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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2002 (03.01.2002)

PCT

(10) International Publication Number
WO 02/00001 A2

(51) International Patent Classification: Not classified

(21) International Application Number: PCT/EP01/07533

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

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/215,777 30 June 2000 (30.06.2000) US

(71) Applicant *(for all designated States except US)*: **INGENIUM PHARMACEUTICALS AG** [DE/DE]; Fraunhoferstrasse 13, 82152 Martinsried (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WATTLER, Frank** [DE/DE]; Bennostrasse 11a, Stockdorf 82131 (DE). **WATTLER, Sigrid** [DE/DE]; Bennostrasse 11A, 82131 Stockdorf (DE). **TROMMLER, Paul** [DE/DE]; Randeckstrasse 4, 81375 München (DE). **NEHLS, Michael, C.** [DE/DE]; Paul-Keller-Strasse 6, 82131 Stockdorf (DE).

(74) Agent: **DÖRRIES, FRANK-MOLNIA & POHLMAN**; Postfach 221661, 80506 Munich (DE).

(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, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

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

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

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

WO 02/00001 A2

(54) Title: HUMAN G PROTEIN-COUPLED RECEPTOR IGPcR20, AND USES THEREOF

(57) Abstract: A novel human G protein-coupled receptor (GPcR) protein, IGPcR20, is identified and characterized. IGPcR20-encoding nucleotides, IGPcR20 proteins and fusion proteins, antibodies to the receptor, host cell expression systems, animal models in which the IGPcR20 gene is mutated, recombinant knock-out animals that do not express IGPcR20 and transgenic animals that express an IGPcR20 transgene are encompasssed by the invention, as are compounds that modulate gene expression or receptor activity of IGPcR20 and their use for drug screening, and diagnosis or treatment of diseases and disorders, particularly cancer, reproductive disorders and infertility related to epididymal dysfunction.

Human G protein-coupled receptor IGPcR20, and uses thereof

Field of the Invention

The present invention relates to the field of cellular and molecular biology, protein biochemistry, and pharmacology. The invention relates particularly to the identification of the polynucleotide sequence of a novel G protein-coupled receptor (GPcR) and the characterization of nucleic acids that encode this G protein-coupled receptor, which is referred to herein as IGPcR20. The invention further relates to animal orthologs of the human gene encoding IGPcR20, to expression of both human and animal proteins, to the function of the gene product and to uses for the receptor, and its ligands in drug screening and in diagnosing, preventing and treating disease, particularly pain, cancer, metabolic and inflammatory disorders and reproductive disorders and infertility, especially those related to testicular and epididymal dysfunctions. Animal models of such diseases and dysfunctions, in which the IGPcR20 gene is mutated, knocked-out or present in the form of a transgene, are also incorporated within the invention.

Background of the Invention

It is well established that many medically significant biological processes are mediated by proteins that participate in signal transduction pathways involving G proteins and second messengers; *e.g.* cAMP, diacylglycerol and inositol phosphates (Lefkowitz, 1991, Nature, 351:353-354). Herein these proteins are referred to as proteins participating in pathways with G protein-coupled receptors, either as the receptors themselves, such as those for adrenergic agents and dopamine (Kobilka, BK, *et al.*, 1987, P.N.A.S., USA, 84:46-50; Kobilka BK *et al.*, 1987, Science,

238:650-656; Bunzow JR, *et al.*, 1988, Nature, 336:783-787), or as the G proteins to which the receptors are coupled, or as effector proteins, *e.g.* adenylate cyclase, protein kinase A and protein kinase C (Simon MI, *et al.*, 1991, Science, 252:802-808).

Upon hormone binding to a GPcR the receptor interacts with the heterotrimeric G protein and induces the dissociation of GDP from the guanine nucleotide-binding site. At normal cellular concentrations of guanine nucleotides, GTP fills the site immediately. Binding of GTP to the alpha subunit of the G protein causes the dissociation of the G protein from the receptor and the dissociation of the G protein into alpha and beta-gamma subunits. The GTP-carrying form then binds to the generator of an intracellular second messenger: in one common form of signal transduction, activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the intrinsic GTPase activity of the G protein alpha subunit, returns the G protein to its basal, inactive form. The GTPase activity of the alpha subunit determines the time period during which the G protein is active. The GDP-bound form of the alpha subunit (alpha.GDP) has high affinity for the beta-gamma subunit complex and subsequent re-association of G protein subunits alpha.GDP with beta-gamma returns the G protein to the basal state. Thus the G-protein serves a dual role: as an intermediate that relays the signal from receptor to effector (in this example adenylate cyclase), and as a timer that controls the duration of the signal.

Examples of members of the G protein-coupled receptor family gene family include acetylcholine, adenosine, adrenergic, bradykinin, cAMP, calcitonin, capsaicic, CCK, CGRP, CRF, cytomegalovirus, dopamine, endothelial differentiation gene-1, endothelin, FSH, galanin, histamine, kinin, motilin, muscarinic, neurokinin, neuropeptideY, neurotensin, nociceptin, odorant, opsin, rhodopsin, serotonin, somatostatin, thrombin, TSH and VIP receptors. Alteration of GPcR genes and gene products can cause medical disorders, dysfunctions, or diseases hereafter generally referred to as "diseases". The mechanism of disease may be due to a loss of receptor function or by constitutive receptor activation (reviewed by Coughlin *et al.*, 1994,

Curr. Opin. Cell Biol., 6:191-197). For example, activating mutations of rhodopsin receptor have been found in *retinitis pigmentosa* and congenital night blindness (Rao *et al.*, 1994, Nature 367:639-642); mutations of TSH receptor have been detected in sporadic and inherited hyperthyroidism (Parma *et al.*, 1993, Nature 365:649-651) and nephrogenic *diabetes insipidus* (Holtzman *et al.*, 1993, Hum. Mol. Genet. 2:1201-1204); mis-sense mutations in the luteinizing hormone receptor (LHR) gene, leading to constitutive activation of the LHR, have been shown to be associated with a condition in boys called familial male-limited precocious puberty (Cocco *et al.*, 1996, Hum. Mut., 7:164-166; Kosugi *et al.*, 1995, Hum. Mol. Genet., 4:183-188). Moreover, dopamine receptors are known to bind neuroleptic drugs used for treating disorders of the central nervous system (CNS).

As a characteristic feature, G protein-coupled receptors exhibit seven transmembrane domains which are connected by three hydrophilic extracellular loops alternating with three intracellular loops. Most G protein-coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. The seven transmembrane domains or regions are designated as TM1, TM2, TM3, TM4, TM5, TM6 and TM7. The cytoplasmic loop which connects TM5 and TM6 may be a mayor component of the G protein binding domain.

Most G protein-coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxyl terminus. For several GPcRs, such as the beta-adrenergic receptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

It has also been shown that certain G protein-coupled receptors, *e.g.* the calcitonin receptor-like receptor, might interact with small single pass membrane proteins called receptor-activity-modifying-proteins (RAMPs). This interaction of the GPcR with a certain RAMP determines which natural ligands have relevant affinity for the GPcR-RAMP combination and regulate the functional signaling activity of the

complex (McLathie LM, *et al.*, 1998, Nature, 393:333-339).

For some receptors, the ligand binding sites of the G protein-coupled receptors are believed to comprise hydrophilic sockets formed by several GPcR transmembrane domains, said sockets being surrounded by hydrophobic residues of the G protein-coupled receptors. The hydrophilic side of each GPcR transmembrane helix is thought to face inward and form a polar ligand-binding site. TM3 has been implicated in several G protein-coupled receptors as having a ligand-binding site, such as the TM3 aspartate residues. TM5 serine residues, and TM6 asparagine and TM6 or TM7 phenylalanine or tyrosine residues are also implicated in ligand binding. G-protein coupled receptors bind to a variety of ligands ranging from small biogens to peptides, small proteins and large glycoproteins (Strader CD, *et al.*, 1994, Annu. Rev. Biochem., 63:101-132).

G protein-coupled receptors can be coupled intracellulularly by heterotrimeric G proteins to various intracellular enzymes, ion channels and transporters (see Johnson *et al.*, 1989, Endoc. Rev., 10:317-331). Different G protein alpha-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G protein-coupled receptors has been identified as an important mechanism for the regulation of G protein coupling of some G protein-coupled receptors. G protein-coupled receptors are found in numerous sites within animal, and particularly mammalian hosts.

Evolutionary analyses suggest that the ancestor of G protein-coupled receptors originally developed in concert with complex body plans and nervous systems. With the exception of the visual opsins, the genes for the GPcR family have, in most instances, been characterized by a lack of introns within their coding sequences thus precluding the generation of receptor diversity through alternative splicing. Recent data support the idea that dimerization of G protein-coupled receptors is important in different aspects of receptor biogenesis and function. When considering, for example, the nervous system, the existence of homodimers and heterodimers of

neurotransmitter G protein-coupled receptors offers an attractive explanation of the great diversity and plasticity that is characteristic of such a highly organized and complex system (see Bouvier M, 2001, Nature Rev. Neuroscience, 2:274-286).

In order to understand the role of particular G protein-coupled receptors in normal physiology and disease, knock-out mice have been generated in which the endogenous genes encoding these receptors have been individually targeted. Studies of mas-protooncogene knock-out mice indicate that this GPcR is a determinant of heart rate and blood pressure variability (Walther *et al.*, 2000, Braz J, Med. Biol. Res., 33:1-9). Male mice showed increased anxiety, indicating a function for mas, which is an angiotensin receptor acting in the central nervous system (CNS) (Walther *et al.*, 1998, J. Biol. Chem., 273:11867-11873). Incerti *et al.*, (2000, Hum. Molec. Genet., 9:2871-2788) generated and characterized mice deficient in Oa1 (ocular albinism)-deficient mice by gene targeting. Ophthalmologic examination showed hypo-pigmentation of the ocular fundus in mutant animals compared with wildtype. Also demonstrated was a misrouting of the optic fibers at the chiasm and the presence of giant melanosomes in retinal pigment epithelium, as observed in OA1 patients. Prostaglandin E2 receptor knock-out mice show a mild change in renal water handling, while EP2 receptor knock-out mice display salt-sensitive hypertension (Breyer *et al.*, 2000, Curr. Opin. Nephrol. Hypertens., 9:23-29).

Based on malfunctions discovered in signaling pathways several drugs have been developed, for example, a compound that blocks the farnesylation of ras as a tumor inhibitor, a JAK-2 blocker as an inhibitor of recurrent pre-B cell acute lymphoblastic leukemia, and a platelet-derived growth factor receptor kinase as a blocker of restenosis (Reviewed in Levitzki A, 1996, Curr. Opin. Cell Biol., 8:239-244). G protein-coupled receptors have been identified and successfully used as targets for several existing drugs; for example, dopamine and serotonin G protein-coupled receptors have been targeted for CNS diseases, angiotensin, muscarinic and adrenergic receptor G protein-coupled receptors have been targeted for cardiovascular diseases, histaminic G protein-coupled receptors have been targeted

for respiratory diseases, the prostaglandin GPcR has been targeted for opthalmic purposes, and calcitonin and estrogen for treatment of arthritis.

The following information relating to mammalian reproductive function is provided in relation to the particular human G protein-coupled receptor, and its animal orthologs, that are disclosed by the present invention. Fertilization in mammals depends on a sequence of events that culminates in the activation of an oocyte by a spermatozoon. Mammalian spermatozoa are highly differentiated by the time they leave the testis. Nevertheless, at that stage they do not have the ability to move progressively or to interact with an oocyte: they gain these abilities while passing through the excurrent ducts (vasa efferentia, epididymis, and vas deferens; Robaire B and Hermo L, in: Knobil E, Neill J (eds), 1988, "*The Physiology of Reproduction*", Raven Press, New York, 999-1080).

The functions of the epididymis include absorption of seminiferous fluid, and the maturation, concentration, transport, and storage of spermatozoa. The epididymis is usually divided into three anatomical regions: caput, corpus, and cauda segments. Caput and corpus epididymis are involved in the acquisition of sperm fertilizing ability; the cauda segment is specialized in sperm storage. Human spermatozoa acquire their fertilizing ability when they reach the corpus epididymal portion (Badford JM, 1988, Ann. NY Acad. Sci., 541:284-291). Based upon differential expression of orthologous genes and upon differences in the regionalization of epididymal protein synthesis between species, expression and functional data cannot be converted easily from mouse or rat to human (Turner TT, 1995, J. Androl., 16:292-298).

Furthermore, at least two sperm surface proteins are involved in the fusion of egg and sperm in rodents. One of these is fertilin, the second being the rodent membrane protein AEG (acidic epididymal glycoprotein). However, there is no evidence for AEG in humans.

Several G protein-coupled receptors have been described as being expressed in primary sexual organs: Osterhoff described a human epididymis-specific GPcR, HE6, detected by *in situ* analysis within the epithelial cells lining the epididymal duct (Osterhoff C, *et al.*, DNA Cell. Biol., 1997, 16:379-389). The expression of DAX1, an orphan nuclear hormone GPcR (Zanari E, *et al.*, 1994, Nature, 372:635-641) in Sertoli cells of the testis is regulated during spermatogenesis and may have influence on the development of spermatogenic cells in response to steroid and pituitary hormones (Tamai KT, *et al.*, 1996, Molec. Endocr., 10:1561-1569). Loss of DAX1 results in adrenal hypoplasia and hypogonadotropic hypogonadism; increased DAX1 leads to dosage-sensitive reversal and a female phenoptye or ambiguous genitalia in XY-genotypic males (see McCabe ERB, 1996, J. Clin. Invest., 98:881-882). A lutotropin-choriogonadotropin G protein-coupled receptor, LHCGR, is expressed in testis, placenta and ovary. In males, loss of LHCGR function causes pseudo-hermaphroditism associated with Leydig cell hypoplasia, which supports the concept that a functional receptor is necessary for the early development of Leydig cells (see Kremer *et al.*, 1999, J. Clin. Endocr. Metab., 84:1136-1140). Leydig cells mainly produce the masculinizing hormone testosterone.

Olfactory receptors constitute the largest family among G protein-coupled receptors, with up to 1000 members anticipated (Vanderhaeghen P, *et al.*, 1997, Genomics, 39:239-246). About 10% of mammalian odorant G protein-coupled receptors are transcribed in testis and several testis-expressed odorant receptor (TOR) proteins have been detected in mature sperm cells, suggesting their potential implication in the control of sperm maturation, migration and/or fertilization (Branscomb A, *et al.*, 2000, Genetics, 156:785-797).

Because of the vital role of G protein-coupled receptors in the communication between cells and their environment, such receptors are attractive targets for therapeutic intervention. G protein-coupled receptors have led to more than half of the currently known drugs (Drews, Nature Biotechnology, 1996, 14:1516). Mechanistically, approximately 50% to 60% of all clinically relevant drugs act by

modulating the functions of various G protein-coupled receptors, as either agonist (activating activity) or antagonist (blocking activity) of a GPcR (Cudermann *et al.*, 1995, J. Mol. Med., 73:51-63). This indicates that these receptors have an established, proven history as therapeutic targets.

In consequence, there is a continuing medical need for identification and characterization of further receptors that can play a role in diagnosis, preventing, ameliorating or correcting dysfunctions, disorders, and diseases. Included among such diseases are the visual diseases, particularly those visual diseases related to malfunctions in the nervous system, which in addition to their direct effects in hindering visual acuity, and promoting blindness, may also be associated with debilitating effects upon limb and body movement or with psychological disabilities.

Summary of the Invention

The G protein-coupled receptor of the present invention, IGPcR20, is especially useful for diagnosis, prevention, amelioration or correction of diseases associated with signal processing in male reproductive tissues, such as male infertility. In particular, IGPcR20 satisfies a need in the art for identification and characterization of further receptors that can play an important role in diagnosis, preventing, ameliorating or treatment of, *inter alia*, pain, cancer, inflammatory disorders and metabolic disorders linked to reproductive tissues, especially epididymis.

Embodiments of the invention include an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO:1, or any unique fragment thereof, particularly wherein the nucleotide sequence of the fragment is greater than ten base pairs in length. Embodiments also include an isolated polynucleotide which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or any unique fragment thereof, particularly wherein the amino acid sequence of the fragment is greater than ten amino acids in length.

Embodiments of the invention include any isolated nucleic acid molecule or polynucleotide comprising an allelic variant of a nucleotide sequence or polynucleotide which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein said allelic variant retains at least 70% nucleic acid homology, or in increasing preference at least 80%, 85%, 90%, 95% or 98% nucleic acid homology and hybridizes to the complement of SEQ ID NO:1 under stringent conditions (Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York): also included are such isolated nucleic acid molecules or polynucleotides that comprise a nucleotide sequence which encodes at least one of the group of polypeptides, peptides and fusion proteins, comprising an amino acid sequence at least 70% similar, or in increasing preference at least 75%, 80%, 85%, 90%, 95% or 98% similar, to SEQ ID NO:2.

Vectors comprising an isolated nucleic acid molecule or polynucleotide of the invention as previously described are a further embodiment of the invention.

Additional embodiments include host cells genetically engineered to contain such a vector or genetically engineered to contain such a nucleic acid molecule or polynucleotide of the invention as described above, and particularly wherein the nucleic acid molecule or polynucleotide of the invention is operatively linked with a nucleotide regulatory sequence that controls expression of said nucleic acid molecule or polynucleotide in the host cell. Also included are host cells which are drawn from prokaryotic bacterial cells, or from eukaryotic cells, particularly or yeast, insect or mammalian cells, preferred embodiments employing a mammalian host cell being those in which the host cell is a CHO, BHK, COS, CV1, 293, fibroblast or VERO cell. Embryonic stem cells containing a disrupted endogenous IGPcR20 gene are also preferred embodiments of the invention, the most preferred embryonic stem cells being derived from mice.

Preferred embodiments of the invention include antibodies to the IGPcR20 protein,

polypeptides, peptides, isolated domains and fusion proteins.

Agonists and antagonists of IGPcR20 are preferred embodiments of the invention, including: (a) 'small molecules' of molecular mass less than 6 kDa; (b) molecules of intermediate size, having molecular mass between 5 kDa to 15 kDa; and (c) large molecules of molecular mass greater than 12 kDa; the latter including mutant natural IGPcR20 ligand proteins that compete with native natural IGPcR20 ligand and which modulate IGPcR20 gene expression or gene product activity. Preferred embodiments of the invention are those wherein such molecules bind specifically to the IGPcR20 receptor or to the IGPcR20 gene. Further embodiments are methods of identifying such compounds which modulate the activity of the IGPcR20 receptor or of IGPcR20 gene expression, such as anti-sense and ribozyme molecules that can be used to inhibit IGPcR20 gene expression, or expression constructs that are capable of enhancing IGPcR20 gene expression.

The non-human animal orthologs of the human sequence in SEQ ID NO:1 are preferred embodiments of the invention, particularly ungulate and rodent sequences, and especially those of rat and mouse, and also polynucleotides comprising these sequences or homologous or partially homologous sequences as indicated for the human nucleic acid and polynucleotide. Preferred embodiments include polynucleotides of such non-human animal orthologs comprising a nucleotide sequence which encodes a polypeptide comprising an amino acid sequence at least 70% similar, or in increasing preference at least 75%, 80%, 85%, 90%, 95% or 98% similar, to SEQ ID NO:2; or being at least ten amino acid residues in length and bearing the stated similarity to a unique part of SEQ ID NO:2.

Embodiments of the invention include knock-out animals which are non-human animals and which do not express IGPcR20. Preferred embodiments are those wherein the endogenous animal ortholog is functionally disrupted by homologous recombination methods such as conditional knock-out and/or null allele knock-out of the IGPcR20 gene. Mutated animals that express a non-functional or partially

functional form of IGPcR20 are further embodiments of the invention. Embodiments of the invention also include progeny of the non-human animals described as being embodiments of the invention, the term 'progeny' including both heterozygous and homozygous offspring. Further embodiments are non-human transgenic animal models expressing the human IGPcR20 cDNA sequence as shown in SEQ ID NO:1 or a modification thereof as described above, operatively linked to a nucleotide regulatory sequence that controls expression of the nucleic acid molecule in the host animal. Particularly preferred embodiments are those non-human animals (also termed animal models) in which the human IGPcR20 is encoded by a nucleic acid sequence which is homozygous in the animal model. In each embodiment of the invention comprising a non-human animal, preferable embodiments are those wherein the non-human animal is a mammal, particularly ungulate or rodent, and preferably wherein the non-human animal is from a genus selected from the group consisting of *Mus* (*e.g.*, mice), *Rattus* (*e.g.*, rats), *Oryctologus* (*e.g.*, rabbits) and *Mesocricetus* (*e.g.*, hamsters), mouse being the most preferable of this group.

Embodiments of the invention include primary cells and cell lines derived from any of the non-human animals of the invention, particularly the non-human transgenic animal models of the invention. Further embodiments include the amino acid sequence of those non-human animal orthologs of IGPcR20 that comprise an amino acid sequence at least 70% similar, or in increasing preference at least 75%, 80%, 85%, 90%, 95% or 98% similar, to the sequence of the mouse ortholog provided (SEQ ID NO:7); or a part of said non-human animal sequence which is at least ten amino acid residues in length and bears the stated similarity to a unique part of SEQ ID NO:7.

The use of the non-human animal or animal model of the invention, as described above, for the dissection of the molecular mechanisms of the IGPcR20 pathway, for the identification and cloning of genes able to modify, reduce or inhibit the phenotype associated with IGPcR20 activity or deficiency, constitutes a further embodiment of the invention, as does the use of such non-human animal or animal

model for the identification of gene and protein diagnostic markers for diseases, for the identification and testing of compounds useful in the prevention or treatment of symptoms associated with IGPcR20 activity or deficiency, in particular but not limited to central nervous system disorders, including neurologic, psychiatric and behavioral disorders, metabolic disorders, visual and olfactory disorders, and especially in the case of IGPcR20, visual diseases associated with signal processing in the brain, notably in the cerebrum, and particularly in the occipital lobe.

Additional embodiments of the invention include methods of identifying compounds suitable for modulating the activity of the protein or polypeptide of the invention, as described above, for treatment of diseases characterized by aberrant expression or activity of IGPcR20. Preferred embodiments include methods of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20, by the administration of compounds that bind specifically to the IGPcR20 gene or protein and/or which modulate IGPcR20 expression or IGPcR20 activity; the compounds that that bind specifically to the IGPcR20 gene or protein and/or which modulate IGPcR20 expression or IGPcR20 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20; and the use of compounds that that bind specifically to the IGPcR20 gene or protein and/or which modulate IGPcR20 expression or IGPcR20 activity for prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20. Further preferred embodiments are gene therapy methods of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20, by the administration of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7, that modulate IGPcR20 expression or IGPcR20 activity; the vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR20 expression or IGPcR20 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20; and the use of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR20 expression or IGPcR20 activity for

prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20.

Brief Description of the Figures

Figure 1: Fig. 1 depicts the full-length coding DNA (cDNA) sequence of the human IGPcR20 gene (SEQ ID NO:1).

Figure 2: Fig. 2 depicts the amino acid sequence of the human IGPcR20 protein (SEQ ID NO:2).

Figure 3: Fig.3 depicts UV light-visualized PCR products amplified with a human tissue cDNA panel.

Figure 4: Fig.4 depicts human multi tissue Northern blots hybridized with a human IGPcR20 probe.

Figure 5: Fig. 5a depicts a comparison of the amino acid sequence of the human IGPcR20 to the amino acid sequence of the human GPcR hGPR34;

Fig. 5b depicts a comparison of the amino acid sequence of the human IGPcR20 to the amino acid sequence of the human GPcR hGPR17;

Fig. 5c depicts a comparison of the amino acid sequence of the human IGPcR20 to the amino acid sequence of a human GPcR hP2Y-like protein;

Fig. 5d depicts a comparison of the amino acid sequence of the human IGPcR20 to the amino acid sequence of the mouse ortholog of the human IGPcR20.

Figure 6: Fig. 6a shows a hydropathy plot for the predicted amino acid sequence of the human IGPcR20 protein compared to the human G protein-coupled receptor proteins hGPR34, hGPR17 and hP2Y-like protein.

Fig. 6b shows a hydropathy plot for the predicted amino acid sequence of the human IGPcR20 protein compared to the mouse IGPcR20 protein ortholog.

Figure 7: Fig. 7 depicts the full-length coding DNA (cDNA) sequence of the mouse IGPcR20 gene (SEQ ID NO:6).

Figure 8: Fig. 8 depicts the amino acid sequence of the mouse IGPcR20 protein (SEQ ID NO:7).

Figure 9: Fig. 9 depicts mouse multi tissue Northern blots hybridized with a mouse IGPcR20 probe.

Figure 10: Fig. 10 schematically outlines the construction of a mouse IGPcR20 targeting vector based on the method described by Wattler S & Nehls M, German patent application DE 100 16 523.0, "Klonierungssystem zur Konstruktion von homologen Rekombiationsvektoren", filed April 03, 2000, the major aspects of which are incorporated as Example 9.

Detailed Description of the Invention

The present invention relates to the discovery, identification and characterization of nucleic acids that encode the novel human G protein-coupled receptor IGPcR20. The invention encompasses nucleotide sequences encoding mammalian forms of IGPcR20, including human IGPcR20, nucleotides that encode some or all of its functional domains, such as extracellular domains (ECDs), the transmembrane domains (TMs), and the cytoplasmic domains (CDs); mutants of the IGPcR20 sequences, and fusion proteins of IGPcR20. The invention also encompasses host cell expression systems expressing such nucleotides, the host cells and expression products. The invention further encompasses IGPcR20 proteins, fusion proteins, antibodies to the receptor, antagonists and agonists of the receptor, transgenic

animals that express an IGPcR20 transgene, recombinant knock-out animals that do not express the IGPcR20, and animal models in which the IGPcR20 gene is mutated. The invention also encompasses compounds that modulate IGPcR20 gene expression or IGPcR20 receptor activity that can be used for drug screening, or for diagnosis, monitoring, preventing or treating visual dysfunctions associated with signal processing in the occipital lobe of the brain.

The invention further encompasses the use of IGPcR20 nucleotides, IGPcR20 proteins and peptides, as well as antibodies to IGPcR20, antagonists that inhibit ligand binding, receptor activity or expression, or agonists that increase ligand binding, activate receptor activity, or increase its expression, for the diagnosis and treatment of disorders, including, but not limited to treatment of central nervous system disorders. In addition, IGPcR20 nucleotides and proteins are useful for the diagnosis of an IGPcR20 or pathway abnormality, and for the identification of compounds effective in the treatment of disorders based on the aberrant expression or activity of IGPcR20. The invention also relates to host cells and animals genetically engineered to express the human IGPcR20 (or mutants thereof) or to inhibit or knock-out expression of the animal's endogenous IGPcR20 gene.

IGPcR20, as a new G protein-coupled receptor, can play a role in diagnosis, preventing, ameliorating and correcting diseases. These diseases include, but are not limited to, psychiatric and CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (COD), post traumatic stress disorders (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, multiple sclerosis, Alzheimer's disease/dementia and other neurodegenerative diseases, severe mental retardation, dyskinesias, Huntington's disease, Gille de la Tourette's syndrome, tics, tremor, dystonia, spasms, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorders, epilepsy, migraine, attention deficit/hyperactivity disorder (ADHD), cardiovascular diseases, angina pectoris, including heart failure, angina pectoris, arrythmias, myocardial infarction, cardiac

hypertrophy, hypertension, thrombosis, arteriosclerosis, cerebral vasospasm, subarachnoid hemorrhage, cerebral ischenia, thrombosis, arteriosclerosis, peripheral vascular disease, Raynaud's disease, kidney disease – *e.g.* renal failure; dyslipidemias, obesity, emesis, gastrointestinal disorders, including irritable bowel syndrome (IBS), inflammatory bowel syndrome (IBD), diarrhoea, gastresophageal reflux disease (GERD), motility disorders and conditions of delayed gastric emptying, such as post operative or diabetic gastroparesis, and diabetic ulcers; other diseases including osteoporosis; inflammations; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; chemotherapy induced injury; tumor invasion; immune disorders; autoimmune diseases; urinary retention; asthma, allergies; arthritis; benign prostatic hypertrophy; endotoxin shock; sepsis; complication of diabetis mellitus; and gynaecological and reproductive disorders and male infertility.

In particular the new GPcR IGPcR20 satisfies a need in the art for identification and characterization of further receptors that can play an important role in diagnosis, preventing, ameliorating or correcting of, but not limited to pain, cancer, inflammatory disorders and metabolic disorders linked to reproductive tissues, particularly the testis and epididymis. The GPcR of the present invention, IGPcR20, is especially useful for diagnosis, preventing, ameliorating or correcting of reproductive disorders, especially male infertility.

Definitions

As used herein, the following terms, whether used in the singular or plural, have the meanings indicated.

IGPcR20 nucleotides, sequence or coding sequences – encompass DNA, including genomic DNA (e.g. the IGPcR20 gene), cDNA, RNA and include nucleotide sequences encoding IGPcR20 protein, peptide fragments, or fusion proteins.

IGPcR20 – means natural, or mature, IGPcR20 receptor protein. Polypeptides or peptide fragments of IGPcR20 protein are referred to as IGPcR20 polypeptides or IGPcR20 peptides. Fusions of IGPcR20, or IGPcR20 polypeptides or peptide fragments to an unrelated protein are referred to herein as IGPcR20 fusion proteins. ECD – means "extracellular domain" of the receptor protein; TM – means "transmembrane domain" and CD – means "cytoplasmic domain". A functional IGPcR20 refers to a protein which binds natural IGPcR20 ligand with high affinity and specificity *in vivo* or *in vitro*.

Ligand – a molecule that selectively binds to a receptor.

Receptor – a plasma membrane protein which binds one or more appropriate ligands and propagates their regulatory signals to target cells, either by direct intracellular effects, or by promoting the synthesis and/or release of another regulatory molecule known as a second messenger.

Agonist – a molecule, being a ligand and/or drug, that acts on one or more physiological receptors and mimics the effects of the endogenous regulatory compounds; generally these are compounds that activate the receptor.

Antagonist – a molecule, being a ligand and/or drug that inhibits a receptor, most acting by inhibiting the action of an agonist, for example by competing for agonist binding sites on a receptor. These are generally themselves devoid of intrinsic regulatory activity, but act to block receptor activation.

Transgenic animal – a non-human animal containing one or more additional, often foreign genes or "transgenes", integrated into its genome, that can be used as model systems to determine the phenotypic effects of expressing those genes.

Knock-out or knock-out animal – a non-human animal wherein a transgene is

inserted into the genome to create a partial or complete loss-of-function mutation of an endogenous gene. Endogenous genes are inactivated usually by homologous recombination, using replacement or insertion-type gene targeting vectors.

Gene

Novel GPcR genes may be isolated using expression cloning, by synthesizing specific oligonucleotides based on the sequence of purified proteins, using low stringency hybridization (Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York) and by degenerate PCR using known receptor sequences. GPcR genes may also be identified by large scale sequencing, as in the Human Genome Project, followed by analysis of expressed sequence tags (ESTs), or complete sequences present in databases. Known GPcR sequences or conserved regions thereof may be employed as query sequences to extract novel GPcR sequences from these databases.

The present invention provides IGPcR20, a novel G protein-coupled receptor protein described for the first time herein, and characterized as having seven hydrophobic domains which span the plasma membrane and which are connected by alternating extracellular and intracellular hydrophilic loops. IGPcR20 encodes a protein of 336 amino acids (see Fig. 2; SEQ ID NO:2).

A BLASTP search (Basic Local Alignment Search Tool for Proteins, National Institutes of Health, Bethesda MD, U.S.A.) revealed that the proteins most closely related to human IGPcR20 are GPR34, GPR17 and a P2Y-like GPcR. Human IGPcR20 has 46% similarities with conserved substitutions to the human orphan receptor GPR34 (Schoneberg *et al.*, 1999, Biochim. Biophys. Acta, 7,1446:57-70), isolated from a human fetal brain cDNA library. Human IGPcR20 also has similarities to the mouse receptor GPR34 (45%; protein ID number AAD50531.1). Human IGPcR20 is related to human orphan receptor GPcR17 (43%; SwissProt

accession number Q13304) and to a human P2Y-like GPcR (43%; protein ID number CAA73144.1). The sequence of human orphan receptor GPcR17 indicates that it might be a receptor for chemokines, relatively small peptides with potent chemo-attractant activities (Raport *et al.*, 1996, J. Leukoc. Biol., 59:18-23).

The human P2Y-like GPcR is primarily expressed in the brain and is localized to chromosome 2q21 (Blasius *et al.*, 1998, J. Neurochem 70:1357-1365). The 2 kb mRNA of the GPR34 gene is found in several human and mouse tissues, with an additional 4 kb transcript found only in mouse liver and testis (Schoneberg *et al.*, 1999, Biochim. Biophys. Acta 7,1446:57-70). The human P2Y-like GPcR gene is expressed in the brain as an 2.3 kb and 6.3 kb transcript (Blasius *et al.*, 1998, J. Neurochem 70:1357-1365). The P2 receptors play roles in a variety of processes including neurotransmission, cardiac function, platelet aggregation, vascular tone, muscle contraction and relaxation, hormone secretion, immune response and cell growth (reviewed in Chen *et al.*, 1995, J. Neuroendocrinol., 7:83-96).

In particular, the invention encompasses sequences coding for IGPcR20 polypeptides, or functional domains of the IGPcR20, mutated, truncated or deleted forms of IGPcR20, and IGPcR20 fusion proteins. The invention also encompasses nucleotide constructs that inhibit expression of the IGPcR20 gene, such as anti-sense and ribozyme constructs, or enhance expression of IGPcR20 in combination with regulatory sequences such as promoters and enhancers.

The cDNA sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of human IGPcR20 of this invention are shown in Fig. 1 and Fig. 2. The IGPcR20 nucleotide sequences of the invention include the DNA sequence shown in Fig. 1, nucleotide sequences that encode the amino acid sequence shown in Fig. 2 and any nucleotide sequence that hybridizes to the complement of the DNA sequence shown in Fig. 1 under highly stringent conditions (Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York). Functional equivalents of the IGPcR20

gene product include naturally occurring IGPcR20, mutant and degenerate variants present in humans and other species. Preferred IGPcR20 nucleic acids encode polypeptides that are at least 55% identical or similar to the amino acid sequence shown in Fig. 2. Nucleic acids which encode polypeptides which are at least 70%, and even more preferably, in increasing order of preference, at least 80%, 85%, 90%, 95%, or 98% identical or similar. In a particularly preferred embodiment, the nucleic acid of the present invention encodes a polypeptide having an overall amino acid sequence homology or identity of, in increasing order of preference, at least 70%, 80%, 85%, 90%, 95%, 98%, or at least 99% with the amino acid sequence shown in Fig. 2.

The invention also provides DNA molecules that are the complements of the nucleotide sequences described above and which may act as IGPCR20 anti-sense molecules useful in IGPcR20 gene regulation. Orthologs of the human IGPCR20 gene present in other species can be identified and readily isolated. They can be useful for developing cell and animal model systems for purposes of drug discovery. For example, cDNA or genomic DNA libraries derived from the organism of interest can be screened by hybridization using the nucleotides described above, or by performing PCR using degenerate oligonucleotide primers. (See Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, New York, USA; and Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York). Additionally, expression libraries can be screened using standard antibody screening techniques or by doing database searches for homologues and then cloning them based on the sequence. The identified sequences may be sub-cloned and sequenced. The IGPcR20 gene sequences may additionally be used to isolate mutant IGPcR20 gene alleles, or to detect defects in the regulatory sequences of the IGPcR20 using DNA obtained from an individual suspected of or known to carry the mutant IGPcR20 allele. Mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of disorders arising from the aberrant expression or activity of the IGPcR20 protein.

The isolation of human genomic clones is helpful for designing diagnostic tests and therapeutics. For example, sequences derived from the human gene can be used to design primers for use in PCR assays to detect mutations for diagnostics.

The nucleotides of this invention are also preferred for use in mapping the location of the gene to the chromosome, in a process termed chromosomal mapping. Various techniques known to those skilled in the art, including but not limited to *in situ* hybridization of labeled probes to flow-sorted chromosomes, fluorescence *in situ* hybridization (FISH) and PCR mapping of somatic cell hybrids may be employed. This allows the physical location of gene regions to be associated with genetic diseases, based on a genetic map. Genetic linkage analysis can then be used to identify the relationship between genes and diseases (see Egeland *et al.*, 1987, Nature, 325:783-787). Preferred uses of this map include diagnostic tests and reagents, in pharmacogenetics studies and in monitoring patient responses to drugs in clinical trials.

Proteins and polypeptides

Fig. 2 shows the amino acid sequence of the human IGPcR20 protein. The amino acid sequence of IGPcR20 contains hydrophilic domains located between the transmembrane domains, arranging an alternating location of the hydrophilic domains inside and outside the cell membrane. Polypeptides which are at least 70%, and even more preferably at least 80%, 85%, 90%, 95%, 98% or 99% identical or similar to the amino acid sequence represented by Fig. 2 are encompassed by this invention.

In particular, the invention encompasses IGPcR20 polypeptides, or functional domains of the IGPcR20, mutated, truncated or deleted forms of IGPcR20, and host cell expression systems that can produce such IGPcR20 products. IGPcR20 proteins, polypeptides and peptides, can be prepared for the generation of antibodies, as reagents in diagnostic assays, in the identification of other cellular gene products

involved in regulating IGPcR20, as reagents for screening for compounds that can be used in the treatment of conditions involving IGPcR20, and as pharmaceutical reagents useful in the treatment of related disorders.

The invention also encompasses proteins that are functionally equivalent to the IGPcR20 encoded by the nucleotide sequences, as defined by the ability to bind natural IGPcR20 ligand, the resulting biological effect of natural IGPcR20 ligand binding, *e.g.*, signal transduction, a change in cellular metabolism or change in phenotype. Such functionally equivalent IGPcR20 proteins include but are not limited to additions or substitutions of amino acid residues, which result in a silent change. Also preferred in this invention are mutant IGPcR20 proteins with increased function, and/or greater signaling capacity; or decreased function, and/or decreased signal transduction capacity which may be generated by random mutagenesis techniques and site-directed mutagenesis techniques well known to those skilled in the art. The same strategy can also be used to design mutant forms of IGPcR20 based on the alignment of human IGPcR20 and IGPcR20 orthologs from other species. Highly preferred are other mutations to the IGPcR20 coding sequence that can be made to generate IGPcR20 constructs that are better suited for expression, scale up, *etc.* in the host cells chosen. Host cells may be chosen depending on their varying capacity to modify synthesized proteins.

Peptides corresponding to one or more domains of the IGPcR20 (*e.g.*, ECD, TM or CD), truncated or deleted forms of IGPcR20, as well as fusion proteins are also within the scope of the invention and can be designed on the basis of the IGPcR20 nucleotide and IGPcR20 amino acid sequences disclosed above. Such IGPcR20 polypeptides, peptides and fusion proteins can be produced using techniques well known in the art for expressing protein encoding IGPcR20 sequences. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, N.Y.; and Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green

Publishing Associates, Inc., and John Wiley & Sons, Inc., New York.). A variety of host-expression vector systems may be utilized to express the IGPcR20 nucleotide sequences of the invention. The IGPcR20 peptide or polypeptide may be anchored in the cell membrane and purified or enriched from such expression systems using appropriate detergents and lipid micelles, and methods well known to those skilled in the art. Or, where the IGPcR20 peptide or polypeptide is secreted by the cells, it may be isolated from the culture media. Such host cells themselves may be used to assess biological activity, *e.g.*, in drug screening assays.

The expression systems that may be used for purposes of the invention include, but are not limited to microorganisms such as bacteria (*e.g.*, *E. coli, B. subtilis*); yeast (*e.g.*, *Saccharomyces sp., Pichia sp.*); insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus); plant cell systems infected with recombinant viral or plasmid expression vectors; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing mammalian promoters. Lower amounts of functional protein are expressible in *E. coli* and yeast, particularly as *E. coli* do not contain G proteins or effectors. G proteins may be added to *E.coli* expressing G protein-coupled receptors in cell membrane, in the cell-based assays. Yeast cells may be humanized by co-transfixing human G proteins. The yeast *Pichia pastoris* is preferred over *Saccharomyces cerevisiae* for purification of G protein-coupled receptors for structural studies. The most preferred systems for expression are the baculovirus/insect cell and mammalian cell systems, as they can produce the largest quantities of G protein-coupled receptors in functional form for analysis. Mammalian cells are preferred because they express the necessary G proteins, and vaccinia and Semliki Forest virus are preferred as vectors. (See Tate *et al.*, 1996, Tibtech 14:426-430).

Diagnostic and therapeutic reagents and kits

In one embodiment of the invention, the invention encompasses antibodies directed

against IGPcR20 proteins or peptides, or IGPcR20 fusion proteins, as described above. Such antibodies include but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, anti-idiotypic (anti-Id) antibodies, including Fab fragments. The antibodies may be generated and purified, or conjugated according to methods well known in the art. See for example Harlow E and Lane D, 1988, "*Antibodies: A Laboratory Manual*", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, which is incorporated herein by reference in its entirety.

In another embodiment, the antibodies of the invention may be used, for example, as part of a diagnostic or a prognostic, and as a part of compound screening schemes, for the evaluation of the effect of test compounds on expression and/or activity of the IGPcR20 gene product. Preferably, antibodies may be used in therapeutic regimes as a method for the inhibition of abnormal IGPcR20 activity. Also preferred are antibodies directed against wild type or mutant IGPcR20 gene products or conserved variants or peptide fragments thereof to detect the pattern and level of expression, as well as distribution in tissues, of the IGPcR20 in the body, also by *in situ* detection. The antibodies may be employed as part of an enzyme immunoassay (EIA), a radioimmunoassay, or as an antibody labeled with a chemiluminescent or a fluorescent compound.

In yet another embodiment of the invention, the IGPcR20 proteins or peptides, IGPcR20 fusion proteins, IGPcR20 nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant forms of IGPcR20 or inappropriately expressed forms of IGPcR20, for the diagnosis of disorders including but not limited to central nervous system disorders, neurologic, psychiatric and behavioral disorders, metabolic disorders, visual and olfactory disorders, immune, neuroimmune, neuroendocrine and inflammatory disorders and diseases. DNA encoding IGPcR20 or parts thereof may be used in hybridization or amplification assays of biological samples to detect abnormalities involving IGPcR20 gene structure, including point mutations, insertions, deletions and chromosomal

rearrangements. Such genotyping assays may include, but are not limited to Southern analyses, single stranded conformational polymorphism analyses (SSCP), and PCR analyses (See Mullis KB, U.S. Pat. No. 4,683,202), the use of restriction fragment length polymorphisms (RFLPs), of variable numbers of short, tandemly repeated DNA sequences between the restriction enzyme sites (see Weber, U.S. Pat. No. 5,075,217), and by detecting and measuring IGPcR20 transcription.

Also within the scope of the invention are the IGPcR20 proteins or peptides, IGPcR20 fusion proteins, IGPcR20 nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals. These can be used for screening for drugs effective in the treatment of disorders. The use of engineered host cells and/or animals may offer an advantage in that both compounds that bind to the ECD of the IGPcR20 and compounds that affect the signal transduced by the activated IGPcR20 may be identified.

Screening for receptor modulating agents

In another embodiment of the invention, the invention encompasses the pharmacological testing wherein the cloned IGPcR20 genes are expressed in yeast, insect or mammalian cells and screened for a response to cognate or surrogate agonists. The agonists may be present in, but are not limited to, biological extracts, peptide libraries and/or complex compound collections. The invention provides for screening which may utilize libraries of known compounds, including natural products or synthetic chemicals, and biologically active materials, including proteins, for compounds which are inhibitors or activators. Candidate test compounds include all kinds of combinatorial chemistry derived molecular libraries of amino acids, peptides, soluble peptides, modified peptides, antibodies, small organic and inorganic molecules.

In a further embodiment of the invention, a labeled test compound can be incubated with the receptor to determine whether one binds to the other. Functional assays

including fibroblast and BM transformation assays, cell cycle analysis can be performed; as well as responses using signal transduction assays, including protein phosphorylation, guanylate cyclase activity, ion fluxes (*e.g.* calcium) and pH changes can be measured. High throughput drug screening systems are most preferred and may use assays including, but not limited to, the production of intracellular second messengers, such as cAMP, diacylglycerol and inositol phosphates; the activation of reporter gene transcription, such as luciferase and beta-galactosidase under for example the cAMP-responsive element; receptor-mediated actions on adenylyl cyclase and phospholipase C leading also for example to dispersion or aggregation of frog melanophores. (Reviewed in Tate *et al.*, 1996, Tibtech 14:426-430; included in entirety herein).

In a highly preferred embodiment, a functional genomics approach for protein-protein interaction screening may be employed wherein the GPcR is produced in "humanized yeast cells": expression in yeast along with endogenous or promiscuous mammalian or human G-alpha proteins. Transient expression of cDNA can also be carried out using mammalian CHO, HEK-293 cells or COS-7 cells and receptors can be analyzed for ligand binding and drug interactions (for example as described in Fraser *et al.*, 1995, J. Nucl. Med., 36:17S-21S). Also preferred is site-directed mutagenesis to define regions of IGPcR20 that have functional importance. Site-directed mutagenesis may be used to map ligand-binding pockets and to identify residues important for receptor interaction and activation. Compounds that can be generated using modeling methods to bind these residues are also within the scope of this invention. For example, receptor down-regulation and the development of drug tolerance, such as seen in asthma patients who use bronchial dilators which are beta-adrenergic agonists leading to tachyphylaxis, can be studied in these cell systems. The expression of both intact and hybrid receptors is preferred. The number of receptors, as well as mRNA levels can be measured. Agents for radionuclide imaging to monitor level changes can be developed.

Some of the known receptors and their ligands defined by above techniques are

shown below.

Ligand screening

Ligand Categories	Examples of Ligands	Examples of Receptors
Peptides	Angiotensin, bradykinin, EGF, NPY, neurokinins, PAF, ACTH, C5a, IL8	Leukocyte receptor
Steroids	Testosterone, progesterone, FSH, TSH	Testosterone-R, progesterone-R, FSH-R, TSH-R
Prostaglandins	thromboxane	thromboxane receptor
Neuro-transmitters	Adenosine, adrenergic, dopamine, muscarinic, purinergic, serotonin, opioid	adrenergic receptors, purinergic P2U-R, P2Y1-R
Second messengers	ATP, UDP, cAMP, U, T, adenylyl cyclase, inositol phosphate,	P2X-R, P2Y-R
Ion channels	Calcium, sodium, potassium, chloride	Ligand-gated ion channels
Regulatory sites	Benzodiazepines, glycine, MK-801	Glycine receptors
Uptake sites	Adenosines, choline, dopamine, GABA, glutamate	Dopaminergic receptors, GABA, glutamate receptors
Nucleotide sugars	UDP-glucose, ADP-ribose	KIAA0001

The invention encompasses antagonists and agonists of IGPcR20, as well as compounds or nucleotide constructs that inhibit expression of the IGPcR20 gene (anti-sense and ribozyme molecules), or promote expression of IGPcR20 (wherein

IGPcR20 coding sequences are operatively associated with promoters, enhancers, *etc.*). Highly preferred are the IGPcR20 protein products (especially soluble derivatives of IGPcR20, or truncated polypeptides lacking the TM or CD domains) and fusion protein products, antibodies and anti-idiotypic antibodies, antagonists or agonists (including compounds that modulate signal transduction which may act on downstream targets in the IGPcR20 signal transduction pathway) that can be used for therapy of such diseases, by inhibiting receptor activity.

Nucleotide constructs encoding functional forms of IGPcR20 and mutant forms of IGPcR20 are preferred embodiments of the invention, as their uses include employment in the genetic engineering of host cells. Other preferred embodiments of the invention are anti-sense and ribozyme molecules, preferred for use in "gene therapy" approaches in the treatment of disorders or diseases arising from the aberrant or altered activity of IGPcR20. The gene therapy vector alone or when incorporated into recombinant cells, may be administered in a suitable formulation for intravenous, intra-muscular, intra-peritoneal delivery, or may be incorporated into a timed release delivery matrix.

Transgenic and knock-out animal models

The animal-based and cell-based models can be used to identify drugs, biologicals, therapies and interventions which can be effective in treating disorders with aberrant expression or activity. IGPcR20 sequences can be introduced into, and over-expressed and/or can be disrupted in order to under-express or inactivate IGPcR20 gene expression.

In one embodiment of the invention, the IGPcR20 gene products can also be expressed in transgenic animals. Non-human animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, sheep, cows, goats, may be used to generate IGPcR20 transgenic animals. The present invention provides for transgenic

animals that carry the IGPcR20 transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be expressed in all tissues of the animal, or may be limited to specific tissues. Any technique known in the art may be used to introduce the IGPcR20 transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe PC and Wagner TE, U.S. Patent No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci., USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol. Cell. Biol., 3:1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell 57:717-723); *etc.* For a review of such techniques, see Gordon, 1989, "*Transgenic Animals*", Intl. Rev. Cytol. 115:171-229, which is incorporated by reference herein in its entirety.

The present invention relates to knock-out animals engineered by homologous recombination to be deficient in the production of the IGPcR20. The present invention is directed to a knock-out animal having a phenotype characterized by the substantial absence of IGPcR20, otherwise naturally occurring in the animal. In addition, the invention encompasses the DNA constructs and embryonic stem cells used to develop the knock-out animals and assays which utilize either the animals or tissues derived from the animals. Preferably, these cells, tissues and cell lines are characterized by the substantial absence of IGPcR20 that would otherwise be naturally occurring in their normal counterparts.

Gene targeting is a procedure in which foreign DNA sequences are introduced into a specific locus within the genome of a host cell. In another embodiment of the invention, endogenous IGPcR20 gene expression can be reduced by inactivating or knocking out the IGPcR20 gene or its promoter using targeted homologous recombination. (*e.g.*, see Smithies *et al.*, 1985, Nature 317:230-234; Thomas & Capecchi, 1987, Cell 51:503-512; Thompson *et al.*, 1989, Cell 5:313-321; each of which is incorporated by reference herein in its entirety). For example, a mutant,

non-functional IGPcR20 (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous IGPcR20 gene (either the coding regions or regulatory regions of the IGPcR20 gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express IGPcR20 *in vivo*. Insertion of the DNA construct, via targeted homologous recombination into the genome, results in abolishing IGPcR20 gene function.

One preferred technique for targeted mutagenesis in this invention is based on homologous recombination. The general methodologies of targeting mutations into the genome of cells, and the process of generating mouse lines from genetically altered embryonic stem (ES) cells with specific genetic lesions are well known (Bradley, 1991, Cur. Opin. Biotech. 2: 823-829). See also U.S. patents 5,557,032 by Mak *et al.*, and U.S. Patent No. 5,487,992 by Capecchi *et al.*, included by reference herein. Preferred in this invention is a synthetic recombination vector which contains the genetic information of the targeted chromosomal locus recombines with the genomic DNA after introduction into a cell. A strategy of "positive/negative selection" can be used to enrich the cell population for cells in which targeting vectors have integrated into the host cell genome, and recombination has occurred at the desired gene locus (Mansour, *et al.*, 1988, Nature 336:348). The vector usually contains a positive selection cassette which is flanked by the genetic information of the target locus to enrich for cells where the vector successfully recombines with the chromosomal DNA against the pool of non-recombinant cells.

The likelihood of obtaining an homologous recombination event increases with the size of the chromosomal vector DNA and is further dependent on the isogenicity between the genomic DNA of the vector and the target cell (See te Reile *et al.*, 1992, P.N.A.S. USA 89:5128-5132; Deng *et al.*, 1991, Mol. Cell. Biol., 12, 3365-3371). Also preferred in this invention are large stretches of genomic DNA flanking the IGPcR20 gene ortholog in the target animal species. The cloning of large chromosomal fragments of the target gene, the sub-cloning of this DNA into a bacterial plasmid vector, the mapping of the gene structure, the integration of the

positive selection cassette into the vector and finally, the flanking of one or both homologous vector arms by a negative selection marker are well described in the literature. Also preferred are replacement-type targeting vectors using yeast host cells are described by Storck *et al.*, 1996, Nuc. Acids Res. 24:4594-4596. The use of other vectors such as bacteriophage λ and vectors for phage-plasmid recombination have been described by Tsuzuki *et al.*, 1998, Nuc. Acids Res 26:988-993; transposon-generated gene targeting constructs have also been described by Westphal *et al.*, 1997, Curr. Biol., 7:530-533 and are within the scope of the invention.

The most highly preferred method in this invention is described by Wattler S & Nehls M, German patent application DE 100 16 523.0, "Klonierungssystem zur Konstruktion von homologen Rekombiationsvektoren", filed April 03, 2000, included by reference in whole herein,. and described in part in Example 9. This method reduces the time required for the construction of such vectors from 3-6 months to about 14 days. The vector includes a linear lambda vector (lambda-KO-*Sfi*) that comprises a stuffer fragment; an *E. coli* origin of replication; an antibiotic resistance gene for bacterial selection, two negative selection markers suitable for use in mammalian cells; LoxP sequences for cre-recombinase mediated conversion of linear Lambda phages into high copy plasmids. In a final targeting vector, the stuffer fragment is replaced by nucleotide sequences representing a left arm of homology, an ES cell selection cassette, and a right arm of homology.

The transformation of mouse129 ES cells with the final vector construct is done according to standard procedures. The targeting vector is linearized and then introduced by electroporation into ES cells. Cell clones are positively selected with G418 and negatively selected with GANC (ganciclovir, 0.2 μM). Targeted ES-cell clones with single integration sites are identified, confirmed by hybridization, and expanded in culture for injection.

The invention also encompasses embryonic stem (ES) cells derived from a

developing mouse embryo at the blastocyst stage, that are modified by homologous recombination to contain a mutant IGPcR20 gene allele. The modified ES cells are reintroduced into a blastocyst by microinjection, where they contribute to the formation of all tissues of the resultant chimeric animal, including the germ line (Capecchi, 1989, Trends Genet., 5:70; Bradley, *et al.*, 1984, Nature, 309:255). Modified ES cells may also be stored before reimplantation into blastocysts. The chimeric blastocysts are implanted into the uterus of a pseudopregnant animal, prepared by mating females with vasectomized males of the same species. Typically chimeras have genes coding for a coat color or another phenotypic marker that is different from the corresponding marker encoded by the stem cell genes.

Also within the scope of the invention are chimeric male non-human animals and their heterozygous offspring carrying the IGPcR20 gene mutation which are bred to obtain animals which are homozygous for the mutation, preferred animals being mice. A phenotype selection strategy may be employed, or chromosomal DNA may be obtained from the tissue of offspring, screened using Southern blots and/or PCR amplification for the presence of a modified nucleotide sequence at the IGPcR20 gene locus, liked described in the above section of identifying positively targeted ES cells. Other means for identifying and characterizing transgenic knock-out animals are also available. For example, Northern blots can be used to probe mRNA obtained from tissues of offspring animals for the presence or absence of transcripts coding for either the IGPcR20, the marker gene, or both. In addition, Western blots might be used to assess IGPcR20 expression by probing with antibody specific for the receptor.

These animals are characterized by including, but not limited to, a loss in the ability to bind ligands specific for IGPcR20 and/or by a loss in expression from the IGPcR20 gene locus. Preferably, the animals produce no functional forms of IGPcR20 at all. Once homozygous transgenic animals have been identified, they may preferably be interbred to provide a continual supply of animals that can be used in identifying pathologies dependent upon the absence of a functional IGPcR20 and

in evaluating drugs in the assays described above. Also highly preferred in this invention, are these animals as providing a source of cells, tissues and cell lines that differ from the corresponding cells, tissues and cell lines from normal animals by the absence of fully functional forms of IGPcR20.

The methodology needed to make such animals can be adapted to any non-human animal, preferably rodents such as hamsters, rats or mice, and most preferably, mice. In another embodiment, clones of the non-human transgenic animals can be produced according to methods described in Wilmut *et al.*, 1997, Nature, 385:810-813.

EXAMPLES

Example 1. Identification of a full-length human cDNA coding for a novel GPcR, IGPcR20.

A coding sequence of 1008 basepairs (bp) (SEQ ID NO:1) was identified from the EMBL alert HTGH (High Throughput Genome) database (see Fig. 1). A search was performed using the nucleotide sequence of known G protein-coupled receptors. A potential new GPcR sequence with a statistically significant score was returned and searched for open reading frames. Subsequently a putative coding region was assigned and used in primer design. The tracked human genomic IGPcR20 sequence contains the full-length cDNA sequence, the gene is a single exon coding GPcR. IGPcR20 encodes a protein of 336 amino acids, SEQ ID NO:2 (see Fig. 2).

A BLASTP search (Basic Local Alignment Search Tool for Proteins, National Institutes of Health, Bethesda MD, U.S.A.) revealed that human IGPcR20 has 46% similarity with conserved substitutions to an human orphan receptor GPR34 (Schoneberg *et al.*, 1999, Biochim. Biophys. Acta 7,1446:57-70), isolated from a human fetal brain cDNA library. IGPcR20 also has similarity to the mouse receptor GPR34 (45%; protein ID number AAD50531.1). It is also related to human orphan receptor GPcR17 (43%; SwissProt accession number Q13304) and to a human P2Y-

like GPcR (43%; protein ID number CAA73144.1).

The 2 kb mRNA of the GPR34 gene is found in several human and mouse tissues, with an additional 4 kb transcript found only in mouse liver and testes (Schoneberg *et al.*, 1999, Biochim. Biophys. Acta 7,1446:57-70). The orphan human GPR34 receptor shows a low degree of identity (less than 30%) to the most closely related G protein-coupled receptors. Based on a similar RNA expression pattern described for example for the endothelian-A receptor (Arai *et al.*, 1990, Nature 348:730-732; Hosoda *et al.*, 1992, J. Biol. Chem. 267:18797-18804) or for the anaphytatoxin C3a receptor (Ames *et al.*, 1996, J. Biol. Chem. 267:20231-20234) a function in vascular or lymphatic tissue is speculated. Like human IGPcR20, human GPR34 lacks close structural and functional homologies to other G protein-coupled receptors (Schoeneberg *et al.*, 1999, Biochim. Biophys. Acta, 1446:57-70).

Sequence comparison of orphan receptor GPR17 with known receptors indicates that human GPR17 belongs to the chemokine receptor family of G protein-coupled receptors (Raport *et al.*, 1996, J. Leukoc. Biol., 59:18-23). Chemokines have potent cellular activities towards leukocytes, causing chemotaxis, shape change, increased adhesivity, and release of granule enzymes and other bioactive mediators. The amino acid identity between chemokine receptors ranges from 20-80%. The leukocyte chemo-attractants fMet-Leu-Phe, C5a and platelet-activating factor are each recognized by G protein-coupled receptors that share 21-30% identity (Murphy PM, 1994, Annu. Rev. Immunol. 12:593-633). The identity of human IGPcR20 to human GPR17 is 22%, similarity being 43%.

The human P2Y-like GPcR gene is expressed in the brain as an 2.3 kb and 6.3 kb transcript. The gene is localized to chromosome 2q21 (Blasius *et al.*, 1998, J. Neurochem. 70:1357-1365). Human P2Y-like protein exhibits only 28% identity to P2Y4 and 25% identity to P2Y2 amino acid sequences, its closest homologs (Blaesius *et al.*, 1998, J. Neurochem. 70:1357-1365). Support for P2Y-like protein being a nucleotide binding receptor is provided by the conservation of three of the

four basic residues in transmembrane domains 6 and 7 (TM6 and TM7) that are implicated as being functionally important for binding of the negatively charged phosphate chain of nucleotide ligands of P2Y2 (Erb *et al.*, 1995, J. Biol. Chem. 270:4185-4188). Furthermore, compared to other human P2Y receptors, the P2Y-like GPcR shows 69% amino acid identity in transmembrane domain 3 (TM3), a region particularly well conserved within the P2Y receptor group.

Example 2. Tissue-specific expression of human IGPcR20, analysis by RT-PCR.

A panel of cDNAs derived from total RNA from 29 human tissues (Clontech Laboratories, Inc., Palo Alto CA, USA; Invitrogen Corp., Carlsbad CA, USA) was tested in a reverse transcription-polymerase chain reaction (RT-PCR) assay. The sequence of the primers used to amplify a 501 bp product (SEQ ID No:5) is as follows:

5' - CCAATGGGAATATCAATCTGC (coding sequence position 237-257; SEQ ID NO:3)

5' - GAACTGTAAGTCAAATCTTTCTC (coding sequence position 737-715; SEQ ID NO:4)

The conditions for the PCR were: denaturation at 94°C for 45 seconds, annealing at 56°C for 1 minute, and extension at 72°C for 30 seconds, for a total of 35 cycles, in a Thermocycler (MJ Research, Watertown MA, USA; type PTC-225). The PCR products were analyzed on an 1.8% agarose gel and stained with ethidium bromide to visualize DNA by ultraviolet imaging. The tissues analyzed were: skin, whole brain, fetal brain, cerebellum, thymus, esophagus, trachea, lung, breast, mammary gland, heart, liver, fetal liver, kidney, spleen, adrenal gland, pancreas, stomach, small intestine, skeletal muscle, adipose tissue, uterus, placenta, bladder, prostate, testis, colon, rectum and cervix. Positive (human genomic DNA) and negative (water) controls were included.

Weakly positive PCR products of 501 bp in size were observed observed in RNA prepared from 28 of the 29 tissues assayed, the exception being colon.

The RT-PCR assay was repeated with the same set of primers but another human tissue cDNA panel, comprising 31 different types of cDNA generated from commercially available total RNAs (Clontech Laboratories, Palo Alto CA, USA; Invitrogen Corp., Carlsbad CA, USA; Ambion, Inc, Austin TX, USA)

The conditions of the PCR reaction were as described above, except for the annealing temperature, which was raised from 56°C to 59°C, thereby increasing the specificity of primer annealing in the PCR reaction. The tissues analyzed were: skin, bladder, adipose tissue, esophagus, breast, pancreas, prostate, adrenal gland, uterus, placenta, stomach, kidney, heart, cerebellum, mammary gland, spleen, pericardium, lung, trachea, fetal liver, testis, epididymis, skeletal muscle, thymus, small intestine, salivary gland, rectum, liver, total brain, colon, cervix. A negative control (water) was included.

This second human cDNA panel screen displayed strong band of a 501bp PCR product in epididymis, thymus and colon. A faint band was observed in adipose tissue, uterus, small intestine and cervix, as shown in Figure 3. The correct identity of the 501 bp products amplified was confirmed by sequencing.

Example 3. Tissue-specific expression of human IGPcR20, analysis by Northern hybridization.

Northern hybridization of polyA+ RNAs from several human tissues was carried out using a human IGPcR20 specific DNA-probe. The probe was generated by radiolabeling the purified and sequenced PCR product generated using primers as described in Example 2. The probe spans sequences coding for transmembrane regions 3 to 5 and is 501 bp in length. Commercially available Multiple Tissue Northern Blots (BioChain Institute, Hayward CA, USA) each containing 3 micrograms of poly A + RNA per lane, were hybridized, following the

manufacturer's instructions. These blots are optimized to give best resolution in the 1.0-4.0 kb range, and marker RNAs of 9.5, 7.5, 4.4, 2.4, 1.35 and 0.24 kb were run for reference. Membranes were pre-hybridized for 30 minutes and hybridized overnight at 68°C in ExpressHyb hybridization solution (Clontech Laboratories, Palo Alto CA, USA) as per the manufacturer's instructions. The cDNA probe used was labeled with [α^{32}P] dCTP using a random primer labeling kit (Megaprime DNA labeling system; Amersham Pharmacia Biotech, Piscataway NJ, USA) and had a specific activity of 1 x 10^9 dpm/µg. The blots were washed several times in 2X SSC, 0.05% SDS for 30-40 minutes at room temperature, and were then washed in 0.1X SSC, 0.1% SDS for 40 minutes at 50°C (see Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, New York, USA). The blots were covered with standard domestic plastic wrap and exposed to X-ray film at –70°C with two intensifying screens for four days.

The tissues represented in the BioChain Institute Multiple Tissue Northern Blots are as follows:

Blot #II	Blot #VIII	Blot #V	Blot#I
Stomach	Brain	Uterus	Heart
Jejunum	Kidney	Cervix	Brain
Ileum	Spleen	Ovary	Liver
Colon	Intestine	Testis	Pancreas
Rectum	Uterus	Prostate	Skeletal Muscle
Lung	Cervix	Lung	Lung
---	Placenta	---	---
---	Lung	---	---

The results of this experiment indicate that IGPcR27 is weakly expressed as an approximately 2.8 kb transcript in heart, brain, pancreas, skeletal muscle and brain, (see Figure 4).

Example 4. Characterization of human IGPcR20 protein.

The encoded protein of 336 amino acids was compared to sequences present in public databases EMBL and Genbank. Human IGPcR20 has similarities to human orphan receptor GPR34 (45%) (Schoneberg *et al.*, 1999. Biochim. Biophys. Acta 7,1446:57-70), which codes for a protein of 381 amino acids and to the mouse receptor GPR34 gene (45%), which encodes a protein of 375 amino acids. Human IGPcR20 bears a similarity to human orphan receptor GPcR17 (43%) which codes for a protein of 339 amino acids, and to human P2Y-like GPcR (43%) which codes for a protein of 367 amino acids.

Fig. 5a, 5b and 5c show the amino acid sequence of human IGPcR20 ('query') compared to the amino acid sequence of human GPR34 (Fig. 5a), the amino acid sequence of human GPR17 (Fig. 5b) and the amino acid sequence of human P2Y-like (Fig. 5c) each as ('sbjct'), as abstracted from the SWISSPROT database and analyzed using the BLASTP alignment program. The predicted transmembrane domains (TM) of IGPcR20 are flanked by amino acids 19-39 (TM1), 52-72 (TM2), 98-116 (TM3), 154-172 (TM4), 204-222 (TM5), 250-272 (TM6), 292-310 (TM7), as underlined in Fig. 5a.

Fig. 6a shows a hydropathy plot for the predicted sequence of the human IGPcR20 protein compared to that of human GPR34, human GPR17 and human P2Y-like proteins. The analysis was performed using the method of Kyte and DooLittle (J. Mol. Biol.,1982, 157:105-32), with the DAMBE program (Data Analysis in Molecular Biology and Evolution), University of Hong Kong, version 3.7.49.

Example 5. Identification of mouse ortholog of human IGPcR20.

A cDNA stretch of the mouse ortholog of human IGPcR20 was identified by PCR amplification of 129 mouse genomic ES cell DNA with human IGPcR20 primers SEQ ID NO:6 and SEQ ID NO:8, under the following conditions (termed "touch-up"

PCR):

– denaturation at 94°C for 45 seconds, annealing at 42°C for 1 minute, and extension at 72°C for 45 seconds, for a total of 2 cycles; followed by denaturation at 94°C for 45 seconds, annealing at 48°C for 1 minute, and extension at 72°C for 45 seconds, for a total of 2 cycles; followed by denaturation at 94°C for 45 seconds, annealing at 54°C for 1 minute, and extension at 72°C for 45 seconds, for a total of 35 cycles in a Thermocycler (MJ Research; type PTC-225). The PCR products were analyzed on an 1.8% agarose gel and stained with ethidium bromide to visualize DNA by ultraviolet imaging. A faint band of product was visible. A 1µl aliquot of the original PCR reaction was transferred into a fresh reaction tube, supplemented with a PCR reaction mix as in the first amplification, but with primer SEQ ID NO:7 replacing primer SEQ ID NO:6. Primer SEQ ID NO:7 is located 3' nested to primer SEQ ID NO:6, allowing a half-nested PCR strategy to be performed to selectively boost low amounts of mouse PCR product generated in the first PCR reaction.

The sequence of the human primers used to amplify a mouse 355 bp product (SEQ ID NO:9) was as follows:

5' – GCAGTGCCATGCCTTTCATG (SEQ ID NO:6)

5' – GCAGAGTGGTCAATTTTCTGG (SEQ ID NO:7)

5' – CCAATGAGACCTGCAATCTG (SEQ ID NO:8)

A mouse PCR product of 355 bp in size was sequenced (SEQ ID NO:9), the mouse sequence being used for mouse primer design. Clones from a mouse strain 129 genomic library, containing the full-length cDNA (SEQ ID NO:10; Fig. 7) and flanking genomic sequences are isolated by hybridization, using a mouse IGPcR20 specific DNA probe. Genomic clones are utilized in the construction of targeting vectors.

Example 6. Tissue-specific expression of mouse IGPcR20, analysis by RT-

PCR.

A panel of cDNAs derived from total RNA freshly isolated from 22 mouse tissues was tested in a reverse transcription-polymerase chain reaction (RT-PCR) assay. The sequences of the primers used to amplify a 163 bp product (SEQ ID NO:14) are as follows:

5' - GGAGTCCAAGCAAGAGGC (SEQ ID NO:12)

5' - CTCGACTACTGAGTAGTATAGG (SEQ ID NO:13)

The conditions for the polymerase chain reaction (PCR) were: denaturation at 94°C for 45 seconds, annealing at 56°C for 1 minute and extension at 72°C for 30 seconds, for a total of 35 cycles, in a Thermocycler (MJ Research; type PTC-225). The PCR products were analyzed on an 1.8% agarose gel and stained with ethidium bromide to visualize DNA by ultraviolet imaging. The tissues analyzed were: brain, brain without cerebellum, cerebellum, thymus, eye, tongue, trachea, adipose tissue, lung, heart, spleen, kidney, liver, stomach, small intestine, large intestine, colon, rectum, bladder, uterus, prostate, testis. Positive (mouse genomic DNA) and negative (water) controls were included.

Weakly positive PCR products of 163 bp in size were observed in RNA prepared from brain, brain without cerebellum, cerebellum, eye, trachea, lung, heart, thymus, spleen, kidney, uterus, bladder, prostate, small and large intestine, colon. This is an almost ubiquitous expression pattern.

The RT-PCR assay was repeated using different primers with a different, expanded mouse tissue cDNA panel, comprising 34 different types of cDNA generated from freshly prepared tissue RNAs. The conditions of the PCR reaction were as described above. The tissues analyzed were: lung, kidney, heart, skeletal muscle, total brain, cerebrum, cerebrum left hemisphere, cerebrum right hemisphere, cerebellum, medulla oblongata, olfactory lobe, thymus, adipose tissue, thyroid/ trachea, gall bladder, tongue, esophagus, bladder, eye, salivary gland, stomach, rectum, large intestine, trachea, adrenal gland, spleen, testis, epididymis, prostate, liver, trachea,

embryonic stem cell, ovary, uterus. A negative control (water) was included.

Positive PCR products of 163 bp in size were observed in this mouse panel, with strong expression, in epididymis and testis, and with weak expression in total brain, cerebrum, cerebellum left/right hemisphere, thymus, trachea, spleen, esophagus, bladder, large intestine, uterus, lung, kidney and heart. The correct identity of the 163 bp products amplified was confirmed by sequencing.

To verify the strong PCR signals observed in epididymis and testis, obtained in the RT-PCR assay, a Northern hybridization was performed.

Example 7. Tissue-specific expression of mouse IGPcR20, analyzed by Northern hybridization.

Northern blots against 15 µg aliquots of total RNA isolated from cerebrum, medulla oblongata, olfactory lobe, epididymis, testis, heart, liver, skeletal muscle, total brain, cerebellum, spleen and uterus were hybridized with a mouse IGPcR20 probe, as indicated in Example 6.

The result of the experiment indicates strong expression of mouse IGPCR20 in epididymis and weak expression in testis, with a transcript of 3,8 kb in size. An additional faint band of RNA with about 2,2kb in size is detected (see Fig. 9).

Example 8. Characterization of mouse protein (SEQ ID No. 11; Fig. 8).

Fig. 5d shows the amino acid sequence of human IGPcR20 ('query') compared to the amino acid sequence of mouse IGPcR20 ('sbjct'), analyzed using a BLASTP alignment program. The amino acid identity is 82,8% in an overlap of 328 amino acid residue. The predicted transmembrane domains (TM) of human IGPcR20 are flanked by amino acids 19-39 (TM1), 52-72 (TM2), 98-116 (TM3), 154-172 (TM4), 204-222 (TM5), 250-272 (TM6), 292-310 (TM7), the predicted transmembrane

domains (TM) of mouse IGPcR20 are flanked by amino acids 18-40 (TM1), 54-76 (TM2), 92-114 (TM3), 154-176 (TM4), 202-224 (TM5), 251-273 (TM6), 293-311 (TM7), as underlined in Fig. 5d.

Fig. 6b shows a hydropathy plot for the predicted sequence of the human IGPcR20 protein compared to that of mouse IGPcR20. The analysis was performed using the method of Kyte and DooLittle (1982, J. Mol. Biol., 157:105-32), with the DAMBE program (Data Analysis in Molecular Biology and Evolution) program, University of Hong Kong, version 4.0.41.

Example 9. Generation of ES cells with a modified IGPcR20 allele, produced by homologous recombination.

The most preferred method in this invention is described in Wattler S & Nehls M, German patent application DE 100 16 523.0, "Klonierungssystem zur Konstruktion von homologen Rekombiationsvektoren", filed April 03, 2000. This method reduces the time required for the construction of such a vector from 3-6 months to about 14 days. The vector includes a linear lambda vector (lambda-KO-Sfi) that comprises a stuffer fragment; an *E. coli* origin of replication; an antibiotic resistance gene for bacteria selection, two negative selection markers suitable for use in mammalian cells; LoxP sequences for cre-recombinase mediated conversion of linear lambda phages into high copy plasmids. In a final targeting vector, the stuffer fragment is replaced by nucleotide sequences representing a left arm of homology, an ES cell selection cassette, and a right arm of homology.

To abolish the gene function of mouse IGPcR20 (mIGPcR20) a deletion of approximately 800 bp of the coding region starting approximately 10 bp downstream of the ATG is performed (see Fig. 10). The left arm of homology (hereafter referred as A/C) is PCR amplified with the primers C and A. The primers contain *Sfi I* restriction sites A and C in their 5'-ends, respectively. *Sfi* recognizes and cuts the

nucleotide sequence 5-GGCCNNNNNGGCC-3'. By changing the nucleotides designated N, unique and non-compatible *Sfi* restriction sites are generated. The 3'-end of primer A is homologous to 25 bp of mouse IGPcR20, ending with the 10 bp downstream of the ATG. The 3'-end (25 bp) of primer C is homologous to a position approximately 2500 basepairs upstream of the ATG. The right arm of homology (hereafter referred as B/D) is PCR amplified with primers B and D: B is located approximately 800 bp downstream of the ATG, and D approximately 2000 bp downstream of the stop codon. Both primers contain *Sfi*-restriction sites B or D in their 5'-ends, respectively. To avoid the introduction of point mutations the Expand high fidelity PCR-System, (Boehringer Mannheim / Roche Diagnostics, Basel CH) is used. A ligation of A/C with B/D and a selection cassette leads to an approximately 800 bp deletion of the mIGPcR20 coding region, thereby creating a null allelle. Both PCR-products A/C and B/D are purified using Qiaquick PCR Purification Kit according to the manufacturer (Quiagen, Venlo, NL). The PCR-products are cleaved 3 hours at 50°C with 60 U *Sfi* and subsequently purified (Qiaquick PCR Purification kit). The final volume is 30 µl/product. The ES-cell selection cassette (IRES-ß-lactamase-MCSneo) contains *Sfi*-sites A and B 5'- and 3'-, respectively (Wattler S, *et al.*,1999, Biotechniques, 26:1150-1159). A typical ligation is 50 ng lambda-KO-Sfi-arm (*Sfi*-cleaved), 10 ng selection cassette, 1 ng A/C, 1 ng B/D, 1 x ligation buffer and 1U T4 ligase (Boehringer Mannheim / Roche Diagnostics, Basel CH). The ligation is carried out for 2 hours at room temperature. Two µl of the ligation are used for *in vitro* packaging ('Gigapack plus' from Stratagene, La Jolla CA, USA) for 1.5 hours at room temperature according to the manufacturer's instructions. Aliquots of 10 µl and 50 µl are used to infect C600 bacteria (Stratagene, La Jolla CA, USA) and infection is performed overnight. Single plaques in SM-buffer (Ausubel FM *et al.*, 1994, "*Current Protocols in Molecular Biology*", John Wiley & Sons, New York) are taken to infect BNN 132 bacteria (30 min at 30°C) for plasmid conversion and infection. Bacteria are cultured over 16 hours at 30°C in TB media (Ausubel FM *et al.*, 1994, "*Current Protocols in Molecular Biology*", John Wiley & Sons, New York), containing 100 µg/ml ampicillin (Amersham Pharmacia Biotech, Piscataway NJ, USA; cat. no. US11259-25). Plasmids are harvested using the Qiagen plasmid

kit (Qiagen cat. no. 12143) according to the manufacturer's instructions. To verify plasmid integrity, *Sfi* and *EcoR1*-digests are performed.

The transformation of mouse 129 ES cells with the final targeting vector is performed according to standard procedures. Electroporated 129 mouse ES cells are double-selected with G418 (400 μg/ml) for 7 days and GANC (ganciclovir, 0.2 μM) for 3 days, starting on day 3 after electroporation, for positive and negative selection, respectively, thereby enriching for transformants having the neomycin resistance gene integrated into an endogenous IGPcR20 allele. Single cell clones are propagated, frozen down and expanded for DNA isolation. To identify positively targeted clones, ES cell DNA is isolated from selected clones, incubated with an appropriate restriction enzyme, and the digestion products separated on an agarose gel. Southern blots are hybridized with an 5' external probe and positive targeted candidates are verified by hybridization with a 3' external probe. A single integration is confirmed by hybridization with a probe derived from the neomycin gene. Positive ES cells are isolated and expanded in culture.

Example 10. Mice Deficient in the Expression of the IGPcR20 Gene.

Male chimeric mice are generated by micro-injection of ES cells carrying a recombined allele into 129/SvEv mouse blastocysts, using standard methodology. The chimeric blastocyst is implanted into the uterus of a pseudopregnant mouse, prepared by mating females with vasectomized males of the same species. The chimeras are bred to wild type animals. Tail DNA is isolated from the offspring of these chimeric mice and analyzed by incubation with appropriate restriction enzymes followed by Southern analysis, using the same strategy as outlined above to determine germline transmission. The blots demonstrate the transmission into the mouse genome of the mutation altering the IGPcR20 allele in transformant ES cells. The chimeric male mouse and its heterozygous progeny (+/-) are bred to produce mice homozygous for the mutation (-/-).

Northern blots are used to probe mRNA obtained from tissues of offspring for the presence or absence of transcripts encoding either the IGPcR27, the marker gene, or both. In addition, Western blots are used to assess IGPcR27 expression by probing with antibody specific for the receptor.

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

CLAIMS

1. An isolated nucleic acid molecule, wherein said nucleic acid molecule comprises at least one of:
   (a) the nucleotide sequence of SEQ ID NO:1;
   (b) the nucleotide sequence of SEQ ID NO:10;
   (c) a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length.
   (d) a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:11, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length.

2. An isolated nucleic acid molecule comprising an allelic variant of a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence selected from the group of:
   (a) the amino acid sequence of SEQ ID NO:2; and
   (b) the amino acid sequence of SEQ ID NO:11;
   wherein said allelic variant contains at least 80% nucleic acid homology and hybridizes to the complement of SEQ ID NO:1 under highly stringent conditions equivalent to hybridization in 42°C in a hybridization solution comprising 50% formamide, 1% SDS, 1M NaCl, 10% Dextran sulfate, and washing twice for 30 minutes in a wash solution comprising 0.1xSSC and 1% SDS.

3. The isolated nucleic acid molecule of claims 1 or 2, comprising a nucleotide sequence which encodes at least one of the group of polypeptides, peptides and fusion proteins, comprising an amino acid sequence at least 70% similar to an amino acid sequence selected from the group of:
   (a) the amino acid sequence of SEQ ID NO:2; and

(b) the amino acid sequence of SEQ ID NO:11.

4. The isolated nucleic acid molecule of claims 1 to 3 operatively linked with a nucleotide regulatory sequence capable of controlling expression of the nucleic acid molecule in a host cell or non-human animal.

5. A vector comprising the isolated nucleic acid molecule of any of claims 1 to 4.

6. A host cell genetically engineered to contain at least one of:

(a) the nucleic acid molecule of any of claims 1 to 4; or

(b) the vector of claim 5.

7. The host cell of claim 6 wherein said host cell is a eucaryotic cell, being at least one of:

(a) a yeast cell;

(b) an insect cell; or

(c) a mammalian cell.

8. The human IGPcR20 protein of SEQ ID NO:2, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length, including but not limited to polypeptides, peptides, isolated domains and fusion proteins.

9. The mouse IGPcR20 protein of SEQ ID NO:11, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length, including but not limited to polypeptides, peptides, isolated domains and fusion proteins.

10. Antibodies specifically targeting the IGPcR20 proteins of any of claims 8 or 9, and/or polypeptides, peptides, isolated domains and the the IGPcR20 component of fusion proteins of said IGPcR20 proteins.

11. Agonists and antagonists of IGPcR20 protein that compete selectively with native natural IGPcR20 ligand and which modulate IGPcR20 gene expression or gene product activity, including: (a) 'small molecules' of molecular mass less than 6 kDa; (b) molecules of intermediate size, having molecular mass between 5 kDa to 15 kDa; and (c) large molecules of molecular mass greater than 12 kDa; the latter including mutant natural IGPcR20 ligand proteins that compete with native natural IGPcR20 ligand and which modulate IGPcR20 gene expression or gene product activity.

12. Anti-sense and ribozyme molecules that can be used to inhibit IGPcR20 gene expression or expression constructs used to enhance IGPcR20 gene expression.

13. Methods of identifying compounds of any of claims 11 or 12, which modulate the activity of IGPcR20 or IGPcR20 gene expression.

14. Embryonic stem cells containing a disrupted endogenous IGPcR20 gene.

15. Non-human knock-out animals that do not express IGPcR20, wherein the endogenous animal ortholog of the IGPcR20 gene is functionally disrupted.

16. The non-human knock-out animals of claim 15, wherein the endogenous animal ortholog of the IGPcR20 gene is functionally disrupted by an homologous recombination method.

17. Mutated non-human animals that express a non-functional or partially functional form of IGPcR20.

18. A non-human transgenic animal model expressing the human IGPcR20 cDNA sequence as shown in SEQ ID NO:1 or the nucleic acid molecule of any of claims 1 to 4.

19. The non-human animal model according to any one of claims 17 to 18, whereby the human IGPcR20 is encoded by a nucleic acid sequence which is homozygous in said animal model.

20. Progeny of non-human animals of any of claims 15 to 19, including both heterozygous and homozygous offspring.

21. Non-human animals of any of claims 15 to 20, wherein the animal is from a genus selected from the group consisting of *Mus* (*e.g.*, mice), *Rattus* (*e.g.*, rats), *Oryctologus* (*e.g.*, rabbits) and *Mesocricetus* (*e.g.*, hamsters).

22. Use of the non-human animal according to any one of claims 15 to 21, for the dissection of the molecular mechanisms of the IGPcR20 pathway, for the identification and cloning of genes able to modify, reduce or inhibit the phenotype associated with IGPcR20 activity or deficiency.

23. Use of the animal model according to any of claims 15 to 21 for the identification of gene and protein diagnostic markers for diseases.

24. Use of the animal model according to any of claims 15 to 21 for the identification and testing of compounds useful in the prevention, amelioration or treatment of diseases associated with IGPcR20 activity or deficiency.

25. The use of any of claims 23 or 24 wherein the disease is selected from the group of pain, cancer, metabolic and inflammatory disorders and reproductive disorders and infertility.

26. The use of claim 25 wherein the disease is selected from the group of diseases associated with signal processing in male reproductive tissues, particularly testis and epididymis.

27. A method of identifying compounds suitable for modulating the activity of the protein according to claim 8, for treatment of diseases characterized by aberrant expression or activity of IGPcR20.

28. A method of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20, by the administration of compounds that bind specifically to the IGPcR20 gene or protein and/or which modulate IGPcR20 expression or IGPcR20 activity; the compounds that that bind specifically to the IGPcR20 gene or protein and/or which modulate IGPcR20 expression or IGPcR20 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20; and the use of compounds that that bind specifically to the IGPcR20 gene or protein and/or which modulate IGPcR20 expression or IGPcR20 activity for prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20.

29. A gene therapy method of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20, by the administration of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7, that modulate IGPcR20 expression or IGPcR20 activity; the vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR20 expression or IGPcR20 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20; and the use of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR20 expression or IGPcR20 activity for prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR20.

30. The method of any of claims 27 to 29 wherein the disease selected from the group of pain, cancer, metabolic and inflammatory disorders and reproductive disorders and infertility

31. The method of claim 30 wherein the disease is selected from the group of diseases associated with signal processing in male reproductive tissues, particularly testis and epididymis.

Figure 1. **Human IGPcR20 cDNA sequence.**

ATGAACAACAATACAACATGTATTCAACCATCTATGATCTCTTCCATGGCTTTACCAATCA
TTTACATCCTCCTTTGTATTGTTGGTGTTTTTGGAAACACTCTCTCTCAATGGATATTTTTA
ACAAAAATAGGTAAAAAAACATCAACGCACATCTACCTGTCACACCTTGTGACTGCAAAC
TTACTTGTGTGCAGTGCCATGCCTTTCATGAGTATCTATTTCCTGAAAGGTTTCCAATGGG
AATATCAATCTGCTCAATGCAGAGTGGTCAATTTTCTGGGAACTCTATCCATGCATGCAA
GTATGTTTGTCAGTCTCTTAATTTTAAGTTGGATTGCCATAAGCCGCTATGCTACCTTAAT
GCAAAAGGATTCCTCGCAAGAGACTACTTCATGCTATGAGAAAATATTTTATGGCCATTT
ACTGAAAAAATTTCGCCAGCCCAACTTTGCTAGAAAACTATGCATTTACATATGGGGAGT
TGTACTGGGCATAATCATTCCAGTTACCGTATACTACTCAGTCATAGAGGCTACAGAAGG
AGAAGAGAGCCTATGCTACAATCGGCAGATGGAACTAGGAGCCATGATCTCTCAGATTG
CAGGTCTCATTGGAACCACATTTATTGGATTTTCCTTTTTAGTAGTACTAACATCATACTA
CTCTTTTGTAAGCCATCTGAGAAAAATAAGAACCTGTACGTCCATTATGGAGAAAGATTT
GACTTACAGTTCTGTGAAAAGACATCTTTTGGTCATCCAGATTCTACTAATAGTTTGCTTC
CTTCCTTATAGTATTTTTAAACCCATTTTTTATGTTCTACACCAAAGAGATAACTGTCAGC
AATTGAATTATTTAATAGAAACAAAAAACATTCTCACCTGTCTTGCTTCGGCCAGAAGTA
GCACAGACCCCATTATATTTCTTTTATTAGATAAAACATTCAAGAAGACACTATATAATCT
CTTTACAAAGTCTAATTCAGCACATATGCAATCATATGGT

Figure 2. **Human IGPcR20 amino acid sequence.**

MNNNTTCIQPSMISSMALPIIYILLCIVGVFGNTLSQWIFLTKIGKKTSTHIYLSHLVTANLLVCS
AMPFMSIYFLKGFQWEYQSAQCRVVNFLGTLSMHASMFVSLLILSWIAISRYATLMQKDSSQ
ETTSCYEKIFYGHLLKKFRQPNFARKLCIYIWGVVLGIIIPVTVYYSVIEATEGEESLCYNRQME
LGAMISQIAGLIGTTFIGFSFLVVLTSYYSFVSHLRKIRTCTSIMEKDLTYSSVKRHLLVIQILLIV
CFLPYSIFKPIFYVLHQRDNCQQLNYLIETKNILTCLASARSSTDPIIFLLLDKTFKKTLYNLFTK
SNSAHMQSYG

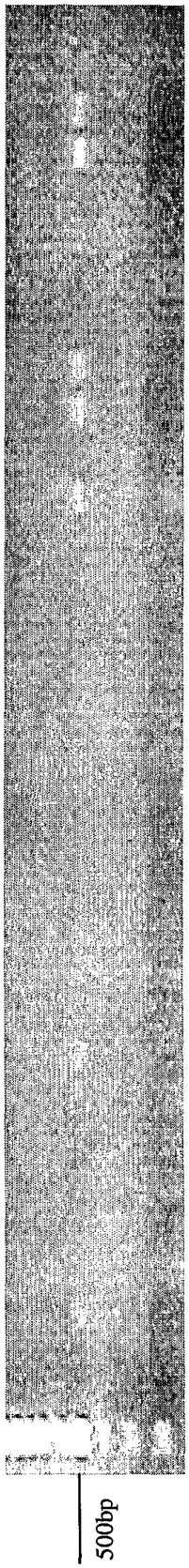


Figure 3.
Human tissue cDNA panel screen for IGPcR20 expression.

1. 100bp DNA ladder
2. Skin
3. Bladder
4. **Adipose tissue**
5. Esophagus
6. Breast
7. Pancreas
8. Prostate
9. Adrenal gland
10. **Uterus**
11. Placenta
12. Stomach
13. Kidney
14. Heart
15. Cerebellum
16. Mammary gland
17. Spleen
18. Pericardium
19. Lung
20. Trachea
21. Fetal liver
22. Testis
23. **Epididymis**
24. Skeletal muscle
25. **Thymus**
26. **Small intestine**
27. Salivary gland
28. Rectum
29. Liver
30. Total brain
31. **Colon**
32. **Cervix**
33. Negative controle

Figure 4. Northern blot analysis of human IGPcR20

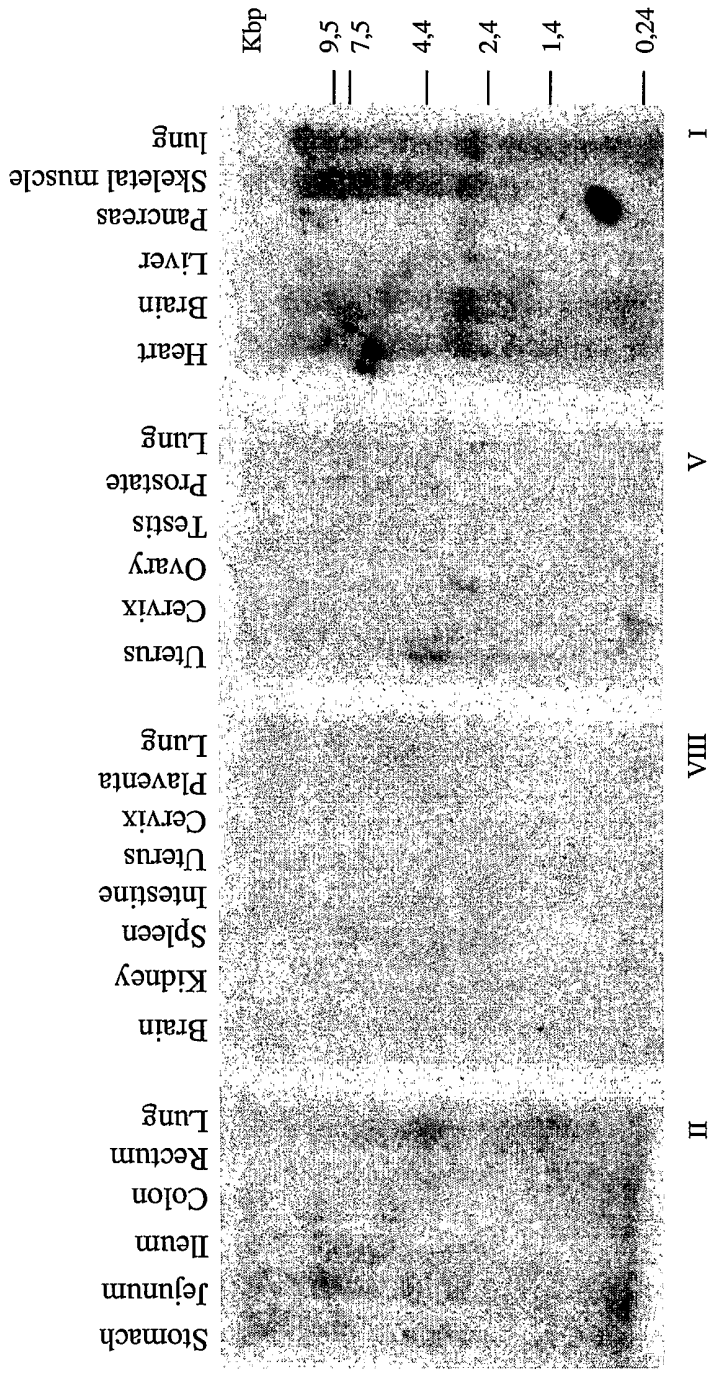


Figure 5a. Amino acid alignment of human IGPcR20 and human GPR34

```
Score =  116 bits (288), Expect = 9e-31
Identities = 77/323 (23%), Positives = 151/323 (45%), Gaps = 25/323 (7%)


Query: 3   NNTTCIQPSMISSMALPIIYILLCIVGVFGNTLSQWIFLTKIGKKTSTHIYLSHLVTANL 62
           N TTC     + S  L   Y ++ IVG+ GN ++ ++FL    K+ S  IYL ++  A+L
Sbjct: 42  NVTTCPMDEKLLSTVLTTSYSVIFIVGLVGNIIALYVFLGIHRKRNSIQIYLLNVAIADL 101

Query: 63  LVCSAMPFMSIYFLKGFQWEYQSAQCRVVNFLGTLSMHASMFVSLLILSWIAISRYATLM 122
           L+   +PF  +Y +   +W      C+VV   GTL  + +M++S+++L +I++ RY  +
Sbjct: 102 LLIFCLPFRIMYHINQNKWTLGVILCKVV---GTL-FYMNMYISIILLGFISLDRYIKIN 157

Query: 123 QKDSSQETTSCYEKIFYGHLLKKFRQPNFARKLCIYIWGVVLGIIIPVTVYYSVIEATEG 182
           +   ++  +  I+                +C  +W V LG  +  T+    ++
Sbjct: 158 RSIQQRKAITTKQSIY---------------VCCIVWMVALGGFL--TMIILTLKKGGH 199

Query: 183 EESLCYNRQMELGAMISQIAGLIGTTFIGFSFLVVLTSYYSFVSHLRKIRTCTSIMEKDL 242
             ++C++ + +  A    I  I        FL+++ SY     +L +I    S
Sbjct: 200 NSTMCFHYRDKHNAKGEAIFNFILVVMFWLIFLLIILSYIKIGKNLLRISKRRSKFPNSG 259

Query: 243 TYSSVKRHLLVIQILLIVCFLPYSIFKPIFYVLHQRD--NCQQLNYLIETKNILTCLASA 300
            Y++  R+  ++ I+  +CF+PY  F+   Y+  Q +  +C     + +T  I+  L+S
Sbjct: 260 KYATTARNSFIVLIIFTICFVPYHAFR-FIYISSQLNVSSCYWKEIVHKTNEIMLVLSSF 318

Query: 301 RSSTDPIIFLLLDKTFKKTLYNL 323
            S  DP+++ L+    +K +  L
Sbjct: 319 NSCLDPVMYFLMSSNIRKIMCQL 341
```

Figure 5b. Amino acid alignment of human IGPcR20 and human GPR17.

```
Score =  116 bits (288), Expect = 8e-31
Identities = 85/322 (26%), Positives = 144/322 (44%), Gaps = 38/322 (11%)


Query: 7   CIQPSMISSMALPIIYILLCIVGVFGNTLSQWIFLTKIGKKTSTHIYLSHLVTANLLVCS 66
           C Q + + +M     Y+L  I+ + GNTL+ W+F+      T  +++L HL  A+L
Sbjct: 23  CGQETPLENMLFASFYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVL 82

Query: 67  AMPFMSIYFLKGFQWEYQSAQCRVVNFLGTLSMHASMFVSLLILSWIAISRYATLMQKDS 126
            +P   +Y   G  W +    CR+  FL  L+M+AS++     L+ I+  R+  ++
Sbjct: 83  VLPTRLVYHFSGNHWPFGEIACRLTGFLFYLNMYASIY----FLTCISADRFLAIVHPVK 138

Query: 127 SQETTSCYEKIFYGHLLKKFRQPNFARKLCIYIWGVVLGIIIPVTVYYSVIEATEGEESL 186
           S                K R+P +A   C ++W VV   + P+ V    ++       L
Sbjct: 139 S----------------LKLRRPLYAHLACAFLWVVVAVAMAPLLVSPQTVQTNHTVVCL 182

Query: 187 -CYNRQMELGAMISQIAGLIGTTFIGFSFLVVLTSYYSFVSHLRK-IRTCTSIMEKDLTY 244
             Y  +    A++S           F F+  +T Y   +  LR+ +R     +EK L
Sbjct: 183 QLYREKASHHALVSLAVAFT------FPFITTVTCYLLIIRSLRQGLR-----VEKRLKT 231

Query: 245 SSVKRHLLVIQILLIVCFLPYSIFKPIFYVLHQRD---NCQQLNYLIETKNILTCLASAR 301
           +V+   +V+ I L VCF+PY + + + YVLH R    +C     L     I +CL S
Sbjct: 232 KAVRMIAIVLAIFL-VCFVPYHVVNRSV-YVLHYRSHGASCATQRILALANRITSCLTSLN 289

Query: 302 SSTDPIIFLLLDKTFKKTLYNL 323
           + DPI++  + + F+  L NL
Sbjct: 290 GALDPIMYFFVAEKFRHALCNL 311
```

Figure 5c. Amino acid alignment of human IGPcR20 and human receptor P2Y-like GPR1.

```
Score = 97.8 bits (231), Expect = 3e-25
Identities = 71/275 (25%), Positives = 122/275 (43%), Gaps = 33/275 (12%)


Query: 7   CIQPSMISSMALPIIYILLCIVGVFGNTLSQWIFLTKIGKKTSTHIYLSHLVTANLLVCS 66
           C Q + + +M     Y+L  I+ + GNTL+ W+F+      T  +++L HL  A+L
Sbjct: 51  CGQETPLENMLFASFYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVL 110

Query: 67  AMPFMSIYFLKGFQWEYQSAQCRVVNFLGTLSMHASMFVSLLILSWIAISRYATLMQKDS 126
            +P   +Y   G  W +    CR+  FL  L+M+AS++     L+ I+  R+  ++
Sbjct: 111 VLPTRLVYHFSGNHWPFGEIACRLTGFLFYLNMYASIY----FLTCISADRFLAIVHPVK 166

Query: 127 SQETTSCYEKIFYGHLLKKFRQPNFARKLCIYIWGVVLGIIIPVTVYYSVIEATEGEESL 186
           S                K R+P +A   C ++W VV   + P+ V    ++       L
Sbjct: 167 S----------------LKLRRPLYAHLACAFLWVVVAVAMAPLLVSPQTVQTNHTVVCL 210

Query: 187 -CYNRQMELGAMISQIAGLIGTTFIGFSFLVVLTSYYSFVSHLRKIRTCTSIMEKDLTYS 245
             Y  +    A++S            F F+  +T Y   +  LR+       +EK L
Sbjct: 211 QLYREKASHHALVSLAVAFT------FPFITTVTCYLLIIRSLRQ----GLRVEKRLKTK 260

Query: 246 SVKRHLLVIQILLIVCFLPYSIFKPIFYVLHQRDN 280
           +V+   +V+ I L VCF+PY + + + YVLH R +
Sbjct: 261 AVRMIAIVLAIFL-VCFVPYHVNRSV-YVLHYRSH 293
```

Figure 5d. Amino acid alignment of human IGPcR20 and mouse IGPcR20.

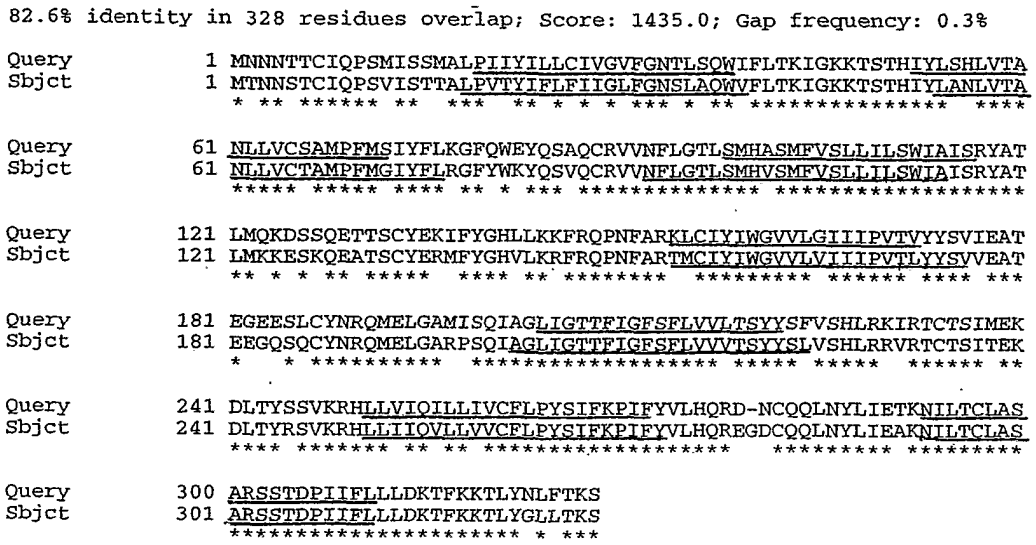


Figure 6a. Hydropathy plots, comparing human IGPcR20, human GPR23, human GPR17 and human P2Y-like GPR1.

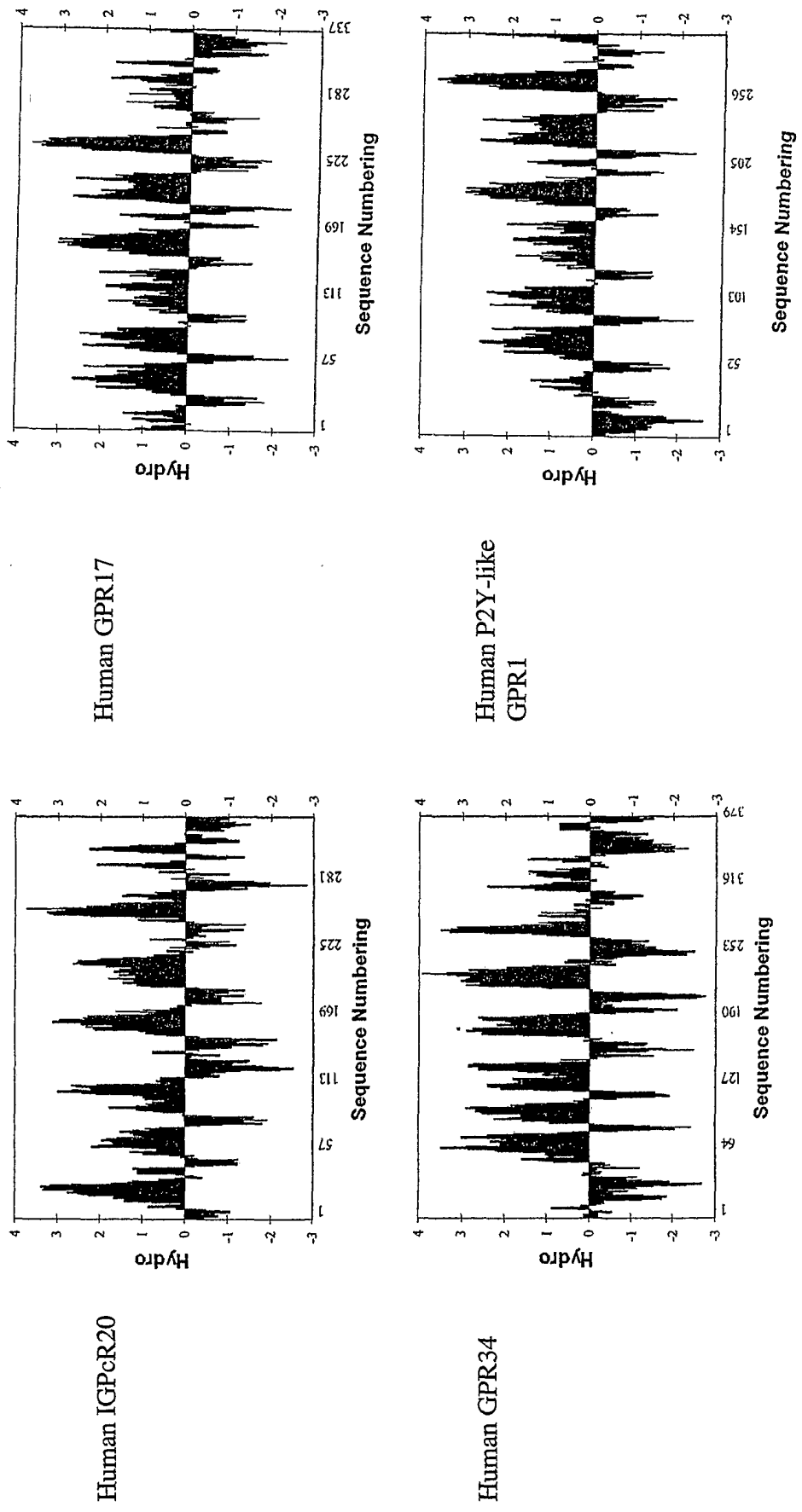
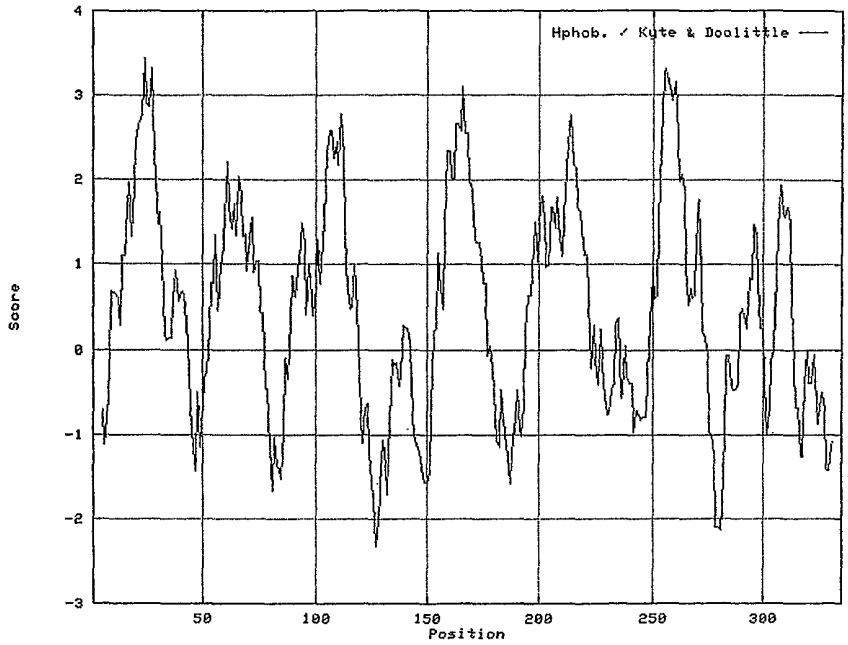


Figure 6b. Hydropathy plots, comparing human IGPcR20 and mouse IGPcR20.

Human IGPcR20



Mouse IGPcR20

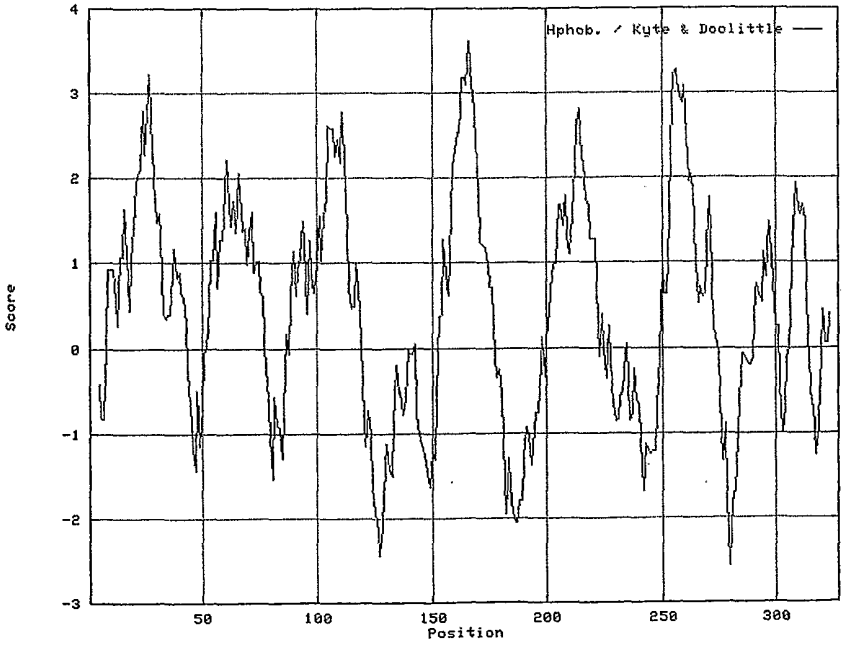


Figure 7. **mouse IGPcR20 cDNA sequence.**

ATGACTAATAACTCGACATGCATTCAACCATCCGTGATTTCTACCACAGCTTTACCGGTCA
CTTACATCTTTTTGTTTATTATTGGTCTCTTTGGAAACTCACTTGCTCAGTGGGTATTTTTA
ACAAAAATAGGTAAGAAAACATCAACACATATCTACTTGGCAAATCTTGTGACTGCAAAC
TTACTTGTGTGCACTGCTATGCCTTTCATGGGTATATATTTCTTGAGAGGTTTTTATTGGAA
ATATCAATCTGTACAATGCAGAGTGGTCAATTTTCTGGGAACTCTTTCTATGCATGTAAGT
ATGTTTGTCAGCCTCTTAATTTTAAGCTGGATTGCCATAAGCCGCTATGCTACCTTAATGA
AAAAGGAGTCCAAGCAAGAGGCCACCTCATGCTACGAGAGAATGTTTTATGGTCACGTA
CTGAAAAGATTTCGCCAGCCCAACTTTGCCAGAACAATGTGTATTTACATATGGGGAGTT
GTGCTTGTCATAATTATTCCAGTTACCCTATACTACTCAGTAGTCGAGGCTACAGAAGAA
GGACAGAGCCAGTGCTACAATCGGCAGATGGAACTGGGAGCCAGACCTTCTCAGATTGC
AGGTCTCATTGGAACCACATTTATTGGATTCTCCTTTTTAGTAGTAGTGACATCATACTAT
TCTCTTGTCAGCCATCTGCGAAGAGTAAGGACCTGTACCTCCATTACCGAGAAAGATTTG
ACCTACAGATCTGTGAAAAGACACCTTTTGATCATCCAGGTCCTCCTAGTAGTTTGCTTTC
TTCCATATAGTATTTTTAAACCCATTTTTTATGTTCTGCACCAGAGAGAAGGAGACTGTCA
GCAGCTGAATTATTTAATAGAAGCAAAAAAACATCCTCACTTGTCTTGCATCAGCCAGAAG
TAGTACAGACCCCATTATATTTCTTTTATTAGATAAAACATTCAAGAAGACACTATATGGT
CTCCTTACAAAATCTTAA

Figure 8. **mouse IGPcR20 amino acid sequence.**

MTNNSTCIQPSVISTTALPVTYIFLFIIGLFGNSLAQWVFLTKIGKKTSTHIYLANLVTA
NLLVCTAMPFMGIYFLRGFYWKYQSVQCRVVNFLGTLSMHVSMFVSLLILSWIAISRYAT
LMKKESKQEATSCYERMFYGHVLKRFRQPNFARTMCIYIWGVVLVIIIPVTLYYSVVEAT
EEGQSQCYNRQMELGARPSQIAGLIGTTFIGFSFLVVVTSYYSLVSHLRRVRTCTSITEK
DLTYRSVKRHLLIIQVLLVVCFLPYSIFKPIFYVLHQREGDCQQLNYLIEAKNILTCLAS
ARSSTDPIIFLLLDKTFKKTLYGLLTKS

Figure 9. Northern blot analysis of mouse IGPcR20.

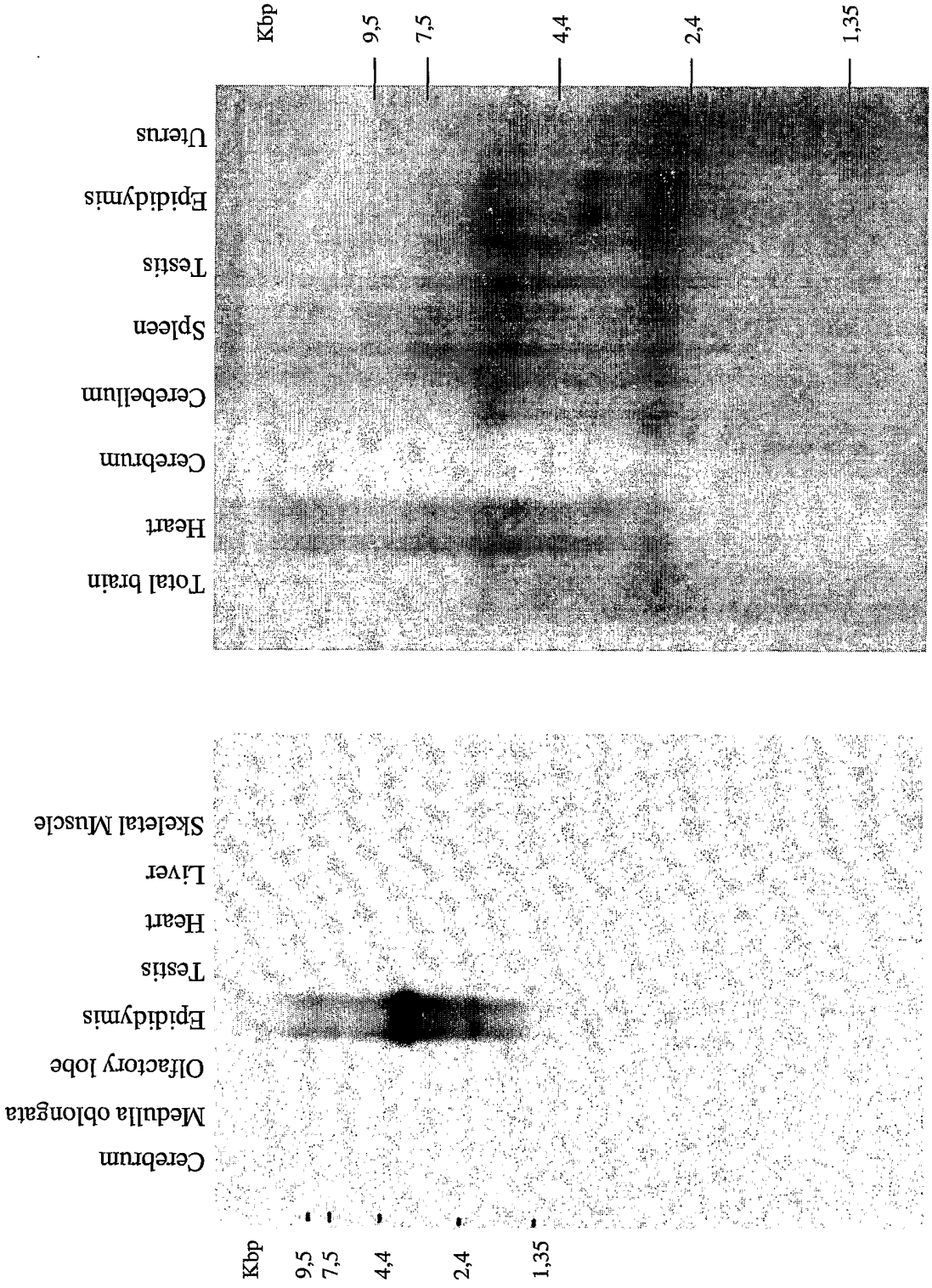
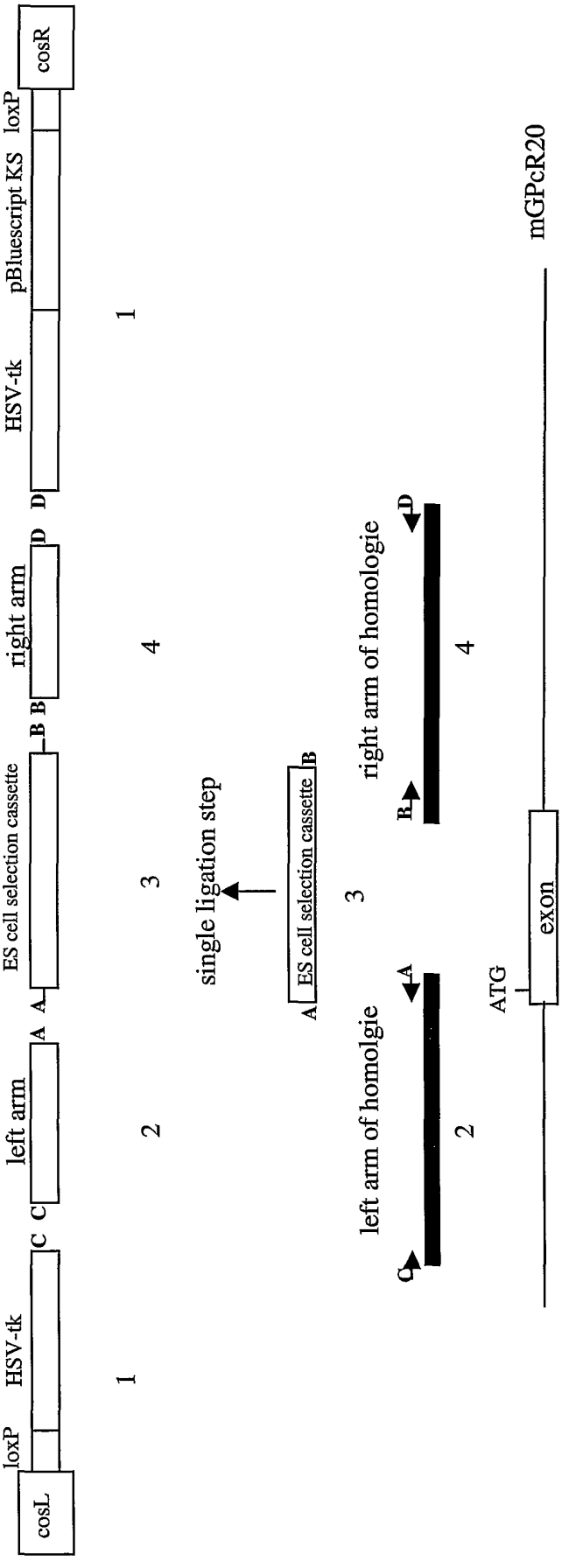


Figure 10. Targeting vector construction.



SEQUENCE LISTING

<110> Ingenium Pharmaceuticals AG

<120> Human G protein-coupled receptor, IGPcR20, and uses thereof

<130> IPA-006-EWO

<140>
<141>

<150> US 60/215,777
<151> 2000-06-30

<160> 14

<170> PatentIn Ver. 2.1

<210> 1
<211> 1011
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1008)

<400> 1

```
atg aac aac aat aca aca tgt att caa cca tct atg atc tct tcc atg   48
Met Asn Asn Asn Thr Thr Cys Ile Gln Pro Ser Met Ile Ser Ser Met
  1               5                  10                  15

gct tta cca atc att tac atc ctc ctt tgt att gtt ggt gtt ttt gga   96
Ala Leu Pro Ile Ile Tyr Ile Leu Leu Cys Ile Val Gly Val Phe Gly
             20                  25                  30

aac act ctc tct caa tgg ata ttt tta aca aaa ata ggt aaa aaa aca   144
Asn Thr Leu Ser Gln Trp Ile Phe Leu Thr Lys Ile Gly Lys Lys Thr
         35                  40                  45

tca acg cac atc tac ctg tca cac ctt gtg act gca aac tta ctt gtg   192
Ser Thr His Ile Tyr Leu Ser His Leu Val Thr Ala Asn Leu Leu Val
     50                  55                  60

tgc agt gcc atg cct ttc atg agt atc tat ttc ctg aaa ggt ttc caa   240
Cys Ser Ala Met Pro Phe Met Ser Ile Tyr Phe Leu Lys Gly Phe Gln
 65                  70                  75                  80
```

```
tgg gaa tat caa tct gct caa tgc aga gtg gtc aat ttt ctg gga act   288
Trp Glu Tyr Gln Ser Ala Gln Cys Arg Val Val Asn Phe Leu Gly Thr
                 85                  90                  95

cta tcc atg cat gca agt atg ttt gtc agt ctc tta att tta agt tgg   336
Leu Ser Met His Ala Ser Met Phe Val Ser Leu Leu Ile Leu Ser Trp
            100                 105                 110

att gcc ata agc cgc tat gct acc tta atg caa aag gat tcc tcg caa   384
Ile Ala Ile Ser Arg Tyr Ala Thr Leu Met Gln Lys Asp Ser Ser Gln
        115                 120                 125

gag act act tca tgc tat gag aaa ata ttt tat ggc cat tta ctg aaa   432
Glu Thr Thr Ser Cys Tyr Glu Lys Ile Phe Tyr Gly His Leu Leu Lys
    130                 135                 140

aaa ttt cgc cag ccc aac ttt gct aga aaa cta tgc att tac ata tgg   480
Lys Phe Arg Gln Pro Asn Phe Ala Arg Lys Leu Cys Ile Tyr Ile Trp
145                 150                 155                 160

gga gtt gta ctg ggc ata atc att cca gtt acc gta tac tac tca gtc   528
Gly Val Val Leu Gly Ile Ile Ile Pro Val Thr Val Tyr Tyr Ser Val
                165                 170                 175

ata gag gct aca gaa gga gaa gag agc cta tgc tac aat cgg cag atg   576
Ile Glu Ala Thr Glu Gly Glu Glu Ser Leu Cys Tyr Asn Arg Gln Met
            180                 185                 190

gaa cta gga gcc atg atc tct cag att gca ggt ctc att gga acc aca   624
Glu Leu Gly Ala Met Ile Ser Gln Ile Ala Gly Leu Ile Gly Thr Thr
        195                 200                 205

ttt att gga ttt tcc ttt tta gta gta cta aca tca tac tac tct ttt   672
Phe Ile Gly Phe Ser Phe Leu Val Val Leu Thr Ser Tyr Tyr Ser Phe
    210                 215                 220

gta agc cat ctg aga aaa ata aga acc tgt acg tcc att atg gag aaa   720
Val Ser His Leu Arg Lys Ile Arg Thr Cys Thr Ser Ile Met Glu Lys
225                 230                 235                 240

gat ttg act tac agt tct gtg aaa aga cat ctt ttg gtc atc cag att   768
Asp Leu Thr Tyr Ser Ser Val Lys Arg His Leu Leu Val Ile Gln Ile
                245                 250                 255

cta cta ata gtt tgc ttc ctt cct tat agt att ttt aaa ccc att ttt   816
Leu Leu Ile Val Cys Phe Leu Pro Tyr Ser Ile Phe Lys Pro Ile Phe
            260                 265                 270
```

```
tat gtt cta cac caa aga gat aac tgt cag caa ttg aat tat tta ata   864
Tyr Val Leu His Gln Arg Asp Asn Cys Gln Gln Leu Asn Tyr Leu Ile
        275                 280                 285

gaa aca aaa aac att ctc acc tgt ctt gct tcg gcc aga agt agc aca   912
Glu Thr Lys Asn Ile Leu Thr Cys Leu Ala Ser Ala Arg Ser Ser Thr
    290                 295                 300

gac ccc att ata ttt ctt tta tta gat aaa aca ttc aag aag aca cta   960
Asp Pro Ile Ile Phe Leu Leu Leu Asp Lys Thr Phe Lys Lys Thr Leu
305                 310                 315                 320

tat aat ctc ttt aca aag tct aat tca gca cat atg caa tca tat ggt   1008
Tyr Asn Leu Phe Thr Lys Ser Asn Ser Ala His Met Gln Ser Tyr Gly
                325                 330                 335

tga                                                               1011
```

<210> 2
<211> 336
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Asn Asn Asn Thr Thr Cys Ile Gln Pro Ser Met Ile Ser Ser Met
  1               5                  10                  15

Ala Leu Pro Ile Ile Tyr Ile Leu Leu Cys Ile Val Gly Val Phe Gly
             20                  25                  30

Asn Thr Leu Ser Gln Trp Ile Phe Leu Thr Lys Ile Gly Lys Lys Thr
         35                  40                  45

Ser Thr His Ile Tyr Leu Ser His Leu Val Thr Ala Asn Leu Leu Val
     50                  55                  60

Cys Ser Ala Met Pro Phe Met Ser Ile Tyr Phe Leu Lys Gly Phe Gln
 65                  70                  75                  80

Trp Glu Tyr Gln Ser Ala Gln Cys Arg Val Val Asn Phe Leu Gly Thr
                 85                  90                  95

Leu Ser Met His Ala Ser Met Phe Val Ser Leu Leu Ile Leu Ser Trp
            100                 105                 110

Ile Ala Ile Ser Arg Tyr Ala Thr Leu Met Gln Lys Asp Ser Ser Gln
```

```
        115                 120                 125

Glu Thr Thr Ser Cys Tyr Glu Lys Ile Phe Tyr Gly His Leu Leu Lys
    130                 135                 140

Lys Phe Arg Gln Pro Asn Phe Ala Arg Lys Leu Cys Ile Tyr Ile Trp
145                 150                 155                 160

Gly Val Val Leu Gly Ile Ile Ile Pro Val Thr Val Tyr Tyr Ser Val
                165                 170                 175

Ile Glu Ala Thr Glu Gly Glu Glu Ser Leu Cys Tyr Asn Arg Gln Met
            180                 185                 190

Glu Leu Gly Ala Met Ile Ser Gln Ile Ala Gly Leu Ile Gly Thr Thr
        195                 200                 205

Phe Ile Gly Phe Ser Phe Leu Val Val Leu Thr Ser Tyr Tyr Ser Phe
    210                 215                 220

Val Ser His Leu Arg Lys Ile Arg Thr Cys Thr Ser Ile Met Glu Lys
225                 230                 235                 240

Asp Leu Thr Tyr Ser Ser Val Lys Arg His Leu Leu Val Ile Gln Ile
                245                 250                 255

Leu Leu Ile Val Cys Phe Leu Pro Tyr Ser Ile Phe Lys Pro Ile Phe
            260                 265                 270

Tyr Val Leu His Gln Arg Asp Asn Cys Gln Gln Leu Asn Tyr Leu Ile
        275                 280                 285

Glu Thr Lys Asn Ile Leu Thr Cys Leu Ala Ser Ala Arg Ser Ser Thr
    290                 295                 300

Asp Pro Ile Ile Phe Leu Leu Leu Asp Lys Thr Phe Lys Lys Thr Leu
305                 310                 315                 320

Tyr Asn Leu Phe Thr Lys Ser Asn Ser Ala His Met Gln Ser Tyr Gly
                325                 330                 335
```

<210> 3
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human
oligonucleotide

<400> 3
ccaatgggaa tatcaatctg c 21

<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human
oliogonucleotide

<400> 4
gaactgtaag tcaaatcttt ctc 23

<210> 5
<211> 501
<212> DNA
<213> Homo sapiens

<400> 5
ccaatgggaa tatcaatctg ctcaatgcag agtggtcaat tttctgggaa ctctatccat 60
gcatgcaagt atgtttgtca gtctcttaat tttaagttgg attgccataa gccgctatgc 120
taccttaatg caaaaggatt cctcgcaaga gactacttca tgctatgaga aaatatttta 180
tggccattta ctgaaaaaat ttcgccagcc caactttgct agaaaactat gcatttacat 240
atggggagtt gtactgggca taatcattcc agttaccgta tactactcag tcatagaggc 300
tacagaagga gaagagagcc tatgctacaa tcggcagatg gaactaggag ccatgatctc 360
tcagattgca ggtctcattg gaaccacatt tattggattt tccttttttag tagtactaac 420
atcatactac tcttttgtaa gccatctgag aaaaataaga acctgtacgt ccattatgga 480
gaaagatttg acttacagtt c 501

<210> 6
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human
oligonucleotide

<400> 6

gcagtgccat gcctttcatg 20

<210> 7
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human oligonucleotide

<400> 7
gcagagtggt caattttctg g 21

<210> 8
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human oligonucleotide

<400> 8
ccaatgagac ctgcaatctg 20

<210> 9
<211> 355
<212> DNA
<213> Mus musculus

<400> 9
gcagagtggt caattttctg ggaactcttt ctatgcatgt aagtatgttt gtcagcctct 60
taattttaag ctggattgcc ataagccgct atgctacctt aatgaaaaag gagtccaagc 120
aagaggccac ctcatgctac gagagaatgt tttatggtca cgtactgaaa agatttcgcc 180
agcccaactt tgccagaaca atgtgtattt acatatgggg agttgtgctt gtcataatta 240
ttccagttac cctatactac tcagtagtcg aggctacaga agaaggacag agccagtgct 300
acaatcggca gatggaactg ggagccagac cttctcagat tgcaggtctc attgg 355

<210> 10
<211> 987
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (1)..(984)

<400> 10

```
atg act aat aac tcg aca tgc att caa cca tcc gtg att tct acc aca   48
Met Thr Asn Asn Ser Thr Cys Ile Gln Pro Ser Val Ile Ser Thr Thr
  1               5                  10                  15

gct tta ccg gtc act tac atc ttt ttg ttt att att ggt ctc ttt gga   96
Ala Leu Pro Val Thr Tyr Ile Phe Leu Phe Ile Ile Gly Leu Phe Gly
             20                  25                  30

aac tca ctt gct cag tgg gta ttt tta aca aaa ata ggt aag aaa aca   144
Asn Ser Leu Ala Gln Trp Val Phe Leu Thr Lys Ile Gly Lys Lys Thr
         35                  40                  45

tca aca cat atc tac ttg gca aat ctt gtg act gca aac tta ctt gtg   192
Ser Thr His Ile Tyr Leu Ala Asn Leu Val Thr Ala Asn Leu Leu Val
     50                  55                  60

tgc act gct atg cct ttc atg ggt ata tat ttc ttg aga ggt ttt tat   240
Cys Thr Ala Met Pro Phe Met Gly Ile Tyr Phe Leu Arg Gly Phe Tyr
 65                  70                  75                  80

tgg aaa tat caa tct gta caa tgc aga gtg gtc aat ttt ctg gga act   288
Trp Lys Tyr Gln Ser Val Gln Cys Arg Val Val Asn Phe Leu Gly Thr
                 85                  90                  95

ctt tct atg cat gta agt atg ttt gtc agc ctc tta att tta agc tgg   336
Leu Ser Met His Val Ser Met Phe Val Ser Leu Leu Ile Leu Ser Trp
            100                 105                 110

att gcc ata agc cgc tat gct acc tta atg aaa aag gag tcc aag caa   384
Ile Ala Ile Ser Arg Tyr Ala Thr Leu Met Lys Lys Glu Ser Lys Gln
        115                 120                 125

gag gcc acc tca tgc tac gag aga atg ttt tat ggt cac gta ctg aaa   432
Glu Ala Thr Ser Cys Tyr Glu Arg Met Phe Tyr Gly His Val Leu Lys
    130                 135                 140

aga ttt cgc cag ccc aac ttt gcc aga aca atg tgt att tac ata tgg   480
Arg Phe Arg Gln Pro Asn Phe Ala Arg Thr Met Cys Ile Tyr Ile Trp
145                 150                 155                 160

gga gtt gtg ctt gtc ata att att cca gtt acc cta tac tac tca gta   528
Gly Val Val Leu Val Ile Ile Ile Pro Val Thr Leu Tyr Tyr Ser Val
                165                 170                 175
```

```
gtc gag gct aca gaa gaa gga cag agc cag tgc tac aat cgg cag atg   576
Val Glu Ala Thr Glu Glu Gly Gln Ser Gln Cys Tyr Asn Arg Gln Met
            180                 185                 190

gaa ctg gga gcc aga cct tct cag att gca ggt ctc att gga acc aca   624
Glu Leu Gly Ala Arg Pro Ser Gln Ile Ala Gly Leu Ile Gly Thr Thr
        195                 200                 205

ttt att gga ttc tcc ttt tta gta gta gtg aca tca tac tat tct ctt   672
Phe Ile Gly Phe Ser Phe Leu Val Val Val Thr Ser Tyr Tyr Ser Leu
    210                 215                 220

gtc agc cat ctg cga aga gta agg acc tgt acc tcc att acc gag aaa   720
Val Ser His Leu Arg Arg Val Arg Thr Cys Thr Ser Ile Thr Glu Lys
225                 230                 235                 240

gat ttg acc tac aga tct gtg aaa aga cac ctt ttg atc atc cag gtc   768
Asp Leu Thr Tyr Arg Ser Val Lys Arg His Leu Leu Ile Ile Gln Val
                245                 250                 255

ctc cta gta gtt tgc ttt ctt cca tat agt att ttt aaa ccc att ttt   816
Leu Leu Val Val Cys Phe Leu Pro Tyr Ser Ile Phe Lys Pro Ile Phe
            260                 265                 270

tat gtt ctg cac cag aga gaa gga gac tgt cag cag ctg aat tat tta   864
Tyr Val Leu His Gln Arg Glu Gly Asp Cys Gln Gln Leu Asn Tyr Leu
        275                 280                 285

ata gaa gca aaa aac atc ctc act tgt ctt gca tca gcc aga agt agt   912
Ile Glu Ala Lys Asn Ile Leu Thr Cys Leu Ala Ser Ala Arg Ser Ser
    290                 295                 300

aca gac ccc att ata ttt ctt tta tta gat aaa aca ttc aag aag aca   960
Thr Asp Pro Ile Ile Phe Leu Leu Leu Asp Lys Thr Phe Lys Lys Thr
305                 310                 315                 320

cta tat ggt ctc ctt aca aaa tct taa                              987
Leu Tyr Gly Leu Leu Thr Lys Ser
                325
```

<210> 11
<211> 328
<212> PRT
<213> Mus musculus

<400> 11

```
Met Thr Asn Asn Ser Thr Cys Ile Gln Pro Ser Val Ile Ser Thr Thr
 1               5                  10                  15

Ala Leu Pro Val Thr Tyr Ile Phe Leu Phe Ile Ile Gly Leu Phe Gly
            20                  25                  30

Asn Ser Leu Ala Gln Trp Val Phe Leu Thr Lys Ile Gly Lys Lys Thr
        35                  40                  45

Ser Thr His Ile Tyr Leu Ala Asn Leu Val Thr Ala Asn Leu Leu Val
    50                  55                  60

Cys Thr Ala Met Pro Phe Met Gly Ile Tyr Phe Leu Arg Gly Phe Tyr
65                  70                  75                  80

Trp Lys Tyr Gln Ser Val Gln Cys Arg Val Val Asn Phe Leu Gly Thr
                85                  90                  95

Leu Ser Met His Val Ser Met Phe Val Ser Leu Leu Ile Leu Ser Trp
            100                 105                 110

Ile Ala Ile Ser Arg Tyr Ala Thr Leu Met Lys Lys Glu Ser Lys Gln
        115                 120                 125

Glu Ala Thr Ser Cys Tyr Glu Arg Met Phe Tyr Gly His Val Leu Lys
    130                 135                 140

Arg Phe Arg Gln Pro Asn Phe Ala Arg Thr Met Cys Ile Tyr Ile Trp
145                 150                 155                 160

Gly Val Val Leu Val Ile Ile Ile Pro Val Thr Leu Tyr Tyr Ser Val
                165                 170                 175

Val Glu Ala Thr Glu Glu Gly Gln Ser Gln Cys Tyr Asn Arg Gln Met
            180                 185                 190

Glu Leu Gly Ala Arg Pro Ser Gln Ile Ala Gly Leu Ile Gly Thr Thr
        195                 200                 205

Phe Ile Gly Phe Ser Phe Leu Val Val Val Thr Ser Tyr Tyr Ser Leu
    210                 215                 220

Val Ser His Leu Arg Arg Val Arg Thr Cys Thr Ser Ile Thr Glu Lys
225                 230                 235                 240

Asp Leu Thr Tyr Arg Ser Val Lys Arg His Leu Leu Ile Ile Gln Val
                245                 250                 255
```

Leu Leu Val Val Cys Phe Leu Pro Tyr Ser Ile Phe Lys Pro Ile Phe
260 265 270

Tyr Val Leu His Gln Arg Glu Gly Asp Cys Gln Gln Leu Asn Tyr Leu
275 280 285

Ile Glu Ala Lys Asn Ile Leu Thr Cys Leu Ala Ser Ala Arg Ser Ser
290 295 300

Thr Asp Pro Ile Ile Phe Leu Leu Leu Asp Lys Thr Phe Lys Lys Thr
305 310 315 320

Leu Tyr Gly Leu Leu Thr Lys Ser
325

<210> 12
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: mouse oligonucleotide

<400> 12
ggagtccaag caagaggc 18

<210> 13
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: mouse oligonucleotide

<400> 13
ctcgactact gagtagtata gg 22

<210> 14
<211> 163
<212> DNA
<213> Mus musculus

<400> 14
ggagtccaag caagaggcca cctcatgcta cgagagaatg ttttatggtc acgtactgaa 60
aagatttcgc cagcccaact ttgccagaac aatgtgtatt tacatatggg gagttgtgct 120
tgtcataatt attccagtta ccctatacta ctcagtagtc gag 163