A Chemogenomic Analysis of the Transmembrane Binding Cavity of Human G-Protein-Coupled Receptors

Jean-Sebastien Surgand, Jordi Rodrigo, Esther Kellenberger, and Didier Rognan*

Bioinformatics of the Drug, CNRS UMR7081, Illkirch, France

ABSTRACT The amino acid sequences of 369 human nonolfactory G-protein-coupled receptors (GPCRs) have been aligned at the seven transmembrane domain (TM) and used to extract the nature of 30 critical residues supposed—from the X-ray structure of bovine rhodopsin bound to retinal—to line the TM binding cavity of ground-state receptors. Interestingly, the clustering of human GPCRs from these 30 residues mirrors the recently described phylogenetic tree of full-sequence human GPCRs (Fredriksson et al., Mol Pharmacol 2003;63:1256–1272) with few exceptions. A TM cavity could be found for all investigated GPCRs with physicochemical properties matching that of their cognate ligands. The current approach allows a very fast comparison of most human GPCRs from the focused perspective of the predicted TM cavity and permits to easily detect key residues that drive ligand selectivity or promiscuity. Proteins 2006;62:509–538.

© 2005 WILEY-LISS, Inc.

Key words: amino acid sequences; human G-protein-coupled receptors; transmembrane domain; bovine rhodopsin

INTRODUCTION

G protein-coupled receptors (GPCRs) constitute a large superfamily of very heterogeneous membrane receptors characterized by a typical heptahelical membrane-spanning fold usually described as a seven-transmembrane (TM) domain.1,2 A striking feature of this protein superfamily is the tremendous chemical diversity of possible ligands including light, small molecular-weight ions (e.g., glutamate, Ca2+), biogenic amines (e.g., dopamine, serotonin), nucleosides and nucleotides (e.g., adenosine, adenosine triphosphate), peptide and protein hormones (e.g., chemokines, glucagon), lipids and eicosanoids (e.g., sphingolipids, prostaglandins).3 Activation of GPCRs upon ligand binding induces a conformational change of the receptor, thereby triggering a specific interaction with intracellular G proteins and subsequent activation/inhibition of secondary messengers.4 Because of the ubiquitous distribution of GPCRs at the surface of many cells, these receptors are regulating a wide array of physiological and pathological processes. As a consequence, GPCRs are particularly attractive targets for therapeutic intervention. Hence, about 30% of top-selling drugs modulate the activity of this family of receptors.5 Until now, few GPCRs (ca. 40) have been targeted by existing drugs. Analyzing human genomic sequences predicts the existence of about 400 nonolfactory GPCRs5 and opens a new avenue for drug discovery, especially with respect to the 100 orphan receptors for which even the endogenous ligand still has not been characterized.6

Historically, GPCRs have been classified according to the chemical nature of their ligands, their specificity for known agonists/antagonists and the pharmacology associated with their activation/inhibition. The knowledge of the amino acid sequence of cloned GPCRs widened the classification to more receptors and resulted in a standard classification into three different families of mammalian GPCRs (from A to C, or I to III) depending on the alignability of the corresponding amino acid sequences.7,8 Because of their typical organization in seven membrane-spanning α-helices, discriminating GPCRs and more generally 7-TM receptors from other genomic targets is rather straightforward.9 Classifying GPCRs into subfamilies has been addressed by several methods including simple distance-based neighbor joining,10 support vector machines,11 hidden Markov models,12 amino acid fingerprints,13 covariant-discriminant analysis of amino acid composition,14 and alignment-independent extraction of principal chemical properties of amino acid sequences.15 However, previous clustering analyses, despite methodological merits, did not address the full dataset of human GPCRs. Recently, an exhaustive phylogenetic analysis of 342 mono-functional nonolfactory GPCRs9 led to a revised classification consisting of five main families named glutamate (G family), rhodopsin (R), adhesion (A), frizzled/taste2 (F), and secretin (S). The classification can be used to search for similarities/dissimilarities among selected receptors. This information is crucial in drug discovery for many reasons: (1) it can be used to prioritize binding studies of a given ligand to GPCRs in order to address the ligand...
selectivity as early as possible. (2) It helps to locate poorly studied or orphan targets in the GPCR universe by predicting the druggability of selected GPCRs in the light of known data on already investigated receptors. However, establishing a direct link between such phylogenetic trees and structure-based drug design is not straightforward. It requires an exact alignment of amino acid sequences, an accurate identification of amino acids lining the binding cavity and a 3-D model for every GPCR. Looking at similarities or differences at the binding sites would provide information that can be easily translated into ligand structures. Up to now, this chemogenomic approach has only been applied to a small subset of GPCRs for which enough experimental information is available.

In the current study, we present an easy and straightforward classification scheme based on the alignment of 30 critical GPCR positions supposed to line the TM-binding cavity for inverse agonists or antagonists. The analysis, applied to a set of 369 nonredundant nonolfactory human GPCRs, provides a binding site-driven phylogenetic tree, as well as precise 2D and 3D structural determinants for each cluster that can be translated into structure-based ligand design.

MATERIALS AND METHODS

Input Data

Receptor sequences were extracted from the UniProt release 3.1 and the NCBI Entrez Protein database and parsed in plain text format, and further in XML format. First, only human proteins were selected by the field name “OS” (plain text version) or “organism” (XML version) containing the value “Human.” Only GPCRs were retained, by checking the field “KW” or “keyword” with the value “G protein-coupled receptor.” Olfactory receptors were then suppressed from the receptors list by checking the same field with the value “Olfaction.” Last, receptors fragments were removed from the dataset, by looking whether the case-independent word “fragment” appeared in the description field (“DE” or “protein name”). Throughout this study, the UniProt entry name has been used to describe each receptor. Therefore, the entry name described herein does not necessarily map the gene name coding for the corresponding receptor, especially for orphan receptors whose official gene’s name usually begins with GPR (e.g., GPR110). In the UniProt nomenclature, entries registered in the Swiss-Prot database are given a four or five character identifier: GPRx (x ≤ 9), GPRxx (10 ≤ xx ≤ 99) or GPxxx (xxx ≥ 100).

Amino Acid Alignment of the 7-TMs

The GPCRmod program was used to align the 7-TMs of selected human GPCRs, as recently described. Briefly, GPCRmod first predicts the rough location of TMs using the TMHMM algorithm and then, in each isolated TM, looks for family and TM-specific amino acid patterns. Upon family detection (rhodopsin, secretin, glutamate, frizzled), the query TM is then aligned with that of the corresponding family template (rhodopsin: bovine rhodopsin, secretin: human calcitonin receptor, glutamate: human calcium-sensing receptor, frizzled: human frizzled type 1) assuming TM lengths similar to that depicted in the X-ray structure of bovine rhodopsin. No attempts were made to align either N- or C-terminal domains or intra- and extracellular loops. A filtering procedure was then applied to suppress duplicate 7-TM sequences, usually keeping the Swiss-Prot entry and removing the TrEMBL duplicates. The final GPCR dataset contained 369 entries when the manuscript was prepared (see Annex 1 in Supplementary Data).

Clustering Human GPCRs From 30 Discontinuous Amino Acids

The selection of the 30 amino acids was performed as follow: in the X-ray structure of the bovine rhodopsin–retinal complex (pdb entry 1f88), we defined the receptor cavity as the collection of 81 residues enclosed within a 10-Å sphere centered on the bound ligand. The solvent-accessible surface was then computed for the 81 selected amino acids. We finally picked out the 30 residues (Fig. 1) that present at least 25% of their surface accessible to a putative ligand and whose side chain is pointing inward the 7-TM bundle. Each GPCR was then described by an ungapped sequence of 30 residues. The resulting 30-aminocid-long sequences were hierarchically clustered using the UPGMA method to yield a rooted phylogenetic tree. Protein pair-wise distances were measured by computing sequence identity. The statistical significance of the obtained binary tree was assessed by bootstrapping. The starting dataset (369 sequences of 30 residues) was bootstrapped 1,000 times by randomly selecting the sequences input order and allowing substituting not more than 10% of the starting dataset by randomly selected entries. The 1,000 bootstrapped datasets were clustered as previously described to yield 1,000 binary trees, which after normalization were used to derive a consensus tree using the program CONSENSE, from the PHYLIP package (http://cmgm.stanford.edu/phylip/). Trees were visualized with the NJplot program.

3D Model Building and Cavity Detection

A 3D model of each GPCR in its ground state was obtained using the GPCRmod program as recently described. Briefly, the GPCRmod alignment of the 7 TMs was converted into a 3D model using eight backbone templates originating from previously described homology models and two rotamer libraries for side chain positioning. For each model, all possible channels and cavities were detected using the MOLCAD module (Multi channel function) of the SYBYL package. Connolly surfaces and channels were calculated using a 1.4-Å radius probe and a dot density of six points/area. The biggest cavity was finally selected and its surface and volume computed.

RESULTS AND DISCUSSION

Classifying GPCRs According to the Discontinuous Sequence of Amino Acids that Form the TM Ligand Binding Site

The current study proposes a straightforward classification of most druggable human GPCRs from the point of
view of a putative ligand. To achieve this task, 30 critical amino acids disseminated on the 7-TM domain were extracted for each receptor (Fig. 1).

Our clustering approach implies two assumptions: (1) the overall fold of the 7-TM domain around the binding crevice has been conserved along evolution; (2) critical hotspots spread over the 7-TM domain repeatedly account for ligand binding. Although very few structural information is available for the three most important GPCR classes (Classes A, B, C), numerous experimental data do provide evidence in favor of strong structural similarities among many GPCRs: (1) most of the residues known to affect small molecular-weight ligand binding to unrelated GPCRs are included in the selected 30 residues, suggesting a common architecture of the TM pocket, (2) many known ligands are promiscuous for even unrelated GPCRs and are usually anchored through so-called privileged structures to common features of different GPCRs. Of course, class B and class C GPCRs exhibit an additional orthosteric site located outside the 7-TM bundle. Therefore, conclusions drawn in the present paper only apply to the 7-TM binding site.

**TM-Cavity Based Clustering of 369 Human GPCRs Recalls Classification Based on Full TM Sequences**

Almost all nonredundant druggable human GPCRs have been analyzed in the present study. Forty-six putative GPCRs (see Annex 2 in supplementary data) were rejected from our dataset for three main reasons: (1) seven TMs could be detected but no satisfactory alignment with other GPCRs could be found (e.g., Taste 2 receptors), (2) less/more than seven TMs could be detected (e.g., GPR172A), (3) no TM domain at all could be detected (e.g., CRCP_HUMAN).

The TM cavity-derived phylogenetic tree obtained from 369 human nonredundant nonolfactory GPCRs is shown in Figure 2. A total of 22 clusters were defined in order to encompass the maximum number of related entries within a branch characterized by the highest possible bootstrap value. Thirty-four out of 369 entries could not be assigned to any of the existing 22 clusters (Annex 1). Instead of generating very small-sized additional clusters, we prefer to define them as singletons unrelated to any of the current clusters. The resulting tree is very similar to the most complete phylogenetic tree (GRAFS classification) known to date, although the latter has been obtained from full TM sequences. In both classifications, GPCRs of the Frizzled, Glutamate, Secretin, and Adhesion families cluster in well separated groups. The large Rhodopsin family is split into 18 different clusters. Remarkably, all known GPCR subfamilies (e.g., receptors for biogenic amines, purines, and chemokines) are reproduced with high bootstrap support. The five main families (Glutamate, Rhodopsin, Adhesion, Frizzled, Secretin) reported in the GRAFS classification are recovered with no overlaps between the corresponding clusters. The single exception is a rhodopsin-like GPCR (GPR88) which clusters with class C GPCRs. Interestingly, receptors for which the orthosteric binding site is not located in the TM domain (Adhesion, Secretin, and Glutamate families) are nevertheless grouped into homogeneous clusters. Furthermore, we have detected a putative binding site in the TM cavity of each of these
receptors. We assume that a strong evolutionary pressure has been applied to the 7-TM domain of class B and C GPCRs to maintain a TM cavity whose function, besides triggering G-Protein coupling, could be to define a putative binding site for endogenous inverse agonists or allosteric regulators. To ensure the statistical significance of class B and C GPCRs clustering, a new tree was computed by selecting the same number of amino acids (30), the same proportion of residues per TM, but from amino acids whose side chains do not point inward the TM cavity. No homogeneous clusters could be found in that case suggesting that the herein selected 30 residues do have a clear functional role.

In the next sections, the physicochemical properties of the predicted TM cavity will be described for each cluster. Where possible, relationships were established between the TM cavity and their respective ligands. For sake of clarity, GPCRs are identified by their UniProt entry names (see Annex 1 for an exhaustive description of all entries) and TM residues by the Ballesteros numbering.32 As expected from the reduced set of templates used for comparative modelling, the modeled TM cavity of most GPCRs share significant similarities, notably the presence of two subsites, a first one delimited by TMs 1, 2, 3, and 7 (site 1, Fig. 1), and a second one delimited by TMs 3, 4, 5, and 6 (site 2, Fig. 1). Throughout this study, subsites 1 and 2 will refer to this definition.

The Frizzled Receptor Cluster (11)

The frizzled receptor cluster regroups 11 receptors (10 frizzled receptors, smoothened) basically involved in cell growth and proliferation [Fig. 3(A)]. Known ligands for the frizzled receptors are Wnt proteins33 and are supposed to bind to the N-terminal cystein-rich domain of frizzled receptors. The cluster is identical to the frizzled family in the GRAFS classification. The alignment shown in Figure 3(B) indicates six fully conserved residues (Phe3.29, Tyr3.33, Phe5.38, Val5.39, Pro5.42, and Ile7.45). The predicted 3D structure is characterized by a large TM cavity (about 1,500 Å³) extending very deep toward the intracellular side between TMs 5 and 6. Interestingly, TMs 1–4 greatly contribute to the cavity through hydrophilic and charged residues whereas TMs 5–7 bring small aliphatic amino acids [Fig. 3(B)]. Although TM-binding endogenous ligands of this GPCR family are still elusive, the herein described topology of the TM cavity perfectly matches the chemotype of known smoothened synthetic antagonists34,35 which consists in bulky hydrophobic molecules with a cationic head.

The Secretin Receptor Cluster (15)

The secretin receptor cluster is made of 15 GPCRs belonging to the class of secretin-like receptors [Fig. 4(A)]. This very homogeneous cluster has recently been proposed to form a homogeneous GPCR family.16 Ligands of GPCRs belonging to this cluster are peptide hormones supposed to bind to the long N-terminal extracellular tail via their C-terminal residues and to contact the TM cavity through their N-terminal residues.36,37 However, a typical GPCR from this cluster (e.g., CALCR) presents a narrow and hydrophobic TM cavity (ca. 1,000 Å³) with a set of residues (Ile/Val1–39, Phe/Leu7–58, Asp2–61, Ile/Leu/Val13–36, Phe/
Tyr3.40, Trp4.60, Ile/Val5.42, Ile5.43, Pro5.46, Phe6.51, Ile/Leu/Met7.35, Ser7.43 and Gln7.45) rather conserved throughout the cluster [Fig. 4(B)]. The TM cavity is predicted to be large enough to accommodate small molecular-weight ligands (notably though a salt bridge to Asp2.61) that could inhibit the action of the natural hormone at the remote binding site.

It remains difficult to probe 3D models of this GPCR cluster because of the paucity of known nonpeptide antagonists. For the glucagon receptor (GLR) however, there are several small molecular weight antagonists available.3 8–4 0 The GLR cavity presents a rather hydrophilic centre (Gln3.33, Ser7.43, Gln7.45) surrounded by two hydrophobic subsites. Consensus positions between TMs 1 and 2 (Val1.39, Ser1.42, Gly1.46, Val2.57, Leu2.58) as well as residues not taken into account by our clustering scheme (Phe 2.50, Ala2.54) define a narrow and deep subsite 1. Subsite 2 is a large pocket formed by TMs 4, 5 and 6 (Tyr3.40, Trp4.60, Phe5.39, Ile5.42, Pro5.46, and Phe6.51). The TM cavity of a few class B GPCRs (CRFR1, GLR, CGRP, GLP1) is probably an allosteric binding site for nonpeptide antagonists that precludes for the activation of the receptor by endogenous peptide agonists.41 Interestingly, site-directed mutagenesis suggests that Phe184 2.50 and Tyr239 3.40 are

### Fig. 3. The frizzled receptor cluster: (A) Phylogenetic tree. Branch lengths are proportional to the distance between two receptors. Number in italic represent bootstrap values. (B) Cavity-based alignment of 30 discontinuous positions colored according to the degree of similarity (white foreground/black background, 100%; white foreground/grey background, > 80%, black foreground/grey background, > 60%). Receptors are labeled according to their UniProt entry names and amino acid positions identified by the Ballesteros numbering scheme.32 SMO: Smoothened.

### Fig. 4. The secretin receptor cluster: (A) Phylogenetic tree. (B) Cavity-based alignment. CALCR: Calcitonin receptor; CALRL: Calcitonin gene-related peptide type 1 receptor; CRFR1: Corticotropin releasing factor receptor; CRFR2: Corticotropin releasing factor receptor 2; GHRHR: Growth hormone-releasing hormone receptor; GIPR: Gastric inhibitory polypeptide receptor; GLP1R: Glucagon-like peptide 1 receptor; GLP2R: Glucagon-like peptide 2 receptor; GLR: Glucagon receptor; PACR: Pituitary adenylate cyclase activating polypeptide type I receptor; PTHR1: Parathyroid hormone receptor 1; PTHR2: Parathyroid hormone receptor 2; SCTR: Secretin receptor; VIPR1: Vasoactive intestinal polypeptide receptor 1; VIPR2: Vasoactive intestinal polypeptide receptor 2.
involved in the binding of the nonpeptide GLR antagonist L-168049 whose binding site does not overlap that of the endogenous hormone.\(^3\) Point mutation of position 3.36 in the CRF\(_1\) receptor has also been shown to affect the binding of a nonpeptide CRF\(_1\) antagonist.\(^4\) Despite some warnings about the druggability of the allosteric TM binding site, notably regarding its hydrophobicity,\(^31\) it is very likely that the potential of this GPCR cluster has not been fully addressed yet.

**The Adhesion Receptor Cluster (33)**

Members of the adhesion receptor cluster [Fig. 5(A)] are related to the secretin receptor cluster as being part of the former secretin-like GPCRs, but are unambiguously clustered with high bootstrap values into a separate family, in agreement with the GRAFS classification.\(^36\) These GPCRs exhibit specific repeats at the N-terminal extracellular domain (EGF, cadherin, mucin) likely to be involved in cell adhesion processes.\(^43\) Many of these GPCRs are expressed at the cell surface as heterodimers consisting of a large N-terminal domain associated with the 7-TM domain. A unique feature of some of these GPCRs is the intracellular location of their ligands (e.g., CD55, chondroitin sulfate) that at first associate with the extracellular domain.\(^43\)

The binding cavity of adhesion GPCRs resembles that of the former cluster of secretin-like hormone receptors. It is of limited volume (around 1,000 Å\(^3\)) and centered on a conserved histidine residue [His\(^3.33\), Fig. 5(B)]. Side chains of conserved aromatic amino acids at 3.36 and 3.40 limit the width of the cavity center. Medium-sized amino acids of both TM1 and TM2 (Ala, Val, Leu, and Ile) participate to the formation of a deep hydrophobic subsite. Depending on position 5.43, the cavity extents towards TMs 5 and 6. In about a half of these receptors, a Phe at 5.43 restricts the depth of the cavity between TMs 5 and 6. Smaller residues are found at position 5.43 in the rest of adhesion-like GPCRs and permit the extension of the TM cavity towards TMs 5 and 6. By analogy to most rhodopsin-like GPCRs, an important Trp is often present at position 6.48. In rhodop-
sin-like GPCRs, the residues at position 6.48 of the TM cavity has been proposed to act as a molecular switch for locking the GPCR either in a ground state or in an activated form.\(^{32}\) It is tempting to speculate the involvement of the conserved Trp\(^{6.48}\) in the conformational changes occurring upon activation of this cluster of receptors as well.

No small molecular-weight ligands have ever been described for this GPCR cluster. Therefore, it is currently impossible to match the proposed cavity description to known ligand chemotypes.

### The Glutamate Receptor Cluster (23)

The glutamate receptor cluster regroups members of the formerly known class C GPCRs. Class C RCPGs are characterized by a long N-terminal domain which folds into a bilobed venus-fly trap domain delimiting the orthosteric binding site of the endogenous ligand (GABA, glutamate, Ca\(^{2+}\)).\(^{45}\) Their 7-TM domain is responsible for receptor activation and G protein coupling\(^{46}\) and contains a cavity that is known to bind positive or negative allosteric modulators.\(^{47,48}\) Bootstrap values clearly indicate the presence of three separated branches [Fig. 6(A)].

A typical 7-TM cavity of a GPCR from this cluster has a volume of ca. 1,500 Å\(^3\) and can be viewed as two hydrophobic subsites linked by charged and/or hydrophilic residues at positions 3.28 and 3.29. Looking at the eight metabotropic glutamate receptors, the ligand-accessible residue at position 3.28 is a glutamine in MGR1 and MGR5 whereas it is an arginine in the six others receptors. This may account for the functional similarity of MGR1 and MGR5. Conserved hydrophobic residues in TM1 (position 1.42), TM2 (position 2.58), TM3 (position 3.32), and TM7 (positions 7.43 and 7.45) delimit subsite 1. Subsite 2 is lined by conserved aromatic residues in TM3 (position 3.40) and TM6 (positions 6.48, 6.51 and 6.55). Many of these amino acids have been experimentally found to map the binding site of both positive and negative allosteric modulators of GPCRs belonging to this cluster (MGR1, MGR2, MGR5, CASR).\(^{47–50}\) Moreover, recent findings proved that the MGR5 receptor does not require the N-terminal venus fly-trap module for the recognition of allosteric modulators.\(^{46}\) Hence, receptors of the current cluster probably comprise at least two binding sites; the extracellular binding site for endogenous ligands and the transmembrane binding site for allosteric modulators. It has been shown that the truncated MGR5 receptor which lacks the N-terminal extracellular domain, exhibits constitutive activity attributable to the 7-TM domain only.\(^{46}\) Interestingly, Trp\(^{6.48}\) is also found in class C GPCRs which provides support to the hypothesis that the overall fold of the 7-TM domain of most GPCRs in their ground states has been rather conserved during evolution.

The known chemotypes of noncompetitive agonists/antagonists for either metabotropic glutamate receptors (mGluR) or calcium-sensing receptor includes the above-described cavity description. mGluR allosteric modulators usually bear two aromatic groups separated by an amide/sulfonamide moiety\(^{39,47,50}\) that is likely to bind to one of the 7-TM cavity. The chemotypes of the known allosteric modulators for mGluR, which nicely fit the above-described cavity, are shown in Table 6(B). The cavity-based alignment of the glutamate receptor cluster is shown in Fig. 6(B).
the three hydrophilic residues at the center of the cavity (positions 3.28, 3.29, and 7.39). In CaSR allosteric ligands, a positively charged secondary amine is anchored to Glu8377.39 and is surrounded by two aromatic moieties filling the hydrophobic subsites.48,49

By contrast to glutamate receptors, the GABA-B receptor is a heterodimer consisting in two subunits; GABR1 is responsible for ligand binding at the venus fly-trap domain whereas GABR2 participates in G protein coupling.51 The 7-TM domain of GABR2 has recently been shown to bind a positive allosteric regulator of the GABA-B receptor.52 The 7-TM binding cavity resembles that of the metabotropic glutamate receptors but is more hydrophobic and significantly smaller. Another striking difference is the presence of small hydrophilic amino acids at position 6.48 and 6.51 that could be anchoring points for small-sized allosteric regulators of the GABA-B receptor53,54 although their binding site has not been precisely mapped up to now.

The above-cited data suggest that the 7-TM cavity of class C receptors has a functional role in triggering receptor activation, and can be viewed as an allosteric site for putative ligands. Whether this role may be linked to the existence of naturally occurring ligands targeting the allosteric subsite for class C GPCRS (as well as for members of the previously defined adhesion and secretin clusters) still has to be demonstrated.

The Prostanoids Receptor Cluster (8)

The prostanoids receptor cluster contains eight receptors classified in two main subgroups [Fig. 7(A)]. Prostanoids are cyclooxygenase metabolites and consist in C-20 unsaturated fatty acids. They comprise prostaglandins, prostacyclins, and thromboxanes, and exert a variety of actions, including the relaxation and contraction of various types of smooth muscles. This branch is separated in two parts according to their G protein coupling.55,56 Strong homology between the four Gs-coupled prostaglandin receptors (PD2R, PE2R2, PE2R4, and PI2R) and the four Gq/Gi-coupled receptors (PE2R1, PE2R3, PF2R, and TA2R) suggests that these receptors evolved from an ancestral receptor. Receptors for prostanoids present a large TM cavity (1,840 Å³ for PI2R) clearly composed of two separate hydrophobic subsites. Subsite 1 is delimited by conserved small or medium-sized residues at positions 1.46, 2.58, 2.61, and 7.43 [Fig. 7(B)], thereby extends relatively deep between TMs 1 and 2. Subsite 2 located between TMs 3, 5, and 6 is delimited by positions 3.29, 3.36, 5.38, 5.39, 5.42, 5.43, 6.51, and 6.55. This apolar subsite extends deep toward the intracellular side of the cavity between small and medium side chains at TMs 5 and 6. Main differences between the two prostanoid receptor subgroups occurs at two positions (6.48 and 7.35) of subsite 2, the latter being polar for the first subgroup and hydrophobic for the second one. Experimental evidences suggested that Arg7.40, a conserved amino acid throughout prostanoid receptors but not taken into account to define the current cluster, interacts with the carboxylate moiety of prostanoids.57 According to our 3D model of prostanoid receptor, the side chain of Arg7.40 is accessible and takes part in the first subsite of the TM cavity. More generally, site-directed mutagenesis mapping of prostanoid binding sites58–60 agrees well with the herein predicted TM cavities. The cyclopentane ring of prostanoids is proposed to interact with small aliphatic residues of TMs 1, 2, and both alkyl chains are directed towards TMs 3 and 7 with an ionic bond between Arg7.40 and the acidic moiety of the ligand.59,60 The recent mapping of the prostacyclin receptor (PI2R) binding pocket61 revealed four important anchoring residues: Tyr2.65, Phe7.39, Phe7.39, Arg7.40. Three of them belong to our subset of 30 positions. Synthetic agonists and antagonists of prostanoid receptors are likely to bind in a similar way because of their high structural similarity to endogenous prostanoids.61,62

The Glycoproteins Receptor Cluster (8)

The well-defined glycoproteins receptor cluster regroups eight members of the rhodopsin family [Fig. 8(A)]. All representatives are receptors for glycoprotein hormones. Their structure consists of a N-terminal domain containing several leucine-rich repeats (LRR), followed by the
7-TM domain. The first domain is responsible for the binding of the natural hormone and the second one is involved in triggering receptor activation. Although the natural hormone binds to the N-terminal ectodomain, these GPCRs exhibit a clear TM cavity (ca. 1,230 Å³ for FSHR) composed by two subsites (hydrophobic site between TMs 1, 2, and 7; hydrophilic site between TMs 3, 5, and 6) linked by an hydrophilic channel [Ser/Thr3.36, Asn7.45; Fig. 8(B)]. Several conserved polar side chains (Ser/Thr3.36, Asp/Asn6.44) could constitute excellent anchoring points for any ligand lying in the TM cavity.

The MAS-Related Receptor Clusters (11)

This group includes the MAS proto-oncogene and related receptors [Fig. 9(A)] expressed in specific subpopulations of sensory neurons that detect painful stimuli and may regulate nociceptive function and/or development. Few data are available on these receptors. However, several endogenous signaling molecules have recently been reported for MAS-related (MRG) receptors. Angiotensin(1-7) is a ligand for the MAS receptor, thereby counteracting effects of Angiotensin II. β-Alanine is a signaling molecule for MRGRD, a receptor known to module neuropathic pain. Adenine and neuropeptide RF-amide receptors in rodents likely to be involved in nociception are close in sequence to human SNS/MRG receptors. Last, MRGX2 has been shown to be a receptor for a new neuropeptide, corticotstatin, which exerts a sleep regulation effect.
The predicted TM cavity of the MAS receptor has a volume of ca. 1,250 Å³. Subsite 2 is significantly more polar than subsite 1 for most entries of this cluster [Fig. 9(B)]. Interestingly, several receptors of this cluster share a negatively charged spot in subsite 2 (Glu4.60, Asp5.38) which could be, like in chemokine receptors (see Chemokines cluster), an important anchoring point for ligands of that subfamily.

The SREB Cluster (6)

This cluster encloses six orphan receptors [Fig. 10(A)] out of which three entries (GPR27, GPR85, GP173) are subtypes of super-conserved receptors expressed in the brain (SREBs) and highly conserved in vertebrates. Three additional receptors (GP101, GP161, and Q14439) whose function is currently unknown are proposed to belong to this cluster. Members of the SREB family exhibit a small apolar TM cavity (ca. 950 Å³) with conserved aliphatic residues at TMs 1, 2 and 6 [Fig. 10(B)]. A charged residue is found for four out of the six GPCRs at position 1.35 (Arg/Lys) and represents a putative target for ligands of these poorly characterized receptors.

The Opsins Receptor Cluster (10)

The opsins receptor cluster regroups ten opsin and related receptors [Fig. 11(A)] activated by light. It includes the only GPCR (OPSD) for which a high-resolution X-ray structure is available. 11-Cis retinal covalently binds to the conserved Lys7.43 via the formation of a Schiff base with the aldehyde moiety of retinal. For many opsins, a negatively charged residue at 3.28 helps stabilizing the protonated Schiff base and closes the cavity between TMs 3 and 7. The cavity is about 1,100 Å³ large and mainly consists in a very hydrophobic subsite 2 (Gly/Ala/Val3.32, Val/Leu/Met5.42, Phe/Tyr/Leu6.44, Trp/Tyr6.48) that serves to anchor the /H9252 ionone moiety of retinal. Activation of the GPCR upon light induces an isomerization of 11-cis to all-trans retinal and subsequent translational/rotational TM motions triggering G protein coupling. Strikingly, RGR binds preferentially all-trans retinal despite the large similarity with other opsins in the TM cavity [Fig. 11(B)].

The Lipids Receptor Cluster (14)

This homogeneous cluster [Fig. 12(A)] groups 11 receptors for sphingosine-1-phosphate and lysophosphatidic...
Our clustering approach manages to unambiguously separate EDG receptor subtypes for sphingosine-1-phosphate or S1P (EDG1, 3, 5, 6, and 8) from receptors for lysophosphatidic acid or LPA (EDG2, 4, and 7). Interestingly, three constitutively active GPCRs predominantly expressed in the brain (GPR3, GPR6, and GPR12) known to bind S1P with nanomolar affinities cluster just beneath EDG receptors. Similarly, the clustering of cannabinoid receptors with EDG receptors can be related to the chemically similar nature of their respective endogenous ligands (anandamide, S1P, LPA) consisting of a long alkyl chain and a polar head (ethanolamine, phosphate). Four of the 30 analyzed residues are very well conserved throughout the cluster (Val/Ile/Leu/Met, Pro, Trp, and Asn) If GP119, GPR3, GPR6, and GPR12 are discarded from the analysis, three additional positions are well conserved (Phe/Tyr, Lys/Arg, Tyr). Looking at the eight EDG receptors, receptors for LPA present a unique combination of six conserved residues (Gly, Gln, Asp, Ser, Trp, Gly) that differentiate them from S1P receptors. Notably, physicochemical properties of position 3.29 (Gln for EDG2, 4, 7; Glu for EDG1, 3, 5, 6, 8) is used as a discriminant to select the most appropriate endogenous ligand through the establishment of specific hydrogen bonds with either a hydroxyl group of LPA or a protonable amino group of S1P.

The cavity of the 14 GPCRs can be clearly divided into a hydrophilic subsite 1 (positions 1.35, 2.65, 3.28, 3.29) and a long hydrophobic channel extending from the polar subsite towards the center of the TM bundle between TMs 4 and 5 (positions 3.33, 4.56, 4.60, 5.38, 5.39, 5.42). Residues not taken into account in the present clustering also contribute to the TM binding cavity (e.g., position 7.36 for the hydrophilic subsite 1 and 4.59 for the hydrophobic channel).

Our modeled cavities are in good agreement with site-directed mutagenesis experiments on EDG1 and EDG2 receptors as well as on the cannabinoid receptors. As an example, the positively charged residue at 3.28 represents the main anchoring atom for both receptor agonists and antagonists. Although GPR3, GPR6, and GPR12 also bind to S1P, it is very unlikely that their binding modes resemble that proposed for S1P/EDG receptors. Indeed, the two key residues in EDG receptors (Arg, Glu) are both absent in GPR3, GPR6, and GPR12 receptors.

The Peptides Receptor Cluster (26)

The peptides receptor cluster includes 26 receptors for peptide ligands that can be classified in four main branches (Fig. 13[A]). The low conservation of amino acids in the 7-TM cavity (Fig. 13[B]) reflects the large heterogeneity of the cognate peptide ligands and the shifted location of the peptide binding site towards the extracellular loops. Besides the quasi invariant location of an aromatic residue at positions 1.39, 6.44, 6.48 and the presence of aliphatic amino acids at positions 1.42 and 3.36, the TM cavity of peptide ligands receptors show a very diverse pattern. Interestingly, many of the receptors of this cluster (20 out of 26) exhibit a cysteine at position 2.57. Throughout 369 human GPCRs, 40 entries share a Cys at 2.57. Out of these 40 entries, there are 16 orphan receptors. For 23 of the 24 liganded GPCRs, the endogenous ligand is a peptide (the exception refers to the histamine H3 receptor which belongs to the cluster of biogenic amine receptors). Site-directed mutagenesis data on cholecystokinin receptors (GASR, CCKAR) have previously identified Cys as important for either nonpeptide ligand binding or G Protein coupling.

A typical cavity of a peptide receptor comprises following subites; site 1 between TMs 1-2, 3, and 7 is either
hydrophobic (e.g., galanine, NPFF, and NPY receptors) or more hydrophilic (endothelin, bombesin, and orexin receptors). Subsite 2 delimited by TMs 4-5-6 is a large aromatic cage for most receptors (aromatic residue often found at positions 4.56, 5.38, 5.43, 5.46, 6.51, 6.53, and 6.55 in addition to conserved Phe⁶.⁴⁴ and Trp/Tyr⁶.⁴⁸). Since most peptides bind to a surface delimited by extracellular loops and the upper part of the TM bundle, our selected 30 residues only partially account for peptide recognition. For some peptides (e.g., galanine), the N-terminal tail is buried in the TM region and the C-terminus contacts extracellular loops at the receptor surface. Other peptides (e.g., neuropeptide Y, cholecystokinin, tachykinins) probably bind in a reverse orientation. Many studies agree to conclude that peptide and nonpeptide ligand binding sites although sharing some key residues significantly differ. Even the same nonpeptide antagonist can use distinct binding sites for two subtypes of the same receptor (e.g., NK1R, NK2R). Due to the very large heterogeneity of possible binding mode to peptide receptors, it is impossible to draw general conclusions about the TM cavity. It is also very difficult to match target and ligand chemical spaces in this cluster. We will therefore analyze few key molecular determinants of peptide receptors on an individual basis.

Receptors for bombesin peptides (GRPR, NMBR, BRS3) are characterized by a conserved triad Asp².6¹/Glu⁶.⁴⁰/Arg⁷.₃⁹ out of which the first two residues have been shown to be of crucial importance for activation of the gastrin-releasing peptide receptor. Bombesins as well as peptide antagonists are believed to mainly bind to the third extracellular loop of bombesin receptors and also to some TM residues (e.g., 3.32, 6.55, 7.43). Furthermore, position 5.42 which is variable within the subfamily [Fig. 13(B)] has been demonstrated to account for the specific recognition of a peptoid NMBR antagonist.

The two receptors for endothelins (EDNRA, EDNRB) share 27 out of 30 identical residues. Thus their TM cavities are very similar. This explains why many endothelin receptor antagonists exhibit a poor selectivity among the two subtypes. Subtype selective antagonists have nevertheless been reported and are likely to take advantage of the variability at position 3.28 (Phe for EDNRA, Val for EDNRB). The remarkable feature of EDNRA and EDNRB TM cavities is their high hydrophilicity: out the 30 amino acids studied, seven are conserved charged residues [Asp².5⁷, Lys³.₃⁳, Glu⁴.⁶⁰, Lys⁵.₃⁸, Asp⁵.₃⁹, Arg⁶.₅⁵, Asp⁷.₃⁵; Fig. 13(B)]. Accordingly, nonpeptide ligands of these receptors are very polar and usually share an acidic moiety. Distinct binding sites for peptide agonists and nonpeptide antagonists have been located to subsites 1 and 2, respectively. The four charged amino acids Asp².5⁷, Lys³.₃⁳,
Arg<sup>6.55</sup>, Asp<sup>7.35</sup> have been found to delimit the nonpeptide antagonist binding site. However, because of the spatial proximity of some charged residues (e.g., Arg<sup>6.55</sup>/Asp<sup>7.35</sup>, Lys<sup>3.33</sup>/Glu<sup>4.60</sup>/Lys<sup>5.38</sup>), it is difficult to ascertain whether these amino acids contact nonpeptide ligands and/or form intramolecular salt bridges for stabilizing the receptor. It should be noticed that the Endothelin B receptor-like protein-2 (ETBR2) does not cluster with known endothelin receptor subtypes, nor with any other GPCR cluster.

Receptors for tachykinins (NK1R, NK2R, and NK3R) present a much more hydrophobic TM cavity [Fig. 13(B)]. Substance P and Neurokinin A binding sites in their cognate receptors (NK1R, NK2R, respectively) are deeply buried in the TM cavity. The substance P binding site strongly involves conserved residues at TM2 (Asn<sup>2.57</sup>, Asn<sup>2.61</sup>), TM7 (Tyr<sup>7.35</sup>) whereas neurokinin A mainly interacts with TM3 (Gln<sup>3.28</sup>), TM5 (His<sup>5.39</sup>, Ile<sup>5.43</sup>), and TM6 (Tyr<sup>5.51</sup>, His<sup>5.52</sup>, Phe<sup>5.55</sup>). Since the peptide binding sites are significantly buried in the TM cavity, binding areas for peptidic and nonpeptide ligands largely overlap, especially at the aromatic cage between TMs 5, 6, and 7. Conformational adaptation of the ligand to its receptor is likely to play an important role for tachykinin receptors since mutation of residues conserved among the three subtypes (e.g., His<sup>5.52</sup>, Tyr<sup>7.35</sup>) differentially affects the binding the same nonpeptide antagonist.

Receptors for cholecystokinin (CCKAR, GASR) exhibit a rather apolar TM cavity. The binding site of the biologically important C-terminal tail of the cholecystokinin CCK8 has been mapped to the herein predicted TM cavity. Nine out of the 10 residues shown to line the peptide binding site are included in the 30 critical positions taken into account in our study. Interestingly, the peptide binding site, which is deeply buried in the TM cavity, involves residues from five TMs (2, 3, 5, 6, and 7) and is shared with that of a nonpeptide agonist. Like for neurokinin receptors, it seems that the binding modes of peptide/nonpeptide agonists and nonpeptide antagonists to both CCK receptors (CCKAR, GASR) are different.

Last, site-directed mutagenesis data on NPY receptors have mapped binding site for the C-terminal tail of neuropeptide Y to a hydrophobic pocket between TMs 1, 2, 6, and 7. More especially Tyr<sup>2.64</sup> (not taken into consideration in our analysis) appears to be a critical anchoring point. And once more, binding sites for peptidic and nonpeptide ligands are distinct although some overlap exists depending on the ligand.

The Melatonin Receptor Cluster (7)

The melatonin receptor cluster contains seven receptors predominantly expressed in the brain and divided in two branches [Fig. 14(A)]. Members of the melatonin cluster present in their TM cavity (ca. 1,050 Å<sup>3</sup> for MTR1A) a clear hydrophobic-aromatic subsite 2 (hydrophobic/aromatic side chains frequently observed at position 3.36, 3.40, 5.42, 5.43, 5.46, 6.44, 6.48, and 6.51) with a possible polar contribution of two residues (Asn/Ser<sup>4.60</sup> and Tyr<sup>5.38</sup>) at one end of the cavity [Fig. 14(B)]. The concomitant presence of bulky side chains at 2.58 (Tyr/Met), 7.39 (Tyr/Leu), and 7.43 (Tyr) restricts the width of the cavity between TMs 2 and 7. At the other side of the cavity, several small side chains at 3.44 and 3.47 (two positions not considered for the phylogenetic tree) allow the formation of a hydrophobic needle between TMs 3 and 5.

The proposed cavity for melatonin receptors (MTR1A, MTR1B) is in perfect agreement with experimental data delineating the crucial role of few residues (Ser<sup>3.39</sup>, Ser<sup>3.39</sup>, Asn<sup>4.60</sup>, His<sup>5.44</sup>, Glu<sup>6.55</sup>) for endogenous and synthetic agonist binding. Receptor antagonists (e.g., Luzindole, 2-Phenylmelatonin, and 4-P-ADOT) are structurally and chemically related to synthetic agonists yet are usually bulkier with an extra aromatic ring. Their binding mode to melatonin receptor resembles that of synthetic agonists but also involves additional interactions with Trp<sup>6.48</sup> and the hydrophobic needle between TMs 3 and 5. The latter area may nicely fit the extra phenyl ring often found in melatonin receptor antagonists.

Interestingly, some residues proved to be essential for ligand binding in several GPCR clusters (e.g., 3.28 and 5.38) and may play a crucial role in the specificity of agonists and antagonists.
3.32) are not anchoring points in melatonin receptors. The orphan MTR1L receptor is closely related to MTR1A and MTR1B receptors with respect to sequence alignment [Fig. 14(B)] but is not a receptor for melatonin. Main differences observed in the TM cavity resides in five positions (2.61, 3.28, 3.32, 5.38, 6.55, 7.35). Site-directed mutagenesis suggests that only the Gly/Thr6.55 and the Phe/Tyr5.38 mutations located between two critical residues (Asn4.60 and His5.46) are likely to be responsible for the singularity of MTR1L whose endogenous ligand still remains to be identified.

The Vasopeptides Receptor Cluster (7)

The vasopeptides receptor cluster is small group of only seven receptors [Fig. 15(A)]; the orphan Q6W5P4 receptor (GPR154) recently proposed to be involved in the pathogenesis of atopy and asthma, four receptors for vasoconstrictive peptides (oxytocin, vasopressin), and two receptors for prokineticins. The TM cavity of these GPCRs is about 1,500 Å³ large and uses overlapping but distinct binding sites for agonists and antagonists. The endogenous peptidic ligands and analogs fit the rather hydrophilic subsite 1 at the top of the 7-TM bundle and nonpeptide antagonists bind to subsite 2 between TMs 3, 5, 6, and 7. A clear difference between vasopressin and prokineticin receptors lies in the physicochemical properties of the second subsite which is quite polar in prokineticin receptors and much more hydrophobic in vasopressin receptors [see positions 3.32, 5.46, and 7.39; Fig. 15(B)]. This explains why typical vasopressin receptor antagonists are very hydrophobic. Due to the high similarity between vasopressin and oxytocin receptors, it is difficult to find selective nonpeptide agonists for these receptors, especially for the oxytocin receptor (only moderately selective oxytocin receptor agonists have been described yet). Fine subtype selectivity for antagonists is easier to tackle and may be regulated by single amino acid differences at the receptor level (e.g., positions 5.42 and 7.39 for vasopressin V1a and V1b receptors). The close proximity of GPR154 and vasopressin/oxytocin receptors in our tree suggests that vasopressin/oxytocin receptor antagonists may represent good starting points for identifying GPR154 antagonists.

The Adenosine Receptor Cluster (6)

The adenosine receptor cluster groups six receptors in two well separated branches [Fig. 16(A)]. The first one encloses two receptors for gonadotropin-releasing hormones GnRH and GnRH II, the second branch comprises four receptors for adenosine. Clustering GnRH with adenosine receptors is unexpected with respect to the GRAFS classification. However, a conserved pattern in the TM cavity of GnRH and adenosine receptors can be clearly identified [Fig. 16(B)]. Similarities mainly involve hydrophobic residues. As an example, four of the five 100% conserved residues in the TM cavity are hydrophobic. By contrast, polar side chains undergo a clearly different distribution [Fig. 14(B)] and may be responsible for receptor selectivity. Indeed two charged residues (Asp2.61, Lys3.32) conserved only in GnRH receptors are crucial anchoring amino acids for the GnRH hormone. GnRH binding site has been extensively investigated by site-directed mutagenesis and has been mapped to TMs 2, 5, and 6. A remote aromatic pocket, delimited between TMs 5 and 6 (Tyr5.38, Phe5.42, Tyr6.45) experimentally shown to line the hormone binding site is comprised in the 30 critical residues analyzed herein. Three of the four other important residues (2.50, 2.53, 2.64, and 5.41) are either accessible to a putative ligand (2.53, 2.64) or involved in the structural integrity of the receptor (Asn2.50). The TM cavity of GnRH receptors (1,500 Å³ for GNRHR) is surprisingly hydrophilic, notably at subsite 1 (Arg1.35, Glu2.50, Asp2.61, Asn2.65, Lys3.32). A remote aromatic pocket, delimited between TMs 5 and 6 (Tyr3.38, Phe4.44, Trp5.46, Tyr6.51, Tyr6.52) interacts with aromatic amino acids of the GnRH peptide. The chemotypes of known nonpeptide GnRH antagonists match rather well the properties of the TM cavity. The heterocyclic scaffold (quinolone, indole) could H-bond to the numerous polar side chains at TMs 1 and 2, the basic amine (piperidine, guanidine) may inter-
The Amines Receptor Cluster (42)

The amines receptor cluster is the largest one of our classification. Although it includes 42 representatives [Fig. 17(A)], it shows a remarkable homogeneity as indicated by the high bootstrap value (Fig. 2). Because of the extraordinarily vast literature on this receptor cluster, we will not discuss in details the binding site cavity of each receptor but rather concentrate on main general features explaining either selectivity or promiscuity.

Receptors from the amines receptor cluster present a well-defined TM cavity (e.g., 1070 Å³ for 5HT2A) with two hydrophobic/aromatic subsites on both side of the conserved Asp³.32. Subsite 2 between TMs 4, 5, and 6 (positions 4.56, 4.60, 5.38, 5.42, 5.43, 5.46, 6.44, 6.48, 6.51, 6.52, 7.35) has a strong aromatic character whereas subsite 1 contributes mainly by aliphatic residues [Val/Ile1.35, Leu/Ile/Met/Thr1.39, Val/Ile/Leu/Gly¹.4², Met/Leu/Val/Ile².5₈, Leu/Phe/Tyr/Trp².7₅, Val/Phe,Tyr/Asn².7₉, Tyr/Trp².4₃ and Ser/Asn².7₄₅, see Fig. 17(B)].

A typical hallmark of this cluster is a conserved aspartic acid at position 3.32 (excepted for the two orphans GPR61 and GPR62 which present a serine) and a cluster of aromatic residues at TM6 (Phe/Tyr₆.₄₄, Trp₆.₄₈, Phe/Tyr₆.₅₁, Phe₆.₅₂) and TM7 (Tyr/Trp².₇₄). The carboxylate of Asp³.3₂ is the counter ion of the basic amine,¹⁷ and all above-cited aromatic amino acids are known to be crucial for recognition of the aromatic moiety common to biogenic amines.¹²⁵–¹²⁷ For the 5-HT4 receptor, a network of intramolecular interactions between Asp³.3₂, Trp₆.₄₈, and Phe₆.₅₁ has been proposed to stabilize the receptor in a silent state suitable for inverse agonist binding.¹²₈ The aromatic cluster (notably Trp₆.₄₈) seems to act as a molecular switch in triggering receptor activation by a cascade of changes. The modification of the rotameric states of these aromatic residues is followed by the alteration of the Pro₆.₅₀ kink¹²₉ and the disruption of an ionic lock between TMs3 and 6.¹₃₀

The nature of three residues on top of TM5 (5.42, 5.43, 5.46) allows a clear differentiation of amine receptors. The 5-HT1 branch (including 5HT5A and 5HT7R) presents a “STA” fingerprint whereas 5-HT2 subtype presents a “GSA/S” sequence [Fig. 17(B)]. Receptors for catecholamines (dopamine, adrenaline) are enriched in Ser residues (e.g., “SSS” for dopamine and adrenergic receptors). Receptors for acetylcholine exhibits a “TAA” sequence. Receptors for trace amine present a more diverse pattern of residues. This variability reflects the variety of substituents on the aromatic rings of the endogenous ligands (phenyl, phenol, catechol, imidazole, indole).¹³¹ Some additional specific amino acids also account for the selectivity of ligand recognition. Hence, Asn⁶.₅₂ is typical of acetylcholine receptors. It is important for both agonist and antagonist binding, with a more pronounced effect for antagonist recognition.¹³₂ Likewise, Asn⁵.₅₅ found mainly in β-adrenergic receptors has been shown to account for the stereospecific recognition of β-adrenergic receptor agonists.¹³₃ In the case of histamine receptors, although residues at positions 5.42, 5.43, and 5.46 of TM5 are not strictly conserved, they do participate to the selective
recognition of the histamine imidazole ring. Asp$^{5.42}$ and Glu$^{5.46}$ are responsible for the selective binding of the histaminergic agonists to HRH2, and HRH3/HRH4 subtypes respectively, by establishing an ion pair with the protonated imidazole ring. The HRH1 subtype lacks a negatively charged residue at TM5 yet utilizes an asparagine side chain (Asn$^{5.46}$) for H-bonding the imidazole ring of histamine. Thus, the example of histamine receptors demonstrates that the endogenous ligand must adapt its conformation and binding mode to the local TM5 environment. The HRH1 subtype is unique among histamine receptors in possessing a Lys$^{5.39}$ which has been shown to be a selective anchoring point for second-generation HRH1 antagonists sharing an acidic moiety (e.g., acrivastine, cetirizine). It is now widely accepted that agonists and antagonists of biogenic amine receptors share overlapping but not identical binding sites; both agonists and antagonists occupy subsite 2 whereas subsite 1 is predominantly filled by antagonists. Out of the four TM2 residues used in our clustering scheme, three positions (2.57, 2.61, and 2.65) have been demonstrated to be occluded upon antagonist binding. The pseudo-symmetrical distribution of two hydrophobic cavities around the central Asp$^{3.32}$ explains why many biogenic amine receptor ligands also exhibit nonspecific chemotypes (a symmetrical distribution of aromatic rings around a central basic amine). However, subsite 1 can be used to design selective antagonists since it has been shown to be responsible for the selective binding of dopamine D2 versus D4 antagonists. Interestingly, in the two orphan receptors GPR61 and GPR62, the prototypical Asp$^{3.32}$ is replaced by a serine but a conserved Glu$^{7.35}$ might be an alternative anchoring residue for putative ligands of these receptors.

The Melanocortins Receptor Cluster (5)

The melanocortins receptor cluster comprises five related receptors for melanocortins [Fig. 18(A)]. A unique feature of this cluster is the availability of an endogenous

---

**Fig. 17.** The amines receptor cluster: (A) Phylogenetic tree, (B) Cavity-based alignment. SHT1A: 5-hydroxytryptamine 1A receptor; SHT1B: 5-hydroxytryptamine 1B receptor; SHT1D: 5-hydroxytryptamine 1D receptor; SHT1E: 5-hydroxytryptamine 1E receptor; SHT1F: 5-hydroxytryptamine 1F receptor; SHT2A: 5-hydroxytryptamine 2A receptor; SHT2B: 5-hydroxytryptamine 2B receptor; SHT2C: 5-hydroxytryptamine 2C receptor; SHT2R: 5-hydroxytryptamine 2 receptor; SHT5A: 5-hydroxytryptamine 5A receptor; SHT5D: 5-hydroxytryptamine 5D receptor; 5HT5R: 5-hydroxytryptamine 5 receptor; ACM1: Muscarinic acetylcholine receptor M1; ACM2: Muscarinic acetylcholine receptor M2; ACM3: Muscarinic acetylcholine receptor M3; ACM4: Muscarinic acetylcholine receptor M4; ACM5: Muscarinic acetylcholine receptor M5; ADA1A: Alpha-1A adrenergic receptor; ADA1B: Alpha-1B adrenergic receptor; ADA1D: Alpha-1D adrenergic receptor; ADA2A: Alpha-2A adrenergic receptor; ADA2B: Alpha-2B adrenergic receptor; ADA2C: Alpha-2C-adrenergic receptor; ADRB1: Beta-1 adrenergic receptor; ADRB2: Beta-2 adrenergic receptor; ADRB3: Beta-3 adrenergic receptor; DRD1: D1A dopamine receptor; DRD2: D2 dopamine receptor; DRD3: D3 dopamine receptor; DRD4: D4 dopamine receptor; DRD5: D1B (D5) dopamine receptor; GPR61: G protein-coupled receptor 61; GPR62: G protein-coupled receptor 62; HRH1: Histamine H1 receptor; HRH2: Histamine H2 receptor; HRH3: Histamine H3 receptor; HRH4: Histamine H4 receptor; O14804: Putative neurotransmitter receptor; TAR01: Trace amine receptor 1; TAR03: Trace amine receptor 3; TAR04: Trace amine receptor 4; TAR05: Trace amine receptor 5.
inverse agonist (Agouti-related protein) for two receptor subtypes (MC3R and MC4R). The sequence alignment shown in Figure 18(B) is characterized by a high proportion of hydrophobic or aromatic residues among conserved amino acids (Phe1.39, Ile/Leu1.42, Val/Leu3.28, Val/Met3.36, Ile/Leu/Met3.40, Met/Phe4.60, Phe/Leu/Ser4.63, Ser/Leu/Met5.46, Leu/Phe5.48, Leu/Met6.48, Phe6.51, Val/Phe6.52, Val/Leu6.55, Phe7.35, Ile/Met7.43). The predicted TM cavity (1,280 Å³ for MSHR) is formed by two subsites around a conserved aspartic acid residue (Asp3.29). Subsite 1 is delimited by hydrophobic side chains of TMs 1, 2 (at positions 1.39, 1.42 and 2.58) and two hydrophilic residues of TMs 2, 7 (Asn/Ser/Lys2.57, Asn7.45). Subsite 2 is more voluminous and defined by aromatic and aliphatic side chains (Leu/Val/Met3.36, Ile/Leu/Met3.40, Phe/Leu/Met4.60, Leu/Phe5.43, Trp6.48, and Val/Phe6.52). The proposed cavity agrees very well with site-directed mutagenesis data. It demonstrates the importance of both the acidic residues (Asp3.29, as well as Asp5.25 not used for the classification yet conserved in the receptors of the cluster) and the numerous hydrophobic/aromatic side chains (3.28, 3.40, 4.56, 6.51, 7.35) for peptide agonist binding and fine subtype selectivity. It is also compatible with the chemotypes of known nonpeptide agonists or antagonists of melanocortin receptors (arylpirperazines, phenylguanidines, tetrahydroisoquinolines) that are likely to bind through an H-bond assisted salt bridge to one of the two above-cited important acidic residues.

The Brain–Gut Peptides Receptor Cluster (10)

The brain–gut peptides receptor cluster is defined by ten receptors organized into two branches [Fig. 19(A)]. Our classification for this cluster significantly differs from that proposed by Fredriksson. Hence, GPR38 is assigned to be the ghrelin receptor in the GRAFS classification and not the motilin receptor as herein or in the Swiss-Prot (the ghrelin receptor being the growth-hormone secretagogue receptor GHSR). The related GPR39 receptor is not classified in the GRAFS classification whereas we find it close to GPR38 (MTLR) in agreement with a previous report. Last, the receptors for hypothalamic peptides (MCHR1, MCHR2, TRFR) are far from each other and from the neurotensin receptors in the GRAFS classification. This cluster is characterized by conserved medium-sized and small aliphatic residues [Fig. 19(B)] in TM 1 (Ile/
Leu\textsuperscript{1.42}, Gly\textsuperscript{1.46}) and a cluster of aromatic residues in TM6 (Phe/Tyr\textsuperscript{6.44}, Trp\textsuperscript{6.48}, and Phe/Tyr\textsuperscript{6.51}, His/Tyr\textsuperscript{6.52}). Their cavity (e.g., 1250 Å\textsuperscript{3} for GHSR) is typically formed by two hydrophobic subsites; subsite 1 between TMs 1, 2, and 7 involves positions 1.35, 1.42, 2.57, 2.65 and subsite 2 between TMs 5 and 6, centered around position 3.40 (generally apolar) involves positions 5.39, 5.42, 6.44, 6.48, and 6.52. At the center of the cavity, several polar residues (Asp/Arg\textsuperscript{3.32}, Asp/Glu\textsuperscript{3.33}, and Gln/Arg\textsuperscript{6.55}) connect both hydrophobic subsites and provide anchoring atoms for putative ligands.

Position 3.33 is indeed an anchoring amino acid to the basic amine found in most GHSR ligands\textsuperscript{147}. In agreement with our cavity model, a hydrophobic subsite involving Met\textsuperscript{5.39} and His\textsuperscript{6.52} has been confirmed by side-directed mutagenesis\textsuperscript{147}. Our predictions are also consistent with the experimentally-determined binding site between TMs 6 (Tyr\textsuperscript{6.51}) and 7 (Tyr\textsuperscript{7.35}) for a neuretinin-1 receptor antagonist\textsuperscript{148}. It should be noted that a charged residue (Arg\textsuperscript{6.54}) not taken into account herein but vicinal to the above-described Arg\textsuperscript{6.55}, is proposed to neutralize the negative charge of the NTR1 agonist SR-48692\textsuperscript{148}.

Several key residues (3.33, 5.46, 6.48, 6.51, 6.52, 7.39) have been found to map the thyrotropin-releasing hormone (TRH) binding site to its receptor (TRFR).\textsuperscript{149}–\textsuperscript{152} They are all among the 30 critical positions used for the current analysis. The binding mode of TRH (PyroGlu-His-Pro-NH\textsubscript{2}) has been thoroughly investigated and is in perfect agreement with our TRFR predicted cavity. Tyr\textsuperscript{1063.33} interacts with the pyroglutamyl carbonyl group, Tyr\textsuperscript{2826.51} faces the His aromatic ring, and Arg\textsuperscript{3067.39} is H-bonded to the C-terminal carboxamide moiety of the peptide. Amino acid at position 3.33 plays a noticeable role in ligand binding.\textsuperscript{153,154} It is an acidic residue in seven receptors [GHSR, GPR39, MTLR, NTR1, NTR2, Q9GZQ4, and Q9HB89; Fig. 19(B)]. In two other receptors (MCHR1 and MCHR2), an aspartate is observed at the preceding position (Asp\textsuperscript{3.32}). The concomitant presence of an acidic residue in TM3 and an aromatic cluster in TMs 5 and 6 (excepted for TRFR) explains why many receptor antagonists from this cluster (e.g., MCHR1) can be derived from biogenic amine receptor ligands.\textsuperscript{155}

The Acids Receptor Cluster (5)

The acids receptor group comprises five receptors [Fig. 20(A)] for which few information is available, excepted that known ligands share a carboxylic acid moiety. G109B (HM74) and Q8TDSS4 (HM74a) have recently been shown to bind nicotinic acid. Their function could be to decrease lipolysis in adipose tissues, although nicotinic acid is unlikely to be the endogenous ligand.\textsuperscript{156} GPR31 and GPR81 are two orphan receptors of unknown function. They are all among the 30 critical positions used for the current analysis. The binding mode of TRH (PyroGlu-His-Pro-NH\textsubscript{2}) has been thoroughly investigated and is in perfect agreement with our TRFR predicted cavity. Tyr\textsuperscript{1063.33} interacts with the pyroglutamyl carbonyl group, Tyr\textsuperscript{2826.51} faces the His aromatic ring, and Arg\textsuperscript{3067.39} is H-bonded to the C-terminal carboxamide moiety of the peptide. Amino acid at position 3.33 plays a noticeable role in ligand binding.\textsuperscript{153,154} It is an acidic residue in seven receptors [GHSR, GPR39, MTLR, NTR1, NTR2, Q9GZQ4, and Q9HB89; Fig. 19(B)]. In two other receptors (MCHR1 and MCHR2), an aspartate is observed at the preceding position (Asp\textsuperscript{3.32}). The concomitant presence of an acidic residue in TM3 and an aromatic cluster in TMs 5 and 6 (excepted for TRFR) explains why many receptor antagonists from this cluster (e.g., MCHR1) can be derived from biogenic amine receptor ligands.\textsuperscript{155}

The Chemokines Receptor Cluster (23)

The chemokines receptor cluster contains 23 receptors [Fig. 21(A)] whose endogenous ligands are small-sized proteins (chemokines, adrenomedullin). The present cluster is likely receptors for anionic ligands due to the conserved presence of two arginine residues in the center of the cavity (positions 3.36 and 6.55). It is worth noting that receptors for carboxylic acids are also found in adjacent clusters (chemokines, chemotactant, purinoreceptors, Fig. 2).
As in several other clusters (Opiates, Chemoattractants) a conserved Gly<sup>1.46</sup>/Leu<sup>2.57</sup>/Pro<sup>2.58</sup> motif is likely to induce a kink at the C-terminal part of TM2 [Fig. 21(B)]. In the center of the cavity, four aromatic side chains (Phe/Tyr/His<sup>3.32</sup>, Phe/Tyr<sup>3.36</sup>, Phe/Tyr<sup>6.44</sup>, Trp<sup>6.48</sup>, and Tyr/Phe<sup>6.51</sup>) form an aromatic cage, typical of this cluster. Last, a much more polar subsite 2 is formed between TMs 3, 4 and 5 involving notably positions 3.33 (Glu, Lys), 4.60 (Asp, Glu, Gln), 5.42 (Glu, Arg, Lys, Gln), and 5.43 (Asn). The proposed TM cavities are in good agreement with experimental data. Likewise, many chemokine receptors (e.g., CCR2, CCR4, CCR5) have been shown to use a negatively-charged residue (Asp/Glu<sup>4.60</sup>, Glu<sup>7.39</sup>) to recognize the protonable amine of typical chemokine receptor antagonists. Notably a negative charge at 7.39 is close in space to the prototypical position 3.32 used by biogenic amine receptors, thus explaining why chemokine receptor antagonists may also bind to biogenic amine receptors. Besides the conserved Asp/Glu at 7.39, the above-described aromatic cage between TMs 3 and 6 is also an important anchoring site for aromatic moieties of chemokine receptor antagonists.

The somatostatin receptor cavity (ca. 1,070 Å<sup>3</sup> for SSR1) exhibits two hydrophobic subites. Subsite 1 is delimited by Val<sup>1.42</sup>, Val/Leu<sup>2.57</sup>, Pro<sup>2.58</sup> and Tyr<sup>7.43</sup>. Subsite 2 is lined by Leu/Met<sup>3.29</sup>, Ala/Gly<sup>3.33</sup>, Val/Met<sup>4.56</sup>, Tyr/Val<sup>4.60</sup>, Phe<sup>5.38</sup> and Tyr/Phe<sup>6.52</sup>. Phe<sup>3.37</sup>, which is not included in the 30 selected amino acids, is conserved in all the somatostatin receptors and contributes to the second subsite. Both subites are lined by a hydrophilic channel formed by two polar side chains conserved in almost all the cluster members (Asp<sup>3.32</sup> and Asn<sup>7.45</sup>). Structure–function relationships studies indicate that the somatostatin (SST-14) core residues (spanning Phe<sup>6</sup> to Phe<sup>12</sup>) interact with a binding pocket located between TMs.

**The Opiates Receptor Cluster (13)**

The opiates receptor cluster contains 13 receptors that can be divided into four main branches [Fig. 22(A)]. Endogenous ligands for receptors of this cluster are all short peptides (from 5 to 30 amino acids) usually rich in basic residues likely to interact with a conserved Asp<sup>3.32</sup> in TM3. The proposed classification is in agreement with that of Fredriksson. This large cluster is characterized by conserved hydrophobic residues at TMs 1 and 2 [Ile/Leu/Val<sup>1.35</sup>, Tyr<sup>1.39</sup>, Ile/Val<sup>1.42</sup>, Gly<sup>1.46</sup>, Val/Leu/Met<sup>2.57</sup>, Pro<sup>2.58</sup>; Fig. 22(B)] and hydrophilic amino acids at TM 7 (His/Tyr<sup>7.43</sup>, Asn<sup>7.45</sup>). A conserved proline at 2.58, which might induce a kink at the TM2, is typical of this cluster and also present in chemokines and purinoreceptors clusters.

The somatostatin receptor cavity (ca. 1,070 Å<sup>3</sup> for SSR1) exhibits two hydrophobic subites. Subsite 1 is delimited by Val<sup>1.42</sup>, Val/Leu<sup>2.57</sup>, Pro<sup>2.58</sup> and Tyr<sup>7.43</sup>. Subsite 2 is lined by Leu/Met<sup>3.29</sup>, Ala/Gly<sup>3.33</sup>, Val/Met<sup>4.56</sup>, Phe<sup>5.38</sup> and Tyr/Phe<sup>6.52</sup>. Phe<sup>3.37</sup>, which is not included in the 30 selected amino acids, is conserved in all the somatostatin receptors and contributes to the second subsite. Both subites are lined by a hydrophilic channel formed by two polar side chains conserved in almost all the cluster members (Asp<sup>3.32</sup> and Asn<sup>7.45</sup>). Structure–function relationships studies indicate that the somatostatin (SST-14) core residues (spanning Phe<sup>6</sup> to Phe<sup>12</sup>) interact with a binding pocket located between TMs.
Several studies agree to define positions 3.32, 6.55, and 7.35 as key residues for agonist binding. Although some nonpeptide somatostatin receptor agonists and antagonists have been described recently, residues lining their binding pocket have still not been elucidated.

The opioid-receptor cavity resembles that of somatostatin receptors; it is centered around the conserved aspartic acid (Asp3.32) although its role might not be as crucial as for somatostatin and biogenic amine receptors. The subsite located between TMs 3, 5, and 6 is mostly composed of aromatic residues (Tyr3.33, Phe5.43, and Trp6.48). All these residues are conserved across different opioid receptors (OPRD, OPRK, OPRM) and the nociceptin receptor (OPRX) at position 5.39 [Lys vs. Ala, Fig. 22(B)] has been shown to be related to the selective binding of opioid alkaloids (e.g., naltrexone, bremazocine) to classical opioid receptors only. The presence of a conserved aromatic acid at 3.32 and of aromatic residues at 6.48 and 6.51 for other poorly-studied members of that family (e.g., GPR7, GPR8) suggests that biogenic amine ligands sharing an aromatic moiety and a protonated amine could bind to these two orphan receptors as well.

The Chemoattractants Receptor Cluster (17)

The chemoattractants receptor cluster groups receptors for chemoattractants, angiotensin II, and bradykinin into three branches (Fig. 23(A)). This cluster is nearly identical to that proposed in the GRAFS classification. A typical hallmark of this cluster is the very hydrophobic nature of the transmembrane cavity (positions 1.39, 2.57, 2.58, 3.32, 3.36, 3.40, 6.44, 6.48, 6.51, 7.43) topped by charged amino acids (e.g., Asp/Lys3.33, Arg5.38, Arg/Lys/His5.42, Asp/Glu7.35; see Fig. 23(B)). Few site-directed mutagenesis studies have been undertaken to map the binding site of natural anaphylactic peptide ligands. All conclude that the N-terminal part of these ligands is recognized by charged residues of the receptor N-terminal domain whereas the C-terminus interacts with charged residues at the top of the TM cavity (Glu5.35, Arg5.42, Asp7.35). A negatively-charged anchoring amino acid at 7.35 is specific of anaphylactic peptide receptors and may ensure the absolute specificity of the activation. Interestingly, the chemotypes of nonpeptidic C5a receptor ligands (agonists, antagonists) perfectly match observed physicochemical properties of the computed TM cavity. C5a agonists usually share a positively charged substituent that is likely to mimic one of the two important basic residues of C5a (Lys68, Arg74) that interacts with the conserved negatively charged side chains at the top of the 7-TM bundle (Glu5.35, Asp7.35). C5a receptor antagonists are much more hydrophilic and probably interact with the hydrophilic core of the cavity as suggested by one site-directed mutagenesis study. Ile3.32 and Val7.39 are important anchoring residues, in addition to Glu5.35 and Asp7.35 that may also contact the positively charged group of the ligand.

The modeled TM cavity of formyl peptide receptors (FPR1R, FPR2L, FPR3L) is also in agreement with experimental data. Charged residues, either conserved throughout the cluster (Arg5.42) or among formyl peptide receptors (Asp5.33, Arg5.38) have been demonstrated to directly interact with both the N-formyl and the C-terminal carboxylate moieties of endogenous peptide ligands. The peptide backbone is probably parallel to the main axis of TM5 with hydrophobic side chains contacting the numerous hydrophobic residues of the cavity, especially at TM2. Few antagonists of formyl peptide receptors (small peptides, bile acids) have been described up to
Like previously described C5a receptor antagonists, they all share a very hydrophobic core and a negatively-charged group likely to interact with Arg5.38 in the TM cavity.

The predicted cavity for angiotensin II type 1 receptors (AG2R, AG2S) perfectly agrees with the site-directed mutagenesis used to map the binding sites for nonpeptide antagonists; for instance, insurmountable antagonism180,181 is induced by the interaction of biphenyltetrazole substructure (e.g., Losartan) with both Lys5.42 and the neighboring aromatic cage at TM6.

**The Purine Receptors Cluster (35)**

The purine receptors cluster is one the largest in our classification [Fig. 24(A)] and contains 35 members with a significant proportion of orphan targets (nine in total). All known endogenous ligands are anionic molecules (carboxylates or phosphates). The major difference between the GRAFS classification and ours resides in the location of monocarboxylic acids receptors (GPR40, 41, 42, 43) which could not been classified by Fredriksson et al.16 whereas we cluster them with known receptors for dicarboxylic acids (GPR80, GPR91). The large purinoreceptors cluster is characterized by basic residues [less than five per sequence; Fig. 24(B)] found at various spots in the TM cavity (2.65, 3.29, 3.32, 3.33, 4.56, 5.38, 5.39, 5.42, 6.52, 6.55, 7.39). Side chains of these arginine, lysine, and histidine residues are likely to neutralize the negative charges of the ligands. TMs 1, 2, and 3 contribute mainly to the hydrophobic part of the cavity (Phe/Tyr/Leu1.39, Ile/Val1.42, Gly1.46, Leu/Phe2.57, Pro2.58, Phe/Tyr/Leu3.33, Leu/Met/Ile3.36, and Val/Ile/Met3.40). A bunch of aromatic residues in TM6 (Phe/Tyr6.44, Phe/Tyr6.48, Phe/Tyr/His6.51) is much conserved throughout the cluster. Interestingly, position 6.48, which is known to play a well-defined role in molecular activation of rhodopsin-like GPCRs is a phenylalanine for most members of this cluster.

In the case of carboxylic acid receptors, the distribution of charges at the protein surface is consistent with the structure of their ligands. Receptors for dicarboxylic acids (GPR80, GPR91) present five basic residues organized into two positively charges areas whereas receptor for monocarboxylic acids (GPR40, GPR41, GPR42, GPR43) present only four basic residues all involved in a single positive spot around Arg5.39 and Arg7.35. The first basic area of dicarboxylic acid receptors is located at TM3 (Arg 3.29, His3.33) and the second one involves TMs 6 and 7 (His6.52, Arg6.55, Arg7.39). Four of these residues have been shown to be anchoring points for succinic acid (GPR80/GPR91).185 The modeled TM cavity of carboxylic acid receptors is also coherent with respect to ligands’ size and shape. Hence, the cavity of the long chain fatty acid receptor (GPR40) extends much deeper into an hydrophobic subsite between TMs 3, 5, and 6 than the
corresponding cavity of short chain fatty acid receptors (GPR41, GPR43). The bulkiness of side chains at positions 6.44 and 6.48 is the suggested molecular explanation for this feature [Fig. 24(B)].

Receptors for cationic glyco- and phospholipids (G2A, GPR4, SPR1, and PSYR) share a conserved Arg5.42 that is likely the anchoring residue of the phosphate group of endogenous ligands (LPC: lysophosphatidylcholine, SPC: sphingosylphosphorylcholine, psychosine). No negatively-charged residues are present in the TM cavity for counter-balancing the choline positive charge. The choline moiety therefore probably interacts with either the aromatic cluster at TM6 (Phe6.44, Phe/Tyr6.48, Phe/Tyr6.51, His6.55) or the very acidic second extracellular loop of SPC/LPC receptors. Last, the fatty carbon chain of SPC/LPC is proposed to fill a shallow hydrophobic needle between TMs 1 and 2 terminating at conserved positions between TM1 (Val1.42, Gly/Ser1.46) and TM2 (Tyr2.53, Gly7.46).

Recent studies demonstrated that a patch of basic residues (Arg3.29, Arg/Lys6.55, Arg7.35, Arg7.39) in nucleotide receptors participate to ligand recognition. Interestingly, Gq-coupled receptor subtypes (P2RY1, P2RY2, P2RY4, P2RY6) are well separated from Gi-coupled receptor subtypes (GPR86, P2Y12, P2Y14) and present slightly different basic patches. The first subgroup share basic residues at positions 3.29, 6.55, and 7.39. The second subgroup share basic amino acids at positions 6.55 and 7.35. A thorough study of purinergic receptors recently agrees with our conclusion and additionally proposes that a Lys residue in the second extracellular loop to replace Arg3.29 for Gi-coupled subtypes. Other purinoreceptor subtypes (P2Y10, P2RY5, P2RY9), which do not recognize nucleotides present much less accessible basic residues in the TM cavity. For example, P2RY9 has recently been shown to be a receptor for lysophosphatic acid (LPA). Interestingly, it presents together with other atypical P2Y subtypes and related GPCRs (GP174, GP92, P2Y10) a conserved glutamic acid at position 5.43. P2Y9 is not related to EDG receptors (EDG2, EDG4, and EDG7) which also recognize LPA but use Arg3.28 and Gln3.29 as main anchoring residues (see Lipids cluster). These findings suggest that a single ligand can be recognized by unrelated...
GPCRs using quite different binding modes. It is likely that P2RY5 which shares 70% sequence identity with P2RY9 on our 30 consensus positions is also a phospholipid receptor.

PAR receptors exhibit a much more hydrophobic cavity, notably at the center between TMs 3 and 6, with a few conserved small-sized polar residues (Ser/Thr$^{6.51}$, Asn$^{6.52}$, and Ser/Thr$^{7.43}$). Accordingly, known PAR antagonists are quite hydrophobic with several aromatic rings along a pseudopeptide main chain.$^{190}$

**CONCLUSIONS**

We have reported an exhaustive classification of human G protein-coupled receptors based on the analysis of 30 critical positions supposed to delimit the binding cavity of the typical transmembrane domain of ground-state receptors. The proposed phylogenetic tree is coherent and compatible with numerous experimental data, notably in identifying important residues for small molecular-weight ligand recognition. The present study does not claim that only the 30 selected residues are of interest for explaining ligand binding. It simply attempts to draw general structure-binding relationships over the entire family of human receptors on a minimal set of common positions.

We are confident about the alignment of members of a particular family (e.g., rhodopsin-like receptors) because the alignment method matches conserved fingerprints. The alignment of members from different families (e.g., rhodopsin-like versus glutamate-like receptors) is more questionable. Recent mutagenesis studies on glutamate-like receptors suggest that the proposed alignment of class C to class A receptors is compatible with many experimental data.$^{46,47,53,54}$ However, we admit that the comparison of class A/class C with class B receptors is still speculative although many of the consensus positions used herein for the clustering have been experimentally shown to be of importance for antagonist binding. The current study pinpoints relationships between a few critical positions of the 7-TM domain and ligand binding. The conformational dynamics of the cavity that is likely to play a key role in ligand recognition has not been investigated here for many reasons (irrelevance of unrestrained MD simulations regarding the paucity of 3D information about the membrane-embedded receptor, unknown status of oligomerization, unknown membrane regulatory proteins). Studying the plasticity of the ligand–receptor interactions should thus be realized on a case-by-case basis. Our high-throughput 3D models provide a reasonable start for such studies at the condition that extracellular loops are build and joined to the 7-TM helices.

**Privileged Structures of GPCR Ligands Can Be Matched with TM Cavity Hotspots**

Reducing the complexity of chemogenomic data is likely to facilitate the rational design of ligands or focused libraries and to better predict selectivity towards a family of therapeutically relevant macromolecular targets. Relating cluster members to precise molecular features is here greatly facilitated by the analysis of a small subset of amino acids. For each of the 22 clusters, there is often a clear relationship between known ligand chemotypes (e.g., amines, carboxylic acids, phosphates, peptides, eicosanoids, and lipids) and the cognate TM cavities. For example, receptors for bulky ligands (e.g., phospholipids, prostanoids) have a TM cavity significantly larger than that for smaller compounds (e.g., biogenic amines, nucleotides). Receptors for charged ligands (cationic amines, phosphates, mono and di-carboxylic acids) always present, among the 30 critical residues, one or more conserved amino acid exhibiting the opposite charge (e.g., Asp$^{3.32}$ for biogenic amines; Asp$^{4.60}$/Glu$^{7.39}$ for chemokines; Arg$^{3.29}$/Lys$^{6.55}$/Arg$^{7.35}$ for nucleotides). Such complementarity has already been pointed out by several previous studies on subsets of the GPCR proteome for which numerous data exist for both receptors and cognate ligands.$^{19–22,31,186}$ However, the present report exemplifies for the first time such considerations to the entire collection of nonolfactive human GPCRs.

GPCR ligands sharing a common substructure (“privileged structure”) and exhibiting promiscuous binding to unrelated GPCRs are a current important source for GPCR library design.$^{3,31}$ Assuming that conserved moieties of the ligands are likely to bind to conserved subsites of the targets, matching privileged structures with TM hotspots can be achieved very easily without biasing the match by a manual or automated 3D docking. What is required is access to a GPCR ligand database$^{191–193}$ in which both ligand structures and known GPCR targets are listed.

As an example, biphenyltetrazoles and biphenylcarboxylic acids are known to bind to at least six GPCRs (AG22, AG2R, AG2S, GHSR, LT4R1, LT4R2).$^{181,194,195}$ Fine details of 3D recognition of this privileged substructure by GPCR hotspots have been recently investigated by a thorough mutagenesis-guided manual docking of several GPCR ligands.$^{31}$ We propose here a much simpler approach leading to the same outcome; looking at the 30 residues lining the TM cavity of the later six GPCRs allows us to clearly identify putative TM residues able to interact with this substructure (Fig. 25). Conserved aromatic residues are likely to interact with the biaryl moiety cluster between TMs 6 and 7 (Phe$^{6.44}$, Trp$^{6.48}$, Phe/Tyr/His$^{6.51}$, Phe/Tyr$^{7.43}$). A positively-charged residue that probably interacts with the bioisosteric tetrazole and carboxylate groups should be located nearby the aromatic cluster. Hence, three basic residues (Lys$^{5.42}$, Arg$^{6.55}$, and Arg$^{7.35}$) fulfill this requirement. Last, a polar side chain at position 6.52 (His/Gln) is conserved for the six investigated GPCRs and might H-bond to the acidic moiety of the privileged structure. By simply looking at sequence alignments of TM cavity-lining amino acids, and without relying on any 3D docking data, we managed to find out the same important anchoring residues than Bondensgaard et al.$^{32}$ Searching our TM cavity database for additional GPCRs fulfilling the above-described requirements (Phe$^{6.44}$, Trp$^{6.48}$, Phe/Tyr/His$^{6.51}$, Phe/Tyr$^{7.43}$, and Lys$^{5.42}$ or Arg$^{6.55}$ or Arg$^{7.35}$ and His/Gln$^{6.52}$) extracts 17 new GPCRs that might accommodate biphenyl-tetrazoles and biphenyl-carboxylic acids (Fig.
Among putative targets are ten chemoattractant receptors (APJ, C3AR, C5AR, C5ARL, CML1, FPR1, FPRL1, GPR15, GPR44, and GPR1), three brain-gut peptide receptors (MTLR, NTR1, and Q9GZQ4), two cationic phospholipid receptors (G2A, SPR1) and two peptide receptors (GALR1, GALR2). This target list contains receptors recently identified by Bondensgaard et al. (e.g., APJ, NTR1). It also suggests totally new putative targets for the investigated privileged structure that might serve as a common scaffold for small-sized combinatorial libraries targeting the new receptors list. Several ligand-based approaches based on known privileged structures have been reported for designing GPCR focused libraries.

However, the use of “generic” privileged structures (e.g.,
aryl-4-piperazine, biphenyl) is conceptually incompatible with the design of innovative compounds. It is therefore of outmost importance to generate novel templates by incorporating structure-based knowledge. The herein proposed approach may contribute to identify such substructures by an exhaustive prioritization of fragments according to their likelihood to bind predicted TM cavities of particular clusters.

Prioritizing Ligand/Library Selection and Design for Receptor Deorphanization

Matching TM hotspots to privileged structures may be useful for finding out ligands of orphan targets. Considering not only the target orphan receptor but also other receptors present in the same cluster enables to derive sufficient information to generate compound libraries. Receptors very close in the current phylogenetic tree present significant similarities in their TM binding cavity. A rationale source for putative ligands of an orphan receptor is thus to evaluate first known ligands of GPCRs which are the closest to the orphan target. There are still numerous orphan GPCRs spread over nearly all clusters presented in the current study (see a nonexhaustive list in Table I).

GPR88 represents an interesting example. Although rather close to dopamine D1 and D5 receptors when considering the 7-TM domain, this receptor clusters with class C GPCRs [Fig. 6(A)] when looking at the 30 residues lining the putative TM binding site. Finding starting hits for this receptor could then be addressed by evaluating first, known noncompetitive ligands for class C GPCRs, especially allosteric ligands of the GABA-B receptors. Likewise, known ligands for Angiotensin II receptors (Fig. 23) should represent good starting points for identifying putative ligands of closely related orphan targets (APJ, GPR15, GPR25, GPR44, GPR1). An experimental validation of this approach has been recently reported by the identification of GPR44 ligands by evaluating known Angiotensin II receptor ligands.198 In the absence of any strongly conserved anchoring residues within the GPCR subfamily under investigation, a more systematic approach could be to dock a library of preselected scaffolds to pick out the best scored structures and prioritize scaffolds enriched among the top scorers when considering the target receptor or even better the whole cluster branch to which this receptor belongs to.

As a conclusion, the herein presented phylogenetic tree can be used to study the selectivity profile of either a ligand or a receptor under investigation. Knowing whether the ligand (receptor) is likely to be permissive for numerous receptors (ligands) is of crucial importance for drug discovery. The location of a precise receptor on the tree quickly indicates the number and identity of close GPCR neighbors. Looking then at the TM cavity alignment of this GPCR subset may help in the identification of selective or permissive residues that may favor or hinder design of selective ligands. Conversely, designing promiscuous ligands for either related or unrelated receptors may be guided by the analysis of their TM cavity and the identification of common anchoring amino acids that direct library design.

ACKNOWLEDGMENTS

This work is supported by grants from the French Ministry of Research and Technology, the European Union (Marie Curie fellowship HPMF-CT-2002-02141) and the Centre National de la Recherche Scientifique (CNRS, Programme Protéomique et génie des protéines). Prof. H. Schioth (Uppsala University, Sweden) is sincerely acknowledged for critical reading of the manuscript. All data (full sequences, sequence of TM residues, sequence of cavity-lining residues, and phylogenetic tree in Newick format) on 369 human GPCRs can be obtained from the following URL: http://bioinfo-pharma.u-strasbg.fr/gpcrdb/gpcrdb_form/html

REFERENCES

2. Kristiansen K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled


10. NYDYL 6.91, TRIPoS, Assoc., Inc., St.-Louis, MO.


31. Kniazeff J, Galvez T, Labesse G, Pin JP. No ligand binding in the...
CHEMOGENOMIC ANALYSIS OF A TRANSMEMBRANE BINDING CAVITY


Binet V, Braxon C, Le Corre L, Acher F, Pin JP, Prezeau L. The human GABA B receptor (GABA(B)2) is activated directly by CPGP7930, a positive allosteric modulator of the GABA(B) receptor. J Biol Chem 2004;279:29085–29091.


98. Gubitz AK, Reppert SM. Chimeric and point-mutated receptors reveal that a single glycine residue in transmembrane domain 6 is critical for high affinity melanotropin binding. Endocrinology 2000;141:1236–1244.


CHEMOGENOMIC ANALYSIS OF A TRANSMEMBRANE BINDING CAVITY


somatostatin receptor. Biochem Biophys Res Commun 2004;316:
1059–1064.
166. Poutoul L, Roubert P, Contour-Galcera MO, Moint C, Lannoy J.
173. Cain SA, Coughlan T, Monk PN. Mapping the ligand-binding site
172. Gerber BO, Meng EC, Dotsch V, Baranski TJ, Bourne HR. An
171. DeMartino JA, Konteatis ZD, Siciliano SJ, Van Riper G, Under-
170. Chen Z, Zhang X, Gonnella NC, Pellas TC, Boyar WC, Ni F.
The conserved asparagine residue in the third putative transmem-
177. Chen X, Yang D, Shen W, Dong HF, Wang JM, Opppenheim JJ,
176. Mills JS, Miettinen HM, Cummings D, Jesaitis AJ. Characteriza-
166. Poitout L, Roubert P, Contour-Galcera MO, Moinet C, Lannoy J,
165. Rohrer SP, Birzin ET, Mosley RT, Berk SC, Hutchins SM, Shen
163. Poitout L, Roubert P, Contour-Galcera MO, Moinet C, Lannoy J,
162. Frimurer T, Ulven T, Elling CE, Gerlach LO, Kostenis E,
170. Chen Z, Zhang X, Gonnella NC, Pellas TC, Boyar WC, Ni F.
169. Fathy DB, Mathis SA, Leeb T, Leeb-Lundberg LM. A single
168. Costanz S, Mamedova L, Gao ZG, Jacobsen KA. Architecture of
171. DeMartino JA, Konteatis ZD, Siciliano SJ, Van Riper G, Under-
163. Poitout L, Roubert P, Contour-Galcera MO, Moinet C, Lannoy J,
162. Frimurer T, Ulven T, Elling CE, Gerlach LO, Kostenis E,
The endogenous opioid spinorphin blocks Met-Leu-Phe-induced neutrophil chemotaxis by acting as a specific antagonist at the N-formylpeptide receptor subtype FPR. J Immunol 2001;167:
6609–6614.
166. Poitout L, Roubert P, Contour-Galcera MO, Moint C, Lannoy J.
166. Poitout L, Roubert P, Contour-Galcera MO, Moinet C, Lannoy J,
169. Fathy DB, Mathis SA, Leeb T, Leeb-Lundberg LM. A single
168. Costanz S, Mamedova L, Gao ZG, Jacobsen KA. Architecture of
171. DeMartino JA, Konteatis ZD, Siciliano SJ, Van Riper G, Under-
163. Poitout L, Roubert P, Contour-Galcera MO, Moinet C, Lannoy J,
162. Frimurer T, Ulven T, Elling CE, Gerlach LO, Kostenis E,
The endogenous opioid spinorphin blocks Met-Leu-Phe-induced neutrophil chemotaxis by acting as a specific antagonist at the N-formylpeptide receptor subtype FPR. J Immunol 2001;167:
6609–6614.
166. Poutoul L, Roubert P, Contour-Galcera MO, Moint C, Lannoy J.