Minireview

G Protein-coupled Receptors

I. DIVERSITY OF RECEPTOR-LIGAND INTERACTIONS

Tae H. Ji§, Mathis Grossmann¶, and Inhae Ji‡

From the 3Department of Molecular Biology, University of Wyoming, Laramie, Wyoming 82071-3844 and *Walter and Eliza Hall Institute, Melbourne, Australia

Nearly 2000 G protein-coupled receptors (GPCRs)1 have been reported since bovine opsin was cloned in 1983 (1) and the β-adrenergic receptor in 1986 (2). They are classified into over 100 subfamilies according to the sequence homology, ligand structure, and receptor function. A substantial degree of amino acid homology is found among members of a particular subfamily, but comparisons between subfamilies show significantly less or no similarity.

Mutations have been observed that relate to a wide spectrum of hereditary and somatic disorders and diseases from cancer to infertility. These mutant receptors are incapable of binding ligand or generating normal signals, constitutively generate signals, or are not appropriately expressed on the cell surface. On the other hand, some mutations are beneficial. For example, a mutation in a chemokine (CCR5) receptor, which is a co-receptor for human immunodeficiency virus (HIV), prevents binding of HIV to target cells and consequently prevents HIV viral infection among the majority of homozygotes with this mutation (3).

Although the majority of GPCRs mediate signal transduction via G proteins, emerging evidence indicates that some of these receptors are also capable of sending signals via alternative signal molecules, e.g. Jak2 kinase, phospholipase Cγ, or protein kinase C. These alternative pathways are an indication of the overall diversity occurring in the GPCR superfamily. Furthermore, there are putative seven transmembrane molecules, which do not appear to be coupled to a G protein.

The most striking difference has been observed in the sites and modes of ligand binding and signal generation, which not only manifests the diversity but also indicates the availability of numerous alternative approaches to clinical and industrial applications. In this review, we describe the general structure and ligand interactions of the receptors, and in the following review, Koblika's group (4) focuses on the conformational changes during receptor activation.

General Structure: N-terminal Segment, Seven TMs, Three Exoloops, Three–Four Cytoloops, and C-terminal Segment

As shown in Fig. 1A, all GPCRs have an extracellular N-terminal segment, seven TMs, which form the TM core, three exoloops, three cytoloops, and a C-terminal segment. A fourth cyttoplasmic loop is formed when the C-terminal segment is palmitoylated at Cys. Each of the seven TMs is generally composed of 26–27 amino acids. On the other hand, N-terminal segments (7–595 amino acids), loops (5–230 amino acids), and C-terminal segments (12–359 amino acids) vary in size, an indication of their diverse structures and functions. Interestingly, there is a weak positive correlation between an N-terminal segment’s length and ligand size (5), suggesting a role in ligand binding, in particular for large polypeptides and glycoprotein hormones. A notable exception is the ~600-amino acid N-terminal segment of neurotransmitter receptors such as the calcium receptor.

Why Seven TMs?—The ubiquitous adoption of a seven TM structure raises the inevitable question concerning its structural and functional merits. Odd numbers of TMs place the N- and C-terminal segments at opposite membrane surfaces. It allows glycosylation and ligand binding at the N-terminal segment, and phosphorylation and palmitoylation at the C-terminal segment for desensitization (6) and internalization. Seven TMs may be the minimum necessary to form six loops and a TM core with a sufficient size and versatility to offer a prodigious number of specificities, regulatory mechanisms, and contact sites for G protein and other signal molecules such as Jak2 kinase (7), phospholipase Cγ (8), GPCR kinases (6), arrestin (6), calmodulin (9), and/or protein kinase C (9). In contrast, five TMs may be insufficient to form a stable yet flexible TM core, whereas nine would be more than enough. For example, ion-gated channel proteins are pentamers of four TM subunits, which form a central hydrophilic channel (5) or a 20-TM monomer (10). In either case, they comprise 20 TMs, and the channel is narrow, excluding ions >5 Å (10). Apparently, four TMs by themselves are not sufficient to form a functional core.

Kinked and Tilted TM α-Helices of Unequal Length and Hydrophobicity—TMs are likely to assume diverse structures. The TMs of bacteriorhodopsin (11) and animal rhodopsin (12) form α-helices although some TM of other membrane proteins such as porins have β structures (13). TM α-helices vary in length and can extend beyond the lipid bilayer. Therefore, the boundaries between TMs and loops are likely to be uneven and may be dynamic. TMs 1, 4, and 7 are significantly more hydrophobic than TMs 2, 3, 5, and 6 that contain several ionic and/or neutral residues. In fact, some of the more hydrophobic TMs 1, 4, and 7 have only one hydrophilic residue such as Asn or Ser. Pro residues are frequently found in TM α-helices, unlike α-helices in globular proteins. A Pro can kink the helix backbone by ~26° and impact the global structure. Some TM α-helices are tilted in the membrane. For example, TM 2 of animal rhodopsin is most tilted by ~30°, whereas TMs 4, 6, and 7 are the least tilted (12). TMs 2, 3, 5, and 6 are suggested to protrude more than the others from the lipid bilayer toward the extracellular surface. TMs often contain Cys residues, and some are believed to be apposed although there is no convincing evidence for a TM disulfide linkage.

Counterclockwise Orientation and Closed Loop Formation of Seven TMs—The seven TMs of bacteriorhodopsin (11), animal rhodopsin (12), and adrenergic receptors (14) are arranged as a closed loop in the counterclockwise direction from TM 1 to TM 7 when viewed from the extracellular surface (Fig. 1A). The orientation of the TMs imposes a stereo- and geometric specificity on a ligand’s entry into and binding to the TM core. In this arrangement, the core is primarily comprised of TMs 2, 3, 5, 6, and perhaps 7, whereas TMs 1 and 4 are peripherally sequestered. In general, TMs 1, 2, and 7 are apposed (15). This arrangement is consistent with the view that the more hydrophobic TMs 1, 4, and 7 are exposed more to the lipid bilayer than the less hydrophobic TMs 2, 3, 5, and 6. It is unknown whether there are GPCRs that are open-looped or not closed between TMs 1 and 7. Different arrangements can provide interesting mechanisms for ligand binding and receptor activation. For example, if TM 7 associates more closely with TM 2 than TM 1, there will be a long crevice between TMs 1 and 7. Such a crevice offers a binding site for the fatty chain of lysophosphatidic acid. Also, TMs 1 and 7 of a GPCR may associate with TMs 1 and 7 of another GPCR to form a dimer.

TMs are entropically driven into the lipid bilayer primarily by water molecules present outside of the membrane. However, it is unclear what forces are primarily responsible for arranging TMs.
Ligand Binding and Receptor Activation

Several distinct modes (Fig. 1, B–F) have been observed for high affinity ligand binding to the TM core exclusively (photon, biogenic amines, nucleosides, eicosanoids, and moieties (lysophosphatidic acid and sphingosine 1-phosphate) of lipids), to both the core and exoloops (peptides ≤40 amino acids), to exoloops and N-terminal segment (polypeptides ≤90 amino acids), or exclusively to the N-terminal segment (glycoproteins ≥30 kDa). The distinction between ligand binding and receptor activation is supported by the existence of antagonists that competitively inhibit agonist binding. Further, glycophospholine hormones bind to the N-terminal segment whereas receptors are activated at the membrane-associated domain (5). Interactions of ligands and receptors seem to involve hydrogen bonds, ion pairs, and hydrophobic contacts.

Ligand Binding and Signal Generation in the TM Core

To define better the mechanics of receptor activation it may be divided into at least three steps, signal generation, TM signal transduction, and signal transfer to cytoplasmic signal molecules (40, 5). The simplest way to activate receptors is for a ligand to bind, generate a signal in, and transduce it through the TM core (Fig. 1B). In such a case, changes in TM hydrogen bonds are likely to be the underlying mechanism for signal generation and TM reorganization. For example, phototransactivation of retinal straightens its bent configuration and rearranges hydrogen bonds and the packing of TMs 3 and 5–7. In biogenic amine receptors, TM 3 is believed to be the primary site for ligand binding whereas TMs 5 and 6 are the signal generation site. Although initial ligand binding and signal generation are related yet separate phenomena, it is technically difficult to distinguish them. However, a few receptor systems such as glycophospholine hormone receptors and protease-activated receptors are available as good models to distinguish experimentally ligand binding from signal generation and thus investigate the transition between these two steps.

Rhodopsin—The family of photoreceptors comprises several isoforms specific for red, blue, and green colors. These color characteristics stem from variations in their amino acid sequences. All of them have a retinal chromophore, which is attached through for...
mation of a Schiff's base between the aldehyde moiety of retinal and the e-amino of the Lys in the middle of TM 7 (18). The protonated Schiff's base is paired with conserved Glu\textsuperscript{113} present at the boundary between TM 3 and exoloop 1, thus bringing TMs 3 and 7 into close apposition. Photoaffinity labeling, mutational analysis, and modeling show that the \( \beta \)-ionone ring of retinal associates with TMs 3, 5, and 6, particularly TM 6 Trp\textsuperscript{265} and Tyr\textsuperscript{268} (19). The C\textsubscript{12} methyl group of 11-cis-retinal appears to associate with TM 3 Gly\textsuperscript{121} (19). Light absorption causes a bent 11-cis to linear all-trans isomerization and part of the TM peptide backbone to be exposed. As a result at least one water hydrogen bond is formed (20) and the salt bridge constraint between TM 3 Glu\textsuperscript{113} and TM 7 Lys\textsuperscript{296} may break. Because each hydrogen bond energy is worth 3–7 kcal/mol, these changes are likely and sufficient to rearrange TMs, particularly TMs 3, 6 (19), and 7, thereby generating a signal.

**Biogenic Amine Receptors**—Biogenic amines, aminephrine, nor-epinephrine, dopamine, and histamine enter and bind the TM core (4, 18). The amine of catecholamines pairs with the carboxyl group of TM 3 Asp\textsuperscript{113} of the receptor, 9 Å deep in the core. TM 3 Asp\textsuperscript{113} is crucial for ligand binding but not for signal generation. The fact that Asp\textsuperscript{113} could not be replaced with Glu underscores the tight space and importance of the size and orientation of the carboxyl chain. The catechol ring of the ligand is thought to dock the pocket consisting of TMs 5 and 6. Specifically, the \textit{meta} and \textit{para} OH groups of the catechol ring appear to hydrogen bond, respectively, to TM 5 Ser\textsuperscript{204} and Ser\textsuperscript{207}, which lie on a \( \alpha \)-helical turn apart on the same side of the TM 5 \( \alpha \)-helix. Because this double hydrogen bonding would constrain the TM 5 \( \alpha \)-helix (18) and/or the ligand and the linked TM 3 \( \alpha \)-helix, the TM structure and packing are expected to be adjusted. In fact, the TM 6 structure, particularly Cys\textsuperscript{285}, changes in activated receptors (21), and Phe\textsuperscript{298} is thought to stabilize the catechol ring (18). Because the interactions of the catechol ring are more important for receptor activation than ligand binding, these changes could generate a signal. In addition, the interaction of TM 3 Asp\textsuperscript{113} and TM 7 Lys\textsuperscript{296} may break upon receptor activation (22). For histamine and dopamine receptors, Cys\textsuperscript{285} interacts with the TM 3 Asp, whereas the imidazole ring associates with TM 5 Asp and Thr, which correspond to TM 5 Ser\textsuperscript{204} and Ser\textsuperscript{207} of catecholamine receptors (18). Acetylcholine also appears to bind similarly the TM core, particularly TMs 5 and 6 for signal generation (18).

**Nucleoside and Nucleotide Receptors**—Extracellular adenosine, adenosine nucleotides, and uridine nucleotides are capable of binding \( G \) protein-coupled metabotropic receptors (P2Y receptors). Studies of metabotropic adenosine receptors suggest that ligands bind primarily in the TM core and exoloops. For the adenosine \( \text{A}_1 \) receptor, ligand binding requires TMs 1–4 (23), 6, and 7 as well as exoloop 2 (24). On the other hand, TMs 5–7, particularly Ser\textsuperscript{277} and His\textsuperscript{278} of TM 7, are important for ligand binding by the adenosine \( \text{A}_2 \) receptor (25). Signal generation by adenosine receptors appears to require additional unknown ligand contacts, because agonists and antagonists bind the receptors differently. Four subclasses of chemotactic rat cAMP receptors are found in Dictyostelium. These receptors are coupled to a G protein that activates adenyl cyclase. For ligand binding by the cAMP receptor, TMs 3, 6, and 6 exoloop 3 are important, and exoloop 2 appears to modulate ligand access to the TM core binding site (26).

**Eicosanoid Receptors**—Eicosanoids are derived from arachidonic acid and have a carboxyl group. This family of lipid hormones, which includes leukotrienes and prostanoids (prostaglandins, prostacyclins, and thromboxanes), binds specific subfamily receptor members of high homology. The ligands bind the TM core, and interactions with TMs 3, 6, and 7 are crucial (27, 28). A conserved TM 7 Arg near the extracellular surface, in the eight-helix bundle with the carboxyl group of the ligand (27). The structure and orientation of the cyclopentane ring of prostanoids influence receptor specificity (28), suggesting variations in the TM core structure.

**Short Peptide Receptors with Ligand Binding Partially in the TM Core and Exoloops**

**Formyl Receptor**—Tripeptide N-formyl-Met-Leu-Phe is one of the smallest peptide ligands for GPCRs. Spectroscopic and mutational studies (50) show that the N-formyl moiety of the ligand binds in the TM core around TMs 2 and 3, whereas the C-terminal region of the ligands associates with the N-terminal segment and exoloops 1 and 2 (29). Recently, the Val\textsuperscript{83}Arg\textsuperscript{84}Lys\textsuperscript{85} sequence that joins TM2 and exoloop 1 has been shown to be photoaffinity labeled with the ligand (30).

**Glycoprotein Hormone-releasing Hormone Receptors**—Mutational analyses of the thyrotropin-releasing hormone (pyroEHP-NH\textsubscript{2}) receptor show that several amino acids in TM 3 and one in exoloop 2 are important for ligand binding. Mutation and affinity-labeling studies of gonadotropin-releasing hormone suggest the N terminus of GnRH binds the TM core and the C terminus around exoloops 2 and 3. His\textsuperscript{2} of GnRH is believed to be close to TM 3 Lys\textsuperscript{121}, Arg\textsuperscript{6} of GnRH to exoloop 3 Asp\textsuperscript{296}, and the C-terminal glycinamide to exoloop 2 Asp\textsuperscript{165} (15). Also, the N-terminal segment appears to be near the ligand (31). Because TM 3 Lys\textsuperscript{121} is ~7.5 Å away from the extracellular surface and its side chain could extend to near the surface, GnRH enters, but not deeply into, the TM core.

**Angiotensin Receptor**—The hydrophobic C-terminal region of angiotensin II (DRVYIHPF) appears to enter the TM core of the receptor, and the C-terminal carboxyl group pairs with TM 5 Lys\textsuperscript{199} (32), 7–14 Å from the extracellular surface. On the other hand, Asp-Arg of the DRVYIHPF sequence seems to ion pair with exoloop 2 His\textsuperscript{163} and exoloop 3 Asp\textsuperscript{296} of the receptor, respectively (33). The interaction with Asp\textsuperscript{296} is more crucial for signal generation.

**Multistep Contacts at the N-terminal Segment and Exoloops**

**Receptors for Glucagon, Calcitonin, and Vasocoactive Intestinal Peptides**—This class of peptide hormones is 30–40 amino acids long. On the other hand, the N-terminal segments of their receptors are 116–147 amino acids long and primarily responsible, but not sufficient, for high affinity ligand binding, as exoloops also appear to be required (34). Partial entry of the ligand into the TM core is a possibility. Gly\textsuperscript{145}-Arg\textsuperscript{146} of calcitonin is essential for interaction with the N-terminal segment of the receptor. The resulting transient hormone-receptor complex, including particularly the N-terminal 10 residues of calcitonin, may make secondary contact with the membrane-associated domain (34). In the glucagon receptor, Asp\textsuperscript{126}-Lys\textsuperscript{137} of the 145-amino acid N-terminal segment and exoloop 1 are crucial for ligand binding (35).

**Parathyroid Hormone (PTH) Receptors**—The C-terminal region of 84-amino acid PTH is important for receptor binding, whereas the N-terminal region is crucial for receptor activation. For the receptor, ligand specificity resides primarily in the N-terminal segment whereas signal is generated in the membrane-associated domain, particularly involving TM 3 His\textsuperscript{37} and exoloop 2 Tyr\textsuperscript{316} (36). The two-step interaction of PTH and the receptor, first to form the transient PTH-N-terminal segment complex and next for the complex to interact with the membrane-associated domain to generate a signal.

**Initial Ligand Binding at the N-terminal Segment and Subsequent Secondary Contact of the N-terminal Segment with Exoloops**

**Protease-activated Receptors**—This family of receptors for thrombin and other proteases is activated by proteolysis of the N-terminal segment (37). The protease thrombin recognizes the sequence L\textsuperscript{38}DPRSRLLRPNDKYPFP\textsuperscript{55} in the 74-amino acid N-terminal segment of the thrombin receptor and cleaves it at the thrombin cleavage sequence, L\textsuperscript{38}DPR\textsubscript{39} S\textsuperscript{42}. The N-terminal segment is sufficient for the proteolysis by thrombin, and the released peptide, Met\textsuperscript{39}Arg\textsuperscript{44}, is a strong agonist to stimulate platelet aggregation (38). At the same time, the resulting new 33-amino acid N-terminal segment acts as a tethered ligand and intramolecularly binds to exoloops of the remaining receptor; in particular Arg\textsuperscript{6} of the N-terminal segment interacts with exoloop 2 Gln\textsuperscript{275} to generate the signal (39). Protease-activated receptor 2 is structurally similar to thrombin receptors and is cleaved at R\textsuperscript{46} S\textsuperscript{47} in the N-terminal segment. The resulting new N-terminal segment contacts exoloops, most likely between the conserved Arg of the N-terminal segment and Glu in exoloop 2 (40).

**Glycoprotein Hormone Receptors**—The glycoprotein hormone receptors consist of two roughly equal domains, a 350–400-residue N-terminal segment and a ~300-amino acid membrane-associated domain. The N-terminal segment alone is capable of high affinity ligand binding whereas the membrane-associated domain is the
site of receptor activation (5). The N-terminal segment has 8–9 Leu-rich repeats. These Leu-rich repeats are thought to form a crescent with the concave inner surface consisting of β-sheets, which may bind ligands (41, 42). In addition to the Leu-rich crescent, the N-terminal region of the N-terminal segment contacts the hormone (43). Although the N-terminal segment alone is capable of high affinity ligand binding, this is modulated by exoloops of the receptor (44).

Glycoprotein hormones, LH, FSH, CG, and TSH, are the largest (30–40 kDa) and most complex GPCR agonists. They are heterodimers of a common α-subunit and a hormone-specific β-subunit. These hormones initially bind exclusively to the N-terminal segment and both subunits of the hormone (41, 42). They involve the α C-terminal region, the N-oligosaccharide at Asn65, and a unique α-helix in the α-subunit, as well as an unusual loop (seat belt) in the β-subunit and the peripheral β-hairpin loops of both subunits (41). Secondary contacts occur between the liganded N-terminal segment and exoloops 1–3 (5, 44, 46). Signal is, therefore, likely generated at exoloops.

In fact, some mutations in exoloops 1–3 result in constitutive activation (47). Secondary conformational changes (45) and making secondary interactions with the membrane-associated domain, thus generating a signal (5). The initial high affinity interaction includes multiple contacts between the N-terminal segment and both subunits of the hormone (41, 42). They involve the α C-terminal region, the N-oligosaccharide at Asn65, and a unique α-helix in the α-subunit, as well as an unusual loop (seat belt) in the β-subunit and the peripheral β-hairpin loops of both subunits (41). Secondary contacts occur between the liganded N-terminal segment and exoloops 1–3 (5, 44, 46). Signal is, therefore, likely generated at exoloops. In some cases, the structure of the GABΑ<sub>B</sub> and calcium receptor are similar to those of the metabotropic glutamate receptors. Therefore, they are expected to similarly bind ligands and be activated.

Conclusions and Future Directions

An increasing number of GPCR subfamilies shows diverse modes of ligand binding, signal generation, TM signal transduction, and signal transfer to various cytoplasmic signal molecules other than G protein, such as Jak2 kinase (7), phospholipase C (8), and protein kinase C (9). Therefore, the question arises whether these distinct responses are induced by a single signal originating from a single receptor or by separate signals generated by a single receptor or different receptors. If a single receptor is capable of activating one or more effectors, does it activate them simultaneously or only one at a time? The general underlying mechanisms appear to involve changes in interactions and conformation, particularly hydrogen bonds, salt bridges, TM rearrangement, and perhaps receptor oligomerization (4).

Acknowledgments—We thank Drs. Hanlee Ji, Kurt W. Miller, Brian K. Koblika, Martha Vaughan, Peter Nollert, Ehab M. Landau, and Lee E. Limbird for helpful comments.

REFERENCES

1. Nathans, J., and Hoggess, D. S. (1983) Cell 34, 807–814
2. Dixon, R. A., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., Rhee, S. P., Domalewski, M. D., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., and Stuhlmann, H. (1996) Science 277, 1076–1081
3. Vassart, G. (1995) J. Biol. Chem. 270, 13725–13728
4. Roth, P., Delange, F., Degrip, W., and Rothchild, K. (1998) Biochem. J. 329, 713–717
5. Avruch, J., DiRocco, C., and Jack, J. (1998) Cell 97, 31–40
6. Lefkowitz, R. J. (1998) J. Biol. Chem. 273, 23382–23388
7. Ali, M., Sayeski, P., Dirksen, L., Hayzer, D., Marrero, M., and Bernstein, K. (1997) J. Biol. Chem. 272, 23382–23388
8. Mastroeni, P., Alexiou, V., and De Camilli, P. (1995) J. Biol. Chem. 270, 13739–13746
9. Tanaka, A., Kobayashi, S., and Ueda, M. (1995) J. Biol. Chem. 270, 13747–13751
10. Stryer, L. (1995) Biochemistry, 4th Ed., pp. 300–302, Freeman & Co., New York
11. Pibey-Peyroula, E., Rummel, G., Rosenbusch, J., and Landau, E. (1997) Science 277, 1076–1081
12. Unger, V., Hargrave, P., Baldwin, J., and Schertler, G. (1997) Nature 389, 203–206
13. Limbird, L., Dohlman, H., and Stuhlmann, H. (1996) Science 273, 512–514