Leu | Ala | Asp | Val | Lys | His | Thr | Phe | Glu | Lys | Met | Thr | Thr | Ser |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Val | Thr | Pro | Gly | Thr | Thr | Leu | Pro | Ser | Ile | Leu | Ser | Gly | Ala | Thr | Ser |
|     |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |
| Gly | Ser | Val | Ile | Ser | Lys | Ser | Pro | Ile | Leu | Thr | Trp | Leu | Leu | Ser | Ser |
|     |     |     | 130 |     |     |     | 135 |     |     |     |     | 140 |     |     |     |
| Leu | Pro | Ser | Gly | Ser | Pro | Pro | Ala | Thr | Val | Ser | Asn | Ala | Pro | His | Val |
| 145 |     |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |
| Met | Thr | Ser | Ser | Thr | Val | Glu | Val | Ser | Lys | Ser | Thr | Phe | Leu | Thr | Ser |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Asp | Met | Ile | Ser | Ala | His | Pro | Phe | Thr | Asn | Leu | Thr | Thr | Leu | Pro | Ser |
|     |     |     |     | 180 |     |     |     |     | 185 |     |     |     | 190 |     |     |
| Ala | Thr | Met | Ser | Thr | Ile | Leu | Thr | Arg | Thr | Ile | Pro | Thr | Pro | Thr | Leu |
|     |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |
| Gly | Gly | Ile | Thr | Thr | Gly | Phe | Pro | Thr | Ser | Leu | Pro | Met | Ser | Ile | Asn |
|     |     |     | 210 |     |     |     | 215 |     |     |     |     | 220 |     |     |     |
| Val | Thr | Asp | Asp | Ile | Val | Tyr | Ile | Ser | Thr | His | Pro | Glu | Ala | Ser | Ser |
| 225 |     |     |     |     |     | 230 |     |     |     | 235 |     |     |     |     | 240 |
| Arg | Thr | Thr | Ile | Thr | Ala | Asn | Pro | Arg | Thr | Val | Ser | His | Pro | Ser | Ser |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Phe | Ser | Arg | Lys | Thr | Met | Ser | Pro | Ser | Thr | Thr | Asp | His | Thr | Leu | Ser |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Val | Gly | Ala | Met | Pro | Leu | Pro | Ser | Ser | Thr | Ile | Thr | Ser | Ser | Trp | Asn |
|     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |
| Arg | Ile | Pro | Thr | Ala | Ser | Ser | Pro | Ser | Thr | Leu | Ile | Ile | Pro | Lys | Pro |
|     |     |     | 290 |     |     |     | 295 |     |     |     |     | 300 |     |     |     |
| Thr | Leu | Asp | Ser | Leu | Leu | Asn | Ile | Met | Thr | Thr | Thr | Ser | Thr | Val | Pro |
| 305 |     |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     | 320 |
| Gly | Ala | Ser | Phe | Pro | Leu | Ile | Ser | Thr | Gly | Val | Thr | Tyr | Pro | Phe | Thr |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
| Ala | Thr | Val | Ser | Ser | Pro | Ile | Ser | Ser | Phe | Phe | Glu | Thr | Thr | Trp | Leu |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| Asp | Ser | Thr | Pro | Ser | Phe | Leu | Ser | Thr | Glu | Ala | Ser | Thr | Ser | Pro | Thr |
|     |     |     | 355 |     |     |     |     | 360 |     |     |     | 365 |     |     |     |
| Ala | Thr | Lys | Ser | Thr | Val | Ser | Phe | Tyr | Asn | Val | Glu | Met | Ser | Phe | Ser |
|     |     |     | 370 |     |     |     | 375 |     |     |     |     | 380 |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Phe | Val | Glu | Glu | Pro | Arg | Ile | Pro | Ile | Thr | Ser | Val | Ile | Asn | Glu |
| 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |
| Phe | Thr | Glu | Asn | Ser | Leu | Asn | Ser | Ile | Phe | Gln | Asn | Ser | Glu | Phe | Ser |
|     |     |     | 405 |     |     |     |     |     | 410 |     |     |     |     | 415 |     |
| Leu | Ala | Thr | Leu | Glu | Thr | Gln | Ile | Lys | Ser | Arg | Asp | Ile | Ser | Glu | Glu |
|     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |
| Glu | Met | Val | Met | Asp | Arg | Ala | Ile | Leu | Glu | Gln | Arg | Glu | Gly | Gln | Glu |
|     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
| Met | Ala | Thr | Ile | Ser | Tyr | Val | Pro | Tyr | Ser | Cys | Val | Cys | Gln | Val | Ile |
|     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
| Ile | Lys | Ala | Ser | Ser | Ser | Leu | Ala | Ser | Ser | Glu | Leu | Met | Arg | Lys | Ile |
| 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |
| Lys | Ser | Lys | Ile | His | Gly | Asn | Phe | Thr | His | Gly | Asn | Phe | Thr | Gln | Asp |
|     |     |     | 485 |     |     |     |     |     | 490 |     |     |     |     | 495 |     |
| Gln | Leu | Thr | Leu | Leu | Val | Asn | Cys | Glu | His | Val | Ala | Val | Lys | Lys | Leu |
|     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |
| Glu | Pro | Gly | Asn | Cys | Lys | Ala | Asp | Glu | Thr | Ala | Ser | Lys | Tyr | Lys | Gly |
|     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |
| Thr | Tyr | Lys | Trp | Leu | Leu | Thr | Asn | Pro | Thr | Glu | Thr | Ala | Gln | Thr | Arg |
|     | 530 |     |     |     |     | 535 |     |     |     |     |     | 540 |     |     |     |
| Cys | Ile | Lys | Asn | Glu | Asp | Gly | Asn | Ala | Thr | Arg | Phe | Cys | Ser | Ile | Ser |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| Ile | Asn | Thr | Gly | Lys | Ser | Gln | Trp | Glu | Lys | Pro | Lys | Phe | Lys | Gln | Cys |
|     |     |     | 565 |     |     |     |     | 570 |     |     |     |     |     | 575 |     |
| Lys | Leu | Leu | Gln | Glu | Leu | Pro | Asp | Lys | Ile | Val | Asp | Leu | Ala | Asn | Ile |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| Thr | Ile | Ser | Asp | Glu | Asn | Pro | Glu | Asp | Val | Ala | Glu | His | Ile | Leu | Asn |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |
| Leu | Ile | Asn | Glu | Ser | Pro | Ala | Leu | Gly | Lys | Glu | Glu | Thr | Lys | Ile | Ile |
|     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |
| Val | Ser | Lys | Ile | Ser | Asp | Ile | Ser | Gln | Cys | Asp | Glu | Ile | Ser | Met | Asn |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |
| Leu | Thr | His | Val | Met | Leu | Gln | Ile | Ile | Asn | Val | Val | Leu | Glu | Lys | Gln |
|     |     |     | 645 |     |     |     |     | 650 |     |     |     |     |     | 655 |     |
| Asn | Asn | Ser | Ala | Ser | Asp | Leu | His | Glu | Ile | Ser | Asn | Glu | Ile | Leu | Arg |
|     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |
| Ile | Ile | Glu | Arg | Pro | Gly | His | Lys | Met | Glu | Phe | Ser | Gly | Gln | Ile | Ala |
|     |     | 675 |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |
| Asn | Leu | Ala | Val | Ala | Gly | Leu | Ala | Leu | Ala | Val | Leu | Arg | Gly | Asp | His |
|     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |
| Thr | Phe | Asp | Gly | Met | Ala | Phe | Ser | Ile | His | Ser | Tyr | Glu | Glu | Gly | Pro |
| 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |
| Asp | Pro | Asp | Ile | Phe | Leu | Gly | Asn | Val | Pro | Val | Gly | Gly | Ile | Leu | Ala |
|     |     |     | 725 |     |     |     |     | 730 |     |     |     |     | 735 |     |     |
| Ser | Ile | Tyr | Leu | Pro | Lys | Ser | Leu | Thr | Glu | Arg | Ile | Pro | Leu | Ser | Asn |
|     |     | 740 |     |     |     |     |     | 745 |     |     |     |     | 750 |     |     |
| Leu | Gln | Pro | Ile | Leu | Phe | Asn | Phe | Phe | Gly | Gln | Thr | Ser | Leu | Phe | Lys |
|     |     | 755 |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |
| Thr | Lys | Asn | Val | Thr | Lys | Ala | Leu | Thr | Thr | Tyr | Val | Val | Ser | Ala | Ser |
|     | 770 |     |     |     |     | 775 |     |     |     |     | 780 |     |     |     |     |
| Ile | Ser | Asp | Met | Phe | Ile | Gln | Asn | Leu | Ala | Asp | Pro | Val | Val | Ile | Thr |
| 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |     |     |     | 800 |
| Leu | Gln | His | Ile | Gly | Gly | Asn | Gln | Asn | Tyr | Gly | Gln | Val | His | Cys | Ala |
|     |     |     | 805 |     |     |     |     |     | 810 |     |     |     |     | 815 |     |
| Phe | Trp | Asp | Phe | Glu | Asn | Asn | Asn | Gly | Leu | Gly | Gly | Trp | Asn | Ser | Ser |
|     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |

|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--|--|
| Gly  | Cys  | Lys  | Val  | Lys  | Glu  | Thr  | Asn  | Val  | Asn  | Tyr  | Thr  | Ile  | Cys  | Gln  | Cys  |  |  |
|      |      | 835  |      |      |      |      | 840  |      |      |      |      | 845  |      |      |      |  |  |
| Asp  | His  | Leu  | Thr  | His  | Phe  | Gly  | Val  | Leu  | Met  | Asp  | Leu  | Ser  | Arg  | Ser  | Thr  |  |  |
|      |      | 850  |      |      |      | 855  |      |      |      |      | 860  |      |      |      |      |  |  |
| Val  | Asp  | Ser  | Val  | Asn  | Glu  | Gln  | Ile  | Leu  | Ala  | Leu  | Ile  | Thr  | Tyr  | Thr  | Gly  |  |  |
| 865  |      |      |      |      | 870  |      |      |      |      | 875  |      |      |      |      | 880  |  |  |
| Cys  | Gly  | Ile  | Ser  | Ser  | Ile  | Phe  | Leu  | Gly  | Val  | Ala  | Val  | Val  | Thr  | Tyr  | Ile  |  |  |
|      |      |      |      | 885  |      |      |      |      | 890  |      |      |      |      |      | 895  |  |  |
| Ala  | Phe  | His  | Lys  | Leu  | Arg  | Lys  | Asp  | Tyr  | Pro  | Ala  | Lys  | Ile  | Leu  | Ile  | Asn  |  |  |
|      |      |      | 900  |      |      |      |      | 905  |      |      |      |      | 910  |      |      |  |  |
| Leu  | Cys  | Thr  | Ala  | Leu  | Leu  | Met  | Leu  | Asn  | Leu  | Val  | Phe  | Leu  | Ile  | Asn  | Ser  |  |  |
|      |      | 915  |      |      |      |      | 920  |      |      |      |      | 925  |      |      |      |  |  |
| Trp  | Leu  | Ser  | Ser  | Phe  | Gln  | Lys  | Val  | Gly  | Val  | Cys  | Ile  | Thr  | Ala  | Ala  | Val  |  |  |
|      |      | 930  |      |      |      | 935  |      |      |      |      |      | 940  |      |      |      |  |  |
| Ala  | Leu  | His  | Tyr  | Phe  | Leu  | Leu  | Val  | Ser  | Phe  | Thr  | Trp  | Met  | Gly  | Leu  | Glu  |  |  |
| 945  |      |      |      |      | 950  |      |      |      |      | 955  |      |      |      |      | 960  |  |  |
| Ala  | Val  | His  | Met  | Tyr  | Leu  | Ala  | Leu  | Val  | Lys  | Val  | Phe  | Asn  | Ile  | Tyr  | Ile  |  |  |
|      |      |      | 965  |      |      |      |      |      | 970  |      |      |      |      | 975  |      |  |  |
| Pro  | Asn  | Tyr  | Ile  | Leu  | Lys  | Phe  | Cys  | Leu  | Val  | Gly  | Trp  | Gly  | Ile  | Pro  | Ala  |  |  |
|      |      | 980  |      |      |      |      |      | 985  |      |      |      |      | 990  |      |      |  |  |
| Ile  | Met  | Val  | Ala  | Ile  | Thr  | Val  | Ser  | Val  | Lys  | Lys  | Asp  | Leu  | Tyr  | Gly  | Thr  |  |  |
|      |      | 995  |      |      |      | 1000 |      |      |      |      |      | 1005 |      |      |      |  |  |
| Leu  | Ser  | Pro  | Thr  | Thr  | Pro  | Phe  | Cys  | Trp  | Ile  | Lys  | Asp  | Asp  | Ser  | Ile  | Phe  |  |  |
|      | 1010 |      |      |      |      | 1015 |      |      |      |      | 1020 |      |      |      |      |  |  |
| Tyr  | Ile  | Ser  | Val  | Val  | Ala  | Tyr  | Phe  | Cys  | Leu  | Ile  | Phe  | Leu  | Met  | Asn  | Leu  |  |  |
| 1025 |      |      |      |      | 1030 |      |      |      |      | 1035 |      |      |      |      | 1040 |  |  |
| Ser  | Met  | Phe  | Cys  | Thr  | Val  | Leu  | Val  | Gln  | Leu  | Asn  | Ser  | Val  | Lys  | Ser  | Gln  |  |  |
|      |      |      |      | 1045 |      |      |      |      | 1050 |      |      |      |      | 1055 |      |  |  |
| Ile  | Gln  | Lys  | Thr  | Arg  | Arg  | Lys  | Met  | Ile  | Leu  | His  | Asp  | Leu  | Lys  | Gly  | Thr  |  |  |
|      |      |      | 1060 |      |      |      |      | 1065 |      |      |      |      | 1070 |      |      |  |  |
| Met  | Ser  | Leu  | Thr  | Phe  | Leu  | Leu  | Gly  | Leu  | Thr  | Trp  | Gly  | Phe  | Ala  | Phe  | Phe  |  |  |
|      |      | 1075 |      |      |      |      | 1080 |      |      |      |      | 1085 |      |      |      |  |  |
| Ala  | Trp  | Gly  | Pro  | Met  | Arg  | Asn  | Phe  | Phe  | Leu  | Tyr  | Leu  | Phe  | Ala  | Ile  | Phe  |  |  |
|      | 1090 |      |      |      |      | 1095 |      |      |      |      | 1100 |      |      |      |      |  |  |
| Asn  | Thr  | Leu  | Gln  | Gly  | Asn  | Trp  | Cys  | Phe  | Phe  | Ala  | Phe  | Ser  | Val  | Ala  | Ser  |  |  |
| 1105 |      |      |      |      | 1110 |      |      |      |      | 1115 |      |      |      |      | 1120 |  |  |
| Tyr  | Thr  | Cys  | Ser  | Lys  | Ala  | Phe  | Val  | Ala  | Leu  | Glu  | Asn  | Asn  | His  | Leu  | Leu  |  |  |
|      |      |      |      | 1125 |      |      |      |      | 1130 |      |      |      |      | 1135 |      |  |  |

Glu Thr

&lt;210&gt; 7

&lt;211&gt; 1988

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 7

|            |            |            |             |             |            |     |
|------------|------------|------------|-------------|-------------|------------|-----|
| atgagaaaaa | tcaaaagtaa | aatacatggc | aacttcacac  | atggaaactt  | cacacaagat | 60  |
| caattgacgt | tattagtaaa | ctgtgaacac | gttgacagtga | aaaaactaga  | gcctggaaat | 120 |
| tgcaaagctg | atgaaacagc | ctctaaatac | aaagggacct  | ataagtggct  | attaaccaac | 180 |
| cctacggaga | cagcccaaac | cagatgcata | aaaaatgagg  | atggaaatgc  | cacaagattc | 240 |
| tgttcaatca | gcatcaacac | gggcaaattc | cagtgggaaa  | agccaaagtt  | taaacaatgc | 300 |
| aaattgcttc | aagaacttcc | tgacaagatt | gtggatcttg  | ctaataattac | cataagtgat | 360 |
| gagaatcctg | aggatgttgc | agagcatatt | ttaaatttga  | taaatgaatc  | cccagccctg | 420 |
| ggtaaagaag | agacaaagat | tattgtttct | aaaatatcag  | atatttcaca  | atgtgatgag | 480 |
| ataagtatga | acctaactca | tgttatgtta | caaataatca  | acgttggttt  | ggaaaagcaa | 540 |

```

aacaattccg cctctgatct gcatgaaata agcaatgaaa ttctgaggat aattgagcgt      600
cctggtcaca agatggagtt ttctgggcag atagcaaate tggcgggtggc cgggctggct      660
ttggctgtgc tgcgggggga ccacacgttt gatggcatgg ctttcagcat tcactcctat      720
gaagaaggcc cagaccctga cattttccta ggcaatgtcc ctgtgggagg gattttggct      780
tccatatatt tgcctaaatc actgacggag agaattcctc ttagcaactt acaaccgatc      840
ttgtttaatt tctttggcca aacttcactc tttaagacca aaaatgtcac taaagcatta      900
accacatatg ttgtgagtgc cagcatttca gatatgttca ttcaaaactt agctgacca      960
gtggttatca ctctgcagca tattggagga aaccagaatt atgggtcaagt tcactgtgcc     1020
ttttgggatt ttgagaataa taatgggctg ggtggatgga attcgtcagg ctgtaaagta     1080
aaggaaacaa atgtaaatta cacaatctgt cagtgtgacc acctcaccca ttttggagtc     1140
ttaatggatt tatccagggtc tacagtggat tcagtgaatg aacagatatt agcgcttata     1200
acatacaccg gatgtggaat ctctccatt ttctgggag ttgcagtggg gacatacata     1260
gcttttcaca aacttcgaaa agattatcct gccaaaattc tgatcaacct gtgcacagca     1320
ctactgatgc taaacctggg atttttgatc aattcttggg tgtcatcatt tcagaaagtg     1380
ggagtttgta tcacagctgc agtggcactt cattacttcc tgcttgtttc ttttacttgg     1440
atgggcctgg aggcagtcca catgtatttg gctctagtca aagtcttcaa catatacatt     1500
ccaaattata tccttaaatt ttgtctagtt ggttggggaa tcccggctat catggtggca     1560
atcacagtca gtgtgaaaaa agatctgtat ggaactctga gcccaacaac tccgttttgt     1620
tggtataaag atgattctat cttttacatc tcagtgggtg cttatttttg cctcatattt     1680
ctcatgaatc tctccatgtt ctgcactgtt cttgttcaac tgaattctgt gaaatcccaa     1740
atccagaaga ctcggcggaa gatgatcctg catgacctca aaggcacaat gagcctgaca     1800
ttcttacttg gcctcacctg ggggtttgca ttttttgctt ggggacccat gaggaacttt     1860
ttcttgtatt tgtttgccat ttttaacact ttgcaaggta actggtgctt ttttgccttt     1920
tctgtggcca gctacacatg cagcaaagct tttgttgctt tggaaaataa tcacctgttg     1980
gaaacatt

```

&lt;210&gt; 8

&lt;211&gt; 662

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 8

```

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

```

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Ile | Leu | Arg | Ile | Ile | Glu | Arg | Pro | Gly | His | Lys | Met | Glu | Phe | Ser |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Gly | Gln | Ile | Ala | Asn | Leu | Ala | Val | Ala | Gly | Leu | Ala | Leu | Ala | Val | Leu |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Arg | Gly | Asp | His | Thr | Phe | Asp | Gly | Met | Ala | Phe | Ser | Ile | His | Ser | Tyr |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Glu | Glu | Gly | Pro | Asp | Pro | Asp | Ile | Phe | Leu | Gly | Asn | Val | Pro | Val | Gly |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Gly | Ile | Leu | Ala | Ser | Ile | Tyr | Leu | Pro | Lys | Ser | Leu | Thr | Glu | Arg | Ile |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Pro | Leu | Ser | Asn | Leu | Gln | Pro | Ile | Leu | Phe | Asn | Phe | Phe | Gly | Gln | Thr |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Ser | Leu | Phe | Lys | Thr | Lys | Asn | Val | Thr | Lys | Ala | Leu | Thr | Thr | Tyr | Val |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Val | Ser | Ala | Ser | Ile | Ser | Asp | Met | Phe | Ile | Gln | Asn | Leu | Ala | Asp | Pro |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Val | Val | Ile | Thr | Leu | Gln | His | Ile | Gly | Gly | Asn | Gln | Asn | Tyr | Gly | Gln |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
| Val | His | Cys | Ala | Phe | Trp | Asp | Phe | Glu | Asn | Asn | Asn | Gly | Leu | Gly | Gly |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| Trp | Asn | Ser | Ser | Gly | Cys | Lys | Val | Lys | Glu | Thr | Asn | Val | Asn | Tyr | Thr |
|     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
| Ile | Cys | Gln | Cys | Asp | His | Leu | Thr | His | Phe | Gly | Val | Leu | Met | Asp | Leu |
|     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |
| Ser | Arg | Ser | Thr | Val | Asp | Ser | Val | Asn | Glu | Gln | Ile | Leu | Ala | Leu | Ile |
| 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |
| Thr | Tyr | Thr | Gly | Cys | Gly | Ile | Ser | Ser | Ile | Phe | Leu | Gly | Val | Ala | Val |
|     |     |     | 405 |     |     |     |     |     | 410 |     |     |     |     | 415 |     |
| Val | Thr | Tyr | Ile | Ala | Phe | His | Lys | Leu | Arg | Lys | Asp | Tyr | Pro | Ala | Lys |
|     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |
| Ile | Leu | Ile | Asn | Leu | Cys | Thr | Ala | Leu | Leu | Met | Leu | Asn | Leu | Val | Phe |
|     | 435 |     |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
| Leu | Ile | Asn | Ser | Trp | Leu | Ser | Ser | Phe | Gln | Lys | Val | Gly | Val | Cys | Ile |
|     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
| Thr | Ala | Ala | Val | Ala | Leu | His | Tyr | Phe | Leu | Leu | Val | Ser | Phe | Thr | Trp |
| 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |
| Met | Gly | Leu | Glu | Ala | Val | His | Met | Tyr | Leu | Ala | Leu | Val | Lys | Val | Phe |
|     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |
| Asn | Ile | Tyr | Ile | Pro | Asn | Tyr | Ile | Leu | Lys | Phe | Cys | Leu | Val | Gly | Trp |
|     |     | 500 |     |     |     |     |     | 505 |     |     |     |     | 510 |     |     |
| Gly | Ile | Pro | Ala | Ile | Met | Val | Ala | Ile | Thr | Val | Ser | Val | Lys | Lys | Asp |
|     | 515 |     |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |
| Leu | Tyr | Gly | Thr | Leu | Ser | Pro | Thr | Thr | Pro | Phe | Cys | Trp | Ile | Lys | Asp |
|     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
| Asp | Ser | Ile | Phe | Tyr | Ile | Ser | Val | Val | Ala | Tyr | Phe | Cys | Leu | Ile | Phe |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| Leu | Met | Asn | Leu | Ser | Met | Phe | Cys | Thr | Val | Leu | Val | Gln | Leu | Asn | Ser |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |
| Val | Lys | Ser | Gln | Ile | Gln | Lys | Thr | Arg | Arg | Lys | Met | Ile | Leu | His | Asp |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| Leu | Lys | Gly | Thr | Met | Ser | Leu | Thr | Phe | Leu | Leu | Gly | Leu | Thr | Trp | Gly |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |
| Phe | Ala | Phe | Phe | Ala | Trp | Gly | Pro | Met | Arg | Asn | Phe | Phe | Leu | Tyr | Leu |
|     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |
| Phe | Ala | Ile | Phe | Asn | Thr | Leu | Gln | Gly | Asn | Trp | Cys | Phe | Phe | Ala | Phe |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |

WO 02/36633

PCT/US01/50938

Ser Val Ala Ser Tyr Thr Cys Ser Lys Ala Phe Val Ala Leu Glu Asn  
 645 650 655  
 Asn His Leu Leu Glu Thr  
 660

<210> 9  
 <211> 4185  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 caactttggg catattatct gggattacta acagggtccct atctactgtg aacagtggta 60  
 caggggtagc tctcacagat acttattcca gaatcactgt tcctgaaaat atgctttcac 120  
 ctactcatgc agatagtctc catacttcct tcaatattca ggtttcccca tctctgacta 180  
 gctttaagag tgcttctgga cccacaaaaa atgttaaaac aaccaccaat tgcctttctt 240  
 ctaatactag aaagatgact tccttgtttag aaaagacttc cttaacaaac tatgccacat 300  
 ctttgaatac ccctgtttca taccctccat ggaccccatc cagtgcact ctaccttctt 360  
 tgacatcatt tgtttattca cctcatagta ctgaagctga gatctctact ccaaagacct 420  
 ctctcctcc cacaacccaa atgggtgaat ttccagttct gggaacaaag aatgacatct 480  
 agtaataccc aacctctgct tatgacttcc tggaacatac ccacagctga aggttctcag 540  
 tttccaattt ccaccactat taatgtacct acatccaatg agatggaaac agagactcta 600  
 caccttggtc ctgggccttt gtcaacattc acagcctctc agactgggtc agtatctaaa 660  
 gatgtcatgg caatgtcatc aattcctatg tcaggaattc ttcctaacca tgggcctttct 720  
 gagaaccctt cattatcaac atctttaaga gctatcactt ccacattggc tgacgttaag 780  
 cacacatttg agaaaatgac cacatctgta actcctggga ccacactccc atcaattctt 840  
 tctgggtgcca cttcaggatc tgtaatttca aagtcaccca ttctgacatg gctcttatct 900  
 agtctccctt ctgggtcccc tccggcaact gtatctaattg cccctcatgt tatgacttcc 960  
 tctacagtag aggtgtcaaa atcaacattt ctgacatctg acatgatatc agcgcacca 1020  
 ttcactaact tgacaacact accctctgct actatgagca ccatactcac ccgaaccatt 1080  
 cctacaccta cactgggtgg tatcactact ggcttcccaa cttctctccc tatgtctata 1140  
 aatgtcacag atgacattgt gtacatttcc acacaccctg aggcacctc cagaaccaca 1200  
 ataactgcca accccaggac tgtgtctcat ccttcactct tcagcagaaa gactatgtca 1260  
 ccttctacaa ctgaccacac tctatctgtt ggtgccatgc ctctgcctag ctctacaata 1320  
 acatcttcat ggaacagaat tccaactgca tcatcaccct ctactttaat tattcctaag 1380  
 cccacactgg actcccttct aaatataatg actactacat ccactgttcc tggagcctca 1440  
 tttccactca tatccactgg ggtgacatat ccttttacag caactgtgtc ttcaccaata 1500  
 tcgtcccttt ttgaaacaac ttggctggac tccacacctt cctttctatc tacggaagca 1560  
 tcgacttcgc ctactgccac caagtccaca gtttccttct acaatgttga aatgagcttc 1620  
 tctgtctttg ttgaagagcc aaggatccct attaccagtg ttataaatga atttacggaa 1680  
 aattcgttga attctatatt tcagaacagt gaattttctc ttgctactct ggaaacccaa 1740  
 attaaaagca gggacatttc agaggaagag atgggtcatgg atcgagctat tttggaacag 1800  
 agagaaggac aagaaatggc tacaatttcc tatgtaccat acagttgtgt ttgtcaggtc 1860  
 atcataaaag ccagctcttc cttagcatcc tctgaattga tgagaaaaat caaaagtaaa 1920  
 atacatggca acttcacaca tggaaacttc acacaagatc aattgacgtt attagtaaac 1980  
 tgtgaacacg ttgcagtga aaaactagag cctggaaatt gcaaagctga tgaaacagcc 2040  
 tctaaataca aagggaccta taagtggcta ttaaccaacc ctacggagac agcccaaacc 2100  
 agatgcataa aaaatgagga tggaaatgcc acaagattct gttcaatcag catcaacacg 2160  
 ggcaaatctc agtgggaaaa gccaaagttt aaacaatgca aattgcttca agaacttctt 2220  
 gacaagattg tggatcttgc taatattacc ataagtgatg agaatcctga ggatgttgca 2280  
 gagcatattt taaatttgat aaatgaatcc ccagccctgg gtaaagaaga gacaaagatt 2340  
 attgtttcta aaatatcaga tatttcacaa tgtgatgaga taagtatgaa cctaactcat 2400  
 gttatgttac aaataatcaa cgttggtttt gaaaagcaaa acaattccgc ctctgatctg 2460  
 catgaaataa gcaatgaaat tctgaggata attgagcgtc ctgggtcacia gatggagttt 2520  
 tctgggcaga tagcaaatct ggcggtggcc gggctggctt tggctgtgct gcgggggggac 2580  
 cacacgtttg atggcatggc tttcagcatt cactcctatg aagaaggccc agaccctgac 2640  
 attttcctag gcaatgtccc tgtgggaggg attttggctt ccatatattt gcctaaatca 2700

|             |             |             |            |            |             |      |
|-------------|-------------|-------------|------------|------------|-------------|------|
| ctgacggaga  | gaattcctct  | tagcaactta  | caaccgatct | tgtttaattt | ctttggccaa  | 2760 |
| acttcactct  | ttaagaccaa  | aatgtcact   | aaagcattaa | ccacatatgt | tgtgagtgcc  | 2820 |
| agcatttcag  | atatgttcat  | tcaaaactta  | gctgacccag | tggttatcac | tctgcagcat  | 2880 |
| attggaggaa  | accagaatta  | tgggtcaagtt | cactgtgcct | tttgggattt | tgagaataat  | 2940 |
| aatgggctgg  | gtggatggaa  | ttcgtcaggc  | tgtaaagtaa | aggaaacaaa | tgtaaattac  | 3000 |
| acaatctgtc  | agtgtgacca  | cctcacccat  | tttggagtct | taatggattt | atccagggtct | 3060 |
| acagtggatt  | cagtgaatga  | acagatatta  | gcgcttataa | catacacccg | atgtggaatc  | 3120 |
| tcctccattt  | tcctgggagt  | tgcagtgggtg | acatacatag | cttttcacaa | acttcgaaaa  | 3180 |
| gattatcctg  | ccaaaattct  | gatcaacctg  | tgcacagcac | tactgatgct | aaacctggta  | 3240 |
| tttttgatca  | attcttgggt  | gtcatcattt  | cagaaagtgg | gagtttgtat | cacagctgca  | 3300 |
| gtggcacttc  | attacttctt  | gcttgtttct  | tttacttggg | tgggcctgga | ggcagtcac   | 3360 |
| atgtatttgg  | ctctagtcaa  | agtcttcaac  | atatacattc | caaattatat | ccttaaattt  | 3420 |
| tgtctagtgt  | gttgggggaat | cccggctatc  | atggtggcaa | tcacagtcag | tgtgaaaaaa  | 3480 |
| gatctgtatg  | gaactctgag  | ccaacaact   | ccgttttgtt | ggattaaaga | tgattctatc  | 3540 |
| ttttacatct  | cagtgggtgg  | ttatTTTTgc  | ctcataattc | tcatgaatct | ctccatgttc  | 3600 |
| tgcactgttc  | ttgttcaact  | gaattctgtg  | aaatcccaa  | tccagaagac | tcggcggaag  | 3660 |
| atgatcctgc  | atgacctcaa  | aggcacaatg  | agcctgacat | tcttacttgg | cctcacctgg  | 3720 |
| gggtttgcat  | tttttgcttg  | gggacccatg  | aggaactttt | tcttgtattt | gtttgccatt  | 3780 |
| tttaacactt  | tgcaaggatt  | cttcattttt  | gtgtttcact | gtgtgatgaa | ggagagtgtg  | 3840 |
| cgggagcagt  | ggcagataca  | cctctgctgt  | gggtgggtgc | gattggataa | ctcttctgat  | 3900 |
| gggagcagcc  | ggtgtcagat  | aaaggttgga  | tataaacagg | agggactaaa | gaaaatcttt  | 3960 |
| gagcaciaaac | tgttgacgcc  | atctctcaag  | tcaactgcaa | ctagctccac | tttcaaattc  | 4020 |
| ttaggctctg  | cacaaggcac  | accttcagaa  | ataagctttc | caaatgatga | cyttgacaaa  | 4080 |
| gaccttact   | gttcctctcc  | ttgatttgtg  | aagttgygcc | taattatgta | aaaagaatat  | 4140 |
| aatacctgtg  | gaaataaaaa  | tgaattccaa  | gtggaaaaaa | aaaaa      |             | 4185 |

## WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising the nucleotide sequence drawn from the group consisting of SEQ ID  
5 NO:1, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7.
2. An isolated nucleic acid molecule comprising a nucleotide sequence that:
  - (a) encodes the amino acid sequence shown in SEQ ID  
10 NO:2; and
  - (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:1 or the complement thereof.
- 15 3. An isolated nucleic acid molecule according to Claim 1, wherein said molecule is a cDNA.
4. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence drawn  
20 from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8.
5. An isolated expression vector comprising the nucleotide sequence drawn from the group consisting of SEQ ID  
25 NO:1, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7.