The Role of Internal Water Molecules in the Structure and Function of the Rhodopsin Family of G Protein-Coupled Receptors

Leonardo Pardo,*[a] Xavier Deupi,[a] Nicole Dölker,[a] María Luz López-Rodríguez,[b] and Mercedes Campillo[a]

Introduction

Membrane receptors coupled to guanine nucleotide-binding regulatory proteins (commonly known as G protein-coupled receptors, GPCRs) constitute one of the most attractive pharmaceutical targets, as around 40% of clinically prescribed drugs[2,3] and 25% of the top-selling drugs[4,5] act at these receptors. GPCRs are receptors for sensory signals of external origin such as odors, pheromones, or tastes; and for endogenous signals such as neurotransmitters, (neuro)peptides, divalent cations, proteases, glycoprotein hormones, and purine ligands.[6] Phylogenetic analyses of the human genome have permitted GPCR sequences to be classified into five main families: rhodopsin (class A or family 1), secretin (class B or family 2), glutamate (class C or family 3), adhesion, and frizzled/taste2.[7] Specialized databases of GPCRs can be found at http://www.gpcr.org/7tm,[8] http://gris.ulb.ac.be/,[9] and http://www.iuphar-db.org.[10]

Due to the low natural abundance of GPCRs and the difficulty in producing and purifying recombinant protein, only one member of this family, rhodopsin, the photoreceptor protein of rod cells, has been crystallized so far.[11–13] Five structural models of inactive rhodopsin are available at the Protein Data Bank, at resolutions of 2.8 Å (PDB IDs: 1F88 and 1HZX), 2.65 Å (1GZM), 2.6 Å (1L9H), and 2.2 Å (1U19). Structural models of rhodopsin photointermediates such as bathorhodopsin (2G87),[14] lumirhodopsin (2HPY),[15] metarhodopsin I,[16] and a photoactivated deprotonated intermediate reminiscent of metarhodopsin II (2I37)[17] are also available. Rhodopsin is formed by an extracellular N terminus of four β-strands, seven transmembrane helices (TM1 to TM7) connected by alternating intracellular (I1 to I3) and extracellular (E1 to E3) hydrophilic loops, a disulfide bridge between E2 and TM3, and a cytoplasmic C terminus containing an α-helix (Hx8) parallel to the cell membrane. Statistical analysis of the residues forming the TM helices of the rhodopsin family of GPCRs shows a large number of conserved sequence patterns;[18] this suggests a common TM structure. Thus, the availability of the rhodopsin structure allows the use of homology modeling techniques to build three-dimensional models of other homologous GPCRs.[18] The putative structural homology between rhodopsin and other GPCRs probably does not extend to the extracellular domain, since the extracellular N terminus and loop fragments are highly variable in length and amino acid content.[18]

The class A family of GPCRs contains highly conserved Pro residues in the middle of TMs 5 (P5.50, conserved in 77% of the sequences[19]), 6 (P6.50, 100%), and 7 (P7.50, 96%; residues are identified by the generic numbering scheme of Ballesteros and Weinstein,[20] which allows easy comparison among residues in the 7TM segments of different receptors). Pro residues are normally observed in the TM helices of membrane proteins[21] where they usually induce a significant distortion named a “Pro-kink”.[22] This break arises in order to avoid a steric clash between the pyrrolidine ring of the Pro side chain (at position i) and the carbonyl oxygen of the residue in the preceding turn (position i−4)[23] and leads to a distortion of the helical structure.[24,25] However, TM segments of rhodopsin, either with or without Pro residues in their sequence are far from being standard Pro-kinked or ideal helices, respectively.[26] Their distortions are energetically stabilized through complementary intra- and interhelical interactions involving polar side chains, backbone carbonyls, and, in some cases, specific structural and functional water molecules embedded in the TM bundle. Here we review the role of these water molecules in the structure and function of GPCRs and in building computer-generated homology models of class A GPCRs. We propose that water molecules present in the vicinity of highly conserved motifs are most likely present in the rhodopsin family of GPCRs, being another conserved structural element in the family.

[a] Prof. Dr. L. Pardo, Dr. X. Deupi, Dr. N. Dölker, Prof. Dr. M. Campillo
Laboratori de Medicina Computacional, Unitat de Bioestadística
Facultat de Medicina, Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona (Spain)
Fax: (+34)935-812-344
E-mail: Leonardo.Pardo@uab.es

[b] Prof. Dr. M. L. López-Rodríguez
Departamento de Química Orgánica I
Facultad de Ciencias Químicas, Universidad Complutense
28040 Madrid (Spain)
The Distortion of Transmembrane Helices

Several procedures have been described for measuring the distortion of TM helices.[27–29] Bend and twist angles are two relevant parameters for defining such distortions. Bend angle is defined as the angle between the axes of the cylinders formed by the residues preceding and following the motif, which induces distortion in the helix and is used to measure large-scale structural distortions. A residue–residue twist angle calculated for sets of four consecutive Cα atoms analyzes local helical uniformity.[28] This parameter is interpreted as follows: an ideal α-helix, with approximately 3.6 residues per turn, has a twist angle of approximately 100° (360°/3.6); a closed helical segment, with <3.6 residues per turn, possesses a twist >100°; whereas an open helical segment, with >3.6 residues per turn, possesses a twist <100°. A variation in twist angle will result in a change in the orientation of the amino acid side chain, and, in addition, the local opening or closing will be amplified along the helix length to result in a large-scale change that will relocate the helix end.

Water Molecules in the Environment of the NPxxY Motif in Transmembrane Helix 7

The cytoplasmic end of TM7 contains the highly conserved residues N7.49 (N: 75%; D: 21%), P7.50 (96%), and Y7.53 (92%), which form the NPxxY motif in the rhodopsin family of GPCRs.[18] P7.50 induces a kink in the helix with an estimated bend angle of 14°.[13] Moreover, the helix is opened (>3.6 residues per turn, twist <100°) at the 7.45–7.48 turn and closed (<3.6 residues per turn, twist >100°) at the 7.43–7.46 turn (Figure 1). Opening of the helix at the 7.43–7.46 turn removes the steric clash between the pyrrolidine ring of P7.50 and the carbonyl oxygen at position 7.46 (Figure 2). The unusual conformation of the 7.46 carbonyl oxygen is stabilized by the hydrogen bond interaction with the highly conserved N1.50 (100%) in TM1. The Nα–H atoms of N1.50 act as hydrogen bond donors in the interaction with the carbonyl oxygen of residues at positions 1.46 and 7.46, linking TMs 1 and 7 (Figure 2). In addition, this unusual P7.50 kink removes the intrahelical hydrogen bond between the carbonyl group and the N–H amide at positions 7.45 and 7.49, respectively (Figure 2). Water #2 (water numbering is as in Li et al.,[13] see Table 1 for a comparison of water molecules in the different structures) is located between the carbonyl at position 7.45 and the N–H amide at position 7.49 (Figure 2). As a result the intrahelical hydrogen bond length between N and Oi−4, which in standard α-helices is about 3.0 Å, amounts to 4.9 Å in the crystal structure of rhodopsin. The fact that N1.50 (100%) and P7.50 (96%) are highly conserved amino acids in the rhodopsin family of GPCRs lets us suggest water #2 as another conserved structural element in the family. N1.50 and water #2 stabilize this highly distorted and probably functional P7.50 kink.

We have recently shown for the thyrotropin receptor that the highly conserved N7.49 plays a central role in the process of receptor activation, acting as an on/off switch by adopting two different conformations in the inactive and active states.[30,31] The family-specific Thr at position 6.43 and Asp at position 6.44 of glycoprotein hormone receptors restrain the

---

**Figure 1.** Evolution of the unit twist angles (°) of a standard Pro-kink helix (---) and TMs 6 (top) or 7 (bottom) of rhodopsin (-----). The amino acid sequence of rhodopsin is shown (P6.50 and P7.50 are underlined).

**Figure 2.** Crystal structure of rhodopsin (PDB ID: 1GZM).[13] The α-carbon ribbons are: TMs 1 (red), 2 (yellow), 6 (orange), and 7 and Hx8 (blue). The network of interactions involving highly conserved amino acids within TMs 1, 2, 6, and 7 and Hx8 in rhodopsin in detail is: N1.50 interacts with the carbonyl oxygen of residues at positions 1.46 and 7.46, linking TMs 1 and 7. Water #2 is located between the carbonyl at position 7.45 and the NH amide at position 7.49 and stabilizes an unusual P7.50 kink. Water #9 mediates an interhelical interaction between the side chain of N7.49 and the backbone carbonyl at position 6.40 to maintain the receptor in the inactive state. Water #7 participates in a network of interactions between Y7.53 with N2.40 and F7.60. The figure was created by using MolScript v2.1.1[52] and Raster3D v2.5.1[53].
side chain of N7.49 in the inactive gauche conformation an make it to point towards TM6.\textsuperscript{31} In the rhodopsin-like family of GPCRs, position 6.43 is a hydrophobic residue in 79% of the sequences (A: 20%; L: 13%; V: 34%; I: 9%; M: 2%; F: 1%) and has a polar side chain in another 18% of the sequences (C: 1%; S: 4%; T: 11%; N: 1%; Q: 1%); whereas position 6.44 is a hydrophobic residue in 87% of the sequences (A: 1%; L: 3%; V: 1%; F: 82%) and has a polar side chain in 5% of the sequences (C: 1%; S: 1%; N: 1%; D: 2%).\textsuperscript{18} It seems reasonable to suggest that, in the case of receptors with polar side chains at positions 6.43 or/and 6.44, these side chains restrain N7.49 in the inactive conformation, in a similar way to that found for the thyrotropin receptor. The question arises as to whether the inactivation of N7.49 through interactions with partners in TM6 can be extrapolated to the other, more numerous members of the GPCR family that contain hydrophobic amino acids at these positions. An internal water molecule (water #9) found in the crystal structure of rhodopsin\textsuperscript{12} fulfills the role of the polar amino acids and mediates an interhelical interaction between the side chain of N7.49 and the backbone carbonyl at position 6.40 (Figure 2). This suggests a conserved mechanism in which N7.49 is restrained towards TM6 in the inactive state by interhelical interactions that diverge among GPCR subfamilies or via a water molecule in rhodopsin and possibly in other receptors.

Figure 2 shows the interaction of Y7.53 in TM7 with F7.60 in Hx8. These residues form part of the NPxxYx5,6F motif\textsuperscript{32} and are highly conserved in the rhodopsin family of GPCRs (92% and 68%, respectively). In addition, the hydroxy group of Y7.53 forms hydrogen bonds with the N\textsubscript{a2}–H group and the carbonyl oxygen (via water #7) of the partly conserved N2.40 in TM2 (N: 40%; D: 10%). The Y7.53–F7.60 aromatic-aromatic interaction is disrupted during receptor activation, and leads to a proper realigning of Hx8.\textsuperscript{32,33} Surprisingly, this network of interactions at the intracellular part of the receptor (involving highly conserved amino acids within TM2, TM7, and Hx8) seems to direct the ligand-binding profile at the extracellular domain of the chemokine receptor ORF74.\textsuperscript{44}

Figure 3. The proposed hydrogen-bond network linking D2.50 and W6.48 in the inactive conformation of A) rhodopsin,\textsuperscript{11} B) the histamine H\textsubscript{1} receptor,\textsuperscript{38} and C) delta opioid receptor.\textsuperscript{39} The α-carbon ribbons are: TMs 2 (ochre), 3 (red), 6 (orange), and 7 (blue).
in 2002 (1L9H) was the first in which the water interacting with D2.50 (see Table 1) was described; while the structures published in 2004 (1GZM and 1U19) described the water interacting with W6.48. Water #12 also links TMs 2 and 3, and acts as a hydrogen-bond donor in the interaction between the side chain of D2.50 and the backbone carbonyl at the 3.35 position (Figure 3A). This network of interactions holds the receptor in the inactive conformation.

In addition, 67% of the sequences among rhodopsin-like GPCRs contain N7.45, which is absent in rhodopsin. The polar side chain of N7.45 would allow the formation of a hydrogen bond with the side chains of both C6.47 and W6.48 (Figure 3B). In contrast, rhodopsin contains T7.44, absent in most class A GPCRs (A: 19%; L: 24%; V: 10%; I: 6%; F: 11%), which is able to interact with C6.47 in an exclusive manner (Figure 3A). Thus, the role of N7.45 would be to hold C6.47 and W6.48 pointing towards TM7 in the inactive state of the receptor, as has been shown for the histamine H1 receptor.

Opioid, bradykinin, formyl-met-leu-phe, somatostatin, angiotensin, C5a anaphylatoxin, and proteinase-activated families of GPCRs possess Asn at the 3.35 position (29% of rhodopsin-like GPCRs). N3.35 would mimic the role of water #12 (Figure 3C) by bridging N7.45 and D2.50, acting as a hydrogen-bond donor in the interaction with the side chain of D2.50, and as a hydrogen-bond acceptor in the interaction with N7.45. In the AT1 receptor for angiotensin II, this Asn plays a critical role in stabilizing the inactive conformation of the receptor, as its replacement by Ala or Gly leads to constitutive activity. An additional 36% of GPCRs contain a smaller, but also polar side chain (S: 21%; C: 9%; T: 6%) at this 3.35 position. It is not clear whether this shorter S/C/T3.35 side chain is able to replace water #12 or if its role is to further stabilize water #12 through an extra hydrogen-bond.

Thus, there is a conserved hydrogen-bonding network linking D2.50 and W6.48 (present in 94% and 71%, respectively, of the sequences of family A GPCRs). However, N3.35 and N7.45 are only conserved in 29% and 67% of the sequences, respectively. Thus, we hypothesize that N3.35/N7.45-containing receptors form the D2.50–N3.35–N7.45–W6.48 network of interactions (Figure 3C) of GPCRs lacking N3.35 form a similar network through D2.50–water #12–N7.45–W6.48 interactions (Figure 3B) and if both N3.35 and N7.45 are absent in the sequence, as in rhodopsin, the D2.50–water #12–water #10–W6.48 (Figure 3A) network links D2.50 and W6.48. Therefore, the number of water molecules between D2.50 and W6.48 will depend on the type of side chain at positions 3.35 and 7.45. Disruption of this network leads to constitutive activity of the receptor because it facilitates the reported conformational transition of W6.48 during receptor activation.

A Water-Mediated Pro-kink in Transmembrane Helix 6

The highly conserved P6.50 in TM6 (100% in the rhodopsin family of GPCRs) is considered one of the most structurally and functionally important amino acids, as it is involved in the process of triggering the motions of the TM helices necessary for receptor activation. TM6 has a bend angle of 35° significantly higher than the average Pro-kink induced angle of 20°. In addition, P6.50 opens the 6.46–6.49 turn (> 3.6 residues/turn, twist < 100°, Figure 1), which disrupts the intrahelical hydrogen bond between the carbonyl group at position 6.47 and the N–H amide at position 6.51 (Figure 4). A discrete water molecule located in a small cavity between TMs 6 and 7 energetically stabilizes this unusual conformation. Water #1 acts as a hydrogen-bond acceptor in the interaction with the backbone amide at position 6.51, and as a hydrogen-bond donor in the interactions with the backbone carbonyls at positions 6.47 and 3.8 (Figure 4). Thus, in addition to stabilizing the kink of TM6, this water molecule links TMs 6 and 7. Importantly, 60% of class A GPCRs contain a nonbulky amino acid at position 7.42 (A: 40%; G: 20%); this creates a small cavity between TM6 and TM7 that allows accommodation of a water molecule. An additional 24% of GPCRs (as cannabinoids or acetylcholine receptors) contain a small and polar side chain (S: 13%; C: 7%; T: 4%) at this position. We hypothesize that this side chain further stabilizes the backbone carbonyl oxygen at position 6.47 through an extra hydrogen bond. The highly packed interface between TMs 6 and 7 at the extracellular domain of the helices, the unusual Pro-kink in TM6, and water #1 are probably conserved in most class A GPCRs.

Binding of agonists to the extracellular domain of the receptor modifies the conserved CWxP(F/Y) motif in TM6 and, in particular, the conformation of W6.48. The proposed conformational rearrangement of C6.47 and W6.48 during the process of receptor activation is associated with a structural change in the configuration of the Pro-kink in TM6. Upon receptor activation, the Pro-induced, water-mediated, significant bend angle of TM6 decreases, leading to a movement of
the cytoplasmic end of TM6 away from TM3, and disrupting the ionic interaction between R3.50 and the nearby negative side chains at positions 3.49 and 6.30.\(^\text{[46–49]}\) Clearly, water #1 plays a structural role in defining the inactive conformation of TM6 in the rhodopsin family of GPCRs. The role of this water molecule in defining the active conformation of TM6 remains unknown.

### Conclusions

The recent rhodopsin crystal structures\(^\text{[11–13]}\) have revealed the presence of several structural and/or functional internal water molecules within the TM bundle. It has been proposed that water molecules in the retinal binding pocket regulate spectral tuning in visual pigments and are, thus, specific to the opsin family. The fact that GPCRs interact with an extraordinary diversity of extracellular ligands suggests that each receptor subfamily has specific sequence motifs and structural water molecules to adjust the structural characteristics of their cognate transmitter. Thus, there is a low degree of sequence conservation among GPCR families at these extracellular domains;\(^\text{[18]}\) and the number, position, and role of these family-specific water molecules are difficult to estimate accurately.

In contrast, water molecules present in the vicinity of highly conserved residues regulate the activity of the receptor and they are probably conserved among the rhodopsin family of GPCRs. The most conserved residues in the family of GPCRs are clustered in the central and intracellular regions of the receptor.\(^\text{[18]}\) Therefore, we propose that water molecules in the environment of D2.50 of the (N/S)LxxxD motif in TM2; W6.48 and P6.50 of the CWxP(Y) motif in TM6; and N7.49, P7.50, and Y7.53 of the NPxY motif (see Table 1) are most likely present in the rhodopsin family of GPCRs. These water molecules are, therefore, another conserved structural elements in the family. Agonist binding or constitutively active mutations result in disruption of the interhelical interactions, mediated by these highly conserved motifs or/and these internal water molecules, leading to the stabilization of an active state of the receptor.

### Acknowledgements

This work was supported by the European Community (LSHB-CT-2003-503337), MEC (SAF2004-07103, SAF2006-04966), and AGAUR (SGR2005-00390).

### Keywords:
G-protein-coupled receptors • molecular modeling • Pro-kink • structure–activity relationships • water