DOI: 10.1002/cbic.201200189

Impact of Helix Irregularities on Sequence Alignment and Homology Modeling of G Protein-Coupled Receptors

Angel Gonzalez, Arnau Cordomí, Gianluigi Caltabiano, and Leonardo Pardo*[a]

Comparison of the crystal structures of G protein-coupled receptors (GPCRs) revealed backbone irregularities in the majority of the transmembrane (TM) helices. Among these, wide (π bulge) and tight (3_{10}) helical turns on TM2 and TM5 deserve special attention because of their proximity to the ligand binding site. These irregularities are related to residue insertion or deletion (reflected by inclusion of gaps in sequence alignments) accumulated during the evolution of these two helices. These findings have direct implications for the sequence alignments, phylogeny reconstruction, and homology modeling of class A GPCRs.

Introduction

G protein-coupled receptors (GPCRs) represent the largest family among integral membrane proteins, and constitute about 3% of the human proteome.[1] Their origins are presumed to be ancestral because GPCRs are present in almost every eukaryotic organism, including fungi and plants, and they are highly distributed in mammals.[2] These receptors are responsible for the translation of various endogenous and exogenous signals to cellular responses through their cognate heterotrimeric G proteins.[3] These receptors are essential in cell physiology, and their malfunction commonly results in pathological outcomes.[4] Thus, GPCRs are very important targets in drug discovery.[5] GPCRs are grouped in distinct classes or families, with class A (or rhodopsin-like) being the largest and most studied.[6] Significant advances in crystallization of GPCRs[7] have permitted the elucidation of the structures of several receptors (Table 1; see [8] for a review). All these structures share a common molecular architecture, characterized by the presence of seven α-helical transmembrane (TM) segments connected to each other by three extracellular loops (ECL) and three intracellular loops (ICL), with a disulfide bridge between ECL2 and TM3, and a cytoplasmic C terminus that contains an α helix (Hx8) oriented parallel to the cell membrane.[9]

Table 1. Crystal structures of antagonist-bound G protein-coupled receptors.

Receptor	Year[a]	Ligand	PDB ID	Ref.
bovine rhodopsin (bRho)	2000	11-*cis* retinal	1F88	[17]
squid rhodopsin (sRho)	2008	11-*cis* retinal	2Z73	[18]
β_2-adrenergic (β_2AR)	2007	carbazol	2RH1	[19]
β_1-adrenergic (β_1AR)	2008	cyanopindolol	2VT4	[20]
adenosine A_{2A} (A_{2A}R)	2008	ZM241385	3EML	[21]
chemokine CXCR4	2010	IT1t	3ODU	[22]
dopamine D_3 (D_3R)	2010	eticlopride	3PBL	[23]
histamine H_1 (H_1R)	2011	doxepin	3RZE	[24]
muscarinic M_2 (M_2R)	2012	3-quinuclidinyl-benzilate	3UON	[25]
muscarinic M_3 (M_3R)	2012	tiotropium	4DAJ	[26]
μ-opioid (μOR)	2012	β-funaltrexamine	4DKL	[27]
sphingosine $S1P_1$ ($S1P_1$R)	2012	ML056	3V2Y	[11]

[a] Of publication.

Sequence and Structure Conservation in Class A G Protein-Coupled Receptors

The increased number of available crystal structures permits an initial estimation of the correlation between sequence and structure in class A GPCRs. Figure 1 shows the superimposition of the currently available antagonist-bound crystal structures of GPCRs. Notably, ECLs (ECL2 in particular) are highly variable in sequence, length, and structure.[10] The ECL2 of rhodopsin (white in Figure 1 A), formed by two β-strands, buries the binding site from the extracellular environment, whereas ECL2 of CXCR4 and μOR (orange), also formed by two β-strands, fully exposes the binding site to the extracellular environment. In contrast, a helical segment forms ECL2 of β_1AR and β_2AR (green). This α-helix between TM4 and the disulfide bridge is probably not conserved in the other members of the biogenic amine receptor family, as it was not found in the structure of the dopamine D_3R or muscarinic receptors. In contrast, the N terminus (purple) and ECL2 (red; ECL2 lacks the disulfide bridge to TM3) of $S1P_1$R block the extracellular access of the ligand to the binding pocket. In this case ligand access to the binding site is via an opening between TM helices.[11] Thus, each receptor subfamily has probably evolved a specific ECL2 to accommodate the structural characteristics of its cognate ligands, and to modulate the ligand binding/unbinding events.[12]

[a] *Dr. A. Gonzalez,[+] Dr. A. Cordomí,[+] Dr. G. Caltabiano,[+] Prof. Dr. L. Pardo*
Laboratori de Medicina Computacional
Unitat de Bioestadística, Facultat de Medicina
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona (Spain)
E-mail: Leonardo.Pardo@uab.es
[+] *These authors contributed equally to this work.*

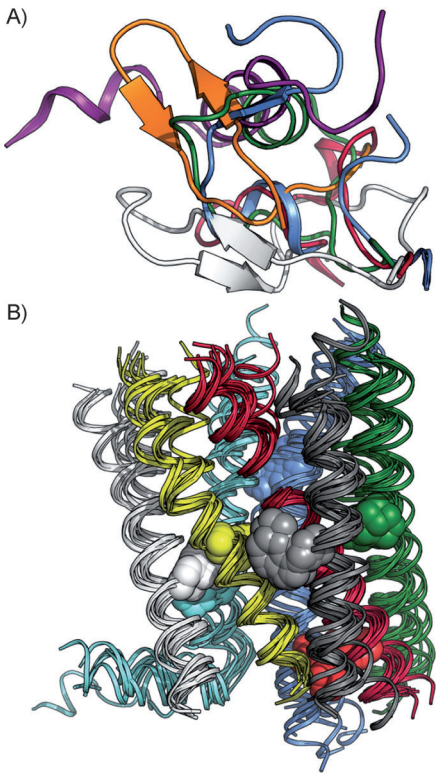

Figure 1. A) Comparison of ECL2 of bRho (white), β_2AR (green), A_{2A}R (blue), μOR (orange), and both the N terminus (purple) and ECL2 (red) of $S1P_1R$. B) Comparison of the TM domain plus Hx8 of the currently available antagonist-bound crystal structures of GPCRs (see Table 1): TM1 (white), TM2 (yellow), TM3 (red), TM4 (gray), TM5 (green), TM6 (dark blue), and TM7 (light blue). The highly conserved N1.50 (white), D2.50 (yellow), R3.50 (red), W4.50 (gray), P5.50 (green), P6.50 (dark blue), and P7.50 (light blue) are shown as spheres.

Conservation of protein sequences is often associated with conservation of protein structures. However, in the case of GPCRs, all the crystal structures retain analogous secondary/tertiary structures at the seven-helical-bundle domain (Figure 1B) despite low sequence identity in the TM segments (Table 2).[13] In contrast to that of other proteins, structure conservation in the GPCR family is associated with the presence of at least one highly conserved amino acid in each helix:[14] N in TM1 (present in 98% of the sequences), D in TM2 (93%), R in TM3 (95%), W in TM4 (96%), P in TM5 (76%), P in TM6 (98%), and P in TM7 (93%). This feature was used by Ballesteros and Weinstein[15] to define a general numbering scheme consisting of two numbers: the first (1–7) corresponds to the helix in which the amino acid of interest is located, while the second indicates its position relative to the most conserved residue (arbitrarily assigned 50) in the helix (see legend of Figure 1). Significantly, the position of these highly conserved amino acids in each helix is the same in the superimposition of the currently available crystal structures (Figure 1B). This finding validates the use of these amino acids as reference points in TM sequence alignments (instead of the common procedure of using substitution matrices and fast sequence-similarity search algorithms), and in the construction of homology models of GPCRs of unknown structure.[16]

However, these TM segments deviate significantly from ideal α helices, with structural anomalies like kinks and bulges at homologous positions. In some cases, these distortions are residue-dependent (both with Pro[28] and non-Pro residues),[29] while in other cases they are stabilized by complementary intra- and inter-helical interactions of polar side chains, backbone carbonyls, and structural and functional water molecules.[30] We have previously summarized these distortions.[30–31] In the following sections, we describe these conformational anomalies, in the form of wide and tight turns, observed in the extracellular parts of TM2 and TM5 in recent crystal structures. These kind of distortions are frequent in TM proteins and modify the polytopic membrane protein architecture[32] and ligand binding specificity,[33] and they have been associated with insertion or deletion (indels) of single residues into regular α-helices.[34]

Sequence and Structure of TM2 and ECL1

Structure-based alignment of TM2 and ECL1

Figure 2A shows the superimposition of TM2 and ECL1 of sRho, CXCR4, β_2AR and D_3R, as representative structures. The extracellular part TM2, which bends towards TM1 and away from TM3, is similar in all structures, despite the strongly divergent amino acid sequence with, for instance, a Pro at positions 2.58 (CXCR4, μOR), 2.59 (β_1AR, β_2AR, D_3R, H_1R, $A_{2A}R$) or 2.60 (sRho). The only exception is TM2 of the adenosine A_{2A} receptor ($A_{2A}R$), which contains Pro at position 2.59 but kinks towards TM3 because of the Cys bridge between ECL1 and ECL2, exclusive to this family (not shown), and TM2 of $S1P_1R$, which lacks Pro in the helix (see below). Unlike $S1P_1R$, bRho and muscarinic M_2R and M_3R receptors (which lack Pro) have a TM2 that bends towards TM1, thus making the shape of TM2 and ECL1 structurally similar to Pro-

Table 2. Sequence identity (%) in TM helices among G protein-coupled receptors of known structure (see Table 1).

	β_1AR	β_2AR	D_3R	H_1R	M_2R	M_3R	$A_{2A}R$	μOR	$S1P_1R$	CXCR4	bRho
β_2AR	69										
D_3R	42	39									
H_1R	37	36	36								
M_2R	30	29	33	37							
M_3R	32	34	30	37	69						
$A_{2A}R$	36	35	32	35	25	29					
μOR	25	27	28	28	26	25	24				
$S1P_1R$	28	26	25	25	25	25	30	19			
CXCR4	23	22	25	23	21	21	20	29	19		
bRho	20	21	27	19	21	18	22	23	20	19	
sRho	23	23	18	20	18	21	19	21	20	18	26

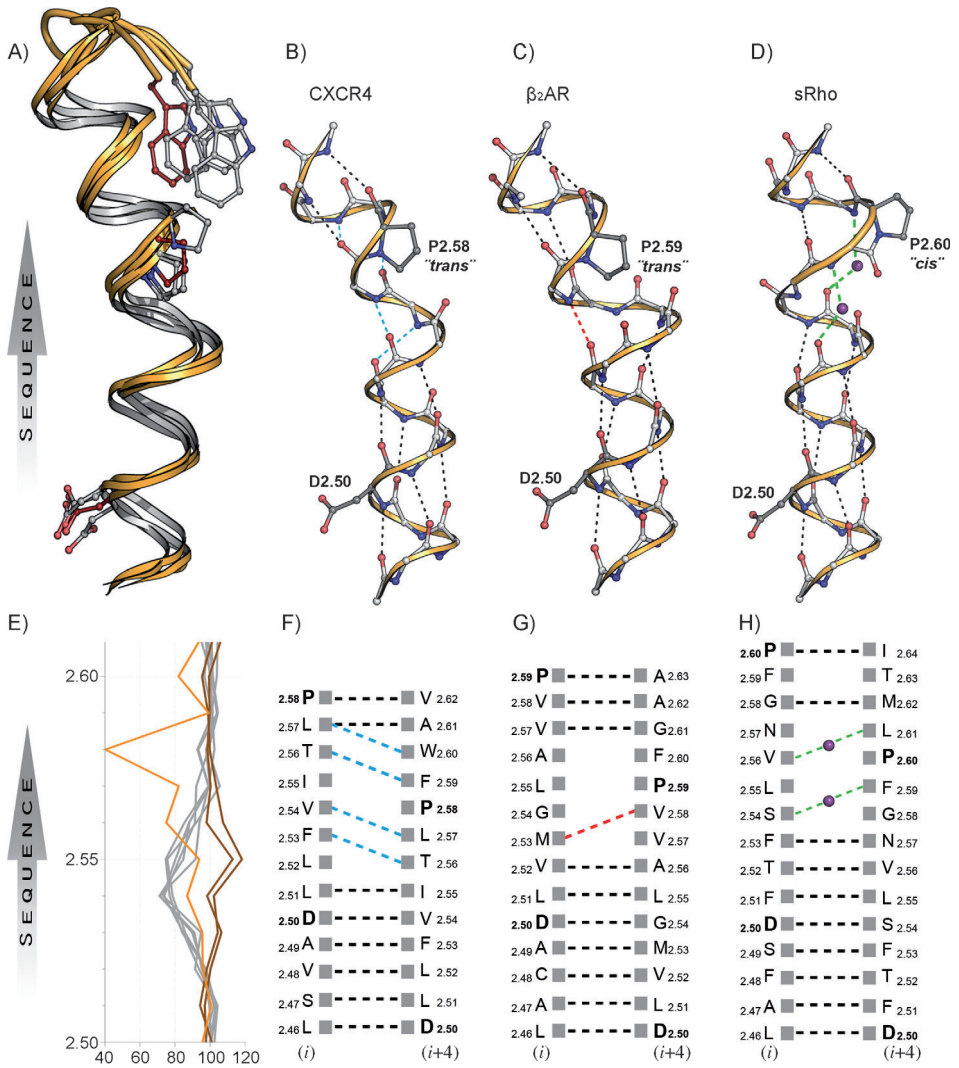

Figure 2. Structure of TM2 in GPCRs. A) Superimposition of TM2 and ECL1 of CXCR4, β_2AR, D_3R, and sRho. Conserved D2.50, Pro (2.58, 2.59, or 2.60), and Trp in ECL1 are shown (red for CXCR4 and grey for the others). B)–D), F)–H) Detailed network of interhelical hydrogen bonds in CXCR4 (B, F), β_2AR (C, G), and sRho (D, H): canonical $i\rightarrow i$+4 H-bonds (black), $i\rightarrow i$+3 (blue), $i\rightarrow i$+5 (red), and water-mediated (green). E) Comparison of the backbone twist angles (*x*-axis) along TM2 (CXCR4 and μOR in brown; β_1AR, β_2AR, D_3R, and H_1R in gray; sRho in orange). Twist angles measure local helix deformation:[35] canonical α helix (~3.6 residues per turn, twist angles ~100° (360°/3.6)), wide turns or π bulges (>3.6 residues per turn, twist angles <100°), tight turns and 3_{10} helices (<3.6 residues per turn, twist angles >100°).

containing receptors (see below). Interestingly, the superimposition reveals that Asp (D2.50) and the Pro residue at position 2.58 (CXCR4), 2.59 (β_2AR, D_3R) or 2.60 (sRho) are perfectly overlaid (Figure 2A). In order to translate this structural observation into the sequence space, a two-residue gap in the CXCR4 and μOR sequences and a one-residue gap in the bRho, β_1AR, β_2AR, D_3R, H_1R, M_2R, and M_3R sequences must be inserted (Figure 3A). Importantly, the conserved Trp residues in ECL1 (pointing toward the helical bundle, between TM2 and TM3) are also superimposed in the crystal structures (Figure 2A), with the exception of S1P_1R (not shown). Also, bRho possesses an aromatic Phe (F103) in ECL1 in place of the aromatic Trp in other receptors (Figure 3A).

Local distortions in the extracellular side of TM2

Because Asp (D2.50) and Pro (either at 2.58, 2.59, or 2.60) superimpose, the backbone helical conformation of the amino acids between these two residues must differ. Local structural changes can be described by backbone twist angles (see legend of Figure 2 for a structural interpretation). Figure 2E shows the pattern of helical twists along TM2. Clearly, in the vicinity of D2.50 of all receptors, TM2 adopts a canonical α helix with twist angles of ~100° (~3.6 residues per turn). In contrast, twist angles vary among receptors towards the extracellular side (from 2.53 to 2.60). In this region, TM2s of CXCR4 and μOR have twist angles of ~120° (~3.0 residues per turn), characteristic of a 3_{10} or tight turn (Figure 2E, brown line); β_1AR, β_2AR, D_3R, and H_1R have twist angles of ~75° (~4.8 residues per turn), which are distinctive of π bulges or wide turn helices (Figure 2E, grey lines), and sRho has a twist angle of ~40° (~9 residues per turn, Figure 2E, orange line). In other words, CXCR4 and μOR adopt a closed helical segment to accommodate seven amino acids between D2.50 and Pro; β_1AR, β_2AR, D_3R, and H_1R have an open helical segment to accommodate eight amino acids between D2.50 and Pro; and sRho adopts an extreme conformation to accommodate nine amino acids between D2.50 and Pro. A different backbone inter-helical hydrogen bond network accomplishes this. CXCR4 contains an arrangement of $i\rightarrow i$+3 hydrogen bond interactions between the backbone carbonyl and the NH amide groups in the 2.53–2.60 region (Figure 2B, F). In contrast, β_2AR displays a characteristic $i\rightarrow i$+5 hydrogen bond interaction pattern between the backbone carbonyl of 2.53 and the NH amide of 2.58 (Figure 2C, G). Finally, sRho presents an extreme distortion in the 2.54–2.61 region, characterized by a *cis* P2.60 backbone conformation that is stabilized by two water molecules (Figure 2C, H).

Pro in TM2 and Trp/Phe in ECL1 are highly conserved in class A GPCRs

Statistical analysis of multiple sequence alignments of TM2 in human non-olfactory class A GPCRs shows that Pro is present in 2% of the sequences at position 2.57 (melatonin and prostaglandin EP1 receptors; family classification adapted from IUPHAR),[36] 39% at position 2.58 (platelet-activating factor, leukotriene B4, purinoceptors, ghrelin, C5a anaphylatoxin, opioid, Fmet-leu-phe, proteinase-activated, urotensin, bradikynin, apelin, adrenomedullin, interleukin, somatostatin, chemokine, and melanin-concentrating hormone receptors), 36% at position 2.59 (gonadotropin-releasing hormone, melatonin, adrenoceptors, histamine, serotonin, dopamine, trace amine, prostacyclin, prostaglandin EP2-4, peropsin, bombesin, cholecystokinin, endothelin, galanin, orexin, prokineticin, prolactin-releasing peptide, neuromedin U, neuropeptide, neurotensin, protease-activated PAR4, somatostatin SST5, purinoceptor P2Y11, and adenosine receptors), and 4% at position 2.60 (Thyrotropin-releasing hormone, β_3-adrenergic, serotonin 5-HT$_6$, short-wave sensitive opsin 1, vasopressin, and oxytocin receptors). However, as Pro is at different positions in the sequences but at the same structural position (Figures 2A and 3A), the true conservation of the Pro residue at the extracellular part of TM2 in non-olfactory human class A GPCRs increase in 78% of the sequences. Interestingly, 63% of the entries containing Pro at 2.58 exhibit Ser (19%) or Thr (44%) at position 2.56, thus forming the S/TxP motif that has been described as a key modulator of TM2.[29c,37] Similarly, all crystal structures contain an aromatic Trp/Phe residue in ECL1 (Figure 2A, 3A), which is part of the previously identified (W/F)x(F/L)G motif.[38] Statistical analysis of the sequences of ECL1 in human non-olfactory class A GPCRs shows that W or F is present in 79% of the sequences (W:73%, F:6%), F/L is in 66% of the sequences (F:47%, L:19%), and G is in 67% of the sequences. We conclude from this conservation pattern that both Pro in TM2 and (W/F)x(F/L)G in ECL1 form a common structural feature in the significantly divergent extracellular domain of GPCRs.

The conformation of TM2 in GPCRs lacking Pro

As noted before, bRho, M$_2$R, M$_3$R, and S1P$_1$R all lack Pro in TM2; however, their conformations differ. Figure 4A shows that bRho, M$_2$R, and M$_3$R have twist angles of ~75° (~4.8 residues per turn, solid gray lines), similar to the values for β_2AR (dotted gray line) and other Pro-containing receptors (gray lines in Figure 2E). Thus, TM2 of bRho, M$_2$R, and M$_3$R possess the distinctive π bulge (or wide turn) conformation and the bend of the helix toward TM1 in Pro-containing receptors (see Figure 4A). The GGxTT motif (positions 2.56–2.60 in bRho) makes the conformation of TM2 structurally similar to that of the other Pro-containing receptors. Mutation of either of the two Gly leads to alterations in the

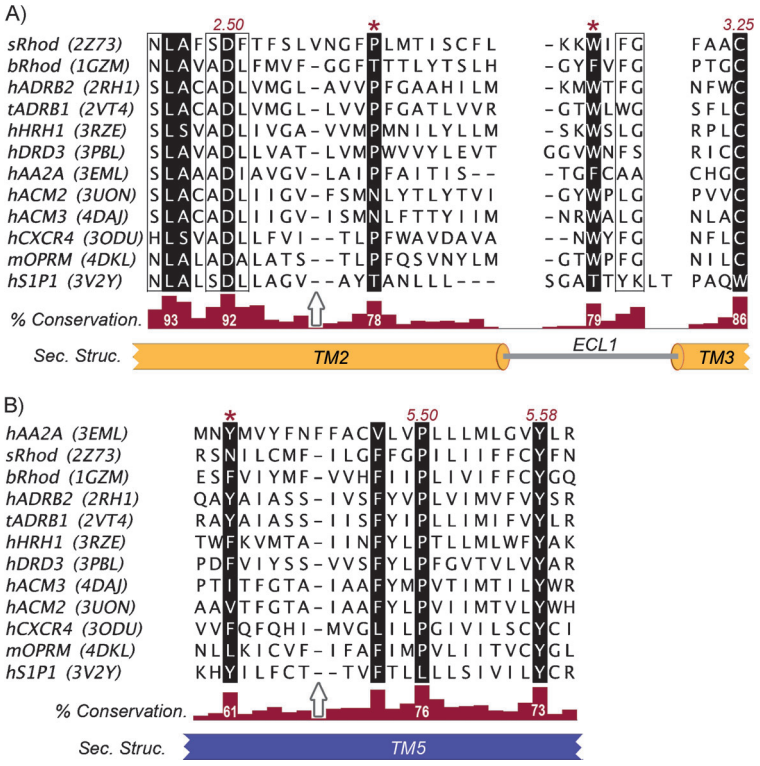

Figure 3. Sequence alignments of A) TM2-ECL1-TM3, and B) TM5 of GPCRs with known structures. Highly conserved residues are shown in black. Conservation in human nonolfactory class A GPCRs is shown by the histogram (red). Arrows show the position in the TM region where gaps must be introduced to correctly aligned the amino acids (marked by *) as observed in the crystal structures (see text). The name of the gene and the PDB ID of each protein are shown: sRhod (2Z73) is sRho in the text, bRhod (1GZM) is bRho, hADRB2 (2RH1) is β_2AR, tADRB1 (2VT4) is β_1AR, hHRH1 (3RZE) is H$_1$R, hDRD3 (3PBL) is D$_3$R, hAA2A (3EML) is A$_{2A}$R, hACM2 (3UON) is M$_2$R, hACM3 (4DAJ) is M$_3$R, hCXCR4 (3ODU) is CXCR4, mOPRM (4DKL) is πOR and hS1P1 (3V2Y) is S1P$_1$R.

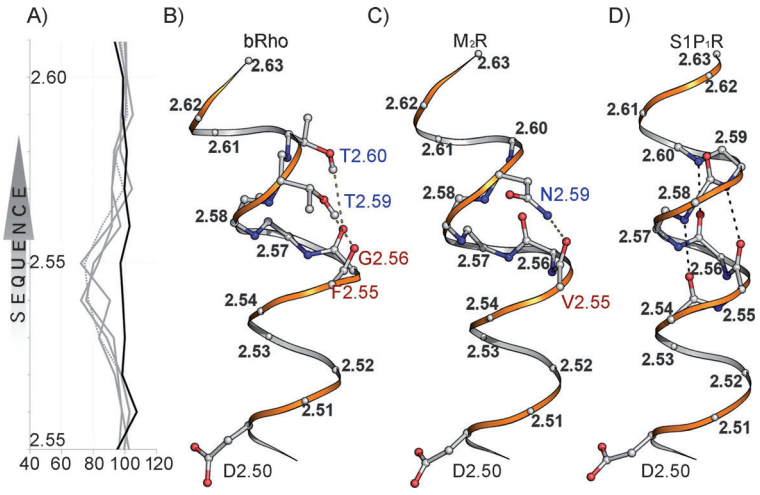

Figure 4. Structure of TM2 in GPCRs lacking Pro. A) Comparison of the backbone twist angles (*x*-axis) along TM2 (bRho, M$_2$R, and M$_3$R in solid gray; S1P$_1$R in black; Pro-containing β_2AR in dotted grey). See legend of Figure 2 for interpretation of twist angles. B)–D) Detailed network of inter-helical hydrogen bonds in bRho, M$_2$R, and S1P$_1$R.

photoactivation pathway of rhodopsin.[39] The opening of the helix by GGxTT is achieved by the characteristic $i \rightarrow i+5$ hydrogen bond interaction pattern (Figure 4B), and, in addition, hydrogen bond interactions between the side chains of T2.59 and T2.60 and the backbone carbonyls at 2.55 and 2.56, respectively, which lack their NH counterpart. The short side chain of Thr is able to hydrogen bond to the backbone carbonyl in the previous turn of the helix, thereby inducing and/or stabilizing distortions in TMs.[29a,b] Similarly, M_2R and M_3R possess, in addition to the $i \rightarrow i+5$ hydrogen bond interactions, a hydrogen bond interaction between the side chain of N2.59 and the backbone carbonyl at 2.55 (Figure 4C). Figure 3A shows the structure-based alignment of the TM2s of bRho, M_2R, and M_3R with the other receptors.

In contrast to bRho, M_2R, and M_3R, the absence of Pro in TM2 of $S1P_1R$ leads to a canonical α helix at the extracellular domain, with twist angles of ~100° (~3.6 residues per turn, black line; Figure 4A and D). This conformation of TM2 moves its extracellular part away from the TM bundle (relative to the other structures), and modifies the orientation of the side chains at the extracellular side. Clearly, the alignment of the $S1P_1R$ sequence to the other receptors requires a two-residue gap relative to sRho, no gaps relative to CXCR4 and μOR, and a one-residue gap relative to all other structures (Figure 3A).

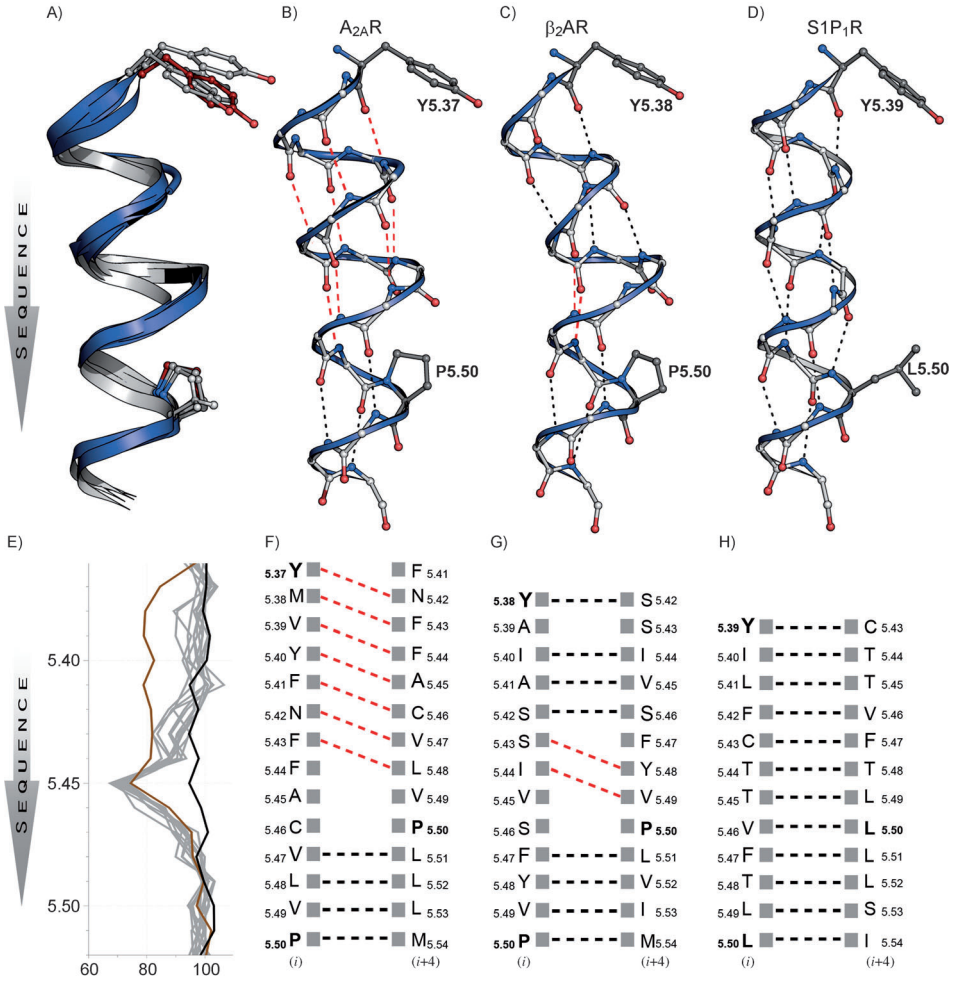

Figure 5. Structure of TM5 in GPCRs. A) Superimposition of TM5 of $A_{2A}R$, β_2R, and $S1P_1R$. Conserved P5.50 and Phe/Tyr at 5.37 ($A_{2A}R$), 5.38 (β_2R), or 5.39 ($S1P_1R$) are shown (red for $A_{2A}R$; gray for others). B)–D), F)–H) Detailed network of inter-helical hydrogen bonds in $A_{2A}R$ (B, F), β_2AR (C, G), and $S1P_1R$ (D, H). See Legend of Figure 2 for the color code of the hydrogen bonds. E) Comparison of the backbone twist angles (*x*-axis) along TM5 ($A_{2A}R$ in brown; $S1P_1R$ in black; others in gray). See legend of Figure 2 for interpretation of twist angles.

Sequence and Structure of TM5

The opening of TM5 at the extracellular part

P5.50 (conserved in 76% of the rhodopsin-like sequences) induces a local opening of TM5 at the 5.43–5.48 turn (Pro-unwinding) in all crystal structures except $S1P_1R$ (see below), and has been proposed to be involved in the mechanism of ligand-induced receptor activation.[31b,40] This leads to twist angles less than 100° (~5 residues per turn), characteristic of a π bulge or wide turn (Figure 5E). The opening of the helix at the 5.43–5.48 turn is achieved by backbone hydrogen bond interactions between the carbonyl at position 5.44 and the NH amide at 5.49 ($i \rightarrow i+5$ H-bond), and between the carbonyl at position 5.43 and both the NH amides at 5.48 ($i \rightarrow i+5$ H-bond) and 5.47 ($i \rightarrow i+4$ H-bond) in the majority of receptors. The extracellular end of TM5 adopts a standard α helix (twist ~100°, ~3.6 residues per turn, $i \rightarrow i+4$ H-bond) in all structures with the single exception of $A_{2A}R$. In the later case, the local opening of the helix is further extended, from 5.35 to 5.48 (Figure 5E, brown line). Accordingly, $A_{2A}R$ contains, in contrast to the other structures, a pattern of seven consecutive $i \rightarrow i+5$ H-bonds from 5.37 to 5.48 (Figure 5B and F). This different opening of the helix in $A_{2A}R$ and the other receptors modifies the orientation of the amino acid side chains at the extracellular domain. Thus, as superimposition of all TM5 structures reveals (Figure 5A), a one-residue gap in the alignment of all other receptors with respect to $A_{2A}R$ must be introduced (Figure 3B). As a consequence, Y5.37 of $A_{2A}R$ overlays F/Y at 5.38 of the remaining structures (Figure 5A).[8]

The conformation of TM5 in GPCRs lacking P5.50

P5.50 is absent in melanocortin, glycoprotein hormone, lysosphingolipid, prostanoid, and cannabinoid receptors. In these, the similarly conserved Y5.58 (73% of the sequences), functionally involved in the stabilization of the active state of the receptor by interacting with R3.50 of the (D/E)RY motif in TM3, as revealed by the crystal structures of β_2AR in complex with G protein[41] and the ligand-free opsin,[42] is used as reference for sequence alignment of TM5 (Figure 3B). Notably, superimposition of the structure of $S1P_1R$ that lacks P5.50 with the other P5.50-containing receptors shows that the position of Y5.58 is the same in all cases (not shown). The absence of Pro in TM5 of $S1P_1R$ leads to a regular α-helical conformation with twist angles of $\sim 100^\circ$ ($\sim$3.6 residues per turn; Figure 5E, black line). Figure 5A shows TM5 of $S1P_1R$ superimposed on those of β_2AR and $A_{2A}R$. Clearly, the aromatic Tyr residues at the extracellular side of TM5 (5.37 in $A_{2A}R$, 5.38 in β_2AR, 5.39 in $S1P_1R$.) are perfectly superimposed. Figure 3B shows the proposed sequence alignment of TM5: $S1P_1R$ needs a two-residue gap relative to $A_{2A}R$ and a one-residue gap relative to all other structures. Importantly, all GPCR members lacking P5.50 contain an aromatic Phe/Tyr/Trp side chain at position 5.39 (e.g., glycoprotein hormone, lysosphingolipid, prostanoid, and cannabinoid receptors) with the sole exception of the melanocortin receptors. Melanocortin receptors together with protease-activated receptors form a unique family of GPCRs because their N-terminal domain is a tethered ligand that interacts with the TM domain for their constitutive activity.[43] This is possible because of a very short ECL2 and a TM4–TM5 interface, exclusive to this family.

We conclude, based on this analysis, that the presence of a highly conserved aromatic Phe/Tyr side chain (present in 61% of human non-olfactory class A GPCRs sequences, Figure 3B) at the extracellular part of TM5 (either at position 5.37 in A_{2A}-like receptors, 5.38 in the other receptors of known structures, or 5.39 in receptors lacking P5.50) together with P5.50 and Y5.58, constitute conserved structural features in TM5 of class A GPCRs.

Outlook

A comparison of the available crystal structures of class A GPCRs has revealed changes in the α-helical scaffold in the form of tight (3_{10} turn) and wide (π bulge) turns at the extracellular ends of TM2 and TM5. These differences are related to residue insertion or deletion events accumulated during the evolution of these two helices. In TM2, P2.58 of CXCR4 and μOR at position i+8 relative to D2.50 is structurally aligned to P2.59 of β_1AR, β_2AR, D_3R, and H_1R (i+9), and P2.60 of sRho (i+10). The GGxTT motif of bRho and N2.59 of M_2R and M_3R (both of which lack Pro in TM2) cause similar conformation of the helix as for the Pro-containing receptors. In contrast, the absence of Pro in TM2 of $S1P_1R$ leads to a regular α helix. These structural alignments require gaps within the 2.50–2.59 region. On the other hand, $A_{2A}R$ displays an extended opening of TM5 (positions 5.35–5.48), in contrast to other P5.50-containing structures in which the opening of the helix is restricted to the 5.43–5.48 range. In contrast, TM5 of $S1P_1R$, which lacks P5.50, adopts a canonical α-helical conformation. Thus, the alignment of the $S1P_1R$ sequence to the other receptors requires a two-residue gap relative to $A_{2A}R$ and a one-residue gap relative to all other sequences, to overlay Y5.37 ($i-13$ relative to P5.50) of $A_{2A}R$ with F/Y5.38 of the P5.50-containing structures ($i-12$ relative to P5.50) and F/Y/W5.39 of the P5.50-lacking structures. The observed anomalies in TMs 2 and 5 could have played an important role in the diversification and evolutionary success of GPCRs, given that indels and amino acid substitutions are the most common events in protein evolution.[38] Moreover, these findings contradict the established paradigm of avoiding gaps in alignments of the TM regions of GPCRs. Importantly, none of the employed numbering schemes for class A GPCRs[15,44] takes into account the gaps in TM helices, thus these schemes do not reflect the position of the amino acids in the structure in these domains of the receptors.[8] These findings have direct implications in sequence alignments, phylogeny reconstruction, and homology modeling of the rhodopsin-like GPCRs.

Acknowledgements

This work was supported by grants from MICINN (SAF2010-22198-C02-02), and ISCIII (RD07/0067/0008). AC is a receiver of a contract grant from ISCIII.

Keywords: G protein-coupled receptors · molecular modeling · sequence alignment · structure elucidation

[1] L. Fagerberg, K. Jonasson, G. von Heijne, M. Uhlén, L. Berglund, *Proteomics* **2010**, *10*, 1141 – 1149.
[2] R. Fredriksson, H. B. Schiöth, *Mol. Pharmacol.* **2005**, *67*, 1414 – 1425.
[3] W. M. Oldham, H. E. Hamm, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 60 – 71.
[4] M. J. Smit, H. F. Vischer, R. A. Bakker, A. Jongejan, H. Timmerman, L. Pardo, R. Leurs, *Annu. Rev. Pharmacol. Toxicol.* **2007**, *47*, 53 – 87.
[5] P. Imming, C. Sinning, A. Meyer, *Nat. Rev. Drug Discovery* **2006**, *5*, 821 – 834.
[6] R. Fredriksson, M. C. Lagerström, L. G. Lundin, H. B. Schiöth, *Mol. Pharmacol.* **2003**, *63*, 1256 – 1272.
[7] a) P. W. Day, S. G. F. Rasmussen, C. Parnot, J. J. Fung, A. Masood, T. S. Kobilka, X.-J. Yao, H.-J. Choi, W. I. Weis, D. K. Rohrer, B. K. Kobilka, *Nat. Methods* **2007**, *4*, 927 – 929; b) M. J. Serrano-Vega, F. Magnani, Y. Shibata, C. G. Tate, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 877 – 882.
[8] V. Katritch, V. Cherezov, R. C. Stevens, *Trends Pharmacol. Sci.* **2012**, *33*, 17 – 27.
[9] G. Liapakis, A. Cordomi, L. Pardo, *Curr. Pharm. Des.* **2012**, *18*, 175 – 185.
[10] M. C. Peeters, G. J. van Westen, Q. Li, A. P. Ijzerman, *Trends Pharmacol. Sci.* **2011**, *32*, 35 – 42.
[11] M. A. Hanson, C. B. Roth, E. Jo, M. T. Griffith, F. L. Scott, G. Reinhart, H. Desale, B. Clemons, S. M. Cahalan, S. C. Schuerer, M. G. Sanna, G. W. Han, P. Kuhn, H. Rosen, R. C. Stevens, *Science* **2012**, *335*, 851 – 855.
[12] a) R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13118 – 13123; b) A. González, T. Perez-Acle, L. Pardo, X. Deupi, *PLoS One* **2011**, *6*, e23815.
[13] J. C. Mobarec, R. Sanchez, M. Filizola, *J. Med. Chem.* **2009**, *52*, 5207 – 5216.
[14] T. Mirzadegan, G. Benkö, S. Filipek, K. Palczewski, *Biochemistry* **2003**, *42*, 2759 – 2767.
[15] J. A. Ballesteros, H. Weinstein, *Methods Neurosci.* **1995**, *25*, 366 – 428.

[16] a) E. Christiansen, C. Urban, M. Grundmann, M. E. Due-Hansen, E. Hagesaether, J. Schmidt, L. Pardo, S. Ullrich, E. Kostenis, M. Kassack, T. Ulven, *J. Med. Chem.* **2011**, *54*, 6691 – 6703; b) I. Marco, M. Valhondo, M. Martín-Fontecha, H. Vázquez-Villa, J. Del Río, A. Planas, O. Sagredo, J. A. Ramos, I. R. Torrecillas, L. Pardo, D. Frechilla, B. Benhamú, M. L. López-Rodríguez, *J. Med. Chem.* **2011**, *54*, 7986 – 7999.
[17] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739 – 745.
[18] M. Murakami, T. Kouyama, *Nature* **2008**, *453*, 363 – 367.
[19] a) V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, R. C. Stevens, *Science* **2007**, *318*, 1258 – 1265; b) D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, X. J. Yao, W. I. Weis, R. C. Stevens, B. K. Kobilka, *Science* **2007**, *318*, 1266 – 1273.
[20] T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. Leslie, C. G. Tate, G. F. Schertler, *Nature* **2008**, *454*, 486 – 491.
[21] V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, R. C. Stevens, *Science* **2008**, *322*, 1211 – 1217.
[22] B. Wu, E. Y. T. Chien, C. D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F. C. Bi, D. J. Hamel, P. Kuhn, T. M. Handel, V. Cherezov, R. C. Stevens, *Science* **2010**, *330*, 1066 – 1071.
[23] E. Y. T. Chien, W. Liu, Q. Zhao, V. Katritch, G. W. Han, M. A. Hanson, L. Shi, A. H. Newman, J. A. Javitch, V. Cherezov, R. C. Stevens, *Science* **2010**, *330*, 1091 – 1095.
[24] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, T. Kobayashi, R. C. Stevens, S. Iwata, *Nature* **2011**, *475*, 65 – 70.
[25] K. Haga, A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga, T. Kobayashi, *Nature* **2012**, *482*, 547 – 551.
[26] A. C. Kruse, J. Hu, A. C. Pan, D. H. Arlow, D. M. Rosenbaum, E. Rosemond, H. F. Green, T. Liu, P. S. Chae, R. O. Dror, D. E. Shaw, W. I. Weis, J. Wess, B. K. Kobilka, *Nature* **2012**, *482*, 552 – 556.
[27] A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka, S. Granier, *Nature* **2012**, *485*, 321 – 326.
[28] a) F. S. Cordes, J. N. Bright, M. S. Sansom, *J. Mol. Biol.* **2002**, *323*, 951 – 960; b) S. Yohannan, S. Faham, D. Yang, J. P. Whitelegge, J. U. Bowie, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 959 – 963.
[29] a) J. A. Ballesteros, X. Deupi, M. Olivella, E. E. J. Haaksma, L. Pardo, *Biophys. J.* **2000**, *79*, 2754 – 2760; b) X. Deupi, M. Olivella, A. Sanz, N. Dölker, M. Campillo, L. Pardo, *J. Struct. Biol.* **2010**, *169*, 116 – 123; c) X. Deupi, M. Olivella, C. Govaerts, J. A. Ballesteros, M. Campillo, L. Pardo, *Biophys. J.* **2004**, *86*, 105 – 115.
[30] a) L. Pardo, X. Deupi, N. Dölker, M. L. López-Rodríguez, M. Campillo, *ChemBioChem* **2007**, *8*, 19 – 24; b) T. E. Angel, M. R. Chance, K. Palczewski, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8555 – 8560.
[31] a) X. Deupi, N. Dölker, M. López-Rodríguez, M. Campillo, J. Ballesteros, L. Pardo, *Curr. Top. Med. Chem.* **2007**, *7*, 991 – 998; b) K. Sansuk, X. Deupi, I. R. Torrecillas, A. Jongejan, S. Nijmeijer, R. A. Bakker, L. Pardo, R. Leurs, *Mol. Pharmacol.* **2011**, *79*, 262 – 269.
[32] R. P. Riek, I. Rigoutsos, J. Novotny, R. M. Graham, *J. Mol. Biol.* **2001**, *306*, 349 – 362.
[33] R. P. Riek, A. A. Finch, G. E. Begg, R. M. Graham, *Mol. Pharmacol.* **2008**, *73*, 1092 – 1104.
[34] R. B. Cooley, D. J. Arp, P. A. Karplus, *J. Mol. Biol.* **2010**, *404*, 232 – 246.
[35] M. Bansal, S. Kumar, R. Velavan, *J. Biomol. Struct. Dyn.* **2000**, *17*, 811 – 819.
[36] J. L. Sharman, C. P. Mpamhanga, M. Spedding, P. Germain, B. Staels, C. Dacquet, V. Laudet, A. J. Harmar, *Nucleic Acids Res.* **2011**, *39*, D534 – D538.
[37] C. Govaerts, C. Blanpain, X. Deupi, S. Ballet, J. A. Ballesteros, S. J. Wodak, G. Vassart, L. Pardo, M. Parmentier, *J. Biol. Chem.* **2001**, *276*, 13217 – 13225.
[38] J. M. Klco, G. V. Nikiforovich, T. J. Baranski, *J. Biol. Chem.* **2006**, *281*, 12010 – 12019.
[39] D. Toledo, E. Ramon, M. Aguila, A. Cordomí, J. J. Pérez, H. F. Mendes, M. E. Cheetham, P. Garriga, *J. Biol. Chem.* **2011**, *286*, 39993 – 40001.
[40] S. G. F. Rasmussen, H.-J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. DeVree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Konetzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis, B. K. Kobilka, *Nature* **2011**, *469*, 175 – 180.
[41] S. G. F. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, B. K. Kobilka, *Nature* **2011**, *477*, 549 – 555.
[42] J. H. Park, P. Scheerer, K. P. Hofmann, H. W. Choe, O. P. Ernst, *Nature* **2008**, *454*, 183 – 187.
[43] S. Srinivasan, C. Lubrano-Berthelier, C. Govaerts, F. Picard, P. Santiago, B. R. Conklin, C. Vaisse, *J. Clin. Invest.* **2004**, *114*, 1158 – 1164.
[44] a) J. M. Baldwin, *EMBO J.* **1993**, *12*, 1693 – 1703; b) T. W. Schwartz, *Curr. Opin. Biotechnol.* **1994**, *5*, 434 – 444.

Received: March 19, 2012

