1
G Protein-coupled Receptors in the Human Genome

Robert Fredriksson and Helgi B. Schiöth

1.1
Introduction

The superfamily of G protein-coupled receptors (GPCRs) is one of the largest families of proteins in the human genome [1, 2] and probably also in most other vertebrate species [3]. GPCRs participate in a diversity of important physiological functions and are targets for many modern drugs. Their ligands are particularly diverse, and include ions, organic odorants, amines, peptides, proteins, lipids, nucleotides and photons, which are all able to activate GPCRs. The main structural characteristic of GPCRs is seven stretches of about 25–35 consecutive amino acid residues that show a relatively high degree of hydrophobicity and represent α-helixes that span the plasma membrane in an anti-clockwise manner. These sequences stretch from the common area or a recognition and connection unit of all GPCRs, enabling an extracellular ligand to exert a specific effect on the cell. This area of the receptors is generally relatively well conserved and is used to identify and classify novel GPCRs as other areas of the receptors are frequently much more diverse. The name GPCRs indicates that these receptors interact with G-proteins. This has however not yet been demonstrated for most of the proteins classified as GPCRs. Moreover, GPCRs are known to have many alternative signaling pathways, interacting directly with a number of other proteins such as arrestins and kinases. Hence, it would perhaps be more technically correct to term this superfamily "seven transmembrane (TM) receptors", but the GPCR terminology has become more established.

Both physiological and structural features have been used to classify GPCRs. The first classification system was introduced in 1994 by Attwood and Findley [4]. They used the term "clans" to designate the different GPCR families. The classified dataset at this time contained over 240 rhodopsin-like GPCRs from different species. Many of these receptors were olfactory and light-recognizing receptors of the opsin type. Independently, but around the same time, Kolakowski presented the well known "Family A–F classification system" [5]. This system included receptors shown to bind G-proteins while the other 7TM receptors were classified as O (other). In conjunction with this classification system the database GPCRdb was

Ligand Design for G Protein-coupled Receptors. Edited by Didier Rognan

ISBN: 3-527-31284-6

developed and included at that time 777 unique GPCRs from various species. Family A contained receptors similar to rhodopsin and biogenic amine receptors. Family B contained receptors similar to the secretin and calcitonin receptors while Family C contained the metabotropic glutamate receptors. Family D and E contained only receptors that were not, and still have not been, identified in mammals, namely the fungal pheromone receptors and the cAMP binding receptors, respectively. Finally, Family F contained archebacterial opsins. The Kolakowski classification system was later extended independently, and differently, by Josefsson and Flower in 1999 [6, 7]. Moreover, another classification system was suggested in 1999 that contained in total five families based on the position of the ligand-binding pocket and the sequence length of the receptors. This system excluded the receptors that are not present in vertebrates [8]. This system used both structural and physiological features to classify the receptors. Recently, we have undertaken large-scale systematic phylogenetic analyses including the majority of the GPCRs in the human genome [9]. This provides us with the GRAFS system showing five main families named *Glutamate* (G; previous family C/3), *Rhodopsin* (R; previous family A/1), *Adhesion* (A; previously part of family B/2), *Frizzled/Taste2* (F; previously O/5 and not included) and *Secretin* (S; previously part of family B/2). Moreover, we subdivided the large Rhodopsin family into 13 subgroups. The grouping was carried out using strict phylogenetic criteria and only a few human receptors did not group into these clusters and these receptors were thus placed into what we called *Other 7TM receptors*. There are several GPCRs that have been discovered since we published this classification [10–13] and here we present an updated version of the human repertoire. In this overview we describe each of the families and groups within the GRAFS classification system and include phylogenetic trees which were derived by Maximum Likelihood and show branch lengths.

1.2 The Adhesion Family

The Adhesion family is the second largest GPCR-family in humans with 33 members. The group is called Adhesion GPCRs according to a recent GPCR classification [9] and this nomenclature seems to prevail. This family has however been assigned various names through the years. These include EGF-TM7 to reflect the presence of epidermal growth factor (EGF) domains in the N-termini [14, 15] and LN-TM7 receptors where LN stands for long N-termini and B2/LNB-7TM to reflect their vague similarity to secretin receptors [16]. The Adhesion family members have several structural features that clearly separate them from all other groups of GPCRs. In a recent article we showed the entire repertoire in human and mouse where the diversity of their N-termini is highlighted [12]. Their long N-termini contain a high percentage of Ser and Thr residues that can create O- and N-glycosylation sites. These N-termini or stalk-like regions are thus thought to be highly glycosylated and act as a mucin-like domain with a rigid erect structure protrud-

ing from the cell surface. The long N-termini are believed to bind various proteins that promote cell-to-cell and cell-to-matrix interactions. All of the Adhesion GPCRs except GPR 123 contain a GPCR proteolytic domain (GPS). Additionally their N-termini can contain a number of different domains that are also found in various other proteins, such as cadherin, lectin, laminin, olfactomedin, immunoglobulin or trombospondin. It is likely that the repertoire of these domains plays an important role in the functional specificity of the receptors.

Phylogenetically, as can be seen in Fig. 1.1, this family forms three main subfamilies with the largest containing lectomedin, EGF-like module, cadherin EGF, EGF lathrophilin, CD97 and GPR 127 receptors. It is interesting to note that all receptors in this group, with the exception of lectomedin receptors, contain EGF domains and that no receptors outside this cluster contain this type of domain [12]. Continuing clockwise in Fig. 1.1 the second group contains 10 receptors termed GPR 110, GPR 111, GPR 113, GPR 115, GPR 116, GPR 123, GPR 124, GPR 125, GPR 133 and GPR 144. The receptors in this cluster have in general very few recognizable domains in their long N-termini, with GPR 123 having no known domains and GPR 110, GPR 111 and GPR 115 having only a GPS domain. The other receptors contain immunoglobulin domains (GPR 124, GPR 125 and GPR 116), hormone-binding domains (GPR 113), leucine-rich repeats (GPR 124 and GPR 125), a pentraxin domain (GPR 144) and a sea urchin sperm domain (GPR 116) [12]. The third family contains brain angiogenesis receptors, human epidymal receptors, the very large GPR 1 (over 6300 amino acids long) and GPR 56, GPR 97, GPR 112, GPR 114, GPR 126 and GPR 128. Also, several receptors in this group are rather sparse in known domains with GPR 56, GPR 97, GPR 114, GPR 126, GPR 128 and human epidymal receptor containing only GPS domains. The three Brain Angiogenesis Inhibitor GPCRs (BAI) contain only hormone-binding and trombospdin domains while GPR 112 contains a pentraxin domain. The very large GPR 1 contains several copies of the sodium–calcium exchange/integrin beta domains [12]. Although several of the more recently discovered Adhesion GPCRs have surprisingly few recognizable functional domains in their N-termini, it is likely that these receptors contain novel domains that are not recognizable using current bioinformatics tools. The majority of the Adhesion GPCRs are orphans and for the few that have been characterized with regard to ligand binding, none has been shown to bind their ligand within the TM regions. CD97 is one of the most studied receptors in this family and is found in several types of blood cell. CD97 interacts with the 312-amino acid membrane protein CD55 (or decay accelerating factor; DAF) which is expressed on most leukocytes [17]. Recently, it was also shown that a glycosaminoglycan (chondroitin sulfate) acts as a cellular ligand specific to the EGF-like domains of the EMR2 [15]. Receptors of the Adhesion family are expressed in various parts of the human body and many of them have prominent expression in the immune system, central nervous system, and in the reproductive organs, suggesting that they might take part in a large variety of physiological functions.

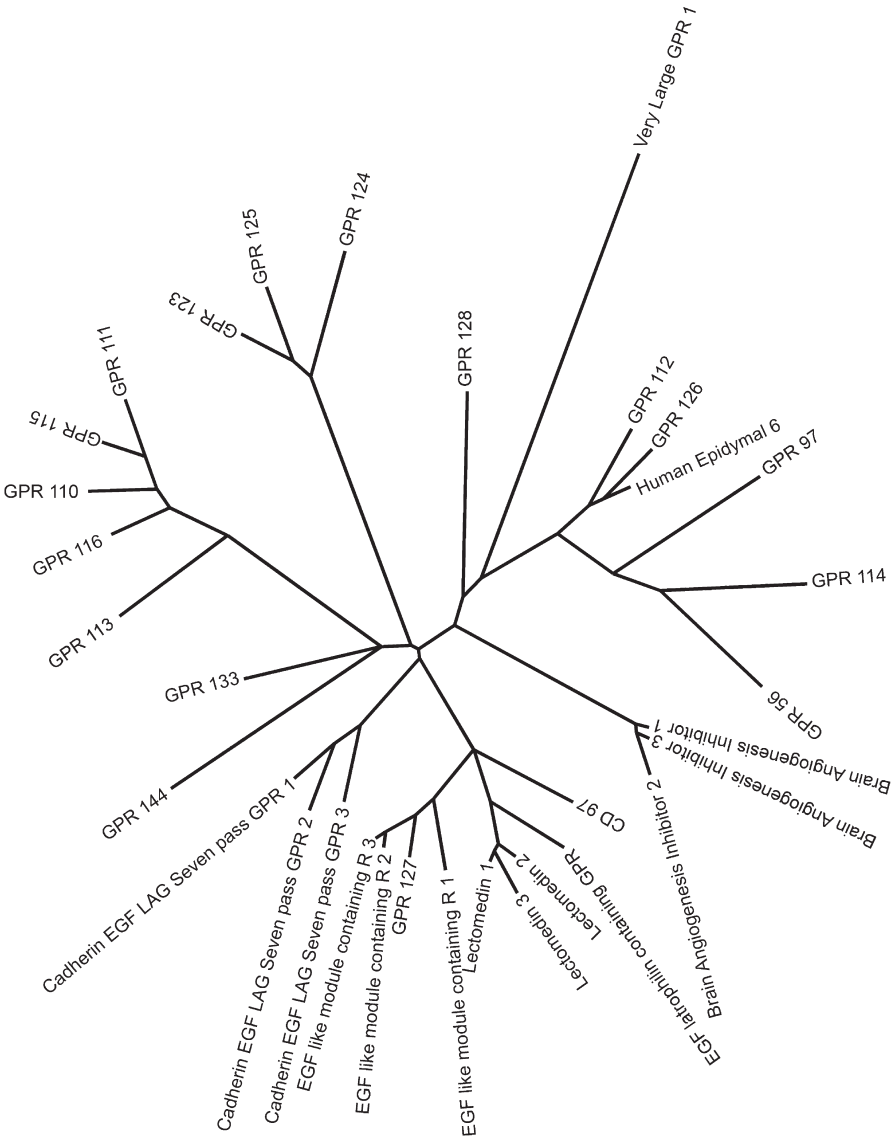

Fig. 1.1 Phylogenetic trees for each family and group of GPCRs. The topology is a consensus tree from 100 bootstrap replicas calculated using ordinary parsimony and the branch lengths are optimized using the Maximum Likelihood method with no assumptions of the presence of a molecular clock and hence the branch lengths correspond to protein distances, not evolutionary time. The calculations are performed on the amino acid sequences from transmembrane region 1 to the end of transmembrane region 7, meaning that no N-termini are used for the calculations as they are very divergent for the different groups. R is the abbreviation for receptors in the annotation on the figures.

1.3
The Secretin Family

The secretin family consists of 15 receptors and occurs widely in all animal species [16, 18]. The N-terminal regions of these receptors share some primary sequence similarity with the Adhesion family of GPCRs and in the A-F classification system both are considered to belong to the B family. Although this sequence similarity is clearly recognizable, it is evident that the Secretin and the Adhesion families are evolutionarily old and that they split into individual groups long ago. Both families are present as multimember families in both insects such as *D. melanogaster* and *A. gambiae* as well as in *C. elegans* [3] as are the other main families, i. e. the Adhesion, Rhodopsin, Frizzled and Glutamate families. The phylogenetic tree of this family has four main subgroups, the largest one consisting of the secretin, Growth Hormone Releasing Hormone, Vasoactive Intestinal Peptide and Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP) receptors. The other groups contain, in clockwise order in Fig. 1.2, the Corticotropin Releasing Hormone/Calcitonin Gene Related Peptide Receptors, Glucagon/Glucagon-like Peptide/Gastric inhibitory peptide receptors and the Parathyroid Hormone Receptors. The receptors in the Secretin family bind rather large peptides and most often act in a paracrine manner. The Secretin family name is related to the fact that the secretin receptor was the first of this family to be cloned and the term secretin-like receptors has also frequently been used in the literature with reference to receptors in this cluster. The N-terminal, between about 60 and 80 amino acids long, contains conserved Cys bridges and is particularly important for the binding of the ligand to these receptors. For example, the N-terminal alone of the VIPR and PACAP receptor constitutes a functional binding site for the ligand. The receptors have a recognizable "hormone binding domain" in the N-termini and these receptors bind rather large peptides that most often act in a paracrine manner.

1.4
The Frizzled/Taste 2 Family

Our phylogenetic studies on the human repertoire have indicated that two very different groups of receptors cluster together. There are few elements in the consensus sequence and the HMM models, such as the consensus sequence of IFL in TM2, SFLL in TM5, and SxKTL in TM7 which are motifs that do not seem to be present in the consensus sequences of the other four families, that could explain why these two groups of receptors cluster together. Further studies are needed to investigate whether these two groups have a common evolutionary history. Below we look at these groups separately.

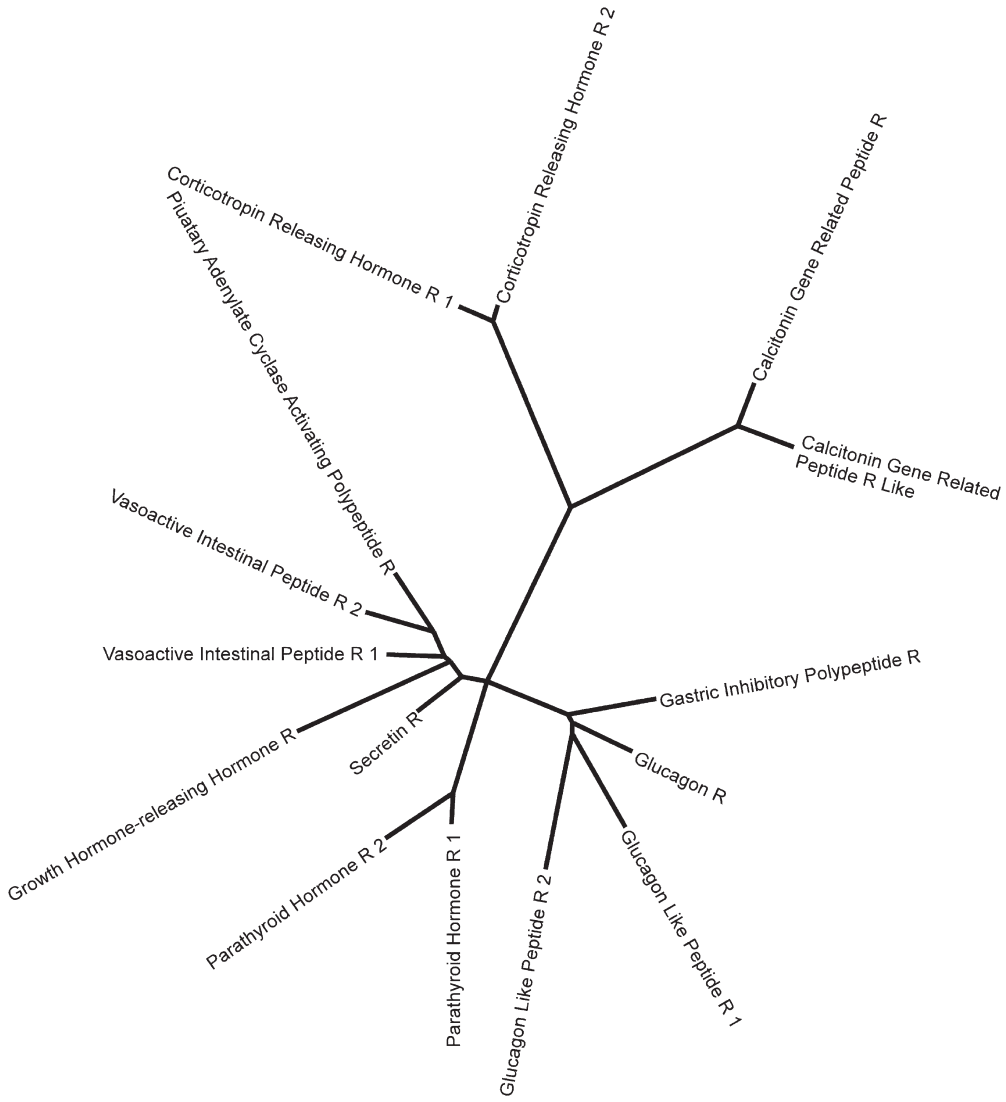

Fig. 1.2 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

1.4.1 The Frizzled Receptor Cluster

The Frizzled group consists of 10 frizzled receptors named Frizzled receptor 1–10 and the single Smoothened receptor. The topology of the tree in Fig. 1.3 shows four main clusters: the cluster containing Frizzled 1, 2 and 7 which have approximately 75% identity to each other; the Frizzled 8 and 5 that have 70% identity, the

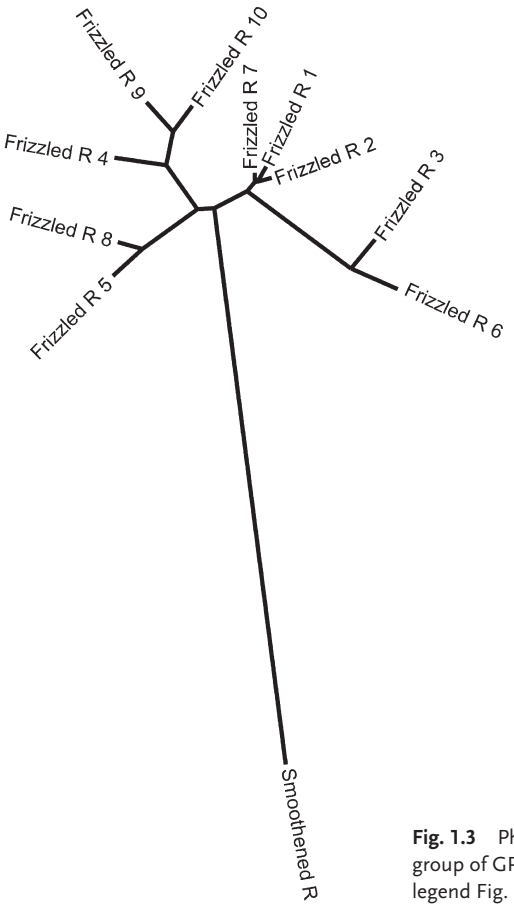

Fig. 1.3 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

Frizzled 10, 9 and 4 that have around 65% identity; and finally Frizzled 6 and 3 that have 50% amino acid identity. The identities shared by receptors from different clusters are between 20 and 40%, indicating that four parental genes from the Frizzled family were initially formed and subsequently the four clusters originated out of these. Smoothened is, as evident in Fig. 1.3, clearly the most divergent of the receptors from the Frizzled family, sharing only 24% amino acid identity to FZD2 and less to the others. The large evolutionary distance between the Smoothened receptor and the other Frizzled receptors also reflects a large evolutionary time of divergence as Smoothened are found as a distinct receptor back in *C. elegans*, which diverged from the lineage leading to mammals more than 600 million years ago [19]. Despite this large sequence divergence between Smoothened and the other Frizzled receptors, all these receptors clearly belong to the same family, which has been shown from phylogenetic analysis of the entire GPCR family [9]. The frizzled receptors control cell fate, proliferation, and polarity during metazoan development by mediating signals from secreted glycoproteins termed Wnt. The

frizzled name was first used for a receptor cloned from *Drosophila*, referring to the curled and twisted Wnt ligand. It was for some time questionable whether the Frizzled receptors were actually true GPCRs, but it has been shown that the Wnt ligand binds to the rat Frizzled receptor 2 and can induce G-protein coupling [20] providing evidence that the frizzled proteins are GPCRs. The frizzled family of receptors has about 200 amino acid-long N-termini with conserved cysteines that are likely to participate in Wnt binding.

1.4.2 The Taste 2 Receptor Cluster

The Taste receptors type 2 (T2Rs) are GPCRs without introns and have very short extracellular N-termini which are believed to be unable to bind ligands and hence it has been suggested that the ligands bind in a pocket within the extracellular parts of the TM regions [21]. The T2Rs show very low sequence similarity to the umami and sweet taste receptors within the Glutamate family (see below), which indicates that the function of recognizing the taste of substances has undergone at least two developmental stages in animals. T2Rs recognizes bitter substances and the relatively large number of T2Rs suggests that mammals have the capability of recognizing many different bitter-tasting substances. Because many poisonous compounds have a bitter taste it has been suggested that this large repertoire of bitter taste receptors has evolved as a key defense mechanism [22]. To date, the majority of T2Rs are orphan receptors without known identified ligands and the only human receptor with a known ligand is T2R16, which has been shown to be activated by salicin [23]. It has also been shown that two mouse and one rat receptor can be activated by bitter compounds and it is highly plausible that the other T2Rs also respond to bitter compounds [22]. Phylogenetically (Fig. 1.4) the relationship between the T2Rs varies considerably with both long branches and clusters of receptors with relatively short branches, such as the cluster of nine receptors containing Taste 2R62. As Taste receptors type 2 are specific for vertebrates [3] and hence have only been around for maybe 450 million years, this suggests that this family is evolving rapidly, perhaps the most rapidly evolving among all GPCR groups. Additionally, it is likely that the clusters of receptors with short branches have arisen rather recently and this is also supported by the fact that these receptors are located in a gene cluster on human chromosome 12 (Bjarnadottir et al., unpublished data).

1.5 The Glutamate Family

The Glutamate family, also known as the clan C receptors, is mostly known for the metabotropic glutamate receptors, which for example mediate glutamate responses in a variety of CNS functions. Other subgroups within this family are the GABA-receptors, the calcium-sensing receptors, a number of orphan GPCRs and the chemosensory receptors, more specifically taste receptors type 1 (TAS1R) and

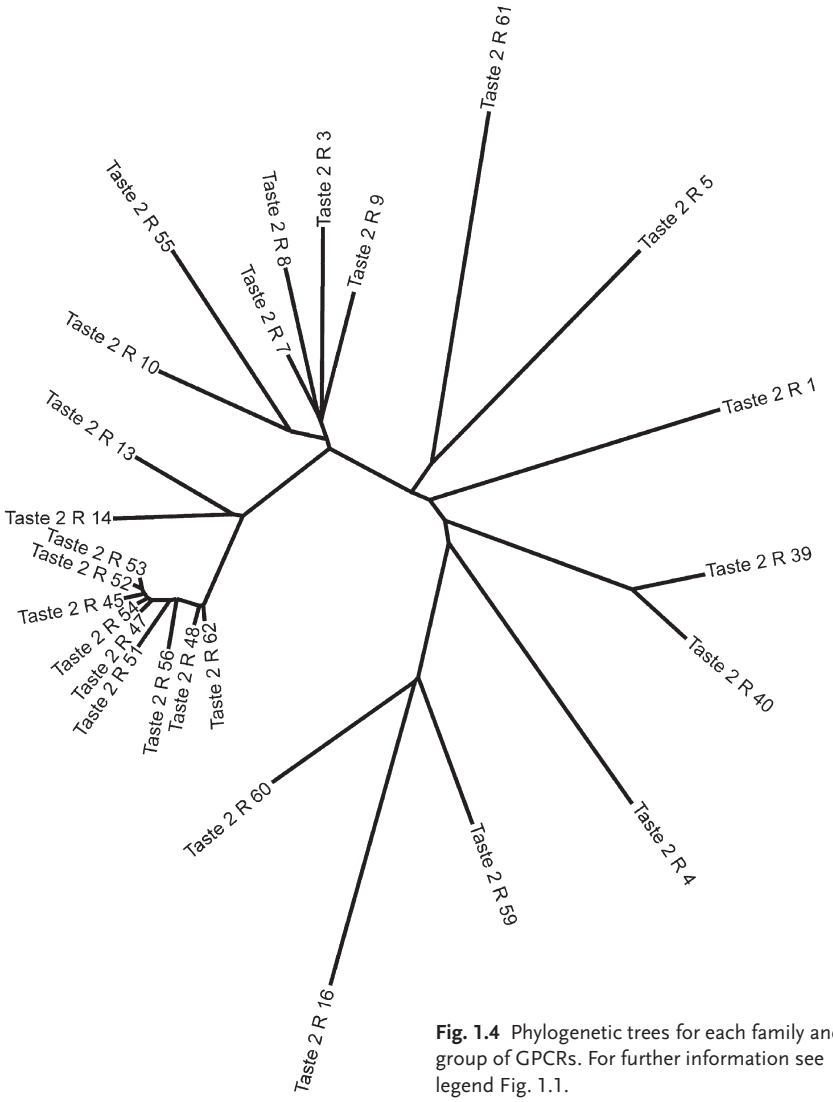

Fig. 1.4 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

the G_0-coupled pheromone receptors (V2Rs; see Fig. 1.5). The V2Rs are specialized in the detection of pheromones related to social and reproductive behavior in most terrestrial vertebrates [24]. Although V2Rs are present in large numbers in all non-primate mammalian genomes investigated, no functional V2Rs are present in the human genome. TAS1R have been identified in mouse, rat and human and are activated by sweet and amino acid taste compounds and these receptors show no close evolutionary relationship to the bitter taste receptors and it is probable that these two families of taste receptors have arisen independently. Surprisingly though, both these types of taste receptors appear to be present only in verte-

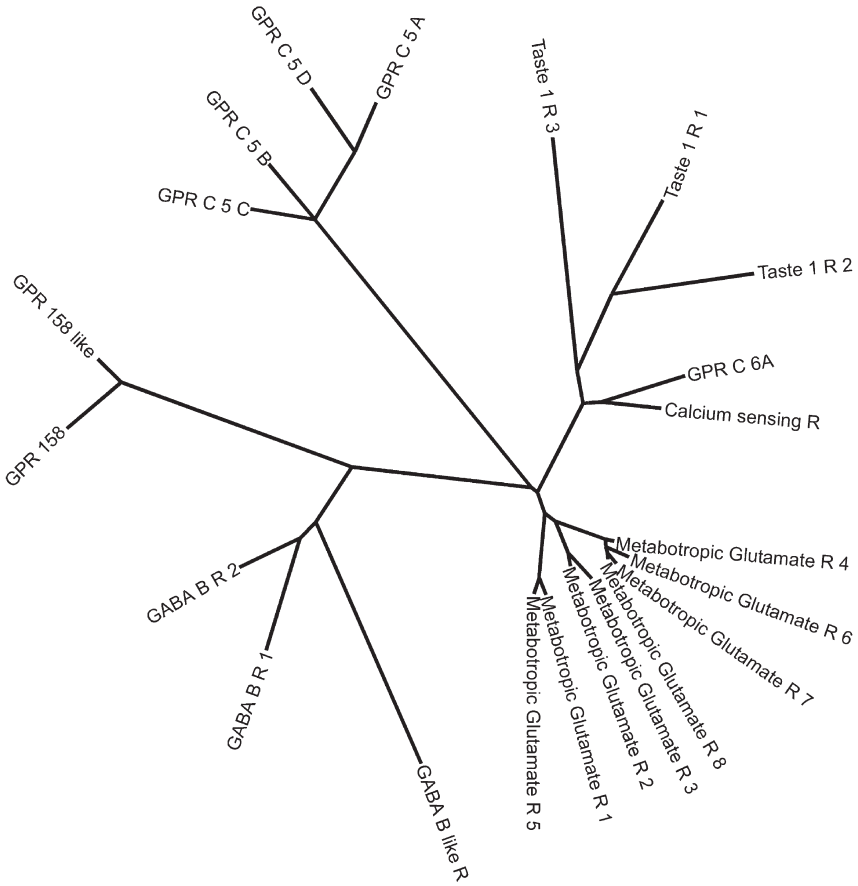

Fig. 1.5 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

brates [3] and might hence have evolved independently but at approximately the same time. This functional diversity of receptors belonging to the same phylogenetic group, the Glutamate GPCRs, is seemingly only matched by the large Rhodopsin family. One of the most important structural features of many of the Glutamate GPCRs is that they possess a large extracellular domain, usually around 600 amino acids. Two-thirds of the extracellular domain has been shown to participate in a two-lobe "Venus flytrap" mechanism which is critical for recognizing and binding ligands among the subgroups of metabotropic glutamate-, GABA- and calcium-sensing receptors. Also the taste receptors and the orphan receptors related to GABA and calcium-sensing receptors have rather long extracellular N-termini. The Glutamate family contains 22 receptors in the human genome (Fig. 1.5) and these are phylogenetically divided into four main groups. The largest is the group of closely related metabotropic glutamate receptors which comprises eight members. Continuing clockwise along the tree, there is a group of

five receptors containing two GABA receptors and three orphans. It is evident that these receptors are clearly distantly related considering the branch lengths of the tree. The next group contains four orphan receptors, GPR C 5A–Ds, which are the only receptors in the Glutamate family that lack long N-termini. The fourth group contains the three taste receptors of type 1, the calcium-sensing receptor and the orphan receptor GPR C 6A.

1.6 The Rhodopsin Family

The Rhodopsin family is made up of the largest number of receptors with about 278 non-olfactory receptors. In addition there are at least 347 olfactory receptors [25] and another 11 receptors currently placed in the "Other" group that are likely to be of Rhodopsin type and this adds up to 636 known Rhodopsin GPCRs. The Rhodopsin family corresponds to what has previously been called either the rhodopsin-like receptors or clan A in the A–E classification system. The crystal structure of bovine rhodopsin has been revealed [26] and this is the only animal GPCR that has had its exact structure determined. Therefore bovine rhodopsin has frequently been used as a template for modeling the structure of other GPCRs from the rhodopsin family [27–29]. It should be noted that bacteriorhodopsin, which has also had its three-dimensional structure determined, has no sequence similarity with the GPCRs in the human genome [6]. The ligands for most of the rhodopsin receptors bind within a cavity between the TM regions [30]. There are however important exceptions to this, in particular for the glycoprotein binding receptors (LH, FSH, TSH and LG), where the ligand-binding domain is in the N-terminal. Our analysis showed four main groups [9] which we have designated α, β, γ and δ. Since our original publication we have identified another 15 non-olfactory rhodopsin GPCR [13] (Gloriam et al., unpublished data) and although these are clearly atypical most seem to belong to one of the four main groups.

1.6.1 The Rhodopsin α-Group

This group has five main branches namely the prostaglandin-, amine-, opsin-, melatonin-, and M.E.C.A. receptor cluster as can be seen in Fig. 1.6. This is the largest of the four main groups in the Rhodopsin family with 101 members in total.

1.6.1.1 The Prostaglandin Receptor Cluster

This cluster contains in total 15 receptors, the seven prostaglandin receptors and the closely related tromboxane receptor together with seven orphan receptors. The orphan receptors are divided into three groups: one containing the super conserved receptor expressed in the brain, one containing group GPR 61 and GPR 62 and one containing GPR 26 and GPR 78.

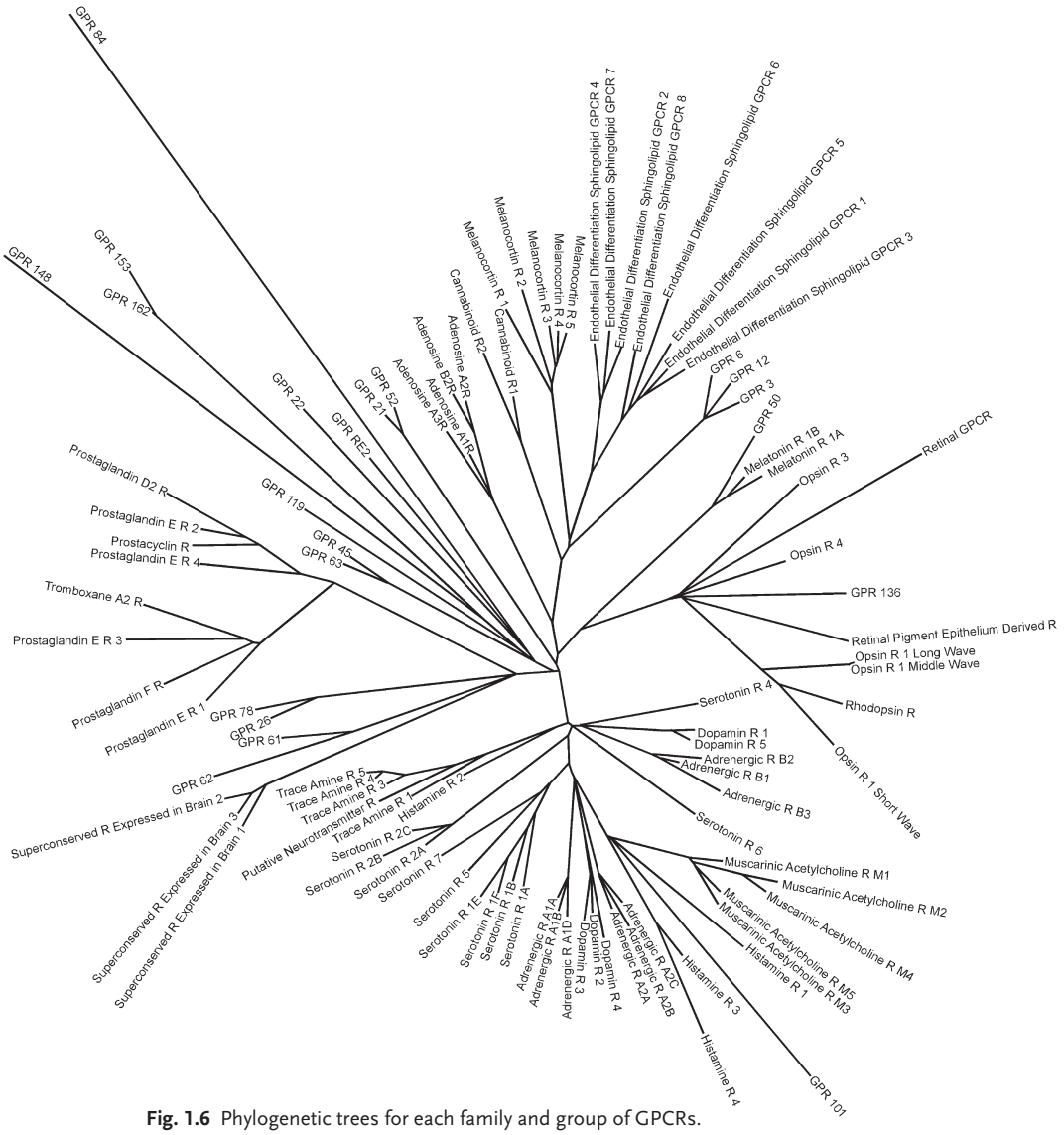

Fig. 1.6 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

1.6.1.2 **The Amine Receptor Cluster**

The Amine receptor cluster contains serotonin, dopamine, muscarinic, histamine, adrenergic, and trace amine receptors. Interestingly this large group contains only one orphan receptor, GPR 101. All the known ligands for the receptors in this group are structurally related small amine molecules with a single aromatic ring. The serotonin receptors display a heterogeneous phylogenetic pattern and two distinct subgroups can be seen, the serotonin receptor 1 and serotonin

receptor 2 branches. Serotonin receptor 5 and 7 cluster basally of the serotonin receptor 1 and 2 but the branches are relatively long which indicates that these receptors are still rather distantly related. Surprisingly serotonin receptors 4 and 6 do not cluster with the other serotonin receptors but rather basally in the Amine group. This indicates that the evolutionary history of the serotonin receptors is peculiar and complex and that the ability to bind the ligand serotonin may have arisen several times in an independent manner during evolution. The trace amine receptor branch contains the trace amine receptors 1 to 5, where trace amine receptor 2 is a pseudogene in humans and hence excluded, and the other four receptors are putative neurotransmitter receptors, also known to bind trace amines. These receptors are closely related to each other with regard to amino acid identity, which indicates that they have arisen recently. The repertoire of trace amine receptors is also very dynamic as species-independent gene family expansions have been shown to have taken place in rat (17 trace amine receptors), mouse (13 trace amine receptors) and zebrafish (57 trace amine receptors; Gloriam et al., unpublished data). The five muscarinic acetylcholine receptors form a homogenous cluster within the amine group, with a relatively high degree of sequence similarity. The adrenergic receptors A and B, also called adrenergic receptors α and β, form separate clusters branching basally in the Amine receptor cluster. This shows that the A and B adrenergic receptors are rather distantly related. The dopamine receptors on the other hand do not cluster together but rather fall into two groups containing two and three receptors each. One of these groups, containing dopamine receptors 2, 3 and 4 is located basally in the adrenergic A receptor branches and the group containing dopamine receptors 1 and 5 is positioned in the adrenergic B receptor branch. This prompts the speculation that dopamine and adrenergic receptors share a relatively recent evolutionary origin. Perhaps these two receptor families originate from two families consisting of promiscuous dopamine- and adrenalin-binding receptors and in each of these families specialization with regard to ligand preferences has evolved. One other heterogeneous group of receptors in the Amine cluster is the histamine receptors, three of these clusters have relatively long branches together with the orphan receptor GPR 101 while histamine receptor 2 is placed in a position basal to the trace amine receptors.

1.6.1.3 **The Opsin Receptor Cluster**

This cluster of Rhodopsin α-receptors comprises the rhodopsin receptor, the three visual cone pigments for long, short and medium wavelength photons, peropsin, encephalopsin, melanopsin, the retinal G-protein coupled receptor and GPR 136. The opsins for long and medium wavelength light are found in the same chromosomal position, Xq28 only 23 700 bases apart, as a result of a gene duplication specific for primates. Unequal crossing over between these genes, resulting in either loss of one of the genes or a hybrid (chimeric) version, is the cause of the commonest form of color blindness [31]. These two proteins are over 96% identical at the amino acid level.

1.6.1.4 The Melatonin Receptor Cluster

The two melatonin receptors cluster on a branch of their own together with orphan receptor GPR 50. Melatonin is a hormone that is mainly produced and secreted at night by the pineal gland. GPR 50 has been linked with a sex-specific risk factor for susceptibility to bipolar disorder [32].

1.6.1.5 The MECA Receptor Cluster

This cluster consists of the melanocortin, endothelial differentiation sphingolipid, cannabinoid, and adenosine receptors. The cluster also contains the three receptors GPR 3, GPR 6 and GPR 12 that recently have been shown to bind lipid ligands [33]. It is interesting to note that the receptors in this group, although clearly closely related phylogenetically, bind structurally very different ligands; melanocortin receptors bind peptides (melanocortins), endothelial differentiation sphingolipid receptors bind lysophosphatidic acid, cannabinoid receptors have anandamide (arachidonylethanolamide) as their endogenous ligand, GPR 3, 6 and 12 bind lipids and adenosine receptors bind adenosine (a purine sugar derivative). The common feature of some of ligands in this group is that both lysophosphatidic acid anadamide and the lipids binding GPR 3, 6 and 12 are derivatives of phospholipids but the adenosine and peptide ligand-binding receptors in this cluster are still peculiar.

1.6.1.6 Other Rhodopsin α-Receptors

A number of more or less unrelated orphan receptors that clearly belong to the Rhodopsin α-cluster do not fall into any of the established groups of receptors. These are the related receptors GPR 45 and GPR 63, GPR 119, GPR 148, the two related GPR 162 and 163 receptors, GPR 84, GPR RE2 and the related receptors GPR 52 and 21. As evident from Fig. 1.6, although they figuratively appear to be related as they are positioned on a common branch, these receptors have long branches to the other receptors in the Rhodopsin α-group, even those on the same branch. Therefore the clustering, and hence the apparent relationship of these receptors, might be a result of a phenomenon known as long-branch attraction. This means that in a dataset that has some long branches leading to the leaves, in addition to some other branches which are short, the long branches will attract each other and appear as a cluster in the tree, even if they are not closely related. This results from the fact that it is more favorable from a scoring point of view, to cluster dissimilar sequences together than to place them separately at the root. The phenomenon of long-branch attraction is known to affect in particular, maximum parsimony algorithms although this type of artefact can present in various types of phylogenetic tree regardless of the algorithm used.

1.6.2 Rhodopsin β-Group

The β-group in the Rhodopsin family is not subdivided further into named sub branches. All the known ligands to the receptors in this cluster are peptides. The group includes the branch containing orexin-, neuropeptide FF-, tachykinin-, cholecystokinin-, prolactin-releasing hormone receptor and the neuropeptide Y receptors. As can be seen in Fig. 1.7, these receptors are relatively closely related as they place on the same branch in the tree. On the same branch, containing 18 receptors in total, the four orphan receptors GPR 72, GPR 73, GPR 73L1 and GPR 103 are also found. Continuing clockwise in Fig. 1.7, we find three orphan receptors, GPR, GPR 19 and GPR 154 that do not cluster with any other receptor but rather branch from the center of the tree. The next main cluster contains the neurotensin, motilin, ghrelin, and neuromedin receptors. This group also contain the orphan receptor GPR 39. Continuing further we find two orphan receptors GPR 75 and GPR 150 that do not clearly group with any other receptors. As can be seen the branches for these receptors are among the longest in the entire tree which shows that they are only distantly related to all the other receptors in the cluster. This is especially pronounced for GPR 75. The next group contains the gonadotropin releasing hormone, angiotensin, vasopressin, and oxytocin receptors. And finally the last group contains the endothelin, neuromedin B, gastrin releasing peptide, and the bombesin-like 3 receptors. Noteworthy in this group is that the neuropeptide Y receptors show a clearly heterogeneous phylogeny. In particular, the neuropeptide Y receptor 2 is not placed together with the other NPY receptors but rather with the tachykinin, prolactin-releasing hormone receptors and orphan receptors GPR 72, 73 and 73L1. It can also be seen that the neuropeptide Y receptor 2 has a higher amino acid identity to prolactin-releasing hormone receptor (36.5%) and GPR 72 (33.9%) than to the other NPY receptors (30.9, 34.2, and 30.9% to neuropeptide receptors 1, 4 and 5 respectively). The neuropeptide Y receptor 5 places equally distant to the neuropeptide Y 1 and 4 pair of receptors and cholecystokinin receptors. The reason for this is probably that neuropeptide Y receptor 5 has a long third extracellular loop due to an insertion, which is also present in the cholecystokinin receptor. The evolutionary history of this group of peptide binding receptors is likely to be much more complicated than it appears at first. For example, the prolactin-releasing hormone receptor is likely to share a recent evolutionary ancestor with the NPY receptors (Lagerstrom et al., unpublished data).

1.6.3 Rhodopsin γ-Group

This group comprises three main clusters termed the SOG, MCH, and Chemokine receptor clusters (Fig. 1.8).

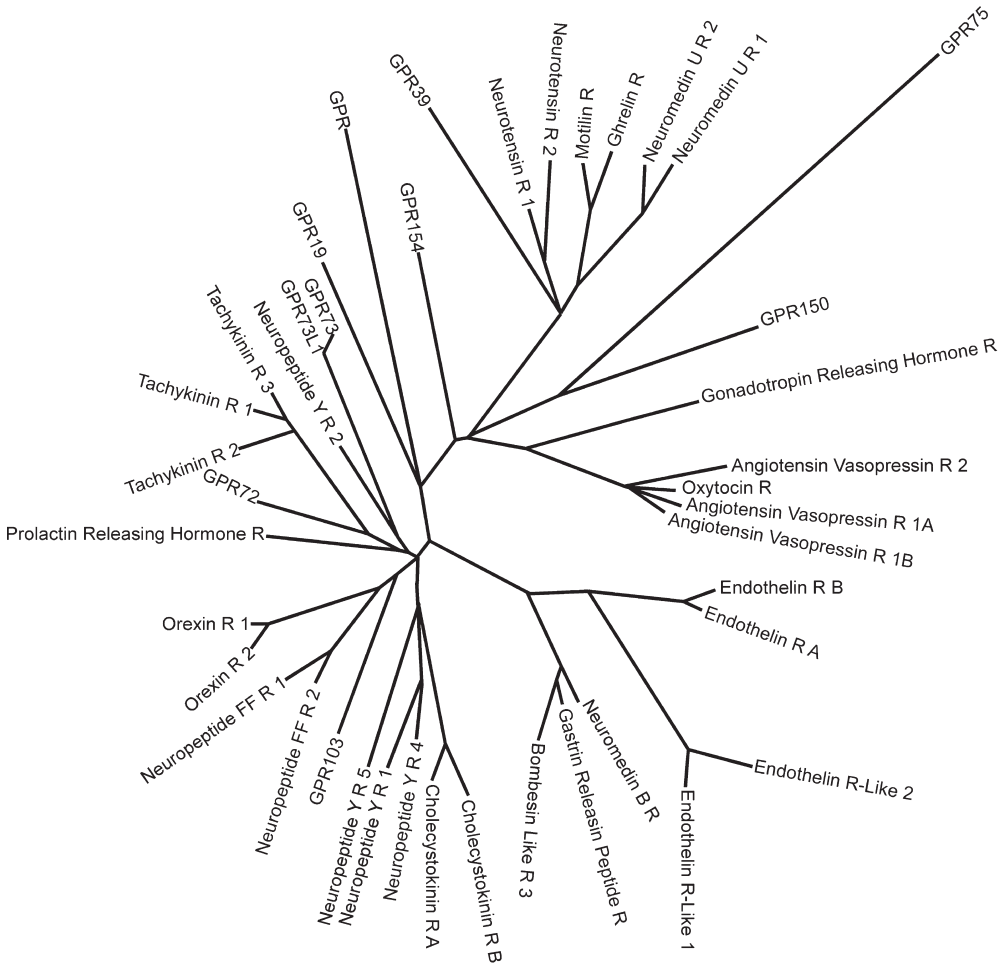

Fig. 1.7 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

1.6.3.1 **The SOG Receptor Cluster**

This cluster of receptors contains the somatostatin, opioid, galanin, and neuropeptide W receptors. It also contains the former orphan receptor GPR 54 which is now known to bind RF-amid. The known ligands to the receptor on this branch are thus all peptides but these ligands have, apart from their peptide nature, no structural similarities. The receptors in this cluster form four phylogenetic clusters (see Fig. 1.8), one containing only the five somatostatin receptors, one containing the two neuropeptide W receptors, one containing the four opioid receptors and finally one containing the three galanin receptors and GPR 54.

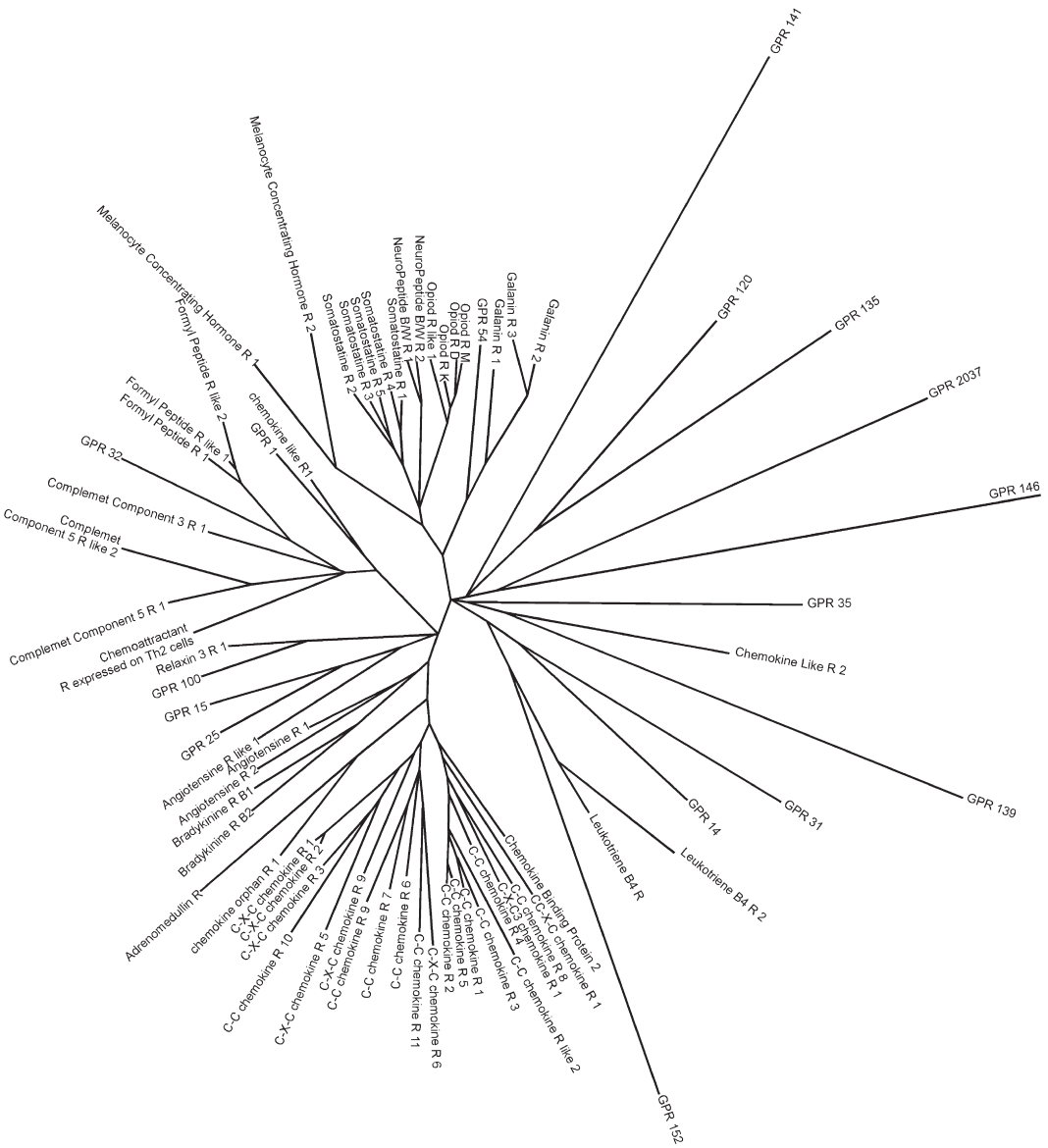

Fig. 1.8 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

1.6.3.2 **The Melanocyte Concentrating Hormone Receptor Cluster**

The two melanocyte concentrating hormone (MCH) receptors branch rather basally in the SOG cluster with relatively long branches. The branches connecting the two MCH receptor leafs are also relatively long which indicates that either these two receptors split quite early from a common ancestor or that they evolved

at a much faster rate than the other receptors in the SOG cluster. The ligand is MCH which is a cyclic neuropeptide of 19 amino acids that is involved in, among other things, the regulation of feeding behavior.

1.6.3.3 The Chemokine Receptor Cluster

This cluster is by far the largest in the Rhodopsin γ-group. The branch with most receptors contains the classical chemokine receptors of the C-C (12 receptors) and C-X-C types (eight receptors). These receptors are known to bind small immunomodulating peptides. At the base of the chemokine branch is a relatively long branch containing the adrenomedulin receptor and one orphan chemokine-like receptor has split off. Continuing clockwise along the tree in Fig. 1.8 is a cluster containing the bradykinin and the two angiotensin receptors. The next branch contains three orphan receptors, the angiotensin-like receptor 1, GPR 15 and GPR 25. Further, the next branch contains the orphan receptor GPR 100 and the relaxin 3 receptor 1. Finally the Chemokine branch contains another large branch containing three complementary component receptors, three formyl peptide receptors, three orphan receptors and one chemoattractant receptor.

1.6.3.4 Other Rhodopsin γ-Receptors

Similar to the Rhodopsin α-group there are a number of mainly orphan receptors connected close to the center of the tree in a loose cluster with long branches. Using reasoning similar to that for Other Rhodopsin α-group receptors these should probably be considered to be more or less unrelated. The receptors in this "branch" are the two leukotriene B4 receptors and 10 orphan receptors namely, GPR 141, GPR 120, GPR 135, GPR 2037, GPR 146, GPR 35, chemokine-like receptor 2, GPR 139, GPR 31, GPR 14 and GPR 152.

1.6.4 The Rhodopsin δ-Group

This group (Fig. 1.9) has five main branches termed the MAS-related receptor cluster, the Glycoprotein receptor cluster, the Purinergic receptor cluster, the Coagulation factor receptor cluster and the Olfactory receptor cluster (not shown in Fig. 1.9).

1.6.4.1 The MAS-related Receptor Cluster

This group contains 10 receptors with the MAS oncogene receptor and the MAS-related receptor branching most basally in the cluster. The four MAS-related GPCR D-G are separated by relatively long branches while the MAS related GPCR X1–4 are clearly highly similar. The MAS related GPCR X receptors have been shown to bind small peptide fragments originating from opioid peptides and these receptors are mainly expressed on small sensory neurons [34].

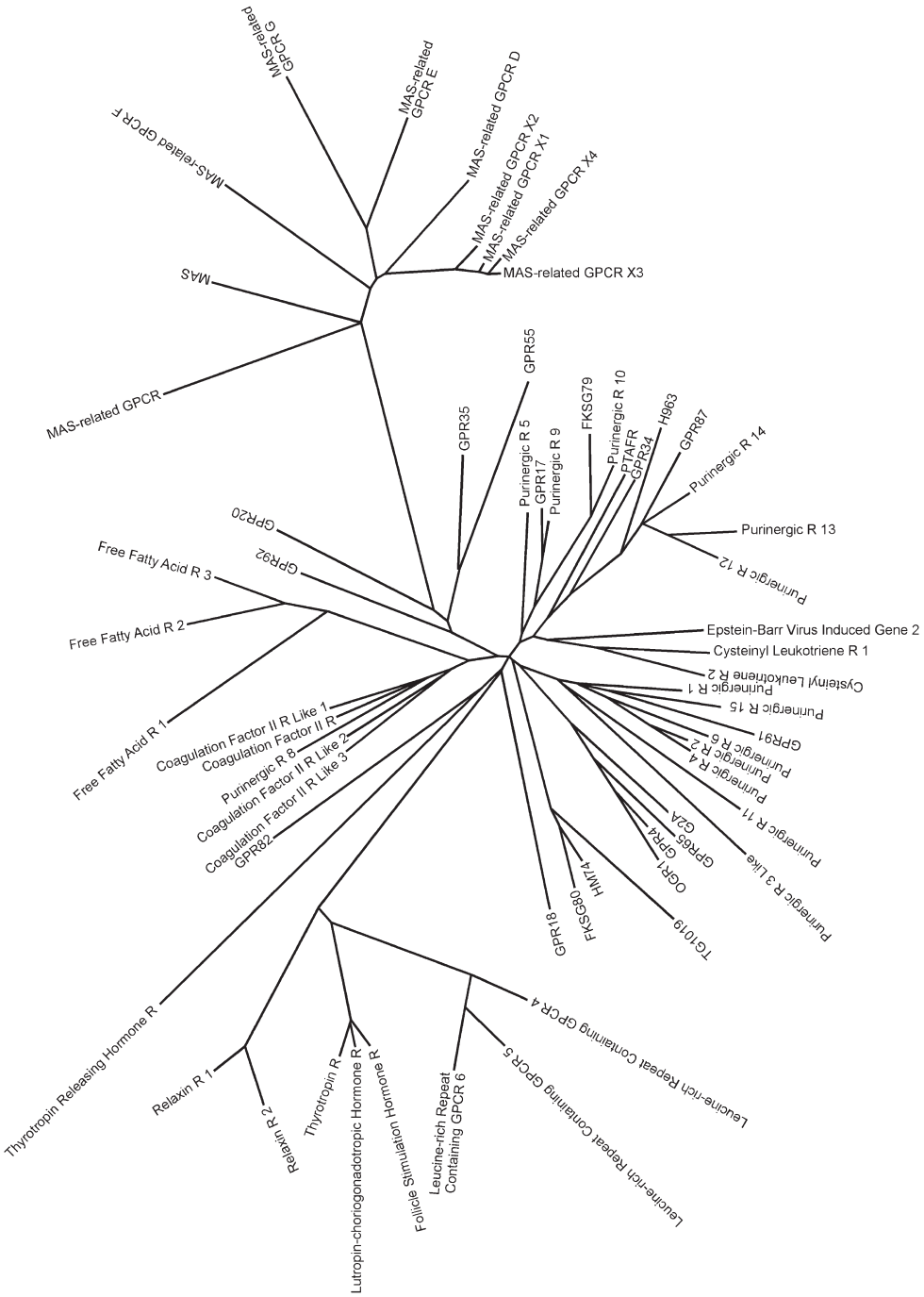

Fig. 1.9 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

1.6.4.2 The Glycoprotein Receptor Cluster

This cluster of receptors contains the classical glycoprotein hormone receptors, follicle stimulating hormone receptor, tyrotropin releasing hormone receptor and the leutropin-choriogonadotropin hormone receptor. It also contains the two relaxin receptors and three orphan receptors. The receptors in this cluster are different from all other Rhodopsin receptors because they have long N-termini which contain the ligand recognition domain, in common with most non-Rhodopsin GPCRs. This cluster also contains the two orphan receptors GPR 18 and GPR 82.

1.6.4.3 The Coagulation Factor Receptor Cluster

This cluster contains four receptors for coagulation factors, one purinergic receptor and a separate branch with the three free fatty acid binding receptors (previously known as GPR 40, 41 and 43) [35–37].

1.6.4.4 The Purinergic Receptor Cluster

This large branch consists of 13 nucleotide binding, or purinergic, receptors (P2Ys), the platelet activating factor receptor, the two cysteinyl leuktriene receptors and 14 orphan GPCRs and hence the known ligands for the receptors in this group are extracellular nucleotides, leukotrienes and thrombin. Considering the close phylogenetic relationship between the receptors with known ligands and the large number of orphan receptors it is likely that additional purinergic receptors will be discovered among the orphan GPCRs in the Purinergic cluster.

1.6.4.5 The Olfactory Receptor Cluster

The olfactory receptors clearly belong to the Rhodopsin γ-group using both phylogenetic and sequence clustering methods and our phylogenetic analysis indicates that the olfactory receptors form a stable phylogenetic cluster, which does not overlap with other groups of the rhodopsin family or with other families. Our previous searches in the human genome databases indicated that there could be over 460 olfactory receptors in the human genome that we consider likely to represent unique functional receptors [9]. A total of 347 putative human full-length odorant receptor genes have previously been identified and physically cloned [25]. It has also been suggested that there are over 900 olfactory receptor-like sequences in the human genome [2]. Further work is needed to identify the total number of olfactory receptors.

1.6.4.6 Other Rhodopsin α-Receptors

As is the case with the Rhodopsin α and γ groups, the Rhodopsin δ group contains four receptors that cannot be positioned phylogenetically on any branch of the tree. These are the orphan receptors GPR 35, GPR 55, GPR 20 and GPR 92.

1.7
Other GPCRs

There are currently 19 GPCRs that either cannot logically be placed into any of the main families (eight) or are likely to belong to the Rhodpsin family but cannot with certainty be placed into any of the four groups (11). Of these putative rhodopsin receptors six (V1RL1, V1RL2, V1RL4, FKSG46, FKSG83 and hGPCR23) are clearly related to each other and hence form a small "cluster". These are all related to vomeronasal receptors type 1, a family of GPCRs with over 100 members in for example the mouse and rat genomes. As humans do not appear to have a vomeronasal system for pheromone detection it is presently unclear whether these five human receptors have a physiological function, although they appear to be functional based on their amino acid sequence, or if they perform some other function in humans unrelated to pheromone detection. The other receptors putatively belonging to the Rhodopsin family are the Duffy receptor, known to be involved in immune system modulation and the four orphan receptors GPR 88, GPR 142, GPR 160 and hGPCR 19. Of the non-Rhodopsin receptors in the "Other" group two pairs of related receptors can be seen. TM7SF1 and C11ORF4 are clearly related to each other as are PERVAR1 and PERVAR2. Interestingly none of these receptors has any detectable sequence identity to any GPCR outside its pair and the same is true for the last four "Other receptors", IEDA, OA1, hGPCR29 and hGPCR43.

1.8
Future Perspective

There are currently 19 GPCRs that either can not be placed clearly into any of the main families (eight) or are likely to belong to the *Rhodopsin* family but can not with certainty be placed into any of the four groups (eleven). Much work has recently been carried out to organize the "functional" protein coding gene repertoire of GPCRs in the human genome. There is however considerable work remaining. Even though most of the genes have been identified, their exact genomic structure, including all alternatively spliced isoforms, needs to be worked on, in particular for those receptors that have recently been identified. Which genes are pseudogenes and which are not? Do many of the pseudogenes have a functional role even though they do not look like GPCRs with intact seven transmembrane regions? Our recent understanding of RNA genes has also emphasized the need to adopt a more "open minded" approach to what is functional and what is not. This is an important problem among those groups of receptors that are either rapidly expanding or shrinking in numbers in the human genome. Some groups such as the olfactory receptors clearly need more work while the exact domain composition of other groups such as the relatively newly established adhesion family also need more fine tuning. Progress is being made on the repertoire of GPCR in other genomes. In a recent article we created Hidden Markov Models

based on the different groups of human GPCRs and added several other models based on receptors not found in mammals [9, 38]. We searched the entire Genscan datasets from 13 species whose genomes are nearly completely sequenced. We reported over 5000 unique GPCRs that were divided into 15 main groups, and that the Rhodopsin family was subdivided into 13 subclasses. This chapter shows that all the five main families in the human genome arose prior to the split of nematodes from the chordate lineage and that several of the subgroups of the Rhodopsin family arose prior to the split of lineage which led to the vertebrates. We are currently working on more detailed analysis in each species with emphasis on the mouse, rat, and chicken genomes. This will provide a better understanding of the origin of the human GPCRs and the subgroup-specific changes that have occurred during the last 300 million years.

Table 1.1 All 19 receptors in the group "Other" were searched against a database, using BLASTP with a cut-off at 0.01, containing only all human GPCRs and the best non-self hit was recorded. The first column shows the overall best hit and the second the best hit when all receptors from the "Other" group of are excluded. E values from the BLASTP searches are presented in parentheses. A putative family assignment based on the hit in the second column is also presented.

Receptor name	Best human BLAST-hit	Best human BLAST-hit excluding "Other""	Putative main family
C11ORF4	TM7SF1 (3.0e-076)	NA	NA
DUFFY	F2RL2 (0.001)	F2RL2 (0.001)	Rhodopsin (δ)
FKSG46	V1RL2 (8.4e-082)	NA	Rhodopsin (δ)
FKSG83	FKSG46 (1.2e-031)	TRHR (2.1e-004)	Rhodopsin (δ)
GPR 142	GPR139 (1.4e-062)	GPR139 (1.4e-062)	Rhodopsin (γ)
GPR 160	TAR5 (1.2e-004)	TAR5 (1.2e-004)	Rhodopsin (α)
GPR 88	OPRL1 (1.7e-005)	OPRL1 (1.7e-005)	Rhodopsin (γ)
IEDA	NA	NA	NA
OA1	VIPR2 (3.0e-005)	GPR144 (0.006)	NA
PERVAR1	PERVAR2 (1e-139)	NA	NA
PERVAR2	PERVAR1 (3e-148)	NA	NA
TM7SF1	C11ORF4 (2.1e-096)	CCRL2 (0.001)	NA
V1RL1	FKSG46 (7.8e-071)	TRHR (1.5e-005)	Rhodopsin (δ)
V1RL2	V1RL4 (9.5e-089)	BRS3 (2.6e-006)	Rhodopsin (β)
V1RL4	V1RL2 (4.0e-089)	EDNRB (6.8e-004)	Rhodopsin (β)
hGPCR 19	EDG6 (6.2e-006)	EDG6 (6.2e-006)	Rhodopsin (α)
hGPCR 23	FKSG46 (4.0e-015)	NA	Rhodopsin (δ)
hGPCR 29	NA	NA	NA
hGPCR 43	NA	NA	NA

Table 1.2 The number of GPCRs in each of the main families. The table also shows the number of receptors in each family that recognize ligands of the types, peptide, biogenic amine, lipids and other. The last column designates the number of orphan receptors in each of the families.

Group	Number	Peptide ligand	Biogenic amine ligand	Lipid ligand	Other ligand	Orphan
Adhesion	33	0	0	0	1	32
Secretin	15	15	0	0	0	0
Frizzled	11	0	0	0	11	0
Taste type 2	26	0	0	0	1	25
Glutamate	22	0	10	0	4	8
Rhodopsin (α)	101	7	40	22	7	25
Rhodopsin (β)	42	32	0	0	0	10
Rhodopsin (γ)	72	54	0	0	2	18
Rhodopsin (δ)	63	17	0	6	14	26
Other	20	1	0	0	0	19
Total	407	126	50	28	40	163

Acknowledgments

We thank David E. Gloriam, Thóra K. Bjarnadóttir and Malin C. Lagerström all at Uppsala University, for providing insight into their preliminary results regarding detailed analysis, assembly and annotation of several subgroups of GPCRs.

Our studies were supported by the Swedish Research Council (VR, medicine), the Swedish Society for Medical Research (SSMF), Svenska Läkaresällskapet, Åke Wikberg Foundation, The Novo Nordisk Foundation, Thurings Foundation, and Magnus Bergwall Foundation.

References

1 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., et al. Initial sequencing and analysis of the human genome. *Nature* **2001**; *409*: 860–921.

2 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., et al. The sequence of the human genome. *Science* **2001**; *291*: 1304–1351.

3 Fredriksson, R., Schioth, H. B. The repertoire of G-protein coupled receptors in fully sequenced genomes. *Mol Pharmacol* **2005**; *67*: 1414–1425.

4 Attwood, T. K., Findlay, J. B. Fingerprinting G-protein-coupled receptors. *Protein Eng* **1994**; *7*: 195–203.

5 Kolakowski, L. F., Jr. GCRDb: a G-protein-coupled receptor database *Receptors Channels* **1994**; *2*: 1–7.

6 Josefsson, L. G. Evidence for kinship between diverse G-protein coupled receptors. *Gene* **1999**; *239*: 333–340.

7 Flower, D. R. Modelling G-protein-coupled receptors for drug design. *Biochim Biophys Acta* **1999**; *1422*: 207–234.

8 Bockaert, J., Pin, J. P. Molecular tinkering of G protein-coupled receptors:

an evolutionary success. *EMBO J* **1999**; *18*: 1723–1729.

9 Fredriksson, R., Lagerstrom, M. C., Lundin, L. G., Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**; *63*: 1256–1272.

10 Fredriksson, R., Lagerstrom, M. C., Hoglund, P. J., Schioth, H. B. Novel human G protein-coupled receptors with long N-terminals containing GPS domains and Ser/Thr-rich regions. *FEBS Lett* **2002**; *531*: 407–414.

11 Fredriksson, R., Gloriam, D. E., Hoglund, P. J., Lagerstrom, M. C., Schioth, H. B. There exist at least 30 human G-protein-coupled receptors with long Ser/Thr-rich N-termini. *Biochem Biophys Res Commun* **2003**; *301*: 725–734.

12 Bjarnadottir, T. K., Fredriksson, R., Hoglund, P. J., Gloriam, D. E., Lagerstrom, M. C., Schioth, H. B. The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics* **2004**; *84*: 23–33.

13 Fredriksson, R., Hoglund, P. J., Gloriam, D. E., Lagerstrom, M. C., Schioth, H. B. Seven evolutionarily conserved human rhodopsin G protein-coupled receptors lacking close relatives. *FEBS Lett* **2003**; *554*: 381–388.

14 McKnight, A. J., Gordon, S. The EGF-TM7 family: unusual structures at the leukocyte surface. *J Leukoc Biol* **1998**; *63*: 271–280.

15 Lin, H. H., Stacey, M., Hamann, J., Gordon, S., McKnight, A. J. Human EMR2, a novel EGF-TM7 molecule on chromosome 19p13.1, is closely related to CD97. *Genomics* **2000**; *67*: 188–200.

16 Harmar, A. J. Family-B G-protein-coupled receptors. *Genome Biol* **2001**; *2*: REVIEWS3013.

17 Eichler, W. CD97 isoform expression in leukocytes. *J Leukoc Biol* **2000**; *68*: 561–567.

18 Chey, W. Y., Chang, T. M. Secretin, 100 years later. *J Gastroenterol* **2003**; *38*: 1025–1035.

19 Blair Hedges, S., Kumar, S. Genomic clocks and evolutionary timescales. *Trends Genet* **2003**; *19*: 200–206.

20 Slusarski, D. C., Corces, V. G., Moon, R. T. Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling. *Nature* **1997**; *390*: 410–413.

21 Kristiansen, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol Ther* **2004**; *103*: 21–80.

22 Chandrashekar, J., Mueller, K. L., Hoon, M. A., Adler, E., Feng, L., Guo, W., Zuker, C. S., Ryba, N. J. T2Rs function as bitter taste receptors. *Cell* **2000**; *100*: 703–711.

23 Bufe, B., Hofmann, T., Krautwurst, D., Raguse, J. D., and Meyerhof, W. The human TAS2R16 receptor mediates bitter taste in response to beta-glucopyranosides. *Nat Genet* **2002**; *32*: 397–401.

24 Ryba, N. J., Tirindelli, R. A new multigene family of putative pheromone receptors. *Neuron* **1997**; *19*: 371–379.

25 Zozulya, S., Echeverri, F., Nguyen, T. The human olfactory receptor repertoire. *Genome Biol* **2001**; *2*: RESEARCH0018.

26 Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., et al. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

27 Lagerstrom, M. C., Klovins, J., Fredriksson, R., Fridmanis, D., Haitina, T., Ling, M. K., Berglund, M. M., Schioth, H. B. High affinity agonistic metal ion binding sites within the melanocortin 4 receptor illustrate conformational change of transmembrane region 3. *J Biol Chem* **2003**; *278*: 51521–51526.

28 Berglund, M., Fredriksson, R., Salaneck, E., Larhammar, D. Reciprocal mutations of neuropeptide Y receptor Y2 in human and chicken identity amino acids important for agonist binding. *FEBS Lett* **2002**; *518*: 5–9.

29 Haitina, T., Klovins, J., Andersson, J., Fredriksson, R., Lagerstrom, M. C., Larhammar, D., Larson, E. T., Schioth, H. B. Cloning, tissue distribution, pharmacology and three-dimensional modelling of melanocortin receptors 4 and 5 in rainbow trout suggest close evolutionary relationship of these subtypes. *Biochem J* **2004**; *380*: 475–486.

30 Baldwin, J. M. Structure and function of receptors coupled to G proteins. *Curr Opin Cell Biol* **1994**; *6*: 180–190.

31 Nathans, J., Thomas, D., Hogness, D. S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **1986**; *232*: 193–202.

32 Thomson, P. A., Wray, N. R., Thomson, A. M., Dunbar, D. R., Grassie, M. A., Condie, A., Walker, M. T., Smith, D. J., et al. Sex-specific association between bipolar affective disorder in women and GPR50, an X-linked orphan G protein-coupled receptor. *Mol Psychiatry* **2004**; *10*: 470–478.

33 Uhlenbrock, K., Gassenhuber, H., Kostenis, E. Sphingosine 1-phosphate is a ligand of the human gpr3, gpr6 and gpr12 family of constitutively active G protein-coupled receptors. *Cell Signal* **2002**; *14*: 941–953.

34 Lembo, P. M., Grazzini, E., Groblewski, T., O'Donnell, D., Roy, M. O., Zhang, J., Hoffert, C., Cao, J., Schmidt, R., Pelletier, M., Labarre, M., Gosselin, M., Fortin, Y., Banville, D., Shen, S. H., Strom, P., Payza, K., Dray, A., Walker, P., Ahmad, S. Proenkephalin A gene products activate a new family of sensory neuron-specific GPCRs. *Nat Neurosci* **2002**; *5*: 201–209.

35 Nilsson, N. E., Kotarsky, K., Owman, C., Olde, B. Identification of a free fatty acid receptor, FFA2R, expressed on leukocytes and activated by short-chain fatty acids. *Biochem Biophys Res Commun* **2003**; *303*: 1047–1052.

36 Kotarsky, K., Nilsson, N. E., Flodgren, E., Owman, C., Olde, B. A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochem Biophys Res Commun* **2003**; *301*: 406–410.

37 Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., et al. The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem* **2003**; *278*: 11312–11319.

38 Perez, D. M. From Plants to Man: The GPCR „Tree of Life" (Relates to article by Fredriksson, et al., FastForward 1 February 2005). *Mol Pharmacol* **2005**; *67*: 1383–1384.