Structured and disordered facets of the GPCR fold
AJ Venkatakrishnan1,3, Tilman Flock1,3, Daniel Estévez Prado1, Matt E Oates2, Julian Gough2 and M Madan Babu1

Introduction

G-protein-coupled receptors (GPCRs) are a large and versatile family of transmembrane signaling proteins. They are expressed in diverse tissue types and mediate the cellular response to a broad spectrum of extracellular signals (e.g. hormones, neurotransmitters) by triggering intracellular signaling cascades via effector proteins such as heterotrimeric G-proteins and β-arrestins. In this manner, GPCRs facilitate a wide variety of important physiological functions such as cardiovascular regulation and neurotransmission. Not surprisingly, GPCRs are among the top protein families that serve as targets for pharmacological intervention [1]. Thus understanding the molecular details of GPCRs is central to our understanding of human physiology and pharmacology.

The functional versatility of GPCRs has been achieved through extensive diversification. There are about 800 different genes that encode GPCRs in humans, which are mainly classified into three categories: class A (rhodopsin-like), class B (secretin-like) and class C (metabotropic glutamate/pheromone) [2]. In addition, Smoothened and Frizzled receptors are categorized as ‘class F’ GPCRs [3]. According to GPCRDB [4], class A forms the largest group with 714 receptors in humans. Class A GPCRs are involved in a plethora of physiological functions including vision (e.g. rhodopsin), olfaction (e.g. olfactory receptors) and immune response regulation (e.g. chemokine receptors). Classes B and C, are made up of 47 and 15 receptors respectively [4], with some of their functions including blood glucose level maintenance (glucagon receptor) and synaptic transmission (metabotropic glutamate receptor). Class F consists of 10 isoforms of Frizzled and one Smoothened receptor, which are involved in Wnt and Hedgehog signaling pathways, respectively [3].

The transmembrane region of GPCRs of all classes is made up of seven alpha helices. However, large parts of the receptor, comprising of the N-terminus, the extracellular loops (ECLs), the intracellular loops (ICLs), and the C-terminus, are solvent exposed and show remarkable diversity not only between, but also within the GPCR classes. In class A GPCRs, which represent >80% of GPCRs, we find that ~40% of the residues, on average, lie outside the membrane, with some GPCRs having >70% in the extramembrane region. In class B and C GPCRs, we observe that over 70% of the residues, on average, are outside the membrane. The most notable structural difference between the classes is observed in the N-terminal region. Unlike class A GPCRs, GPCRs

The seven-transmembrane (7TM) helix fold of G-protein coupled receptors (GPCRs) has been adapted for a wide variety of physiologically important signaling functions. Here, we discuss the diversity in the structured and disordered regions of GPCRs based on the recently published crystal structures and sequence analysis of all human GPCRs. A comparison of the structures of rhodopsin-like receptors (class A), secretin-like receptors (class B), metabotropic receptors (class C) and frizzled receptors (class F) shows that the relative arrangement of the transmembrane helices is conserved across all four GPCR classes although individual receptors can be activated by ligand binding at varying positions within and around the transmembrane helical bundle. A systematic analysis of GPCR sequences reveals the presence of disordered segments in the cytoplasmic side, abundant post-translational modification sites, evidence for alternative splicing and several putative linear peptide motifs that have the potential to mediate interactions with cytosolic proteins. While the structured regions permit the receptor to bind diverse ligands, the disordered regions appear to have an underappreciated role in modulating downstream signaling in response to the cellular state. An integrated paradigm combining the knowledge of structured and disordered regions is imperative for gaining a holistic understanding of the GPCR (un)structure–function relationship.

Addresses
1 MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
2 Department of Computer Science, University of Bristol, Bristol, United Kingdom

Corresponding authors: Venkatakrishnan, AJ (ajv@mrc-lmb.cam.ac.uk), Flock, Tilman (tflock@mrc-lmb.cam.ac.uk) and Madan Babu, M (madanm@mrc-lmb.cam.ac.uk)

3 Contributed equally.
from classes B, C, and F have long and characteristic extracellular structured domains that are important for ligand binding [2]. The N-terminal region of class B receptors is characterized by two pairs of β-sheets and three disulphide bridges [5]; class C contains a Venus fly trap domain [6], and class F contains a cysteine rich domain [3]. While numerous structures of the isolated N-terminal regions of classes B, C, and F had been determined [3,5,6], the structural understanding of the other parts of receptors belonging to these classes was limited until recently. Significant advances in the past two years have allowed the determination of high-resolution structures of the transmembrane regions of GPCRs from classes B [7**,8**], C [9**,10**], and F [11**]. Taken together with the previously determined class A GPCR structures [12**], for the first time, GPCR structures representing all the human GPCR classes (A, B, C and F) are available (Figure 1).

Undoubtedly, crystallography has provided the basis for gaining invaluable molecular insights into the structurally ordered transmembrane region and the extramembrane structured domains of GPCRs. However, large parts of the extramembrane regions, such as the C-terminal tail, have not been amenable to crystallographic studies. These segments are predicted to contain intrinsically disordered polypeptide segments (IDRs; IDRs) [13,14], that is they do not autonomously fold into a stable tertiary structure in the absence of an interaction partner [15,16]. Since the intrinsically disordered regions within the loops and termini of GPCRs form a physical link that allows transmission of extracellular stimuli (extracellular ligands, agonists, antagonists, allosteric modifiers) to the intracellular effector proteins, an understanding of GPCR function based solely on the structured regions will remain incomplete. Clues to a mechanistic understanding of several unresolved functional properties such as the specificity of G-protein coupling, receptor phosphorylation, biased signaling and fine-tuning of receptor activity may lie in investigating the intrinsically disordered regions of GPCRs.

Here, we describe our current understanding of the structural diversity of the transmembrane regions of the GPCR classes based on the most recent crystal structures. We also present our observations by analyzing the disordered regions in the extramembrane region of human GPCRs, experimentally known post-translational modification sites, occurrence of alternative splicing and putative peptide motifs in disordered regions. Finally, we emphasize the need to focus on investigating the intrinsically disordered regions of GPCRs to gain a holistic understanding of GPCR function.

**Structural diversity of the transmembrane region of GPCRs**

Over the past decade and a half, more than 100 structures of over 25 different GPCRs have been determined, with the majority of these structures only being solved in the past three years [17]. The advances in structure determination of GPCRs have been propelled by technological innovations in protein engineering (e.g. T4 lysozyme fusion, thermostabilising mutations, conformation-selective nanobodies) and crystallography of membrane proteins (e.g. lipidic cubic phase) [12**]. Until recently, most of the solved GPCR structures were that of class A receptors [12**]. However, within the past year and a half, crystal structures of class B: corticotropin releasing factor receptor 1 (CRF1) [7**] and glucagon receptor (GCGR) [8**], class C: metabotropic glutamate receptor 1 (mGluR1) [9**] and metabotropic glutamate receptor 5 (mGluR5) [10**], and class F: smoothened (Smo) [11**] have been determined (Figure 1a).

Although there is extensive sequence diversification between different GPCR classes, they share an overall common topology of their transmembrane region [9**]. The relative arrangement of the transmembrane helical axes is conserved across the different classes. Structural superimposition suggests that the similarity of the arrangement of the helical bundle is higher on the intracellular side than on the extracellular side (Figure 1b). Contrary to the extracellular ends of the TM helices, the intracellular ends, particularly those of TMs 1-4, are well aligned. On the extracellular side, for instance, relative to human β2 adrenergic receptor (class A), TM7 of mGluR1 (class C) is shifted inward towards the helical bundle while that of the class B GPCRs CRF1 and GCGR are shifted outward [9**]. This suggests that the variation in the relative distances between helices in the extracellular half of the GPCR fold might be important for achieving the appropriate shape and geometry required for different ligands to access their binding site and for mediating intra-molecular interactions with the extracellular N-terminal domains. On the other hand, the intracellular region is structurally more conserved across classes perhaps to enable coupling with a limited diversity of effector proteins including G-proteins.

As the TM helices serve to mediate communication between extracellular ligands and intracellular effector proteins, the packing between the helices is fundamental to GPCR structure and function. A comparison of class A GPCRs revealed the presence of a conserved structural scaffold, constituted of a network of topologically equivalent residue contacts connecting the TM helices [12**]. It was observed that contacts between structurally equivalent residues are maintained despite the difference in identity of the amino acid across different receptors. Furthermore, it emerged that among all the TM helices, TM3 forms the structural and functional hub in the 7TM GPCR fold wherein every residue in the helix has a role either in making invariant contacts with other TM helices, the ligand or the G-protein [12**]. Consistent with the observation in class A GPCRs, TM3 shares
extensive interactions with other TM helices in class B and class F GPCRs as well [18], thereby ascertaining the role of this helix in the structural and functional integrity of the GPCR fold across multiple classes.

In membrane proteins, kinking of TM helices provides the structural means to achieve functional novelty, for example, as observed in the conformational cycle of K+ channels [19]. Like in several other membrane proteins, GPCRs also harbor proline residues in their TM segments (Figure 1c). Class A GPCRs contain highly conserved prolines in TM5, TM6 and TM7. The activation of class A GPCRs involves rearrangement of TM helices, which includes a Pro-induced distortion of TM5, rotation and translation of TM6 and inward relocation of TM7 [12**]. Classes B and F share prolines at equivalent positions on TM4 and TM5 while class C has two conserved prolines, one each on TM6 and TM7 (Figure 1c). Availability of the active-state structures of receptors from these classes will reveal whether and how the proline-induced kinks contribute to the conformational flexibility of the receptors. It is also noteworthy that none of the classes have conserved prolines in TM1 or TM2, allowing us to surmise that these helices are unlikely to undergo any major movement in the receptor activation of GPCRs.

Despite the low sequence identity of the transmembrane regions among the individual receptors, there exist conserved residues at topologically equivalent positions across different GPCR classes. For instance, Trp4.50 (Ballesteros Weinstein number [20]) on TM4 has its side-chain oriented towards TM3 and is highly conserved across GPCR classes A, B, and F (Figure 1d). In classes B and F, this tryptophan forms non-covalent contacts with another conserved tryptophan in TM3 (Figure 1d). Upon site-directed mutagenesis of this position on TM4, ligand binding is affected in some class A GPCRs, for example, in A1 adenosine receptor [21]. As this position is not in the ligand-binding pocket in known class A GPCR structures [12**], its mutational effect is likely to be indirect. TM4 is proposed to be involved in formation of GPCR oligomers [22] and it remains to be established whether the conserved tryptophan may have a common role across multiple GPCR classes.

In class A GPCRs, the extracellular loops either occlude (e.g. rhodopsin) or provide solvent accessibility (e.g. muscarinic acetyl choline receptors) to the cavities in the 7TM helix bundle [12**]. The structures of class C mGluR1 and class F Smo reveal narrow cavities to the extracellular half of their transmembrane bundle [9**]. Unlike other classes, Class B GPCRs adopt a wide opening on the extracellular side providing access to a solvent accessible cavity [23]. Extracellular ligands of GPCRs penetrate the transmembrane helical bundle to varying depths [12**]. In class A GPCRs, the TM helices provide the majority of orthosteric ligand-binding contacts and residues from TMs 3, 6 and 7 constitute a conserved ligand-binding cradle [12**]. In class B GPCRs, as observed from the CRF1 receptor structure, the ligand penetrates far deeper and has no contacts with the extracellular loops. It is remarkable that GPCR activation can be regulated from varying depths in the TM bundle, and how this is being achieved mechanistically is an open question for future investigation. In this context, molecular dynamics simulations [24,25] and experiments that use chemical probes [26,27] are beginning to provide important molecular insights into how different ligands can activate the receptor. Understanding the principles of ligand binding and uncovering the presence of common mechanisms of activation by diverse ligands can be important for designing engineered GPCRs that allow selective modulation of signaling pathways in vitro and in vivo [28,29].

Intrinsically disordered extramembrane regions of GPCRs

Unlike the transmembrane segments of the GPCR classes and the structured N-terminal domain of classes B, C, and F, large portions of the cytosolic parts of the receptor are intrinsically disordered [13,14,30]. Disordered segments confer a number of functional properties to the receptor (Figure 2a). The intracellular loop length and the C-terminal tail length vary significantly between members of the different GPCR classes [31,32]. In GPCRs, disordered segments mainly occur within the extracellular N-terminus, the intracellular C-terminal region and intracellular loop 3 (ICL3) of the GPCR topology (Figure 2b). The prevalence of disordered regions is variable among individual receptors and between the different classes. This is consistent with the high variability in amino acid composition and length of these regions, which typically ranges from tens of residues to several hundreds. Owing to the importance of these regions in signaling, receptor desensitization, regulation and recycling, it is important to investigate the molecular basis of how disordered regions mediate such functions and facilitate these properties in a cellular context [30,33].

The most studied cytoplasmic interaction partners of GPCRs are heterotrimeric G-proteins, G-protein receptor kinases (GRKs) and β-arrestin. A number of studies have shown that the interaction of these proteins with the receptor involves the disordered third intracellular loop and the C-terminal tail of GPCRs [34–36]. When disordered regions encounter their interaction partner, they frequently undergo disorder-to-order transition and adopt defined conformations to mediate high specificity, low affinity interactions [37,38]. This is one mechanism by which the loop regions and the C-terminal tail of the receptor mediate interactions with their cytosolic partners. For instance, although the phosphopeptide of vasoressin C-terminal tail does not adopt a defined secondary
Figure 1

(a) Milestones in determination of the structures from GPCR classes

class A: rhodopsin
class F: smoothened
class B: CRF1, GCGR
class C: mGluR1

(b) Structural arrangement of the transmembrane helices across diverse GPCR classes

Extracellular view

Intracellular view

(c) Conserved proline-induced kinks in the transmembrane helices of GPCRs

(d) Conserved tryptophan in TM4 of GPCRs
Overview of disordered regions and their functional features in human GPCRs. Disordered regions of GPCRs contain numerous PTMs and peptide motifs important for regulation and fine-tuning of GPCR signaling. (a) Disordered regions within the receptor increase its functional versatility by providing conformational heterogeneity, via linear peptide motifs and by embedding posttranslational modification sites. (b) Average disorder prediction for human GPCRs from the D^2P database of protein disorder predictions [http://d2p2.proj]. y-axis: Average fraction of residues predicted to be disordered across human GPCRs; x-axis: secondary structural elements. TM: transmembrane; ICL: intracellular loop; ECL: extracellular loop; N-term/C-term: N and C terminal regions. (c) Average anchor motif prediction for human GPCRs. y-axis: Average fraction of all residues identified to reside within peptide motifs predicted by the ANCHOR program [http://anchor.enzim.hu] across human GPCRs; y-axis: secondary structural elements. (d) Post-translational modifications in human GPCRs. x-axis: Number of experimentally reported PTMs as documented in the D^2P database across all human GPCRs; y-axis: secondary structural elements.

Besides the interactions with obligate signaling partners, fine-tuning of receptor activity is achieved through interactions with several other proteins such as calmodulin (with class A β1 and β2 adrenergic receptors) and Homer (with class C mGluR1 and mGluR5) [40]. These fine-tuning interactions are predominantly mediated through short segments within the C-terminal tail or ICL3 [40]. The ability of disordered regions in proteins to mediate a large number of protein interactions and act as ‘signaling hubs’ is often driven by the presence of short linear peptide motifs [41–43]. Such motifs are typically 5–10 residues long, occur within disordered regions and are recognized and/or post-translationally modified by structured domains within the interaction partner [41,43,44**]. Using the ANCHOR program [45], we find that >15% of the C-terminal region and parts of ICL3 and the
N-terminal region are predicted to harbor putative peptide motifs (Figure 2c). This suggests two interesting hypotheses: firstly, in addition to known interaction partners of the receptor, several novel uncharacterized or cell-type specific interactions with cytosolic proteins may exist and secondly, common motifs across diverse receptors might serve as evolutionary ‘anchor points’ for mediating interactions with specific cytosolic proteins. Consistent with these ideas, it has been reported that firstly, the polyproline region within the ICL3 of dopamine D4 receptor interacts with the SH3 domain on A-kinase anchoring proteins (AKAP); AKAP in turn serves as a scaffold to recruit protein kinase A to post-translationally modify the residues on the receptor in a cell type specific manner [46–48] and secondly, a number of different GPCRs and other unrelated membrane proteins harbor short peptide motifs in their cytosolic tails that are recognized by the WAVE regulatory complex to nucleate actin polymerization in distinct membrane sites [49].

In addition to providing specific interaction interfaces, disordered regions can form multivalent interactions to form higher order protein assemblies [15,50,51] that can increase local protein concentrations in enzymatic cascades and play a key role in localizing signaling proteins [52**]. An emerging idea in this direction is that cytosolic tails of transmembrane receptors, for instance the LAT protein complex in innate immune receptors, can form higher order assemblies via disordered intracellular regions [53]. Receptor clustering by this mechanism can modulate the agonist-dose response and signaling kinetics in a rapid, non-linear manner and thus has important implications for signaling [54]. Such higher order membrane receptor assemblies are different from defined oligomers in that they are rapidly reversible, with peptide binding motifs and phosphorylation playing an important role in regulating the formation of these reversible assemblies [15,50,51]. New techniques such as single molecule FRET have been used to study GPCR membrane localization [55,56]. In this context, it was recently suggested that there are significant differences in the patterns and dynamics of assembly formation in different GPCRs (GABA-B receptor and β adrenergic receptor) on the cell surface [57*]. Whether disordered regions and post-translational modifications of linear peptide motifs play a role in controlling localization and formation of higher order assemblies involving some GPCRs remains to be investigated.

Intrinsically disordered regions are known to exist in a continuum of conformational states which can mediate transient but highly specific interactions with multiple binding partners [15,37,41] (Figure 2a). Their functionality can be controlled by extrinsic factors such as protein interactions and the alteration of local charge distribution via pH changes to create a rheostat-like regulation of function [52**]. The dynamics and conformational states of intrinsically disordered regions in GPCRs can modulate the accessibility of peptide motifs and facilitate allosteric communication between distal regions [58]. The conformational dynamics of IDR can also be regulated by multiple factors such as posttranslational modifications (PTMs) [59,60] as well as trans-acting factors that interact with the luminal side of the receptor [61]. PTM sites within motifs and disordered regions can act as interaction switches and have important implications for cellular decision-making [62]. In this context, over 800 PTM sites are estimated to occur in human GPCRs (Figure 2d) according to the D3P5 database [63], but their functional role is largely unknown. The presence of multiple phosphorylation sites in the intracellular regions may give rise to combinatorial phosphorylation patterns that can serve as a signaling ‘barcode’, which, for instance can be relevant for the recruitment of β-arrestin [64]. This signaling barcode can also be varied in a cell type-specific manner, as shown by mass spectrometric analysis of the M3 muscarinic receptor [65]. In this way, disordered regions, linear motifs, and PTM sites within a receptor can be modulated by the signaling state of a cell and thus have the potential to modulate downstream signaling in response to the same ligand in a cell type-specific manner.

Intriguingly, several PTM sites can be found within the extracellular N-terminal region of GPCRs (Figure 2d), suggesting that the modification might happen in the ER lumen or during the transport via the Golgi to the plasma membrane. Although the role of glycosylation is well established in ER transport, the role of other modifications in the N-terminal region is not fully understood. Another emerging concept is the endosomal activity of some receptors, following agonist-induced signaling and subsequent receptor endocytosis [66]. Endosomal signaling is potentially important for recycling receptors back to the plasma membrane or for marking them for proteolytic and/or lysosomal degradation [67*]. In this context, ubiquitination after post-endocytic trafficking seems to play a crucial role as an IDR modification. Beyond its key role as a marker for protein degradation, new findings suggest that ubiquitin may function indirectly to modulate ESCRT adaptor protein activity and regulate GPCR sorting through the endosome [68,69]. Thus, understanding the role of diverse PTMs and cracking the “PTM code” within such disordered regions of the receptor has the potential to shed new insights into the regulation of receptor localization, intracellular receptor signaling, protein interaction, receptor turnover and recycling.

Another layer of signaling complexity by GPCRs arises from the consideration that a number of receptors undergo alternative splicing to generate different protein isoforms that differentially include or exclude certain segments in a cell type-specific manner [70]. For instance,
over 800 different isoforms have been documented to arise from about 200 genes [71]. In another study it was shown that alternative splicing is a major mechanism for diversifying the functional pool of GPCR receptors within specialized environments within a tissue, thereby creating complex responses to the same agonist/antagonist [72]. Although a number of splice variants end up being truncated receptors with interesting properties [73], it is likely that alternative splicing in the third intracellular loop and the C-terminal tail [74,75] and tissue-specific splicing of the disordered regions [76–78] may result in a rewiring of protein interactions [79]. This may result in fine-tuned signaling in different cell types and tissues.

Conclusions
The current exponential growth in the number of individual GPCR structures [12‡] is rapidly improving our understanding of the structural diversity of the GPCR family. However, in order to gain a mechanistic understanding of GPCR signaling in classes B, C, and F, several important challenges still remain to be addressed. The currently available structures from classes B, C, and F lack their structured N-terminal domains. In order to understand the mechanism of ligand binding and activation of the receptors, structures of the full-length receptors in the inactive as well as the active conformational states will be required. Furthermore, to understand the specificity of the GPCR–G protein interaction [80] and functional selectivity, we require further structural studies of the complexes of GPCR–G-protein [36] and GPCR–β-arrestin [81]. The recent developments in X-ray free electron laser and cryo-electron microscopy with fast and highly sensitive electron sensors may propel determination of high-resolution structures in different states [82] and in complex with diverse interaction partners [83,84].

Despite our ability to study the structured regions of GPCRs, large portions of the receptor that are in the extramembrane regions are not amenable to crystallography as they are predicted to be intrinsically disordered. The conformational heterogeneity and PTM sites of these regions are important for regulating GPCR signaling. We note a large number of peptide motifs and PTM sites in the putative intrinsically disordered segments of GPCRs are still not understood and this may hold important clues for biased signaling and cell type-specific signaling responses to a ligand. Investigating these peptide motifs and PTM sites is important to understand the fine-tuning and spatio-temporal regulation of GPCR signaling. Finally, we suggest that an integrated paradigm that combines knowledge from structured and disordered regions of GPCRs in the right biological context will allow us to gain a comprehensive molecular understanding of GPCR signaling and regulation.

Acknowledgements
We thank Natasha Latysheva and Louis Maddox for reading this work. AJV acknowledges the MRC: Centenary Early Career Award program. TF acknowledges the MRC and the Boehringer Ingelheim Fonds for funding his PhD. MMB acknowledges the MRC (MC_U105185859) for funding his research program. MO and JG acknowledges Engineering and Physical Research Council (EPSRC) [EP/E501214/1] and Biotechnology and Biological Sciences Research Council (BBSRC) [BB/G02277/1] for funding their research.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


See annotation to Ref. [7]*.


This paper, along with Ref. [10]*, reports the first structures of class C (metabotropic) GPCRs.


See annotation to Ref. [7]*.


This paper reports the first structure of a class F (Frizzled) GPCR.


This paper reports the identification of a conserved structural scaffold and a conserved ligand-binding cradle in class A GPCRs, and highlights the structural and functional importance of the third transmembrane helix in the GPCR fold.


45. This paper proposes the existence of over a million ‘peptide motifs’ in the human proteome. Investigating these peptide motifs in the right biological context may help unravel the molecular functions of the intrinsically disordered loops and termini of GPCRs.


This review provides an excellent summary of the underappreciated role of the intracellular and extracellular loop regions of GPCRs for intracellular and endoecdynic signalling. IDRs in GPCRs are likely to be crucial to allow this complexity of signaling.