

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 November 2001 (01.11.2001)

PCT

(10) International Publication Number
WO 01/81409 A2

(51) International Patent Classification⁷: C07K 14/705

(72) Inventors: WEI, Ming-Hui; Celera, 45 West Gude Drive, Rockville, MD 20850 (US). CRAVCHIK, Anibal; Celera, 45 West Gude Drive, Rockville, MD 20850 (US). DI FRANCESCO, Valentina; Celera, 45 West Gude Drive, Rockville, MD 20850 (US). BEASLEY, Ellen, M.; Celera, 45 West Gude Drive, Rockville, MD 20850 (US).

(21) International Application Number: PCT/US01/13097

(22) International Filing Date: 24 April 2001 (24.04.2001)

(25) Filing Language: English

(74) Agent: CELERA GENOMICS CORP.; Robert A. Millman, Patent Director, 45 West Gude Drive C2-4, Rockville, MD 20850 (US).

(26) Publication Language: English

(30) Priority Data:
60/199,149 24 April 2000 (24.04.2000) US
09/633,146 4 August 2000 (04.08.2000) US

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, 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, MY, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(71) Applicant: PE CORPORATION (NY) [US/US]; Robert A. Millman, Assistant Secretary, 761 Main Avenue, Norwalk, CT 06859 (US).

[Continued on next page]

(54) Title: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogues of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

1 AACAGCAAGA GGTCCTGCTC AGCCAGCTGG GTCCCGAGG ACATGAAAGG
51 ATTATGAGAC CGGAGAGAGC CTCACAGGAA TGGGCGTGG TCTGAGACAA
101 GAGGGCTTG AGAAGACTTG GGGGCGCTC GTCACAGTTC AAAGGGCTCC
151 GCTCTCCDCC ACTGACCCCC CAGGCTGCTC GCGCAGACC CTCTGAGCCG
201 GCGAGATAT TTGGGGTATT TGGGCTCTGG AGAAGTCTCC ACAGTGTGGT
251 CTTCTGCTGG AGCTGATATG TGGGCTCTGG GGGGCGCTC GGTACAGGGC
301 TGGTCTCTGG GAACTCTGGC TTCCGATACA AAGAGGGGCC CTCTCCATC
351 TACTCTCTCC ACTCTGAGCC CCGCAGCTTC CTGTCTCTCT CTGCGGTGT
401 GGGTCTCTCC GGGTCTGGC TTCCGATACA GGGGCGCTC AGCTCTCTCT
451 TCTGCTCTCC CTCTCTGGC TTCCGATACA GGGTCTCTCT GTCGGGGGCC
501 TTGAGGCTGG AGGAGCTGCT CTCGAGCTTC TTCTCGGCT GCTACAGGGC
551 CTCGGGGGCC AGAAGACTTG GGGGCGCTC TTCTCGGCT GGTGAGAGCC
601 CAGCCCTGCC GGGGCTGGC TTCCGATACA AAGGCTGGG CTTGCTGGC
651 AAGAGGGGCT GCGCCCTGCT CTCGAGCTTC TACAGCTGG CAGAGCTCAC
701 GGGTCTCTCC GGTGAGAGCC GGGGCGCTC GAGGCTGGG AGGCTCTCT
751 TTGCTGCTCT GAGCTCTGGC TTCTCTGGC CCGGCGCCG GCTCTAGGGC
801 ATTCTCTGG GGGGCTGCT CTCGAGCTTC TTCTCTGGC TGGCTCTGCT
851 CTCTCTCTCC AGCTGATATG GGGGCGCTC CTCTCTGGC GCGGCTCTCT
901 CCGGCTGGC CAGGCTCTGGC GGGGCGCTC AAGAGACTTC CAGGCGCTCC
951 ATTCTCTGG GGTGAGAGCC AAGGCGGGG AAGGCGAGC CCGTGAAGTC
1001 GGTACTGGG AGGAGCTCTGG GGGGCGCTC CAGGCTGGG GCGGCGGGC
1051 AGTCTCTGCC CATGGGCTCT CTATAGTGG GGTGCGCCG CCGCAGGGC
1101 CTGCGAGGAG GTGGGAGCC CAGGCGGGC TGGCTCTCC CAGAGCTCAC
1151 TCTCTGGG AGGAGAGGCA GCTCTCTGGC TGGGCGCTC AGGCTCTCT
1201 AAGGCTGGG AGGCTCTGGC CTCGAGCTTC CCGGCGCCG CTTCTGAGG
1251 ATGAGGAGG CAGAAATGG AGCTGAGAG AGTCTAGCA AAGAGAGAGA
1301 AAGAGAGAG CTCTCTAGTA GGGGCGCTC GAGGAGCAC GCGGCGGGC
1351 CAGGCGAGG TCGGCTTACC GGGGCGCTC CAGGCTCTCT GAGCTCTGAT
1401 CAGGCGGGG AAGACTGAC ATATAGAGG AGCCAGCAC CAGGCGGGC
1451 TGGGCTCTCC GAGGAGAGG CTCTCTGGC CTCTCTGGC GGGGCGGGC
1501 TCGGCGAGC CTCTCTGGC AGGCTCTCT TCGGCGCTC CAGGCTCTCT
1551 GGTGAGTAC TCTCTCTCC CCGGCTGCT CAGGCTCTCT TGTCTCTCC
1601 ACCGCGGGG GAGGAGAGC AGGCTCTCT CCGGCTCTCT CAGGCGGGG
1651 ACTGAGAGT GGTGATGTC AAGAGAGAG GCGGAGAGT GGTCTCTCT
1701 CTGAGAGAA CAGAGAGAG CCGGCGGGC GGGGAGTGA GAGGCTCTCA
1751 TCGGCTGGT CCGGCTGGT CCGGCTGGC ACTGAGAGC AGGCGGGC
1801 TCGGCTGGT CAGGCGAGT CCGGCTGGT CCGGCTGGT CCGGCTGGT
1851 CCGGCTGGT CAGGCGAGT CCGGCGGGC GAGGCTCTCT GAGGCTCTCT
1901 CAGGCTCTCT CCGGCTGGT CCGGCGGGC GAGGCTCTCT GAGGCTCTCT
1951 AAGAGAGAG AAGAGAGAG AAGAGAGAG AAGAGAGAG AAGAGAGAG
2001 GAGAGAGAG AAGAGAGAG AAGAGAGAG AAGAGAGAG AAGAGAGAG
2051 AAGAGAGAG AAGAGAGAG AAGAGAGAG AAGAGAGAG AAGAGAGAG

(SEQ ID NO: 1)
FEATURES:
Start: 207
Stop: 1074
5'UTR: 1-206
3'UTR: 1074-2999
HOMOLOGOUS PROTEIN:
Top 10 BLAST Hits:
gi134979|sp|P23491|RTA_BAT PROBABLE G PROTEIN-COUPLED RECEPTOR... 102 4e-30
gi1598184|ref|NP_058555.1|| HSI1 oncogene ...|113521|sp|P12512... 101 7e-23
gi1547920|sp|P35470|HRG_HUMAN MAS-RELATED G PROTEIN-COUPLED REC... 101 1e-20
gi1456105|ref|NP_002368.1|| HSI1 oncogene ...|1135920|sp|P04201... 101 1e-20
gi1467804|ref|NP_123778.1|| HSI1 oncogene ...|11265501|sp|P10504... 96 4e-19
gi12118485|pir|I1551001|transforming protein mas - mouse 95 1e-18
gi14758014|ref|NP_004063.1|| chemokine-like receptor 1 ...|16931... 84 2e-09
gi11023627|sp|Q95781|CNIL_HUMAN CHEMOKINE RECEPTOR-LIKE ...|10392... 64 2e-08
gi13642615|sp|AAC36503.1|| (U96378) anaphylatoxin C3a receptor ... 55 6e-08
gi16931498|sp|O88680|C3AR_CAVO C3A ANAPHYLATOXIN CHEMOTACTIC P... 59 6e-08
blast to dbEST:
no significant matches
Tissue expression for modulatory use:
Human Thyroid
Human Kidney
Human Liver
Human Lung
Human Leukocytes
Human Placenta
Human Fetal Brain
Human Testis
Human Heart
Human Pancreas

WO 01/81409 A2



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

Published:

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

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

**ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID
MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF**

RELATED APPLICATIONS

5 The present application claims priority to U.S. Serial Nos. 60/199,149, filed April 24, 2000 (Atty. Docket CL000485-PROV) and 09/633,146, filed August 4, 2000 (Atty. Docket CL00748).

FIELD OF THE INVENTION

10 The present invention is in the field of G-Protein coupled receptors (GPCRs) that are related to the MAS proto-oncogene receptor subfamily, recombinant DNA molecules and protein production. The present invention specifically provides novel GPCR peptides and proteins and nucleic acid molecules encoding such protein molecules, for use in the development of human therapeutics and human therapeutic development.

15 **BACKGROUND OF THE INVENTION**

G-protein coupled receptors

20 G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular
25 inputs.

30 GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest.* 92:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet.* 30:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet.* 26:403-424(1992)), and nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet.* 2:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary

analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the β 2-purinergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem.* 60:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science* 254:1024-1026 (1991); Lin *et al.*, *Science* 254:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science* 258 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science* 241:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem.* 61:1097-1129 (1992)).

There are also a small number of other proteins that present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein that has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, *fz* has not been shown to couple to G-proteins (Vinson *et al.*, *Nature* 338:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α , β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α -subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the $\beta\gamma$ -subunits. The GTP-bound form of the α -subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenylyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α -subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include G_i , G_o , G_q , G_s and G_t . G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

MAS1 oncogene-like receptors

The human MAS1 oncogene was originally detected by its ability to transform NIH 3T3 cells. The MAS1 oncogene was isolated from DNA of that human epidermoid carcinoma cell line

using the cotransfection and tumorigenicity assay (Young et al., 1984). Based on its deduced amino acid sequence, the MAS1 gene product contains 7 potential transmembrane domains typical of the GPCRs. The mas1 protein is, therefore, probably an integral membrane protein. The mas1 encoded protein may be a receptor that, when activated, modulates a critical component in a growth-

5 regulating pathway to bring about its oncogenic effects. Jackson et al. (1988) found that the MAS oncogene shows the greatest sequence similarity to the substance-K receptor. On this basis, they predicted that it would encode a peptide receptor with mitogenic activity which would act through the inositol lipid signaling pathways. Expression in *Xenopus* oocytes and a transfected mammalian cell line demonstrated that the MAS gene product is a functional angiotensin receptor.

10 Ross et al. (1990) identified the RTA receptor, a GPCRs related to the MAS1 oncogene. RTA is expressed abundantly throughout the gut, vas deferens, uterus, and aorta but are only barely detectable in liver, kidney, lung, and salivary gland. In the rat brain, RTA sequences are markedly abundant in the cerebellum. RTA is most closely related to the mas oncogene (34% identity), which has been suggested to be a forebrain angiotensin receptor. However, angiotensin binding to the

15 RTA receptor was not detected after introducing the cDNA or mRNA into COS cells or *Xenopus* oocytes, respectively. Therefore, it was concluded that RTA is not an angiotensin receptor.

Young et al (1988) cloned the rat homolog of the MAS oncogene, determined its DNA sequence, and examined its expression in various rat tissues. A comparison of the predicted sequences of the rat and human mas proteins shows that they are highly conserved, except in their

20 hydrophilic amino-terminal domains. High levels of mas RNA transcripts were detected in the hippocampus and cerebral cortex of the brain, but not in other neural regions or in other tissues. This pattern of expression and the similarity of mas protein to known receptor proteins suggest that mas encodes a receptor that is involved in the normal neurophysiology and/or development of specific neural tissues.

25 The discovery of a new MAS-oncogene like receptor satisfies a need in the art by providing new compositions which are in the prevention, diagnosis, and treatment of cancer and neurodegenerative diseases. (For Example See: 1) Ross PC, et al., RTA, a candidate G protein-coupled receptor: cloning, sequencing, and tissue distribution. *Proc Natl Acad Sci U S A* 1990 Apr;87(8):3052-6; 2) Young D, et al., Characterization of the rat mas oncogene and its high-level

30 expression in the hippocampus and cerebral cortex of rat brain. *Proc Natl Acad Sci U S A* 1988 Jul;85(14):5339-42; 3) Young, D. et al., Isolation and characterization of a new cellular oncogene encoding a protein with multiple transmembrane domains. *Cell* 45: 711-719, 1984; and 4) Jackson, T. R.; et al., The mas oncogene encodes an angiotensin receptor. *Nature* 335: 437-440, 1988.

GPCRs, particularly members of the MAS proto-oncogene receptor subfamily, are a major target for drug action and development, particularly for the prevention, diagnosis, and treatment of cancer and neurodegenerative diseases. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention
5 advances the state of the art by providing a previously unidentified human GPCR.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of nucleic acid sequences that encode amino acid sequences of human GPCR peptides and proteins that are related to the MAS
10 proto-oncogene subfamily, allelic variants thereof and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins and serve as targets for the development of human therapeutic agents.

The proteins of the present inventions are GPCRs that participate in signaling pathways
15 mediated by the MAS proto-oncogene subfamily in cells that express these proteins (see expression information in Figure 1, the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval). As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR protein. Examples of such functions
20 include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP₂), inositol 1,4,5-triphosphate (IP₃) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival

25 The response mediated by the receptor protein depends on the type of cell it is expressed on. Some information regarding the types of cells that express other members of the subfamily of GPCRs of the present invention is already known in the art (see references cited in Background and information regarding closest homologous protein provided in Figure 2 and expression information provided in Figure 1 the GPCR of the present invention is expressed in thyroid, kidney, liver, lung,
30 leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval). For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the

particular GPCR of the present invention, a skilled artisan will clearly know that the receptor protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell thus participating in a biological process in the cells or tissues that express the GPCR (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)).

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP₂) as well as to the activities of these molecules. PIP₂ is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP₂ to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP₃). Once formed IP₃ can diffuse to the endoplasmic reticulum surface where it can bind an IP₃ receptor, e.g., a calcium channel protein containing an IP₃ binding site. IP₃ binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP₃ can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP₄), a molecule that can cause calcium entry into the cytoplasm from the extracellular medium. IP₃ and IP₄ can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-bisphosphate (IP₂) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP₂. The other second messenger produced by the hydrolysis of PIP₂, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-κB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP₂ or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenylyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated

protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

5 By targeting an agent to modulate a GPCR, the signaling activity and biological process mediated by the receptor can be agonized or antagonized in specific cells and tissues (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)). Such agonism and antagonism serves as a basis for modulating a biological activity in a therapeutic context (mammalian therapy) or toxic context (anti-cell therapy, e.g. anti-cancer agent).

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule or transcript sequence that encodes the GPCR of the present invention. In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)).

FIGURE 2 provides the predicted amino acid sequence of the GPCR of the present invention. In addition structure and functional information, such as protein family and function, modification sites, is provided that allows one to readily determine specific uses of inventions based on this molecular sequence as well as significant fragments of the proteins of the present invention.

FIGURE 3 provides genomic sequences that span the gene encoding the GPCR protein of the present invention. In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided that allows one to readily determine specific uses of inventions based on this molecular sequence as well as important fragments for use in probe and primer design and heterologous gene expression control.

DETAILED DESCRIPTION OF THE INVENTION

30 General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural

and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a GPCR protein or part of a GPCR protein, that are related to the MAS proto-oncogene subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human GPCR peptides and proteins, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these GPCR peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the GPCR of the present invention.

10 In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known GPCR proteins of the MAS proto-oncogene subfamily and the expression pattern observed (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the
15
20 Figures, and/or are known within the art for each of the know MAS proto-oncogene family or subfamily of GPCR proteins.

Specific Embodiments

Peptide Molecules

25 The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the GPCR family of proteins (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, such as allelic variants, will be referred herein as the GPCR peptides of the present invention, GPCR peptides, or peptides/proteins of the present invention.
30

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of or are comprised of the amino acid sequences of the GPCR peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA and

Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the GPCR peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated GPCR peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). For example, a nucleic acid molecule encoding the GPCR peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence that such a protein consists of is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that are comprised of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein is comprised of an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the GPCR peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The GPCR peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a GPCR peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the GPCR peptide. "Operatively linked" indicates that the GPCR peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the GPCR peptide.

In some uses, the fusion protein does not affect the activity of the GPCR peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant GPCR peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification

of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A GPCR peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCR peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art know techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the GPCR peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular*

Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*CABIOS*, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the GPCR peptides of the present invention as well as being encoded by the same genetic locus as the GPCR peptide provided herein. The GPCR of the present invention is encoded by

a gene found on chromosome 11, near markers SHGC-32486 (LOD=15.45) and SHGC-20653 (LOD=15.45).

Allelic variants of a GPCR peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by the same genetic locus as the GPCR peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human (the GPCR of the present invention is encoded by a gene found on chromosome 11, near markers SHGC-32486 (LOD=15.45) and SHGC-20653 (LOD=15.45)). As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Paralogs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the GPCR peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the GPCR peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a GPCR peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala,

Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310
5 (1990).

Variant GPCR peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to bind G-protein, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis that identifies critical domains/regions.
10 Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or
15 deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for
20 biological activity such as ligand/effector molecule binding or in assays such as an *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the GPCR peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly fragments identified in Figure 2.
25 The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16 or more contiguous amino acid residues from a GPCR peptide. Such fragments can be chosen based on the ability to retain one or
30 more of the biological activities of the GPCR peptide or could be chosen for the ability to perform a function, e.g. ability to bind ligand or effector molecule or act as an immunogen. Particularly important fragments are biologically active fragments, peptides which are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the GPCR peptide, e.g., active site, a G-protein binding site, a transmembrane domain or a ligand-binding

domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well-known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are
5 provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications
10 that occur naturally in GPCR peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment
15 of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation,
20 racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Accordingly, the GPCR peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature GPCR peptide is fused with another compound,
25 such as a compound to increase the half-life of the GPCR peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature GPCR peptide, such as a leader or secretory sequence or a sequence for purification of the mature GPCR peptide or a pro-protein sequence.

Such modifications are well-known to those of skill in the art and have been described in great
30 detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of*

Proteins, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Protein/Peptide Uses

5 The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures and Back Ground Section; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or receptor) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either
10 constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the binding partner so as to develop a system to identify inhibitors of the binding interaction. Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as commercial products.

15 Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

20 The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, GPCRs isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug to modulate the cells or tissues that express the receptor (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta,
25 fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). Approximately 70% of all pharmaceutical agents modulate the activity of a GPCR. A combination of the invertebrate and mammalian ortholog can be used in selective screening methods to find agents specific for invertebrates. The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention. Such uses can
30 readily be determined using the information provided herein, that known in the art and routine experimentation.

 The receptor polypeptides (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-

related conditions that are specific for the subfamily of GPCRs that the one of the present invention belongs to, particularly in cells and tissues that express this receptor (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)).

5 The receptor polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the receptor protein, as a biopsy or expanded in cell culture (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). In one embodiment, however, cell-based assays involve recombinant host cells expressing
10 the receptor protein.

 The polypeptides can be used to identify compounds that modulate receptor activity of the protein in its natural state, or an altered form that causes a specific disease or pathology associated with the receptor. Both the GPCRs of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the receptor.
15 These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree.

 Further, the receptor polypeptides can be used to screen a compound for the ability to stimulate
20 or inhibit interaction between the receptor protein and a molecule that normally interacts with the receptor protein, e.g. a ligand or a component of the signal pathway that the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). Such assays typically include the steps of combining the receptor protein with a candidate compound under conditions that allow the
25 receptor protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cAMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

30 Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell*

72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

5 One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

10 The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, a cellular process such as proliferation, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade, can be assayed. In one embodiment, the regulatory region of such genes can be operably
15 linked to a marker that is easily detectable, such as luciferase.

Any of the biological or biochemical functions mediated by the receptor can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the
20 information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the receptor can be assayed (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)).

Binding and/or activating compounds can also be screened by using chimeric receptor proteins
25 in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein than that which is recognized by the native receptor. Accordingly,
30 a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be

performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still
5 cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The receptor polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor
10 polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the
15 target receptor region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In
20 one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological
25 conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target
30 molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex

trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting
5 an enzymatic activity associated with the target molecule.

Agents that modulate one of the GPCRs of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

0 Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells or tissues that express the GPCR (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). These methods of treatment include the steps of administering a modulator of the GPCR's activity in a
15 pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the GPCR proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent
20 WO94/10300), to identify other proteins, which bind to or interact with the GPCR and are involved in GPCR activity. Such GPCR-binding proteins are also likely to be involved in the propagation of signals by the GPCR proteins or GPCR targets as, for example, downstream elements of a GPCR-mediated signaling pathway. Alternatively, such GPCR-binding proteins are likely to be GPCR
25 inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a GPCR protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other
30 construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a GPCR-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ)

which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the GPCR protein.

5 This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a GPCR modulating agent, an antisense GPCR nucleic acid molecule, a GPCR-specific antibody, or a GPCR-binding partner) can be used in an animal or other model to determine the
10 efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or insect model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

 The GPCR proteins of the present invention are also useful to provide a target for diagnosing a
15 disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). The method involves contacting a biological sample with a compound capable of interacting with the receptor
20 protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

 One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

25 The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion,
30 insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques

useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagents, such as an antibody or protein binding agent.. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein (the GPCR of the present invention is expressed

in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)). Accordingly, methods for treatment include the use of the GPCR protein or fragments.

Antibodies

5 The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. 10 In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen 15 challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

20 In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods. 25

Antibodies are preferably prepared from regions or discrete fragments of the GPCR proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or receptor/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments. 30

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are

located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various
5 enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include
10 umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Uses

15 The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present
20 invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis,
25 heart, pancreas as determined by cDNA retrieval (Figure 1)). Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function, particularly in cells and tissues that express the receptor (the GPCR of the present invention is
30 expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)). When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific

mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)). Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the GPCR peptide to a binding partner such as a ligand. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a GPCR peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the GPCR peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figures 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ

ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule, for example from about 1-300 additional nucleotides.

The present invention further provides nucleic acid molecules that are comprised of the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule is comprised of a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, human genomic sequences (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the GPCR peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional

non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates
5 purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form of DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding
10 strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the GPCR proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs
15 (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-
20 conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify
25 gene modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing
30 regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene (the GPCR of the present invention is encoded by a gene found on chromosome 11, near markers SHGC-32486 (LOD=15.45) and SHGC-20653 (LOD=15.45)).

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also
5 include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

10 The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods (the GPCR of the present invention is encoded by a gene found on chromosome 11, near markers SHGC-32486 (LOD=15.45) and SHGC-20653 (LOD=15.45)). This is particularly useful in determining whether a particular protein is an allelic variant of one the proteins provided herein

15 The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention as described in detail below.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

20 The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

25 The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)).

30 The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in GPCR protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a GPCR protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)).

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate GPCR nucleic acid expression, particularly in cells and tissues that express the receptor (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)).

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the GPCR gene. The method typically includes assaying the ability of the compound to modulate the expression of the GPCR nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired GPCR nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the GPCR nucleic acid (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determine by cDNA retrieval (Figure 1)) or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for GPCR nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the GPCR protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of GPCR gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of GPCR mRNA in the presence of the candidate compound is compared to the level of expression of GPCR mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When

nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate GPCR nucleic acid expression, particularly to modulate activities within a cell or tissue that expresses the proteins (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)). Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for GPCR nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the GPCR nucleic acid expression in the cells and tissues that express the protein (the GPCR of the present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)).

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the GPCR gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in GPCR nucleic acid, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in GPCR genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally-occurring genetic mutations in the GPCR gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the GPCR gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a GPCR protein.

Individuals carrying mutations in the GPCR gene can be detected at the nucleic acid level by a variety of techniques (the GPCR of the present invention is encoded by a gene found on chromosome 11, near markers SHGC-32486 (LOD=15.45) and SHGC-20653 (LOD=15.45)). Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis (). RNA or cDNA can be used
5 in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids*
10 *Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and
15 comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a GPCR gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

20 Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such
25 as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant GPCR gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996);
30 and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita

et al., *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations
5 include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that, while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's
10 response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the GPCR gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of
15 recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control GPCR gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence
20 production of GPCR protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into GPCR protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of GPCR nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired GPCR nucleic acid expression. This technique involves
25 cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the GPCR protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells
30 that are aberrant in GPCR gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired GPCR protein to treat the individual.

The invention also encompasses kits for detecting the presence of a GPCR nucleic acid in a biological sample, particularly cells and tissues that normally express the protein (the GPCR of the

present invention is expressed in thyroid, kidney, liver, lung, leukocytes, placenta, fetal brain, testis, heart, pancreas as determined by cDNA retrieval (Figure 1)). For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting GPCR nucleic acid in a biological sample; means for determining the amount of GPCR nucleic acid in the sample; and
5 means for comparing the amount of GPCR nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCR protein mRNA or DNA.

Nucleic Acid Arrays

10 The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter,
15 chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods
20 described by Brown et. al., US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in
25 length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides
30 that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the

gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide
5 in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by
10 using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those
15 described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA
20 from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with
25 precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured
30 cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the GPCR proteins/peptides of the present invention and allelic variation within this gene/protein. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes or alleles, at least one of which is a gene and or alleles of the GPCR gene of the present invention.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the GPCR disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a

quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily
5 recognize that the previously unidentified GPCR genes of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

10 The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or
15 artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

20 The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host
25 cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood,
30 however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late

promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include
5 the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and
10 termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors
15 include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage
20 genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. Tissue specific) or may provide for inducible expression in one or more cell types such as by
25 temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an
30 expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells

include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein.

5 Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes
10 include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988))
15 and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of
20 interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*,
25 *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

30 In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are
5 found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.*

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that
10 permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

15 The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include,
20 but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.*)

Host cells can contain more than one vector. Thus, different nucleotide sequences can be
25 introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

30 In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as GPCRs, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with GPCRs, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography; or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a GPCR protein or peptide that can be further purified to produce desired amounts of GPCR protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the GPCR protein or GPCR protein fragments, such as those described above as well as other formats known in the art.

Thus, a recombinant host cell expressing a native GPCR protein is useful for assaying compounds that stimulate or inhibit GPCR protein function.

Host cells are also useful for identifying GPCR protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the
5 mutations are useful to assay compounds that have a desired effect on the mutant GPCR protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native GPCR protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in
10 which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a GPCR protein and identifying and evaluating modulators of GPCR protein activity. Other examples of transgenic animals include non-
15 human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the GPCR protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

20 Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the GPCR protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection,
25 particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its
30 genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

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

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The
15 quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

20 Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, GPCR protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* GPCR protein
25 function, including ligand interaction, the effect of specific mutant GPCR proteins on GPCR protein function and ligand interaction, and the effect of chimeric GPCR proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more GPCR protein functions.

All publications and patents mentioned in the above specification are herein incorporated by
30 reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for

carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.
8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
9. A host cell containing the vector of claim 8.
10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.
13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.
14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.
16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human proteases, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human protease peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.

22. An isolated nucleic acid molecule encoding a human protease peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

1/6
FIGURE 1, sheet 1 of 2

```

1 AACAGCACGA GGCTCCGTCC AGCCAGCTGG GTCCCAGAGG ACATGAAAGG
51 ATTTAGGGAC CGGGAGGAGC CTCCAGGCAA TGAGCGCTGG TCTGAGCACA
101 GAAGGGCTCG AGAGAGCTGG GGGCAGCCTC GTCACAGGTG AGAAGGCTCC
151 GCCTGCCCCC ACTGCACCCC CCAGGTCCTT CGCCAGCACC CTCTGAGCCG
201 GCCAGGATGT TTGGGCTGTT CGGCCTCTGG AGAACCTTCG ACAGTGTGGT
251 CTTCTACCTG ACGCTGATCG TGGGCTCGG GGGACCGGTA GGTAAACGGG
301 TGGTGCTCTG GAACCTCGGC TTCCGCATCA AGAAGGGCCC CTTCTCCATC
351 TACCTGCTGC ACCTGGCCGC CGCCGACTTC CTGTTCTCTT CCTGCCGTGT
401 GGGCTTCTCC GTGGCTCAGG CTGCCCTGGG CGCCCAGGAC ACACTCTACT
451 TCGTGCTCAC CTTCTGTGGT TTCGCGGTGG GGCTCTGGCT GCTGGCGGCC
501 TTCAGCGTGG AGCGCTGCCT CTCCGACCTC TTCCCCGCCT GCTACCAGGG
551 CTGCCGGCCC AGACACGCCT CGGCCGTCTT CTGCGCCCTG GTGTGGACCC
601 CGACCCTGCC GGCCGTGCCG CTGCCCGCCA ACGCCTGCGG CCTGCTGCGC
651 AACAGCGCGT GCCCCTGGT CTGCCCGCGC TACCACGTGG CCAGCGTCA
701 CTGGTCTCTG GTGCTGGCCC GCGTGCCTG GACGGCTGGC GTGGTCTCTT
751 TTGTCTGGGT GACCTGTGTC TCCACTCGCC CGCGGCCAG GCTCTACGGC
801 ATCGTCTGGG GCGCGCTGCT CCTGCTCTTC TTCTGTGGCC TGCCCTCGGT
851 CTTCTACTGG AGCCTGCAGC CCCTGCTGAA CTTCTGCTG CCCGTGTTTT
901 CCCCCTGGC CACGCTGCTG GCCTGCGTCA ACAGCAGCTC CAAGCCCTC
951 ATCTACTCGG GGTGGGGCCG ACAGCCCGGG AAGCGGGAGC CGCTGAGGTC
1001 GGTACTTCGG AGGGCCCTGG GGGAGGGCGC CGAGCTGGGT GCCAGGGGAC
1051 AGTCCCTGCC CATGGGTCTC CTATAAGTGG GCTTGCCCCG CCCACAGGGC
1101 CTGCCAGGAG GTGCCACACC CCACCGACCC TCGCTCACCC CACACCCAGA
1151 TGCTTTCAGG ACCAGGAGGA GCTTTACCC TGAGCACCCA AGGGCTGGAC
1201 ACAGCTGGAG AGACTTGGCC CTGACCACCC CCCAGCCAGG CCTGCTGAGG
1251 ATGAGGGAGG CAGAAAATGG AGCTGGAGAG AGGTCAGGCA AGGAGAGAGA
1301 AAGGAGAAGC CTCCTGAATA GGGGTGAGGA CAGGCACCAC GCCCCAGGC
1351 CCAGCCCAGA TCCCCCTACC CGGCCCTCC CCACCCTGCT GCACCTGAGT
1401 CACAGGGGAG AAAACTGCAC AATAAAACAG AGCCAGCCAC CAGCCCACAG
1451 TGGCCGGATT GGAACCCAGG CTTCGGACT CCTGGGCTAG GTGGGCGCCG
1501 FCCATGCCAC CTGCTGGCTG AGGCTCTGAT TCGCCCCCTA CAGAGTTGAG
1551 GTGGAACTAC TCCTTACTCC CCGCCCTGCT CAGCCATCAT TGTCTTGCCC
1601 ACCCCGCGGG GACCACACC AGGGCTCTGT CCCCCTCTCT GAGGCCCAGG
1651 ACTGGCAGGT GCCTGATGTC ACCAGCAGAG GCCACCAGGT GGTGCTGCTT
1701 CTGCACAGAA CAGACCCAGC CCCGTGGGCC GCGGGATGCA GGAGCCTCCA
1751 TCCCCTCGGT CCCCTCCCAT CCCCTCCAC ACTGGTCCCC ACCCGGCCTC
1801 FCCTGCGTCC CCAGGGCCAT CCGTCTCTG CGGGTTGCTC CCGTCTCCA
1851 CCTCTGCTCT CACCCTCCCT CCTCAGCCCT GGATCATCTG GAGCTTTTGC
1901 CCCAAGTGTG TGCTTTGGAG AGAGAGAGAG AGAGACAGAG AGAGAGAGAG
1951 AGACAGAGAG AGAGAGAAAG AGAGAGAGAG AAAAGAAAAG AAGGAAAGAA
2001 GGAAGGAAGG AAAAAGAAAG AAAGAAAGAA AGAAAGAAAG AAAGAAAGAA
2051 AGAAAGAAAG AAAGGAAGGA AGAAAGGAAG GAAAAAAGAA AAAAAAGAA

```

(SEQ ID NO: 1)

FEATURES:

Start: 207
 Stop: 1074
 5'UTR: 1-206
 3'UTR: 1074-2999

HOMOLOGOUS PROTEIN:

Top 10 BLAST Hits:

```

gi|134079|sp|P23749|RTA_RAT PROBABLE G PROTEIN-COUPLED RECEPTOR... 132 4e-30
gi|6981186|ref|NP_036889.1|| MAS1 oncogene >gi|135921|sp|P12526... 101 7e-21
gi|547920|sp|P35410|MRG_HUMAN MAS-RELATED G PROTEIN-COUPLED REC... 101 1e-20
gi|4505105|ref|NP_002368.1|| MAS1 oncogene >gi|135920|sp|P04201... 101 1e-20
gi|6678804|ref|NP_032578.1|| MAS1 oncogene >gi|266505|sp|P30554... 96 4e-19
gi|2118485|pir||S51001 transforming protein mas - mouse 95 1e-18
gi|4758014|ref|NP_004063.1|| chemokine-like receptor 1 >gi|6831... 64 2e-09
gi|3023627|sp|Q99788|CML1_HUMAN CHEMOKINE RECEPTOR-LIKE 1 (G-PR... 64 2e-09
gi|3642615|gb|AAC36503.1|(U86378) anaphylatoxin C3a receptor [... 59 6e-08
gi|6831498|sp|O88680|C3AR_CAVPO C3A ANAPHYLATOXIN CHEMOTACTIC R... 59 6e-08

```

2/6

FIGURE 1, sheet 2 of 2

blast to dbEST:

no significant matches

Tissue expression for modulatory use:

Human Thyroid
Human Kidney
Human Liver
Human Lung
Human Leukocytes
Human Placenta
Human Fetal Brain
Human Testis
Human Heart
Human Pancreas

FIGURE 2, sheet 1 of 2

1 MFGLEGLWRTFDSVVFYLTIVGLGGPVGNGLVLWNLGFRIKKGFPSIYLLHLAAADFLF
 61 LSCRVGFSVAQAALGAQDTLYFVLTFLWFAVGLWLLAAFSVERCLSDLFPACYQGCRPRH
 121 ASAVLICALVWTPTLPAVPLPANACGLLRNSACPLVCPRYHVASVTWFLVLRVAVTAGVV
 181 LFWVVTCCSTRERPRLYGIVLGAALLLFFCGLPSVFWSLQPLLNFLLPVFSPLATLLAC
 241 VNSSSKPLIYSGLRQPGKREPLRSVLRRALGEGAEELGARGQSLPMGLL
 (SEQ ID NO: 2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
 N-glycosylation site
 242-245 NSSS

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site
 Number of matches: 3
 1 62-64 SCR
 2 189-191 STR
 3 244-246 SSK

[3] PDOC00008 PS00008 MYRISTYL
 N-myristoylation site
 Number of matches: 6
 1 6-11 GLWRTF
 2 25-30 GGPVGN
 3 75-80 GAQDTL
 4 145-150 GLLRNS
 5 198-203 GIVLGA
 6 274-279 GAELGA

[4] PDOC00009 PS00009 AMIDATION
 Amidation site
 257-260 PGKR

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	13	33	1.558	Certain
2	44	64	1.019	Certain
3	80	100	2.370	Certain
4	121	141	1.642	Certain
5	170	190	1.634	Certain
6	196	216	2.261	Certain
7	222	242	1.697	Certain

FIGURE 2, sheet 2 of 2

BLAST Alignment to Top Hit:

```

gi|134079|sp|P23749|RTA_RAT PROBABLE G PROTEIN-COUPLED RECEPTOR...
Query: 13  SVVFYLTFLVGLGGPVGNGLVLWNLGFRIKKGPFPSIYLLHLAAADFLFLSCRVGFSVAQA 72
          +V  Y+ L++ L G VGNGLVLW  GF IK+ PFSIY LHLA+AD ++L    FS A
Sbjct: 45  AVTNYIFLLthe GPCR of the present invention is encoded by a gene found on
chromosome 11, near markers SHGC-32486 (LOD=15.45) and SHGC-20653
(LOD=15.45)GLVGNGLVLWFFGFSIKRTPFSIYFLHLASADGIYL-----FSKAVI 99

Query: 73  ALGAQDT-----LYFVLTEFLWFVAVGLWLLAAFSVERCLSDLFPCYQGCRPRHA 121
          AL   T           + ++   F G+ LL A S+ERC+S +FP Y  RP+
Sbjct: 100 ALLNMGTFLGSFPDYVRRVSRIVGLCTFFAGVSLLPAISIERCVSVIFPMWYRRRPKRL 159

Query: 122 SAVLCALVWTPPTLPVAVPLPANACGLLRNSACPLVCPRYHVA-SVTWFLVLARVAWTAGVV 180
          SA +CAL+W  +   +   C L + A   C   ++ + F +   +   +
Sbjct: 160 SAGVCALLWLLSFLVTSIHNYFCMFLGHEASGTACLNMDISLGILLFFLFCPLMVLPCLA 219

Query: 181 LFFVWVTCSTRPR--PRLYGIVLGALLLFFCGLPSVFYWSLQPLLNFLLPVFSPLATLL 238
          L + V C + R +   +L +VL  + +   +   W L +   P   + L
Sbjct: 220 LILHVECRARRRQRS AKLNHVVLAI VSVFLVSSIYLGIDWFLFWVFQIPAPFPEYVTDLC 279

Query: 239 ACVNSSSKPLIYSGLGRQPGKR--EPLRSVLRRALGEGAEELGARGQSLP 285
          C+NSS+KP++Y  GR   +R  EPLR V +RAL +GAE G   S P
Sbjct: 280 ICINSSAKPIVYFLAGRDKSQRLWEPLRVVFQRALRDGAEPGDAASSTP 328
(SEQ ID NO: 4)
    
```

Hmmer search results (Pfam):

Model	Seq-from	Seq-to	HMM-from	HMM-to	Score	E-value	Description
7tm_1	29	131	1	113	22.3	8.7e-08	7 transmembrane receptor (rhodopsin family)
LGT	10	221	1	324	-157.7	0.51	Prolipoprotein diacylglyceryl transferase
7tm_1	204	250	202	259	0.1	0.86	7 transmembrane receptor (rhodopsin family)

5/6
 FIGURE 3, sheet 1 of 2

1 CGTCCCTCTG GGCCCTGGT GACTGCAGGT GACCGGAGCT ACCAACTAAG
 51 TTAATAGAGA TGCTCCAAGT TTCACCCAAA CAGGAGGCAG GTAAGCAAGG
 101 CTTGGAGGCG CTAGTTGAAA CAGGCCTGGA CGGTGACCTC GAGGAGGAGG
 151 TGGAGCCACG CCGGGAGAGG TCCCGGAGAG CAGCGAGGAC CGGCATCTGG
 201 GGTGGAGGCA GCAGGCAGGG TGGCAGGAGG TGCCCTGGG CCACACGAGC
 251 TGTGCCGTGC GTTTCAAAGC CATGTGCAAC AATGCCGGGA GCTGGGTAGC
 301 CGGCATGCCC CCGCCCAGAG CTGCACAGGC TGCTGGTTTG TGTTCAGAA
 351 GTGATTTCCG AAGCCAAGTC CATTTGAGAA ACGTGAAGAG AAACAGCATC
 401 TTTACGACCG AATTTCTGGC TTGAGCTGAC AACTCTGCA CCCCACCCAG
 451 GAGCTGCGGA ACGCAGTCCG TGGAAGGAAA AACTCTCTTA TCAAGAACAT
 501 CTTACTGTCC AGTGTTCCTG AGAGCAAAAT GCGGAAAACA ATCCGTCAAC
 551 GGCTGGGAGC CCAGGCTCTG AGGCAACAG ATGGACAGTG AGGGGGTCTC
 601 CTTAGGAAGG GGTGGGGGCC GCCCTGTAG ACCACCCACA GACGAGTGG
 651 TCGGCTGACA CATTAGGAGA GCCAGTCCA AGCCTGGGAA CCCAGGCTTC
 701 GAGCTGTCCA TAAACGGTCA GTCCAGTGAG AGGGGAAAAC CAAGGCACAG
 751 AGAGGTTGGG TTATTCTCTC AAGGCCACAC AGCCAGGCC CTGTCCACAG
 801 TTCTAGGCC CCTGTGCCA CAGCCAACCA AAGACTAGCA GGAGCAGCCC
 851 AACATCAGTG GAGTACCCGG AAGGTGGCCC AGCCTCACTG GGCACAGAGG
 901 GTCTCCATTC TCCTTCTCCC TCCCCTCCAT GCACCAGGTG AGGTGGGAGC
 951 GGGTGGAAAG CACTGACCTG GTGCCTCAG TTCCTTCC AGAGGGGCTG
 1001 AGGAGGAAGG AAAGGCCTGC CCACACTGAC AACTTTTTGT TTTAGCTCAT
 1051 TCCTGCAATT TCTGTGAACA GTGCCACACC CTGTCACTTC TAGGGAAGT
 1101 GCCGGGTGCC CTCCGTCCAG CCAGCTGGGT CCCAGAGGAC ATGAAAGGAT
 1151 TTAGGGACCA GGAGGAGCCT CCAGGCAATG AGCGCTGGTC TGAGCACAGA
 1201 AGGGCCTGAG AGAGCTGGGG GCAGCTCGT CACAGGTGAG AAGGCTCCGC
 1251 CTGCCCCCAG TGACCCCCC AGGTGAGGCA CTGGCACCTC CCCCTGCCCT
 1301 TCCGCACTGG CCCCTGCTCT CAGCCTGGCT CCCCGGTGGA TGGGGCTGGC
 1351 TGGGATGTTT GGATTTCTCT TGGGAAGGGA AGGCAGGGCG GGATGCGGGT
 1401 GGAATTCGG GCTAAGTTGA AATGTGTGGA CACCAGAGCT CAAGGGCGTC
 1451 GCTGGAGCGG GCCGAGCAGG TGAACCTGTT CTGATAACAA TGAAGATGCC
 1501 CCCTCCTGCC GCCGGGCCAC TCCTCTCTGT AACTCTCTCT CCAGGAATAG
 1551 AGCCTAACCT CACCCAGCA TGGGGGCTG AGCGCCGGC TCACCCCCAG
 1601 CCTCAGACTC TGAGCCAGAG CCAGAGCACA GCCAGGAGAG TCTGGTTTCC
 1651 TCCTTGAGGA AATTGAGGCT CAGAGAGGGG AGGTGTGTTT CCACCCCACC
 1701 ACCCCACCCC CAGTCACCAC ACTCACCACC ACAAGCCAGG GCCCGGCCGA
 1751 GGTCTCCAC TGCCCCGTGT TGGGTCGGA GGACAGGCCA CCATCACCGT
 1801 GTGCCAGCC ACGGCAAGGC ACAGGAGCCA AGGCCGAGC CGCACTCCTC
 1851 ACCCCCTGAC TAGGGGCAAT GACAAGCACA AAGAAGGTGT CGGGGGATGC
 1901 TCGGCCATGA GGCAGGGGCG TCACGGGGGT CGGAGGAAGC TGCAGGGGGT
 1951 TGCAGGCCCC TGAAAAGGAG CAAAACCTTG AAGTAGGGTG GGACCCTGGG
 2001 TCAGGGGAAG AGGTATTCAG GGCTGTCTGG GCAGGGGACA CAACCCAAGC
 2051 AAAGATGTGA CAGGGAGGGC TCCAGTGTG TGTGGGGCG CTTGATGGCT
 2101 CTCTCTCTGT CTGTCTCTCT CTCTCTCACA CACACACAGT CTGTCTCATC
 2151 TCTGTTTCTC TGCGTCTATC TCTGGCTCTA TATCTCTGTA GCTCTCTGTA
 2201 TCTGACTGTC TGACTCTGTC TATAAATGTA TGTCTACATA TCTGTCTCTC
 2251 CCTGTATATG TGGCTCTCTA TATACTCTTT ATATCTCCAG AGTCTTTTAC
 2301 CGCTGACGCC GCCACTGPTG AGCCTGGTCC TGATGCATCT CCCAGGACT
 2351 AAGAACATGA CCCCATGTGG ATCCTGGGGG AGGAGGGATG GCTGGAAAGG
 2401 GAAGGGCTGC TCTGAGACCC TCTGCCTTTC AGCGCAGCTT CCGGTGGCCA
 2451 ATCTCTCTCG CACTCTGAGA ACATCCTCGG GACTCTCAGC CTCGCTCCTT
 2501 GCAAGTGGCC ACTTGTATCT GCCCTCCCC TTGTGTTAGT GGCTCCTGGG
 2551 AGAGGCCCCC TGGCCCTGCC ACCATCACCA AAGCCCCTCT CTTGGCTTGA
 2601 GGCTTCAGGA GCAGCCTCCT GAGTGCACCG CCGGCAGCAT GAGCCCCGGG
 2651 ATTCAGCTTC TGCTTAGGCC AGGAGTGCAC AGGAAGGCT GGGGGCTCCC
 2701 TCCAGCTCCT CACACCGCAT CTTCCCTCCT CCTCCTCCT CAGACCCT
 2751 GCAGCCTCTG CCTTTGGGCA CCTCCCTGGA CACTCCAGA GACCCTCCGG
 2801 GGAGGGGACT TCGCCAGCAG GCTGGTCTAG GCTCACCAGG TGGGAGCCCC
 2851 ATGAGGGTCC ACCGTGGAGA AGCTCCCACC TCTGGGGGAA GAGGAAGGCG
 2901 CCACACGTGG GTTCCCCATG CACGTTCCCT CTGTCTGCAG ACCCTTGTCT
 2951 GGCACACGGT GGAGGGTGA GCATCACCT CACCACATCC ATGAGGGACA
 3001 TGGACATCTT TTCAGCAAGT TAGTCCACCC TGCCAGGAGC CCAGGGCCTG
 3051 CCCTCAGGTA TGGGGAGGCA TTAGGCACGC CTGACCTGAG GTGCAGAGGG
 3101 CTGAGACCCG GCTGAGGTCC CGCCGCTGGG AAGGTTGGT CAGGGGCTGC
 3151 ACCCAGAGGG CACCTGCAGG GCCATGGATG CCACTCCCTG TGCCCCCAGC

6/6
FIGURE 3, sheet 2 of 2

```

3201 GGGGGCACAA AAGACTCCTG AGGACCATGT CTGTCTCCAT CTTGCAGGTC
3251 CCTCGCCAGC ACCCTCTGAG CCGGCCAGGA TGTTTGGGCT GTTCGGCCTC
3301 TGGAGAACCT TCGACAGTGT GGTCTTCTAC CTGACGCTGA TCGTGGGCCT
3351 CGGGGGACCG GTAGGTAACG GGCTGGTGCT CTGGAACCTC GGCTTCCGCA
3401 TCAAGAAGGG CCCCTTCTCC ATCTACCTGC TGCACCTGGC CGCCGCCGAC
3451 FTCTGTGTCC TCTCCTGCCG TGTGGGCTTC TCCGTGGCTC AGGCTGCCCT
3501 GGGCGCCCAG GACACACTCT ACTTCGTGCT CACCTTCCTG TGGTTCGCGG
3551 TGGGGCTCTG GCTGCTGGCG GCCTTCAGCG TGGAGCGCTG CCTCTCCGAC
3601 CTCTTCCCGG CCTGCTACCA GGGCTGCCGG CCCAGACACG CCTCGGCCGT
3651 CCTCTGCGCC CTGGTGTGGA CCCCAGCCCT GCCGGCCGTG CCGCTGCCCG
3701 CCAACGCCTG CGGCCTGCTG CGCAACAGCG CGTGCCCCCT GGTCTGCCCG
3751 CGTACCACG TGGCCAGCGT CACCTGGTTC CTGGTGTGCG CCCGCGTCGC
3801 CTGGACGGCT GGCCTGTGCC TCTTGTCTG GGTGACCTGC TGCTCCACTC
3851 GCCCGCGGCC CAGGCTCTAC GGCATCGTCC TGGGCGCGCT GCTCCTGCTC
3901 TTCTTCTGTG GCCTGCCCTC GGTCTTCTAC TGGAGCCTGC AGCCCCTGCT
3951 GAACTTCCCTG CTGCCCGTGT TTTCCCGCTG GGCCACGCTG CTGGCCTCGG
4001 TCAACAGCAG CTCCAAGCCC CTCATCTACT CGGGGTGGG CCGACAGCCC
4051 GGAAGCGGG AGCCGCTGAG GTCGGTACTG CGGAGGGCCC TGGGGGAGGG
4101 CGCCGAGCTG GGTGCCAGGG GACAGTCCCT GCCCATGGGT CTCCTATAAG
4151 TGGGCTTGCC CCGCCCACAG GGCCTGCCAG GAGGTGCCCA CCCCACCGA
4201 CCCTCGCTCA CCCACACCC AGATGCTTTC AGGACCAGGA GGAGCTCTTA
4251 CCTTGAGCAC CCAAGGGCTG GACACAGCTG GAGAGACTTG GCCCTGACCA
4301 CCCCCAGCC AGGCCTGCTG AGGATGAGGG AGGCAGAAAA TGGAGCTGGA
4351 GAGAGGTCAG GCAAGGAGAG AGAAAGGAGA AGCCTCCTGA ATAGGGGTGA
4401 GGACAGGCAC CACGCCCCCA GGCCAGCCC AGATCCCCTT ACCCGGCCCC
4451 TCCCCACCCT GCTGCACCTG AGTCACAGGG GAGAAAATG CACAATAAAA
4501 CAGAGCCAGC CACCAGCCCA CAGTGGCCGG ATTGGAACCC AGGCTTCCGG
4551 ACTCCTGGGC TAGGTGGGCG CCGTCCATGC CACCTGCTGG CTGAGGCTCT
4601 GATTCGCCCC CTACAGAGTT GAGGTGGAAC TACTCCTTAC TCCCCGCCCT
4651 GCTCAGCCAT CATTGTCTG CCCACCCCG GGGGACCACA CCCAGGGCTC
4701 TGTCCCCCTC TCTGAGGCC AGGACTGGCA GGTGCCTGAT GTCACCAGCA
4751 GAGGCCACCA GGTGGTGTG CTTCTGCACA GAACAGACCC AGCCCCGTGG
4801 GCCGGCGGAT GCAGGAGCCT CCATCCCCTC GGTCCCCTCC CATCCCCTCC
4851 CACACTGGTC CCCACCCGCG CTCTCCTGCG TCCCCAGGGC CATCCGCTCT
4901 CTGCGGGTTG CTCCCCTCTC CCACCTCTGC TCTCACCTC CCTCCTCAGC
4951 CCTGGATCAT CTGGAGCTTT TGCCCCAAGT GTTTGCTTTG GAGAGAGAGA
5001 GAGAGAGACA GAGAGAGAGA GAGAGACAGA GAGAGAGAGA AAGAGAGAGA
5051 GAGAAAAGAA AAGAAGGAAA GAAGGAAGGA AGGAAAAGA AAGAAAGAAA
5101 GAAAGAAAGA AAGAAAGAAA GAAAGAAAGA AAGGAAGGAA GAAAGGAAGG
5151 AAAAAGAAAG AAAGAAAGCA AGAAAGCAAG AAAGCAAGAA AGAAAGAAAG
5201 AAAGAAAGAA AGAAAGAGAA GAGAAGAGAA GACATAAGGG GAGGGGAGGG
5251 GAAGAAGGAA AGGAAAGAAA ATAATCCTAG GAAGGAAGGA AGGAATGAAG
5301 GGG
    
```

(SEQ ID No: 3)

FEATURES:

Transcript:

Start: 3280

Exon: 3280-4149

Stop: 4147

Map position for allelic variants:

#	SHGCNAME	CHROM#	LOD_SCORE	DIST. (cRs)
1	SHGC-32486	11	15.45	4
2	SHGC-20653	11	15.45	4
3	SHGC-57855	11	11.93	13
4	SHGC-52782	11	11.25	14
5	SHGC-14668	11	10.91	17

SEQUENCE LISTING

<110> WEI et al

<120> ISOLATED HUMAN G-PROTEIN COUPLED
RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
PROTEINS, AND USES THEREOF

<130> CL000748PCT

<140> N/A

<141> 2001-04-24

<150> 60/199,149

<151> 2000-04-24

<150> 09/633,146

<151> 2000-08-04

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2099

<212> DNA

<213> human

<400> 1

```

aacagcacga ggctccgtcc agccagctgg gtccagagg acatgaaagg atttagggac 60
cgggaggagc ctccaggcaa tgagcgctgg tctgagcaca gaagggcctg agagagctgg 120
gggcagcctc gtcacagggt agaaggtccc gcctgcccc actgcacccc ccaggtcctt 180
cgccagcacc ctctgagccg gccaggatgt ttgggctggt cggcctctgg agaaccttcg 240
acagtgtggt cttctacctg acgctgatcg tgggcctcgg gggaccggta ggtaacgggc 300
tggtgctctg gaacctcggc ttccgcatac agaagggccc cttctccatc tacctgctgc 360
acctggccgc cgccgacttc ctgttcctct cctgcctgtg gggcttctcc gtggctcagg 420
ctgccctggg cgcccaggac aactctact tctgtctc ac ctctctgtgg ttcgcggtgg 480
ggctctggct gctggcggcc ttcagcgtgg agcgtgcct ctccgacctc ttccccgect 540
gctaccaggg ctgcccggcc agacacgcct cggccgtcct ctgcccctg gtgtggacc 600
cgacctgccc ggcccgtccc taccacgtgg ccagcgtcac ctggttcctg gtgctggccc 720
gccccctggt ctgcccgcgc taccacgtgg ccagcgtcac ctggttcctg gtgctggccc 720
gcgctgcctg gacggctggc gtggtcctct ttgtctgggt gacctgctgc tccactgcc 780
cgcgcccag gctctacggc atcgtcctgg gcgcgctgct cctgctcttc ttctgtggcc 840
tgccctcggc cttctactgg agcctgcagc ccctgctgaa cttcctgctg cccgtgtttt 900
ccccgctggc cacgctgctg gcctgcgtca acagcagctc caagcccctc atctactcgg 960
ggttgggccc acagcccggg aagcgggagc cgctgaggtc ggtactgcgg agggccctgg 1020
gggagggcgc cgagctgggt gccaggggac agtccctgcc catgggtctc ctataagtgg 1080
gcttgccccg cccacagggc ctgccaggag gtgcccaccc ccaccgacct tcgctcacc 1140
cacaccaga tgctttcagg accaggagg gctcttacc tgagcaccca agggctggac 1200
acagctggag agacttgccc ctgaccacc cccagccagg cctgctgagg atgaggagg 1260
cagaaaatgg agctggagag aggtcaggca aggagagaga aaggagaagc ctccctgaata 1320
ggggtgagga caggcaccac gccccaggc ccagcccaga tccccttacc cggccccctc 1380
ccacctgact gcacctgagt cacaggggag aaaactgcac aataaaacag agccagccac 1440
cagcccacag tggccggatt ggaaccagg cttccggact cctgggctag gtgggcgccc 1500
tccatgccac ctgctggctg aggtctgat tgcgcccta cagagttgag gtggaactac 1560
tccttactcc ccgccctgct cagccatcat tgtcctgccc accccgctgg gaccacacc 1620
agggtctctg ccccctctct gaggcccagg actggcagggt gcctgatgtc accagcagag 1680
gccaccaggg ggtgctgctt ctgcacagaa cagaccagc cccgtgggccc ggcgatgca 1740
ggagcctcca tcccctcggc cccctcccat cccctccac actggtcccc accggcctc 1800
tctgctgctc ccaggccat ccgctctctg cgggttctc ccgtctcca cctctgctct 1860
caccctccct cctcagccct ggatcatctg gagcttttgc cccaagtgtt tgctttggag 1920

```

agagagagag agagacagag agagagagag agacagagag agagagaaaag agagagagag 1980
 aaaagaaaag aaggaaagaa ggaaggaagg aaaaagaaaag aaagaaagaa agaaagaaaag 2040
 aaagaaagaa agaaagaaaag aaaggaagga agaaaggaag gaaaaaaaaa aaaaaaaaaa 2099

<210> 2
 <211> 289
 <212> PRT
 <213> human

<400> 2
 Met Phe Gly Leu Phe Gly Leu Trp Arg Thr Phe Asp Ser Val Val Phe
 1 5 10 15
 Tyr Leu Thr Leu Ile Val Gly Leu Gly Gly Pro Val Gly Asn Gly Leu
 20 25 30
 Val Leu Trp Asn Leu Gly Phe Arg Ile Lys Lys Gly Pro Phe Ser Ile
 35 40 45
 Tyr Leu Leu His Leu Ala Ala Asp Phe Leu Phe Leu Ser Cys Arg
 50 55 60
 Val Gly Phe Ser Val Ala Gln Ala Ala Leu Gly Ala Gln Asp Thr Leu
 65 70 75 80
 Tyr Phe Val Leu Thr Phe Leu Trp Phe Ala Val Gly Leu Trp Leu Leu
 85 90 95
 Ala Ala Phe Ser Val Glu Arg Cys Leu Ser Asp Leu Phe Pro Ala Cys
 100 105 110
 Tyr Gln Gly Cys Arg Pro Arg His Ala Ser Ala Val Leu Cys Ala Leu
 115 120 125
 Val Trp Thr Pro Thr Leu Pro Ala Val Pro Leu Pro Ala Asn Ala Cys
 130 135 140
 Gly Leu Leu Arg Asn Ser Ala Cys Pro Leu Val Cys Pro Arg Tyr His
 145 150 155 160
 Val Ala Ser Val Thr Trp Phe Leu Val Leu Ala Arg Val Ala Trp Thr
 165 170 175
 Ala Gly Val Val Leu Phe Val Trp Val Thr Cys Cys Ser Thr Arg Pro
 180 185 190
 Arg Pro Arg Leu Tyr Gly Ile Val Leu Gly Ala Leu Leu Leu Leu Phe
 195 200 205
 Phe Cys Gly Leu Pro Ser Val Phe Tyr Trp Ser Leu Gln Pro Leu Leu
 210 215 220
 Asn Phe Leu Leu Pro Val Phe Ser Pro Leu Ala Thr Leu Leu Ala Cys
 225 230 235 240
 Val Asn Ser Ser Ser Lys Pro Leu Ile Tyr Ser Gly Leu Gly Arg Gln
 245 250 255
 Pro Gly Lys Arg Glu Pro Leu Arg Ser Val Leu Arg Arg Ala Leu Gly
 260 265 270
 Glu Gly Ala Glu Leu Gly Ala Arg Gly Gln Ser Leu Pro Met Gly Leu
 275 280 285
 Leu

<210> 3
 <211> 5303
 <212> DNA
 <213> human

<400> 3
 cgctccctctg ggcccctggt gactgcaggt gaccggagct accaactaag ttaatagaga 60
 tgctccaagt ttcacccaaa caggaggcag gtaagcaagg cttggaggcg ctagttagaa 120
 caggcctgga cggtgacctc gaggaggagg tggagccacg ccgggagagg tcccggagag 180
 cagcagagac cggcatctgg ggtggaggca gcaggcaggg tggcaggagg tgcccctggg 240
 ccacacgagc tgtgccgtgc gtttcaaagc catgtgcaac aatgccggga gctgggtagc 300
 cggcatgccc cggcccagag ctgcacaggc tgctggtttg tgttcagaa gtgatttccg 360
 aagccaagtc catttgagaa acgtgaagag aaacagcatc tttacgaccg aatttctggc 420

ttgagctgac acactctgca cccaccccag gagctgocga acgcagtcgg tggaaaggaaa 480
 aactctctta tcaagaacat cttactgtcc agtggttctgc agagcaaaaat gcggaaaaca 540
 atccgtcaac ggctgggagc ccaggetctg agggcaaacag atggacagtg aggggggtctc 600
 cttaggaagg ggtggggggc gccctgttag accaccaca gacgcagtg tgggtgaca 660
 cattaggaga gccagtccca agcctgggaa cccaggcttc gagctgtcca taaacggta 720
 gtccagttag aggggaaac caaggcacag agaggttggg ttattctctc aaggccacac 780
 agccaggccc ctgtccacag ttctaggccc cctgtgccc cagccaacca aagactagca 840
 ggagcagccc aacatcagtg gactaccgg aaggtggccc agcctcactg ggcacagagg 900
 gtctccatc tccttctccc tcccctccat gcaccagggt aggtgggagc gggtgaaag 960
 cactgacctg gtgccttcag ttcccctgcc agaggggctg aggaggaagg aaaggcctgc 1020
 ccacactgac aactttttgt tttagctcat tctgtgaaca gtgccacacc 1080
 ctgtcacttc tagggaactg gccgggtgcc ctccgtccag ccagctgggt ccagaggag 1140
 atgaaaggat ttagggacca ggaggagcct ccaggcaatg agcgtggtc tgagcacaga 1200
 agggcctgag agagctgggg gcagcctcgt cacaggtgag aaggctccgc ctgccccac 1260
 tgcaccccc aggtgaggca ctggcacct cccctgccct tccgcagtg ccctgtctc 1320
 cagcctggct ccccggtgga tggggtggc tgggatgttt ggatttctt tgggaaggga 1380
 aggcagggcg gtagcgggt gggaattccg gctaagttga aatgtgtgga caccagagct 1440
 caagggcgtc gctggagcgg gccgagcagg tgaacctgt ctgataacaa tgaagatgcc 1500
 ccctcctgcc gccgggccac tcctctctgt aactctctc ccaggaatag agcctaacct 1560
 caccocagca tgggggctg agcgcggcc tcaccccag cctcagactc tgagccagag 1620
 ccagagcaca gccaggagag tctggtttcc tccctgagga aattgaggct cagagagggg 1680
 aggtgtgttt ccaccccac acccccacc cagtcaccac actcaccacc acaagccagg 1740
 gcccgccga ggtctcccac tgcccctggt tgggctcgga ggacaggcca ccatcacctg 1800
 gtgccagcc acggcaaggc acaggagcca aggcocgagc cgcactcctc acccctgac 1860
 taggggcaat gacaagcaca aagaagggtg cgggggatgc tccggcatga ggcagggggc 1920
 tcacgggggt cggaggaagc tgcaggggtg tgcaggcccc tgaaggag caaaacctg 1980
 aagtaggggt ggacctggg tcaggggaa aggtattcag ggtgtctg gcaggggaca 2040
 caaccaaagc aaagatgtga cagggaggcc tccagggtg tgttgggggg ctgatggct 2100
 ctctctctgt ctgtctctc ctctctcaca cacacacagt ctgtctctc tctgttctc 2160
 tgcgtctatc tctggctcta tatctctgta gctctctgta tctgactgtc tgactgtc 2220
 tataaatgta tgtctacata tctgtctctc cctgtatatg tggctctcta tatactctt 2280
 atatctccag agtcttttac cgtgacgcc gccactgttg agcctggtcc tgatgcatc 2340
 ccccaggact aagaacatga ccccatgtgg atcctggggg aggagggatg gctggaaagg 2400
 gaagggctgc tctgagacc tctgccttc agcgcagctt ccgggtggca atctctctc 2460
 cactctgaga acatcctcgg gactctcagc ctogetcctt gcaagtggcc acttgtatc 2520
 gccctcccc ttgtgttagt ggctcctggg agagggcccc tggccctgcc accatcaaca 2580
 aagccctct ctggcttga ggcttcagga gcagcctcct gagtgaccg ccggcagcat 2640
 gagccocggg atcagcttc tgccttaggc aggagtgcac agggagggtc gggggctccc 2700
 tccagctcct cacaccgat cttcccctc cctcctcct cagaccacct gcagcctctg 2760
 cctttgggca cctcctgga cacctccaga gaccctcgg ggaggggact tccagcagc 2820
 gctggcttag gctcaccagg tgggagcccc atgagggctc accgtggaga agctcccacc 2880
 tctgggggaa gaggaaggcg ccacacgtgg gttcccctg cacgttccct ctgtctgag 2940
 acccttgtct ggcacacggt ggaggggtgca gcateaccct caccacatc atgagggaca 3000
 tggacatctt ttcagcaagt tagtccacc tgcaggagc ccagggectg ccctcaggta 3060
 tggggaggca ttaggcacgc ctgacctgag gtgcagaggg ctgagaccg gctgaggctc 3120
 cgccgctggg aaggtgtgtc caggggctgc accagaggg cacctgcagg gccatggatg 3180
 ccactccctg tgccccagc gggggcacaa aagactcctg aggacctgt ctgtctccat 3240
 ctgacaggtc cctgcagc accctctgag ccggccagga tgtttgggct gttcggcctc 3300
 tggagaacct tgcacagtgt ggtcttctac ctgacgtga tctgtggcct cgggggaccg 3360
 gtaggtaacg ggctggtgct ctggaacctc ggcttccgca tcaagaaggg ccccttctc 3420
 atctacctgc tgcacctgac cgcgcggac ttctgttcc tctcctgccc tgtgggctt 3480
 tccgtggetc aggtgcctt gggcgcccag gacacactc acttctgtct cacttctctg 3540
 tggttcggcg tgggctctg gctgctggcg gccttcagcg tggagcgtg cctctccgac 3600
 ctcttcccc cctgetacca gggctgccc cccagacag cctcgccgt cctctgccc 3660
 ctggtgtaga ccccgacct gccggcctg ccctgccc ccaacgctg cggcctgctg 3720
 cgcaacagc cgtgcccct ggtctgccc cgtaccacg tggccagcgt cacctggtt 3780
 ctggtgctgg cccgctgccc ctggacggct ggctgtgtc tctttgtctg ggtgacctg 3840
 tgcctcact gcccgcgcc caggctctac ggcctctcc tggcgcgct gctcctgctc 3900
 ttcttctgt gcctgccc ggtcttctac tggagcctg agcccctgt gaacttctg 3960
 ctgcccgtgt tttcccctt ggccacgct ctggcctg tcaacagcag ctccaagccc 4020
 ctcatctact cggggttgg ccgacagccc gggagcggg agccgctgag gtcggtagct 4080
 cggagggccc tgggggagg gcgagcgt ggtgcccagg gacagtcct gccatgggt 4140
 ctctataag tgggcttgg ccccccacag ggcctgccc gaggtgccc cccccaccga 4200

```

ccctcgctca cccacacccc agatgctttc aggaccagga ggagctctta ccctgagcac 4260
ccaagggctg gacacagctg gagagacttg gccctgacca cccccagcc aggcctgctg 4320
aggatgaggg aggcagaaaa tggagctgga gagaggctcag gcaaggagag agaaaggaga 4380
agcctcctga ataggggtga ggacaggcac cacgccccca ggcccagccc agatcccctt 4440
accggccccc tccccaccct gctgcacctg agtcacaggg gagaaaactg cacaataaaa 4500
cagagccagc caccagccca cagtggccgg attggaacct aggcttccgg actcctgggg 4560
taggtgggcg ccgtccatgc cacctgctgg ctgaggctct gattcgcccc ctacagagtt 4620
gaggtggaac tactccttac tccccgcctt gctcagccat cattgtcctg cccaccccgc 4680
ggggaccaca cccagggtct tgtccccctc tctgaggccc aggactggca ggtgcctgat 4740
gtcaccagca gaggccacca ggtggtgctg cttctgcaca gaacagacct agccccgtgg 4800
gccggcggat gcaggagcct ccatcccctc ggtcccctcc catcccctcc cacactgggtc 4860
cccacccggc ctctcctgcg tccccagggc catccgctct ctgcgggttg ctcccgtctc 4920
ccacctctgc tctcaccctc cctcctcagc cctggatcat ctggagcttt tgcccccaagt 4980
gtttgctttg gagagagaga gagagagaca gagagagaga gagagacaga gagagagaga 5040
aagagagaga gagaaaagaa aagaaggaaa gaaggaagga aggaaaaga aagaaagaaa 5100
gaaagaaaga aagaaagaaa gaaagaaaga aaggaaggaa gaaaggaagg aaaaagaaag 5160
aaagaaagca agaaagcaag aaagcaagaa agaaagaaag aaagaaagaa agaaagagaa 5220
gagaagagaa gacataaggg gaggggaggg gaagaaggaa aggaaagaaa ataatcctag 5280
gaaggaagga aggaatgaag ggg                                     5303
    
```

```

<210> 4
<211> 285
<212> PRT
<213> rat
    
```

```

<400> 4
Gln Ala Val Thr Asn Tyr Ile Phe Leu Leu Leu Cys Leu Cys Gly Leu
 1          5          10          15
Val Gly Asn Gly Leu Val Leu Trp Phe Phe Gly Phe Ser Ile Lys Arg
 20          25          30
Thr Pro Phe Ser Ile Tyr Phe Leu His Leu Ala Ser Ala Asp Gly Ile
 35          40          45
Tyr Leu Phe Ser Lys Ala Val Ile Ala Leu Leu Asn Met Gly Thr Phe
 50          55          60
Leu Gly Ser Phe Pro Asp Tyr Val Arg Arg Val Ser Arg Ile Val Gly
 65          70          75          80
Leu Cys Thr Phe Phe Ala Gly Val Ser Leu Leu Pro Ala Ile Ser Ile
 85          90          95
Glu Arg Cys Val Ser Val Ile Phe Pro Met Trp Tyr Trp Arg Arg Arg
100          105          110
Pro Lys Arg Leu Ser Ala Gly Val Cys Ala Leu Leu Trp Leu Leu Ser
115          120          125
Phe Leu Val Thr Ser Ile His Asn Tyr Phe Cys Met Phe Leu Gly His
130          135          140
Glu Ala Ser Gly Thr Ala Cys Leu Asn Met Asp Ile Ser Leu Gly Ile
145          150          155          160
Leu Leu Phe Phe Leu Phe Cys Pro Leu Met Val Leu Pro Cys Leu Ala
165          170          175
Leu Ile Leu His Val Glu Cys Arg Ala Arg Arg Arg Gln Arg Ser Ala
180          185          190
Lys Leu Asn His Val Val Leu Ala Ile Val Ser Val Phe Leu Val Ser
195          200          205
Ser Ile Tyr Leu Gly Ile Asp Trp Phe Leu Phe Trp Val Phe Gln Ile
210          215          220
Pro Ala Pro Phe Pro Glu Tyr Val Thr Asp Leu Cys Ile Cys Ile Asn
225          230          235          240
Ser Ser Ala Lys Pro Ile Val Tyr Phe Leu Ala Gly Arg Asp Lys Ser
245          250          255
Gln Arg Leu Trp Glu Pro Leu Arg Val Val Phe Gln Arg Ala Leu Arg
260          265          270
Asp Gly Ala Glu Pro Gly Asp Ala Ala Ser Ser Thr Pro
275          280          285
    
```