G-Protein coupled receptors: structure and function in drug discovery

Chiemela S. Odoemelam, Benita Percival, Helen Wallis, Ming-Wei Chang, Zeeshan Ahmad, Dawn Scholey, Emily Burton, Ian H. Williams, Caroline Lynn Kamerlin and Philippe B. Wilson

The G-protein coupled receptors (GPCRs) superfamily comprise similar proteins arranged into families or classes thus making it one of the largest in the mammalian genome. GPCRs take part in many vital physiological functions making them targets for numerous novel drugs. GPCRs share some distinctive features, such as the seven transmembrane domains, they also differ in the number of conserved residues in their transmembrane domain. Here we provide an introductory and accessible review detailing the computational advances in GPCR pharmacology and drug discovery. An overview is provided on family A-C GPCRs; their structural differences, GPCR signalling, allosteric binding and cooperativity. The dielectric constant (relative permittivity) of proteins is also discussed in the context of site-specific environmental effects.
Family A (rhodopsin-like receptors)

The rhodopsin receptor family (RRF) is the largest of the GPCR families, comprising of approximately 680 members, and accounts for 80% of receptors in humans.\(^4,28\) The RRF is classified into four groups (\(\alpha, \beta, \gamma, \delta\)) and 13 main subdivisions,\(^4,29\) and it has numerous characteristics which indicate a common ancestry.\(^4,29\) These characteristics include the DRY motif situated at the border between TM3 and intracellular loop (IL) 2 and NSxxNPxxY motif in TM7 (Fig. 2).\(^2,29\) The N-terminal region of the family A GPCR receptors are situated extracellularly,\(^2,29\) while the C-terminal is located within the cytoplasm (Fig. 3).\(^2,29,30\) The ligand binding site is located within the extracellular region of the TMD bundle.\(^29\)

According to Palczewski, the arrangement of the seven TMD helices which vary in length from 20 to 30 residues is responsible for the overall elliptical, cylindrical shape of rhodopsin (Fig. 3).\(^30\) The family A GPCRs vary greatly when their ligand preference and primary structure are considered.\(^31\) However, there is homogeneity in the N-termini of family A GPCRs, but heterogeneity within the TMD regions.\(^31\) However, some of the family A GPCRs share specific sequence motifs within the TMD region.\(^31\)

Palczewski reported the dimensions of rhodopsin as an ellipsoid of approximately 35 \(\times\) 48 \(\times\) 75 \(\AA\), with the long axis perpendicular to the membrane in the standard view.\(^26\) The surface area of the section protruding from the membrane is approximately 1200 \(\AA^2\), with cytoplasmic projection being larger in surface area and volume than the extracellular surface (Fig. 2b).\(^26\) The TMD helices of rhodopsin are irregularly shaped due to the conformational changes associated with the Gly–Pro residues; they also incline at several angles in correspondence to the anticipated membrane surface.\(^32\) Teller \textit{et al.} reported that helix 1 tilted from the membrane plane at 25\(^\circ\) and contains a 12\(^\circ\) kink within it as a result of Pro53 residues being present.\(^32\) Helix 2 kinked at an angle of 30\(^\circ\) around Gly89 and Gly90 and the most significant bend being at Helix 6 at angle of 36\(^\circ\) due to the presence of Pro267.\(^32\)

Family B (secretin receptor family)

The family B GPCRs form a small group, and with an extracellular hormone-binding site, they bind to large peptides.\(^28\) The
family name “secretin” derives from the secretin receptor, which was the first to be cloned in this family. In 1975, Sasaki et al. solved the first X-ray crystal structure of glucagon, a family B GPCR. The family corresponds to group B of the A–F system of classification, and comprises 15 members including: vasoactive intestinal peptide receptors (vIPR1, vIPR2), glucagon-like peptide receptors (GLP1R, GLP2R), adenylate cyclase activating polypeptide receptor (PAC1/ADCYAP1R1), growth-hormone-releasing hormone receptor (GHRHR), calcitonin and calcitonin-like receptors (CALCR, CALCRL), gastric inhibitory polypeptide receptor (GIPR), secretin receptor (SCTR), corticotropin-releasing hormone receptors (CRHR1, CRHR2), glucagon receptor (GCGR), and parathyroid hormone receptors (PTH1R, PTHR2). These 15 receptors share between 21 and 67% sequence identity, and a large portion of the dissimilarity is identified in the N-terminal sequence. These receptors contain conserved cysteine residues in the first and second extracellular loops of the TMD regions (Fig. 3). However, the majority of the receptors within this family contain conserved cysteine residues that make up a cluster of cysteine bridges in the N-terminus. The binding profile of the secretin receptors is outlined by three binding domains comprising of the proximal region and the juxtamembrane region of the N-terminus, as well as the extracellular loops, together with TM6 (Fig. 4). The ligand is thought to activate the receptor by spanning the N-terminal and the TMD extracellular loops, this way mediating the active conformation of the receptor, which increases the probability of activation of the signalling units.
In addition to the presence of an extracellular N-terminal domain (ECD) of 120–160 residues, three intracellular (IL) and extracellular (EL) loops interconnect seven TMD (TM1-TM7) of 310–420 residues that are structurally similar and are thus members of the family B GPCR.\textsuperscript{37,38} According to Parthier \textit{et al.} hormonal recognition in family B GPCRs is believed to follow the ‘two-domain’ binding mode, the N- and C-terminal regions of the peptides interact with the J- and N-domains of the receptors respectively, \textit{i.e.} the C terminus of the peptide initiates a peptide recognition with the ECD, thus allowing the peptide N

![Diagram of rhodopsin and its orientation in membranes.](image-url)
terminus to bind the TMD ligand-binding pocket activating the receptor and prompting a downstream signalling cascade.\textsuperscript{34,38–40} The presence of a conserved ECD structure and the ‘two-domain’ binding mode across the family B GPCRs suggest a similar receptor activation across the GPCR family.\textsuperscript{38}

The secretin receptors have immense potential in drug discovery due to their importance in fundamental homeostatic functions.\textsuperscript{31,38} To date, three of these hormones (glucagon, parathyroid hormone and calcitonin) are used clinically for the treatment of hypoglycaemia, osteoporosis and hypercalcaemia individually.\textsuperscript{31} GLP1-R and GLP2-R are particularly relevant targets, as a result of their part in appetite control and the treatment of type 2 diabetes.\textsuperscript{31}

**Family C (metabotropic glutamate receptors)**

The family C GPCRs comprise of the two \(\gamma\)-aminobutyric acid\(\alpha\) (GABA\(\alpha\)) receptors, odorant receptors in fish, eight metabotropic glutamate receptors (mGlu receptors or GRM),
pheromone receptors, Ca$^{2+}$-sensing receptors (CaS receptors or CASR), sweet and umami taste receptors (TAS1R1-3), GPCR Class C Group 6 Member A (GPRC6A) and seven orphan receptors. The taste receptors in this GPCR family are targeted by the taste additives used in the food industry. The CaS, mGlu and GABAB receptors belong to a novel category of drug targets that are essential for considering conditions which affect the central nervous system and calcium homeostasis. Currently, family C GPCRs are targeted by two therapeutic drugs in the market. One is Cinacalcet, the first GPCR allosteric modulator to be marketed, which targets the CaS receptor. The other is Baclofen (now sold under the brand names Lioresal, Liofen, Gablofen, etc.), which is a GABAB agonist used in the treatment of muscle spasms.

The family C GPCRs differ from others by possessing a large extracellular domain, distal to the TMD receptors, and containing the orthosteric sites; they also form constitutive dimers with unique activation systems in comparison with other GPCR families. Similarly, to their related families, family C GPCRs exhibit a typical motif of seven TMD helices however differ structurally from other GPCR families in their possession of an unusually large extracellular domain, an intracellular carboxy-terminal (C-terminal) domain and a heptahelical TMD (Fig. 5A). The family C GPCRs are structurally distinct from other GPCR families as a result of their extracellular domain including a cysteine rich domain (CRD, with the exception of GABAB receptor) and Venus flytrap module (VFT). The TM domain of family C GPCRs contain only the allosteric binding pocket, a CRD and a TMD except for GABAB receptor. The VFT connects to TMD using CRD in the mGlu receptors. The mGlu receptors form homodimers which can potentially offer two other orthosteric binding pocket per dimer.
Table 1  Table showing some characteristics of family A–C GPCRs

<table>
<thead>
<tr>
<th>Feature</th>
<th>Family A</th>
<th>Family B</th>
<th>Family C</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmembrane domains Orthosteric binding site</td>
<td>All families possess seven transmembrane domains TM region</td>
<td>Extracellular loops, extracellular N-terminus, TM6</td>
<td>Extracellular N-terminus (VFTM, SUSHI)</td>
<td>31, 54 and 55</td>
</tr>
<tr>
<td>Number of approved and marketed drugs Motifs</td>
<td>All GPCRs share the D/E-R-Y/W motifs</td>
<td>25</td>
<td>33</td>
<td>94</td>
</tr>
<tr>
<td>Number of conserved residues in TMD regions</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of ligand</td>
<td>Small molecules, proteins, peptides</td>
<td>Proteins, peptides</td>
<td>Small molecules, cations, amino acids</td>
<td>31</td>
</tr>
<tr>
<td>Suitable as drug targets?</td>
<td>Yes, except the sensory receptors</td>
<td>Yes</td>
<td>Yes, except the sensory receptors</td>
<td>31</td>
</tr>
</tbody>
</table>

sites differing from other families with their TM domains conserved while the orthosteric sites are situated in the VFTM module.41,44 Domains present in the family C GPCRs provide numerous ligand sites of action, bar the intracellular C-terminal domain; this is highly variable and plays an essential role in signalling protein coupling and scaffolding.41 The family C GPCRs are unique due to their compulsory dimerization, either as heterodimers (GABA receptor and TIRs) or homodimers (mGlu and CaS receptors) (Fig. 5B).41,47,48

Structural differences

GPCRs share a common structural characteristic, the TMD region, with its intracellular C-terminus and extracellular N-terminus, which exhibits the greatest homology.21,28,49,50 The intracellular loops which span TM5 and 6, the amino terminus and the carboxyl terminus are among the most irregular structures in GPCRs with a substantial variation observed in the amino terminus (N-terminus).21-23 The sequence is relatively short for peptide and monoamine receptors comprising of about 10–50 amino acids,21,23 and larger for glutamate family receptors and glycoprotein hormone receptors (350–600 amino acids).21,23 The largest amino terminal domains were observed in the adhesion family receptors.21,23

Bortolato et al. compared crystal structures of family B and family A GPCRs using receptors in the various classes (glucagon receptors, corticotropin-releasing factor receptor 1 (CRF1) and dopamine D2 receptor).22 The comparison of the CRF1 and glucagon receptor crystal structure to dopamine D2 receptor, a family A GPCR, showed that their cytoplasmic regions superimposed well.25 However, the TM6 regions of both glucagon receptors and CRF1 extend outwardly while the cytoplasmic moieties are situated in proximity to the TM3 regions in sites similar to the dopamine, as well as other class A receptors.22 The family B GPCRs lack the direct connectivity between TM3 and TM6 which is regarded as the classical ‘ionic lock’, playing an important role in family A GPCR activation.52,53 The family C GPCRs structurally differ from family A and B due to their remarkably large extracellular domain which comprises of a cysteine-rich domain and VFTM; an intracellular carboxy-terminal (C-terminal) domain. The TM regions in family A and B GPCRs are conserved however family C GPCRs have the allosteric binding site within the TM3 region (Fig. 4).41 Table 1 shows some of the characteristics of the GPCR families discussed in this review.

Allosteric binding and cooperativity

Allostery is a widespread biological process, which is defined as the ability of interactions occurring at a particular site on a molecule to modulate actions on a different binding site on the same molecule.54,55 For example, the binding of an allosteric modulator on a molecule allosterically changes the conformation of its binding pocket as shown in Fig. 6. Currently, there are two types of marketed pharmaceuticals: allosteric modulators, which bind at the allosteric binding site on the receptor and allosterically change the structural conformation of the receptor binding site, and orthosteric modulators, which bind at the active site of the receptor.41 OrthostERIC-binding drugs must overcome a major challenge in mediating the potential side effects arising from binding to homologous proteins sharing similar binding sites.41 Hence an orthosteric-binding drug must have a very high affinity for its target, in order for a small dose to selectively achieve the goal of target-only binding.41 The binding of transcription factors (TFs) to DNA regulatory elements (REs) provides a good example illustrating the specificity in orthosteric drugs.41

The process of GPCR signaling initiates when an endogenous extracellular signal interacts with the orthosteric binding site of a GPCR, resulting in a conformational change which...
passes on the signal through the plasma membrane traversing the TMD region, and eventually activating intracellular signaling cascades through heterotrimeric G proteins and other adjunct proteins. A different approach, demonstrated for ligand-gated particle channels, is the advancement of allosteric modulators of the receptor subtypes, these small molecules do not bind to the traditional orthosteric binding site, instead interacting with the allosteric binding site to either enhance or inhibit receptor activation.

Allosteric GPCR modulators show at least one of the outlined pharmacological properties (Fig. 6). Agonism/reverse agonism: the allosteric modulator disrupts receptor signaling in either a positive (agonism) or negative (antagonism) manner, notwithstanding the presence or absence of an orthosteric ligand. Efficacy modulation: the effect of allosterism causes changes in intracellular responses, leading to alterations in the inherent efficacy of an orthosteric ligand. Affinity modulation: conformational change influences the orthosteric binding pocket, resulting in dissociation or association rate (or sometimes both) of the ligand being modified (Fig. 6). Some known allosteric modulators of family B GPCRs include NovoNordisk compounds 1–6, T-0632, which blocks the GLP-1 induced...

Fig. 6  Mechanism of action of allosteric modulators. Reprinted with permission from Springer Nature: Springer Nature, Nature Reviews Drug Discovery, Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders, P. J. Conn, A. Christopoulos and C. W. Lindsley, Copyright (2009). (a) Allosteric ligands bind to an alternative binding site on a receptor to modulate the activities of an orthosteric ligand efficacy (blue) and/or affinity (red). A number of allosteric ligands can also directly disrupt signalling in their own right (green). (b) Results from simulation show the effects on the function (right) or binding (left) of an orthosteric agonist mediated by three allosteric potentiators depicted in red, blue and green; red enhanced orthosteric agonist affinity only, blue enhanced only the efficacy, green was observed to modestly enhance both efficacy and affinity, as well as showing allosteric agonism.
G-proteins consist of several families of varied cellular proteins which perform several cellular functions, such as contractility and angiogenesis, learning and memory. These proteins bind to the guanine nucleotides (guanine diphosphate (GDP) and guanine triphosphate (GTP)) and also have inherent GTPase activity. They play a principal role in a many cellular processes, including protein synthesis and cell development, vesicular transport, and cytoskeleton assembly, in addition to signal transduction. G-proteins are trimers comprising of two functional components: a beta-gamma dimer (35 and 8 kDa) which closely relates with the alpha subunit upon binding with GDP, and an alpha subunit (39–52 kDa) which is a catalyst for GTPase activity. Human G proteins are classified into two classes, namely small (monomeric), and heterotrimeric G proteins.

GPCRs are the largest superfamily of cell-surface receptors involved in TMD signalling, usually transmitting signals into cells via their response to a range of extracellular stimuli, such as glycoproteins, polypeptides and ions, and hence regulating a wide variety of physiological and developmental functions. The intracellular-signalling cascades activated by GPCRs have been proven to be remarkably complex. The binding of a ligand to the GPCR binding site leads to a conformational change in the receptor, in turn promoting the binding of the heterotrimeric G proteins, consisting of $G_{\alpha}$-GDP and $G_{\beta\gamma}$ subunits, within the intracellular moiety of the receptor. The exchange of GTP for GDP on the $G_{\alpha}$-subunit results in the reversible dissociation of the G protein subunits, initiating a downstream signalling via $G_{\alpha}$-GTP and $G_{\beta\gamma}$.

### Dielectric constant

The most effective way of correlating the structure and function of macromolecules is through the examination of its electrostatic energies. The intermolecular interactions present are affected by the effective dielectric constant (relative permittivity, $\varepsilon_r$), which differs according to the size and composition of the protein. The accuracy of the method of determination is important in understanding various biochemical interactions such as protein–ligand and protein–protein interactions, charge separation, ion channel selectivity and electron and proton transfer signal transduction and macromolecular assembly. These interactions are influenced by the electrostatic potential of the protein surface. The dielectric constant of dry proteins ranges from 2.5 to 3.5 obtained from direct measurement. The theoretical calculation of local dielectric constant of lone proteins based on their amino acid composition yielded an average of 2.7. The polarity of the residues which make up the structural motifs within a protein have been shown to affect its dielectric constant values, these findings were based on computational studies based on continuum electrostatics and molecular dynamics simulations.

According to Warshel and Aqvist, the value of the dielectric constant of proteins is dependent on the property used to define it. They highlighted several possible ways of defining the dielectric constant in proteins, as outlined in Table 2, where $Q_1$ and $Q_2$ are charges on ionisable groups separated by distance $r$, $\mu$ is a group dipole moment (in units of electron Ångström), $\Delta G$ is the electrostatic Gibbs free energy, $\bar{a}$ is the effective radius.

### Table 2: Some rules for the definition of dielectric constants in proteins

<table>
<thead>
<tr>
<th>Definition</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar = $\varepsilon$ large</td>
<td>$\varepsilon = \text{large}$</td>
<td>Protein sites are always polar near small radii ions. The value of $\varepsilon$ is large for charge–charge interactions. Proteins can provide as much solvation as water for ionised groups with small radii. For functionally important charge-dipole interactions, the value of $\varepsilon$ could be as small as 4. Such a low value, however, requires relatively fixed dipoles with little energy for reorganisation.</td>
</tr>
<tr>
<td>Nonpolar = $\varepsilon$ small</td>
<td>$\varepsilon(r) = 10$ often $\varepsilon(r) \approx 40$</td>
<td></td>
</tr>
<tr>
<td>$\varepsilon(r) = \frac{Q_1 Q_2}{r \Delta G}$</td>
<td>$\varepsilon_b &gt; 10$</td>
<td></td>
</tr>
<tr>
<td>$1 - \frac{1}{\varepsilon_b} = \frac{m \Delta G}{1660 \bar{a}^2}$</td>
<td>$\varepsilon \geq 4$</td>
<td></td>
</tr>
</tbody>
</table>
of charge, and $\varepsilon_0$ is the effective dielectric constant associated with a given interaction.

Li et al. reported that the average dielectric constant inside a protein is relatively low, about 6–7, but this figure reaches about 20–30 on the surface of the protein. The high average local dielectric constant values are often linked to the charged residues while the low values are assigned automatically to the regions comprised of mostly hydrophobic residues.

According to Wilson et al. solvent effects on mechanisms of reactions have been established, but its effect on kinetic isotope effects (KIEs) are rather well less comprehended. A change in solvent can alter the KIE indirectly by changing the transition-state (TS) structure. It can also affect KIE by affecting isotopically sensitive vibrational frequencies directly, notwithstanding the TS structure or identity of the rate-determining step. Wilson et al. investigated the medium effects on KIE for $S_2$2 methyl transfer using UFF or UAO cavity method within the polarized continuum model (PCM) and a hybrid quantum mechanical/molecular mechanical (QM/MM) method. Their findings showed that the majority of variation in the equilibrium isotope effects (EIE) occur within the same range of dielectric constants ($1 \leq \varepsilon \leq 10$) as is considered to occur with enzyme active sites and proteins. There is a possibility that any reaction which involves separation, neutralisation or charge distribution within an enzyme active site could indicate variations in KIEs, between a wildtype and mutant form of an enzyme, which originates as a result of changes in the local dielectric response within the diverse protein environment. The use of UFF or UAO cavity method within the polarized continuum model (PCM) and a hybrid QM/MM method to characterise ligand binding in GPCRs would further assist in understanding the interactions which occur in both the active and inactive states of GPCRs, as well the changes which occur during the transition from inactive state to active state upon ligand activation.

Computational biology techniques in GPCR research

The first major breakthrough in human GPCR structural biology took place in 2007 as the solving of the $\beta_2$-adrenergic receptor ($\beta_2$AR with a diffusible ligand) using a modified lipidic cubic phase (LCP) produce to produce $\beta_2$AR-TCL crystals which diffracted to a resolution of 2.2 Å, the structure was further refined at a 2.4 Å resolution. Presently 64 structures of unique GPCRs with varying resolutions have been solved using spectroscopic methods such as fluorescence, electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR) spectroscopy and structural techniques such as cryogenic electron microscopy (cryo-EM), this provides opportunities in employing computational biology techniques such as molecular modeling, and molecular docking in drug discovery research. The milestones achieved in GPCR structural studies have provided insights on the arrangements of the transmembrane domains, the location of the orthosteric allosteric, bitopic, as well as biased ligand binding sites, the homo- or hetero-oligomerization of receptors and the structural rearrangements associated with conformational changes upon GPCR activation and inactivation. This base of structural information on GPCRs is vital for SBDD, ligand-based drug design (LBDD), and integrated models which complement drug discovery efforts.

In 2012, Sosei Heptares published a detailed account on the use of A2AR structure in identifying series of agents as potential antagonists, this became the first published GPCR SBDD discovery. In a research carried out by de Graaf et al. using structure based virtual screening (SBVS), they identified allosteric modulators of two family B receptors namely; glucagon receptor and glucagon-like peptide receptor. SBDD approaches has also lead to the development of new agonists of the $\alpha_3$ adenosine receptor ($\alpha_3$AR).

Conclusion and future prospects

GPCRs are multifaceted proteins which exist in varying conformations, and that the conformational equilibrium of these group of receptors is influenced both by the bound ligand and the proximity to the related G protein. Their structure is highly conserved comprising of seven TMD. These receptors possess different binding domains, namely; allosteric and orthosteric binding domains. The progress in GPCR structural biology has substantially accelerated our understanding of GPCRs as potential drug targets using SBDD and LBDD approaches. Further computational studies assessing nuclear quantum effects on ligand receptor binding, as well as hybrid QM/MM and empirical valence bond theory in the mechanistic studies of GPCRs would allow for further insight into the interactions which occur in both the active and inactive states of GPCRs, as well the changes which occur during the transition from these states upon ligand activation. This review has aimed to provide an accessible and introductory perspective on advances in GPCR-based drug discovery approaches; many reviews on the topic highlighted herein are indeed highly detailed and authoritative but may not provide as accessible an account for a less specialised or more general audience in the chemical sciences.

Conflicts of interest

There are no conflicts to declare.

References

77 M. Amin and J. Küpper, ChemistryOpen, 2020, 9, 691–694.
87 M. Congreve, C. de Graaf, N. A. Swain and C. G. Tate, Cell, 2020, 181, 81–91.