

Current Topics

Sequence Analyses of G-Protein-Coupled Receptors: Similarities to Rhodopsin[†]

Tara Mirzadegan,^{*,‡} Gil Benkö,[‡] Sławomir Filipek,[§] and Krzysztof Palczewski^{||,⊥,¶}

Roche Bioscience, Inflammatory Disease Unit, Palo Alto, California 94304, Departments of Ophthalmology, Pharmacology, and Chemistry, University of Washington, Seattle, Washington 98195, International Institute of Molecular and Cell Biology, and Faculty of Chemistry, Warsaw University, Warsaw 02093, Poland

Received November 22, 2002; Revised Manuscript Received January 8, 2003

G-protein-coupled receptors (GPCRs)¹ constitute a large superfamily of receptor proteins responsible for signal transduction (see <http://www.gpcr.org/7tm>). Throughout all higher organisms, these receptors mediate recognition of environmental stimuli like light, odor, and taste, but also hormonal and other types of communications across plasma membranes (1). They are also important targets for pharmacological intervention via activating or blocking their action (2). Three families of GPCRs were identified, with family A being by far the largest (reviewed in refs 3–5). Its members are more closely related to each other within a few functional domains than those of the other families. In addition, numerous diseases have been linked to specific mutations within the genes encoding GPCRs, also making these receptors targets for specific therapeutic interventions including gene transfer (6–9).

Recently, two events influenced our thinking about GPCRs. First, the near completion of the draft sequence of the human genome yielded a more complete list of GPCRs

(10, 11) (see also a recent review, ref 12). Second, a high-resolution structure of the prototypical GPCR, rhodopsin, was elucidated (13), providing the first molecular details about how this GPCR binds its ligand, the chromophore 11-*cis*-retinal, and how it may activate cognate G-proteins (14–18). The high-resolution structure of rhodopsin also allowed construction and validation of numerous prediction models of other GPCRs (for example, see refs 3 and 19–21). We combined these two sets of experimental data and performed a comprehensive analysis of the primary sequences of GPCRs from family A. We demonstrate that the extracellular domain is the least conserved, while GPCRs display considerable conservation toward the cytoplasmic side. Interestingly, the lengths of connecting loops also suggest an importance of these elements in retaining a similar general structure of these receptors. These structural similarities may suggest a common mechanism by which GPCRs activate G-proteins and how these receptors may self-organize within native membranes.

Three Structural Models of Rhodopsin. There are three high-resolution models currently deposited in the Protein DataBank (PDB). All three models are based on electron density maps obtained from crystals of rhodopsin generated by the same method. Rhodopsin was isolated from bovine rod outer segments to high purity using a mild detergent (such as nonyl- β -glucoside) and high concentrations of divalent metal ions (e.g., 80 mM Zn²⁺) (22). Vapor diffusion crystallization techniques were used to crystallize rhodopsin, using ammonium sulfate as the precipitant in nonyl- β -glucoside detergent solutions (23). Diffraction quality crystals that were sensitive to visible light (18) were obtained. All

[†] This research was supported by NIH Grant EY08061, a grant from Research to Prevent Blindness, Inc. (RPB) to the Department of Ophthalmology at the University of Washington, and a grant from the E.K. Bishop Foundation. K.P. is a RPB Senior Investigator.

^{*} Corresponding authors. (K.P.) Tel: (206) 543-9074. Fax: (206) 221-6784. E-mail: palczewski@u.washington.edu. (T.M.) Tel: (650) 855-5814. Fax: (650) 354-2442. E-mail: tara.mirzadegan@roche.com.

[‡] Roche Bioscience.

[§] Warsaw University.

^{||} Department of Ophthalmology, University of Washington.

[⊥] Department of Pharmacology, University of Washington.

[¶] Department of Chemistry, University of Washington.

¹ Abbreviations: GPCR, G-protein-coupled receptors; Gt, photo-receptor G-protein transducin.

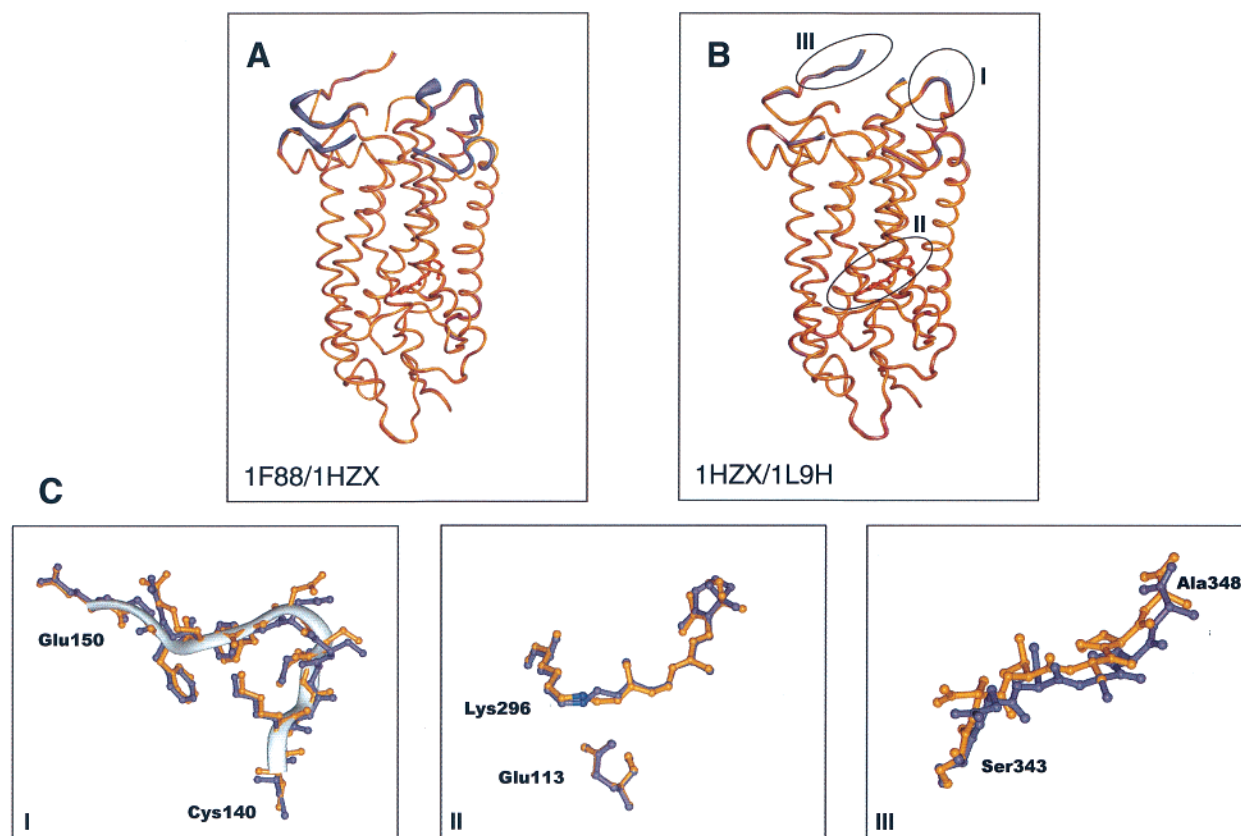


FIGURE 1: Comparison of the structural models of rhodopsin. (A) Comparison of 1F88 and 1HZX, and 1HZX and 1L9H coordinates. 1F88 ribbon is in orange and treated as a reference structure. 1HZX ribbon is colored orange to violet. Color as well as the width of 1HZX ribbon change is proportional to distances between identical amino acids (orange—no difference, violet—maximal difference). (B) Superimposed structures of 1HZX (orange) and 1L9H (orange to violet). Color as well as the width of 1L9H ribbon change proportionally to distances between identical amino acids. Spots with maximal discrepancy between the two structures are denoted as I, II, and III and magnified in panel C. (C) Comparison of the most different parts of 1HZX (orange) and 1L9H (violet) within selected regions. Both structures were superimposed as a whole, and no fitting for particular parts of the structure was made.

of these crystals were merohedrally twinned to different degrees, affecting the quality of the final refinement (23). The initial model, deposited in the PDB under identifier 1F88, showed all major structural features as predicted from years of biochemical and biophysical studies on wild type and mutated proteins (13, 24). Crystallographic refinement of the two molecules in the asymmetric unit generated the model of rhodopsin at 2.8 Å resolution (15) deposited under the identifier 1HZX. The newest structural model was obtained using X-ray diffraction with 2.6 Å resolution (25), using almost identical methods, and deposited in the PDB under the 1L9H identifier. Various models followed, including those of cone pigments (1KPN, 1KPW, and 1KPX) (16) and other receptors based on rhodopsin's structure (e.g., refs 26 and 27).

Here, we compare these structural models to identify differences among them. Differences between 1F88 and refined 1HZX structural models are located mainly in the cytoplasmic side of rhodopsin (Figure 1A; these differences are marked by the color and radius of the ribbon). The root-mean-square distance between C α atoms in 1F88 and 1HZX is equal to 1.05 Å. In 1HZX, there are three additional amino acid residues that are missing in 1F88, Leu³²⁸—Gly³²⁹—Asp³³⁰, which are located shortly after cytoplasmic helix 8. One amino acid (Ser²⁴⁰) was removed from the loop between helices V and VI (loop C-III) because of the ambiguity of its configuration. In 1F88, the C-terminus (334–348) was

filled with Ala residues; now, in 1HZX, all side chains are represented, and the whole configuration of this region has been remodeled. In this region, the most altered positions are residues from both sides of the second gap, Thr³⁴⁰ and Pro³²⁷, which are displaced by 5.5 Å from their initial positions in 1F88. Loop C-III was rebuilt and is most affected around Ala²⁴¹ and Thr²⁴², which are displaced by 11.6 and 6.8 Å from their respective positions in 1F88. Lys¹⁴¹ from C-II connecting helices III and IV was relocated by 2.9 Å. Loops from the extracellular side and the chromophore were changed minimally. The largest changes were for Glu¹⁹⁷ (0.8 Å) and C₁₅ in the Schiff-base region (1.1 Å) and for C₁₇ in the β -ionone ring (0.9 Å) of 11-*cis*-retinylidene. Five amino acids corresponding to residues 236–240 in loop C-III and three amino acids (331–333) in the C-terminus are not yet defined.

The third structure of rhodopsin (1L9H) is changed to a much lesser extent when compared with 1HZX. No additional amino acids were added. The root-mean-square distance between C α atoms is 0.36 Å. These structures differ mainly in the cytoplasmic region: (1) in the C-terminal area (Figure 1B, denoted by an ellipsoid and the number III), the maximal displacement is by 1.7 Å for the last amino acid residue Ala³⁴⁸; (2) in the same region, Asp³³⁰ is displaced by 1.3 Å (the last amino acid before a second break); (3) in the loop between helices III and IV (this region is denoted as I), the maximal movement for two residues Ser¹⁴⁴ and

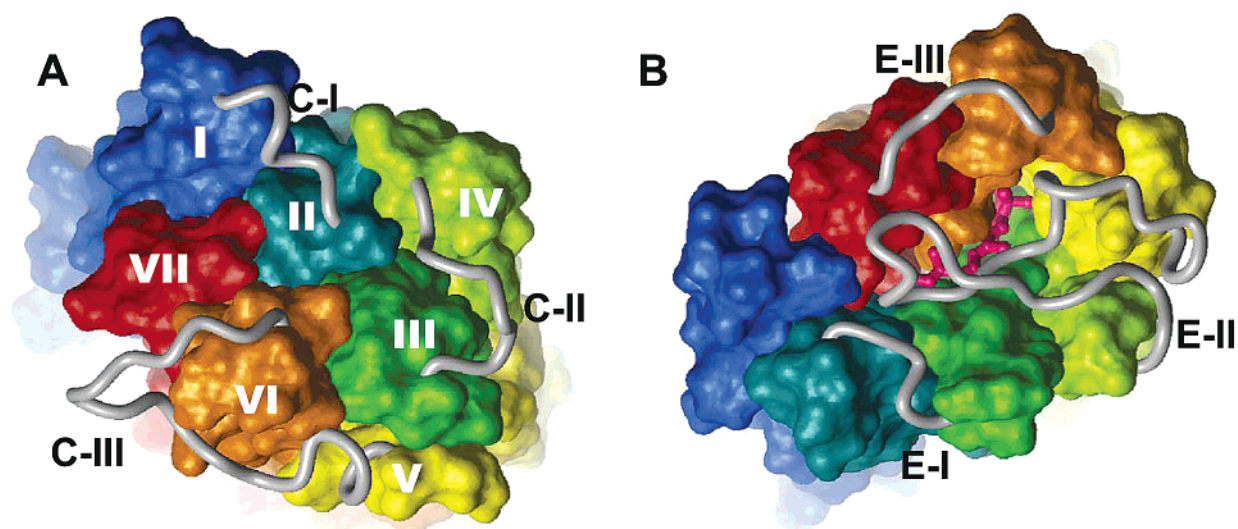


FIGURE 2: Space filling representation of the crystal structure of bovine rhodopsin showing the organization of helices viewed from cytoplasmic (A) and extracellular sides (B). Loops are shown as silver lines, and the chromophore is a stick model in magenta. A break in loop C-III was filled with suitable amino acids and optimized. Figure drawn using MolMol (61).

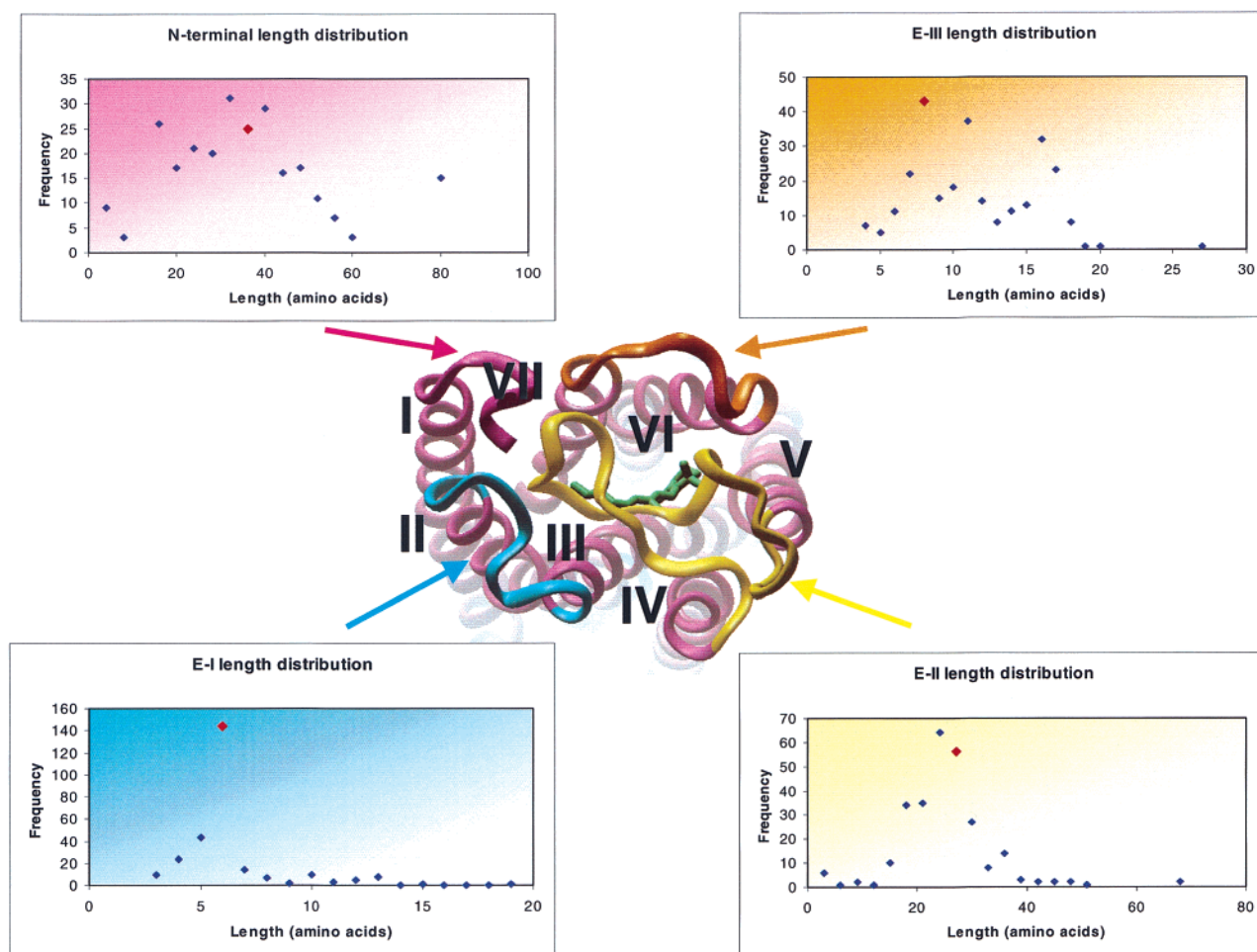


FIGURE 3: Sequence length analysis of the extracellular domains of family A GPCRs. The horizontal axis shows the number of amino acids in each domain. The vertical axis shows the number of receptors, 270 receptors in total. The loop size for rhodopsin is shown in all the graphs as a red diamond.

Asn¹⁴⁵ is, for both, by 0.9 Å; and (4) in extracellular loops, small changes in the N-terminal domain are found around Lys¹⁶, which is located farthest from the center at a distance

of 0.5 Å. For 11-*cis*-retinylidene, the maximal displacement is for the C₁₅ retinal atom by 1.2 Å (note that C₁₅ varied in all three structural models) and in the β-ionone ring for C₁₆

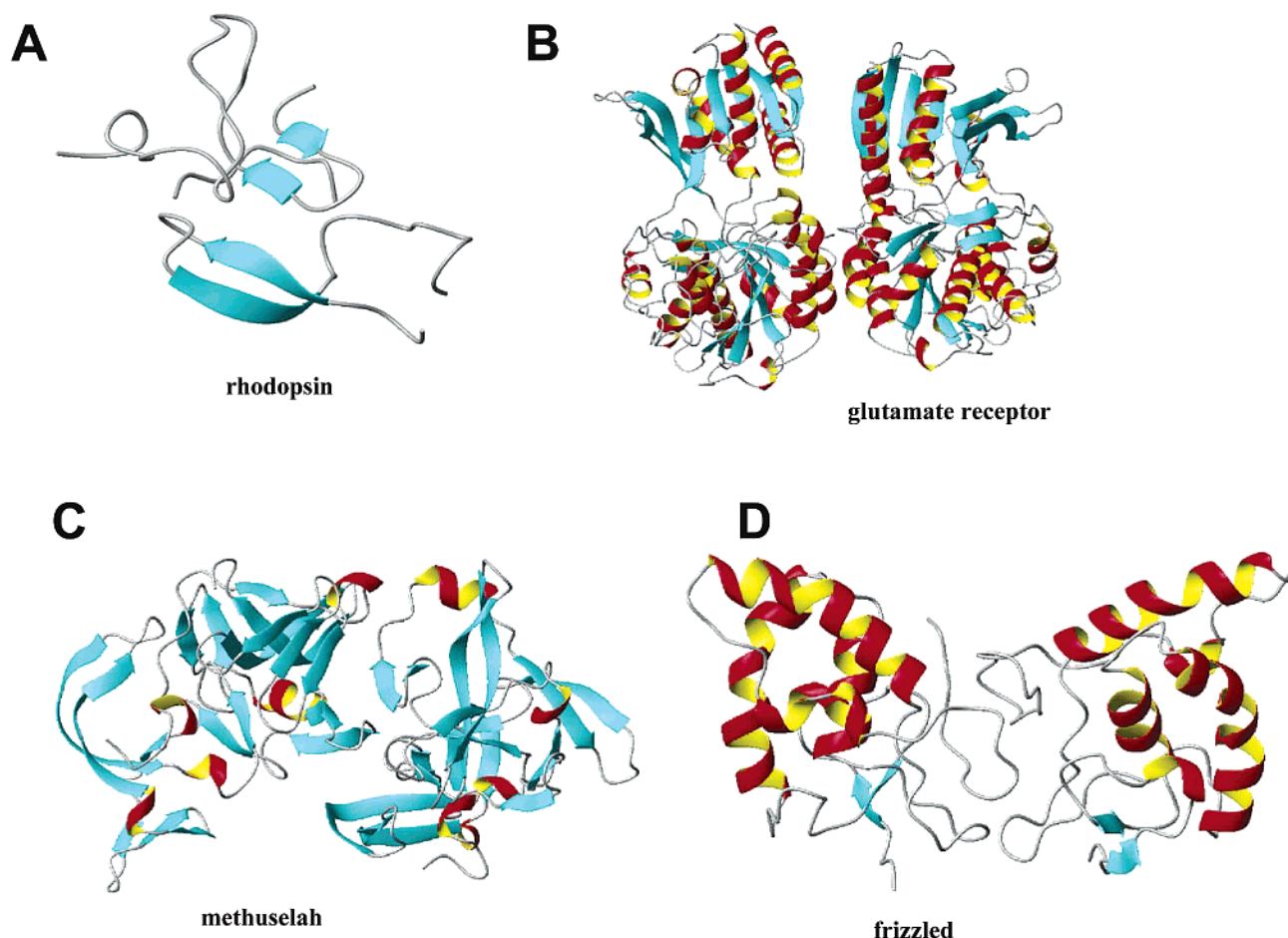


FIGURE 4: Extracellular domains of seven-transmembrane receptors. (A) Rhodopsin (1HZX). (B) Glutamate receptor (1EWK). (C) *Drosophila* Methuselah (1FJR). (D) Frizzled receptor (1IJY). Figure drawn using MolMol (61).

by 1.7 Å. In 1F88, atoms C ϵ –N ζ (Lys²⁹⁶)–(retinal)C₁₅–C₁₄ were in the same plane; however, in both 1HZX and 1L9H this feature is not retained. Although the position of Glu¹¹³ was unchanged (this region is denoted as II), the hydrogen bonds network around these residues is modified by the presence of water molecules in 1L9H (25).

All these models in large part can be superimposed, and no major differences are noted. Taking into account the 2.6–2.8 Å resolution, it is clear that these differences remain within the ambiguity of all three models. However, the higher precision of the positions of water molecules in 1L9H is very valuable (25). Water molecules may be critical during the activation process of rhodopsin (refs 14 and 25 and also reviewed in ref 24).

Comparison of the Extracellular Domains in GPCRs of Family A. GPCR drug design is hampered by the lack of receptor structures. So far, only one GPCR structure, that of bovine rhodopsin, is known (see above). The three-dimensional structure of rhodopsin and the hydrophobicity profile of this receptor support the proposed heptahelical transmembrane bundle model that is shared among all GPCRs (28). The seven helices traverse the membranes in a nonparallel manner toward the plane of the membrane, producing different arrangements of helices at the extracellular and cytoplasmic faces of the receptor as a consequence of the tilted helices in the membrane (Figure 2). This in turn is due to the tilting of helix III, which is tangential to the plane of the membrane. This orientation of helices,

depicted in Figure 2, is most likely conserved among all GPCRs.

Structural information can also be inferred from homology, which is most obvious in the predominant family of GPCRs, termed the rhodopsin-like family or family A. The length of GPCRs from family A varies between 290 and 951 amino acid residues, with the majority of receptors having a length around 310–470 residues (Figure 1, Supporting Information). This family is characterized by a set of conserved residues distributed among the seven helical domains. The conserved residues are in helix I (Gly and Asn), helix II (Leu and Asp), helix III (Cys and AspArgTyr), helix IV (Trp and Pro), helix V (Pro and Tyr), helix VI (Phe, Trp, and Pro), and helix VII (Asn, Pro, and Tyr of the NPXXY motif). These conserved amino acids facilitate the multiple alignments of the GPCR sequences. We found that it is important to use a multiple sequence alignment instead of a pairwise alignment to rhodopsin to avoid misalignments between helices, although an optimal pairwise alignment might produce fewer mismatches, but it can align a loop region of a template sequence with a helical region of a target sequence. A multiple sequence alignment, however, lets a consensus GPCR sequence clearly emerge.

The multiple sequence analyses were done on 270 members of family A, and of those, 153 are orphan GPCRs with unknown ligands. Using the three-dimensional structure of bovine rhodopsin, the N- and C-termini and the three extracellular (EI–III) and three cytoplasmic loops (C I–III)

HELIX –I

HELIX -I																				G		N		L		V	
A	9%	9%	12%	11%	12%	5%	15%	19%	8%	3%	13%	10%	4%	5%	10%	9%	10%	4%	9%	4%	0%	13%	6%	15%	2%	11%	
L	11%	12%	9%	7%	26%	27%	8%	14%	29%	21%	14%	30%	19%	6%	31%	29%	0%	37%	34%	7%	0%	21%	60%	6%	35%	20%	
V	10%	8%	6%	20%	16%	13%	10%	21%	26%	3%	9%	21%	29%	4%	29%	31%	2%	26%	20%	1%	0%	10%	10%	66%	13%	10%	
I	4%	6%	3%	14%	10%	16%	7%	11%	10%	5%	10%	12%	30%	13%	13%	9%	6%	15%	7%	1%	0%	11%	6%	3%	30%	7%	
G	7%	7%	5%	4%	3%	1%	6%	4%	1%	3%	14%	3%	0%	6%	2%	1%	56%	0%	4%	68%	0%	16%	0%	0%	0%	3%	
P	9%	3%	3%	1%	3%	2%	24%	6%	0%	2%	0%	1%	0%	0%	1%	1%	0%	0%	9%	0%	0%	0%	0%	1%	0%	0%	
M	1%	1%	2%	2%	2%	3%	3%	1%	2%	5%	1%	2%	4%	3%	0%	2%	0%	3%	1%	0%	0%	2%	4%	0%	10%	3%	
C	0%	2%	1%	1%	0%	0%	1%	4%	0%	2%	5%	1%	0%	12%	2%	2%	1%	1%	2%	1%	0%	3%	1%	0%	7%	4%	
S	4%	9%	6%	4%	2%	7%	7%	5%	3%	3%	20%	4%	3%	6%	2%	1%	11%	1%	2%	6%	0%	14%	1%	3%	1%	2%	
T	6%	4%	5%	3%	6%	6%	7%	9%	4%	3%	7%	7%	4%	4%	4%	6%	12%	4%	3%	3%	0%	7%	4%	4%	1%	3%	
Q	3%	7%	7%	4%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
N	1%	2%	4%	4%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	
H	1%	1%	4%	3%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	
K	1%	6%	12%	3%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	
R	2%	4%	9%	3%	0%	1%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	
E	5%	7%	2%	1%	0%	2%	0%	1%	0%	1%	0%	0%	3%	0%	0%	0%	0%	0%	0%	7%	0%	0%	0%	0%	0%	0%	
D	3%	3%	3%	1%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
F	12%	3%	3%	3%	14%	4%	3%	3%	12%	7%	2%	9%	3%	38%	5%	9%	0%	9%	9%	1%	0%	2%	6%	1%	1%	7%	
Y	4%	3%	3%	6%	3%	4%	2%	1%	1%	40%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	9%	
W	7%	1%	0%	4%	1%	3%	0%	1%	1%	0%	4%	0%	0%	1%	0%	1%	0%	0%	0%	1%	0%	0%	0%	0%	0%	19%	

HELIX –II

HELIX -II					F					L A					A D L L																			
A	10%	7%	5%	2%	0%	1%	9%	0%	0%	74%	9%	58%	0%	3%	6%	7%	17%	10%	7%	3%	4%	3%	3%	19%	17%	8%	11%	22%	6%	4%				
L	5%	4%	3%	16%	20%	37%	31%	0%	91%	1%	28%	0%	0%	60%	60%	17%	16%	40%	18%	39%	10%	16%	20%	4%	18%	10%	9%	15%	26%	9%				
V	26%	6%	2%	17%	0%	11%	18%	0%	1%	4%	30%	2%	0%	4%	5%	24%	11%	17%	6%	14%	9%	6%	7%	13%	15%	13%	7%	6%	12%	3%				
I	2%	4%	2%	22%	3%	34%	5%	0%	3%	0%	14%	0%	0%	6%	7%	6%	5%	10%	7%	3%	5%	5%	7%	14%	13%	4%	7%	8%	7%	2%				
G	1%	2%	1%	0%	0%	0%	9%	2%	0%	1%	1%	3%	1%	1%	2%	5%	17%	1%	5%	3%	3%	0%	0%	3%	1%	1%	1%	4%	3%	13%				
P	22%	6%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	41%	37%	4%	0%	0%	1%	1%	0%	3%	1%				
M	5%	6%	1%	3%	1%	8%	3%	0%	4%	0%	1%	0%	0%	4%	3%	7%	2%	1%	1%	2%	16%	0%	3%	5%	3%	3%	1%	2%	6%	5%				
C	1%	0%	0%	2%	3%	0%	6%	1%	0%	4%	5%	1%	0%	1%	4%	2%	6%	5%	4%	12%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%				
S	9%	9%	6%	1%	1%	1%	2%	29%	0%	3%	2%	26%	0%	1%	3%	1%	11%	4%	12%	4%	3%	1%	2%	11%	4%	9%	8%	7%	4%	6%				
T	8%	38%	8%	6%	0%	1%	4%	1%	1%	4%	3%	9%	0%	2%	1%	2%	10%	5%	23%	3%	3%	4%	0%	11%	7%	8%	6%	5%	3%	9%				
Q	0%	1%	1%	1%	0%	0%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	2%	1%	0%	3%	3%	0%	3%	2%	2%	3%	6%				
N	1%	5%	40%	1%	0%	0%	0%	51%	0%	0%	0%	0%	3%	0%	0%	0%	1%	0%	1%	4%	0%	2%	3%	4%	0%	4%	5%	2%	7%	4%				
H	0%	0%	5%	1%	1%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	1%	1%	0%	0%	0%	0%	0%	1%	0%	4%	1%	4%	2%	3%	7%				
K	1%	0%	2%	1%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	4%	1%	0%	1%	3%	0%	3%	3%	3%				
R	4%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	7%	0%	0%	1%	2%	1%	5%	9%	9%				
E	3%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	2%	0%	1%	0%	0%	0%	0%	0%	2%	3%	0%	3%	1%	7%	0%	4%	4%				
D	0%	0%	10%	0%	0%	0%	0%	1%	0%	0%	0%	0%	94%	0%	0%	0%	0%	0%	0%	1%	0%	0%	1%	2%	3%	3%	6%	3%	2%	4%				
F	0%	8%	3%	9%	39%	4%	9%	0%	0%	1%	7%	0%	0%	15%	7%	19%	1%	7%	11%	3%	3%	26%	16%	5%	9%	5%	3%	3%	3%	3%				
Y	0%	1%	8%	18%	28%	2%	0%	0%	0%	0%	1%	0%	0%	0%	0%	6%	0%	0%	4%	3%	1%	1%	1%	2%	5%	21%	19%	9%	1%	4%				
W	0%	0%	0%	1%	3%	0%	0%	1%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	15%	0%	0%	0%	3%	1%	1%	1%				

FIGURE 5: Compositions of helices I and II in family A GPCRs (270 receptors). (A) Helix I. The first and last residue numbers (using rhodopsin residue numbers) are 35–60. The color indicates conservation in each position in the helix: green (0–20%), yellow (20–40%), orange (40–60%), red (60–80%), and brown (80–100%). (B) Helix II. Composition of helix II in family A GPCRs. The first and last residue numbers (using rhodopsin residue numbers) are 71–100. The color coding is an indication of residue conservation in each position in the helix: green (0–20%), yellow (20–40%), orange (40–60%), red (60–80%), and brown (80–100%).

were extracted for all 270 receptors. The N-terminal region and the extracellular loops were first used in a length analysis (Figure 3). As it is shown, the amino acid length of the N-terminal region is highly variable, being from as little as four to as many as over 50 amino acid residues in length. Among the three extracellular loops, loop E-I has the most consistent loop size; 144 GPCRs have the same number of amino acids in E-I as rhodopsin (six residues). This loop connects helix II and III, and its short length is perhaps crucial for filling the least amount of space to accommodate other extracellular loops forming the extracellular domain capable of binding ligand. E-I ranges from only three amino acids to as many as 18 amino acid residues. The other two extracellular loops (E-II and E-III) have more variable loop sizes. Loop E-III is located entirely on the surface of the extracellular domain. Only one hydrophobic residue in this loop, Phe²⁸³, intercalates its ring inside rhodopsin. This loop also interacts via hydrogen bonds with the carbohydrate moiety connected to Asn² of rhodopsin. If the positioning of the sugars' attachment is important for recognition of carbohydrate transferases during the maturation of rhodopsin, this loop cannot expand considerably. Therefore, it is possible that few (1–4) additional amino acids in the loop are used

in other GPCRs for increasing its hydrogen bond network with carbohydrates. This interaction would put a limit on the size of this loop, as observed. Note, however, that the position of carbohydrate attachment is not conserved in other members of family A. In general, it is expected that the extracellular domains will be the most variable among GPCRs because this part of the receptor is involved in ligand recognition. The compact structure of the extracellular domain of rhodopsin is built based on four β -strands forming a β -sheet structure and has no α helices (Figure 4A). Lack of structural information on the organization of the extracellular domain does not allow direct comparison between members of family A. However, the structures of the extracellular domain of other seven-transmembrane receptors have been elucidated, and they support the idea of divergences in these regions. The extracellular domain of a glutamate receptor (512 amino acids in monomer) is formed from seven α -helices and two big β -sheets (six and eight β -strands, respectively) (Figure 4B) (29). The receptor is composed of two domains, each of them consisting of separate β -sheets and with α -helices connecting subsequent strands of β -sheets. The ligand binding site is located between these two domains. There are several single turns of 3₁₀

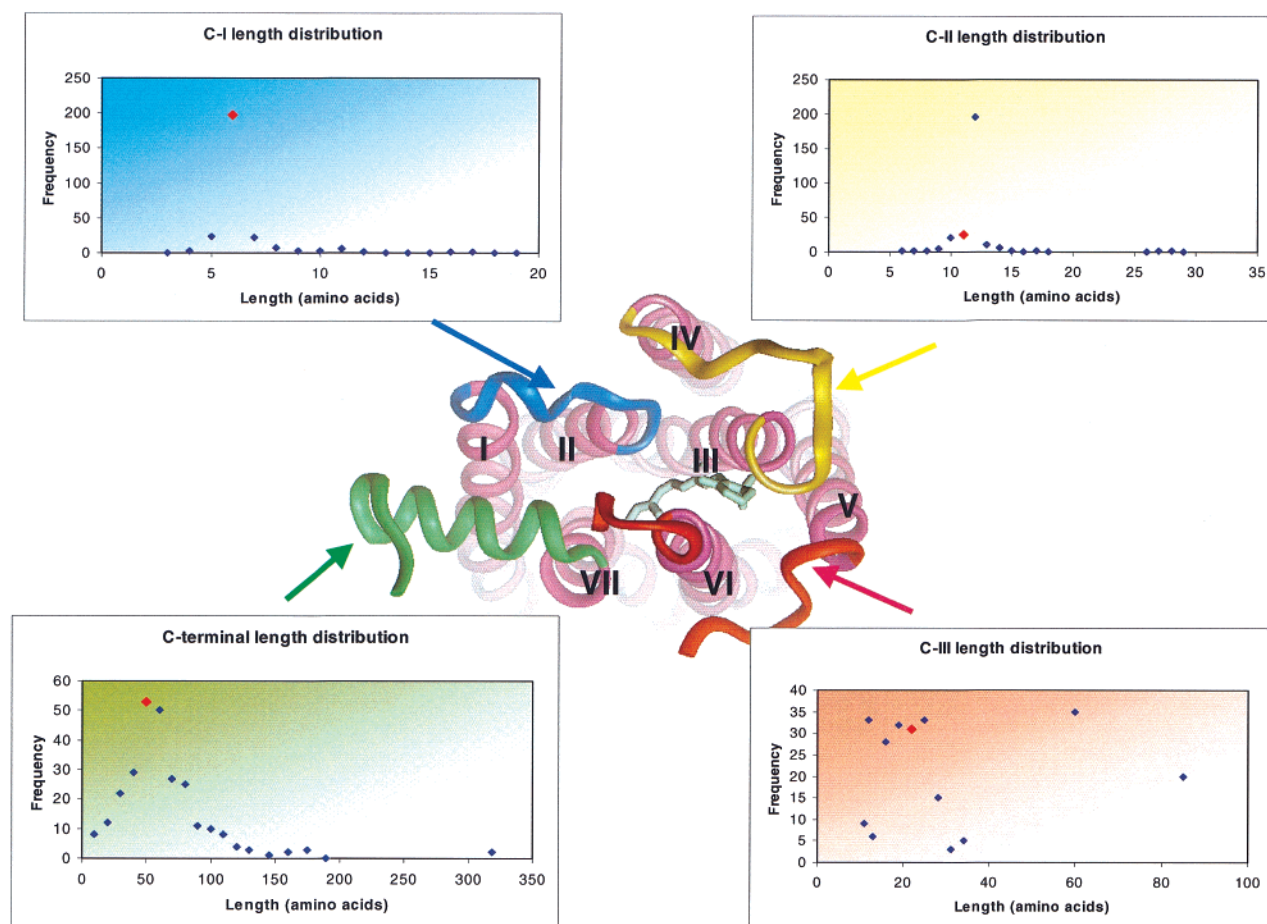


FIGURE 6: Sequence length analysis of the intracellular domains of family A GPCRs. The horizontal axis shows the number of amino acids in each domain. The vertical axis shows the number of receptors, 270 receptors in total. The loop size for rhodopsin is shown in all the graphs by a red diamond.

helices located exclusively in the interface region between two monomers. No disulfide bridges are formed between monomers. The extracellular domain of another putative GPCR of the secretin receptor family from *Drosophila*, Methuselah, also forms a dimer (188 amino acids in the monomer). This domain is composed of seven β -sheets, one short α -helix, and three one-turn 3_{10} helices (Figure 4C) (30). In addition, it was recently determined that the extracellular domain of a member of the Frizzled family of seven-pass transmembrane proteins serving as a receptor in the Wnt signaling pathway (31) displays yet another type of folding. This domain also forms a dimer when in the crystal form (126 amino acids in each monomer). The structure consists of only one small β -sheet composed of two β -strands, three α -helices, and three one-turn 3_{10} helices (Figure 4D). These examples suggest that the dimerization of some seven-transmembrane receptors, but not rhodopsin, is accomplished through distinct extracellular domains that are partly involved in the ligand binding and stabilization of the ligand within the hydrophilic domain of the transmembrane receptor.

Further analyses of the frequencies of individual amino acids in particular positions reveal that helix I has an invariant Asn and that Gly, Leu, and Val are highly conserved (68, 60, and 66%, respectively) (Figure 5). Surprisingly, in several cases, basic residues are found at the beginning of helix I (frequency of 12 and 25% in the second and third positions) and at the end of helix II (frequency of 11 and 19% in the penultimate and the last positions), possibly stabilizing the

interaction of this helix with phospholipids. A lower number of aromatic residues is noticed toward the cytoplasmic region, and frequently, a Pro residue is at the beginning of the helix II, likely without perturbing the helical structure (32). Helix II has six conserved aromatic/hydrophobic residues toward the cytoplasmic surface and invariant Glu, while residues toward the extracellular domain are more divergent. In the longest helix III, two structural features are most noticeable. A Cys residue close to the extracellular domain, which in rhodopsin forms a disulfide bond with the Cys residue of loop E-II, occurs in 90% of the evaluated GPCR cases. Additionally, the residues of the DRY region consist of Asp/Glu, Arg, and aromatic residues present in frequencies of 86, 96, and 83%, respectively (Figure 2, Supporting Information). The DRY region is packed in the conserved hydrophobic pocket. A general observation, with few exceptions, is that the residue immediately following the conserved Cys residue in helix III corresponds to the ligand type for the receptor. If the residue is basic, the ligand for the receptor is most likely a peptide (34% Lys, 19% Arg). If it is acidic (Asp, Glu), it is a biogenic amine. Helix IV contains invariant aromatic residues, mostly Trp (Figure 3, Supporting Information), which in rhodopsin function in the stabilization of the interaction between helix III and II. The analyses of other helices are summarized in subsequent figures in the Supporting Information (Figures 4–6, Supporting Information). Helix V contains two aromatic residues, Phe and Try, that are common in GPCRs (70 and 77%, respectively). Helix

VI has two conserved aromatic residues, Phe and Trp (FxxxW), present in frequencies of 89 and 90%, respectively. In the X-ray structure of rhodopsin, Trp²⁶⁵ is in close proximity of the retinal chromophore, involved in the activation process and release of the chromophore (33), also likely creating a limit as to how far a ligand can penetrate the helical bundle. There are a few exceptions to this highly conserved residue (90% frequency are aromatic residues). In the prostaglandin family of receptors there are four members (PD2R, PE22, PE24, and PI2R) that have a Ser in place of this aromatic residue, yet the other four subtypes (PE21, PE23, PF2R, and TXA2) have a Trp. Interestingly, the members of the prostanoid family with a Ser couple to cause an increase in intracellular cAMP levels, and the others go through a different signaling pathway. It is also interesting to note that in the large family of chemokine receptors there are six members that have a Gln in this position (CCR6, CKR6, CKR7, CKR9, CKR10, and CKR11), while the other family members have Trp. The last exception to this residue conservation is in the large hormone receptor family where human follicle-stimulating hormone (FSHR), lutropin/choriogonadotropin receptor (LHR), and thyrotropin receptor (TSHR) all have a Met in this position. Within helix VII, the most evident conservation is that of Asn and Pro residues occurring in the NPXXY region, which is crucial for receptor activation. This region is highly conserved for Asn (75% Asn, 21% Asp), Pro (96%), and Tyr (92% Tyr, 3% Phe), while X are mostly hydrophobic Leu, Val, and Ile. Similarly to extracellular loops, the loops in the cytoplasmic domain can vary in size, the most conserved loops being C-I and C-II (Figure 6). There are 198 GPCRs that have the same number of amino acids in C-I as rhodopsin (six residues), and the remaining members of GPCR family A have either five or seven amino acids. For C-II, over 150 of the receptors have a loop size of 10–12 amino acids (11 residues in rhodopsin). C-III and the C-terminus have the most variations in amino acid lengths among all the GPCRs. A common activation mechanism that involves the conserved residues in the helical domains and the conserved loop regions is proposed for family A GPCRs.

The amino acid residues of the short helix (VIII) in the C-termini were analyzed (Figure 7, Supporting Information). This analysis was done on only 260 receptors; in the others the C-termini is too short. It is interesting to note that there is a highly conserved aromatic residue (Phe³¹³, 73% similarity). The side chain of this residue is placed between helix I and VII, adjacent to helix II in the helical bundle. This residue stacks against the highly conserved Tyr³⁰⁶ in helix VII and forms a very tight hydrophobic pocket. It is also noted that another residue close to Phe³¹³, residue Val⁶¹ from helix I, is never an aromatic residue (this pocket is too small to accommodate an aromatic residue). The residue immediately following this aromatic residue is a highly conserved basic residue (54% Arg, 17% Lys), facing toward the outside of the helical bundle.

Additional insight into the organization of external domains has been obtained from split receptors, where extracellular or cytoplasmic loops are cut, and the receptors are expressed in two or three fragments. Receptors cut within C-III were functional in ligand binding and coupled to G-proteins (reviewed in ref 34). Also, splitting rhodopsin in C-II and C-III led to production of three-fragment rhodopsins

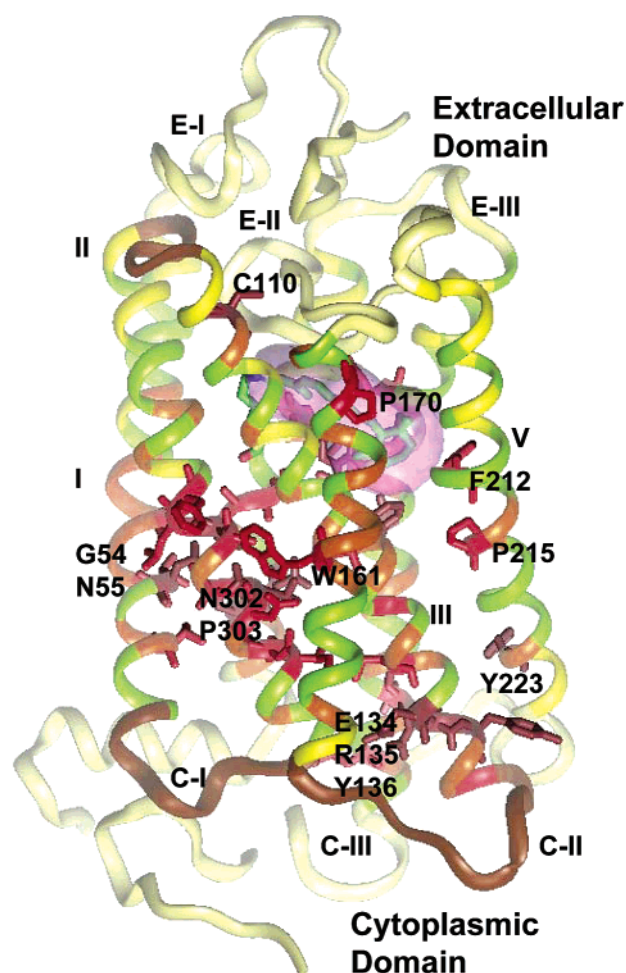


FIGURE 7: Crystal structure of bovine rhodopsin. The conserved residues are color coded (no conservation, pale yellow; 10–20% conservation, light yellow; 20–40%, light green; 40–60%, orange; 60–80%, red; and 80–100%, brown) based on the percent similarity among family A GPCRs (270 receptors). The conserved extracellular loop (E-I) and cytoplasmic loops (C-I and C-II) are colored brown. Residues that are colored red are as follows: helix I—54, 57, 58; helix II—76, 80, 82, 84, 85; helix III—124, 128, 131; helix IV—170; helix V—212, 215; helix VI—249, 252, 264, 268; helix VII—294, 298, 299, 302; and helix VIII—313, 314. Residues that are colored brown are as follows: helix I—55; helix II—79, 83; helix III—110, 134, 135, 136; helix IV—161; helix V—223; helix VI—261, 265, 267; and helix VII—303, 306.

that showed functional and spectral properties similar to the wild type (35, 36). Therefore, loops C-II and C-III have only a small stabilizing effect on these receptors. Helices I-III and V-VII with helix IV connected to either of these two fragments form active rhodopsin, suggesting a tight packing or correct induced folding within these bundles of helices I-III and V-VII. Whereas packing of helical bundle I-III might be aided by the short length of loops E-I, C-I, and C-II, helices V-VII form a structural domain independent of the length of the connecting loops. Nonfunctional receptors were obtained using fragments where helix I, or helices I and II, and the rest of the helical bundle were separated (34). For some receptors, helices I and II are essential for ligand binding; however, five-transmembrane domains appear to be sufficient for the function of chemokine receptors (37). Truncated CCR5 and CXCR4 functioned normally in mediating chemokine-stimulated chemotaxis, Ca²⁺ influx, and activation of pertussis toxin-sensitive G-proteins.

The C-terminus is not strictly conserved, but the most frequent length of this region is ~50 amino acid residues. This region, at least in case of rhodopsin, contains a recognition sequence for the vectorial transport of receptor from the site of synthesis to its destination (38–40). The C-terminus is also the site of functional desensitization upon agonist stimulation (41–47). The specificity of the binding sites of GPCRs for the receptor ligands were recently extensively reviewed (3, 26, 48).

Conservation of Structure and Mechanism of Activation. It is apparent that even without detailed homology modeling, a simple multiple sequence alignment reveals much information about the receptor family. Residue-per-residue alignment highlights the helical domains that are more conserved (Figure 7). Thus, the predicted helical domains in most parts correspond well to the known structure of rhodopsin. By examining the helical domains more closely, a distinct mode of alternating alignment quality appears. In odd helices, homology is higher at the C-terminal side, while it is the opposite for even helices. This pattern corresponds to higher conservation in the lower part of the GPCRs near the cytoplasmic domain, providing evidence for a conserved mechanism of activation and signal transduction. Since GPCRs accept a variety of ligands, the ligand-binding site is the least conserved. These observations support the hypothesis that the receptor activation process is highly conserved for the majority of family A members. A current mechanism of rhodopsin activation, as a model for GPCRs of family A, has been recently reviewed (14, 17, 24).

Conservation of Structure and the Molecular Basis of Oligomerization of GPCRs. Conservation may have an important function in the organization of GPCRs into oligomeric forms. Oligomerization of GPCRs has been deduced from various biophysical studies and even implies a larger diversification of GPCR signaling (49–59). Recently, by the direct method of atomic force microscopy, we showed that rhodopsin forms rows of dimers under native conditions in rod outer segments of photoreceptor cells (60). The importance of GPCR oligomerization may involve cross-talk between different signaling pathways, cooperativity of the signal amplification, vectorial transport, and inactivation. As the cytoplasmic surface of GPCRs is larger than the extracellular domain when projected toward the plane of the membranes, the cytoplasmic ends of the helices would be more critical in the formation of contacts during the assembly of the high-order oligomerized structures.

CONCLUSIONS

Similarity among GPCRs, particularly within family A, is represented by a few structural microdomains containing key residues involved in the interhelical interactions. The extracellular domain involved in ligand binding appears to form a receptor-specific binding site, while the cytoplasmic side and the ends of the transmembrane helices toward the cytoplasm are significantly more conserved. It is reasonable to speculate that the overall fold of these receptors is highly conserved. The changes upon binding of a ligand in a receptor-specific domain of GPCRs evoke a similar cascade of activation events that culminate in the recruitment of G-protein(s) and catalytic nucleotide exchange within these proteins. Perhaps such convergence of the structural similarity is essential for allowing hundreds of GPCRs to activate

dozens of G-proteins, to become phosphorylated by seven G-protein receptor kinases, and to be capped by three arrestins.

ACKNOWLEDGMENT

We are grateful to Drs. Volker Gerke, Oliver Ernst, and Urlik Gether for comments on the manuscript and Yunie Kim and Matthew Batten for help during the manuscript preparation.

SUPPORTING INFORMATION AVAILABLE

Tables of amino acid frequencies among GPCR helices II–VII. This material is available free of charge via the Internet at <http://pubs.acs.org>.

NOTE ADDED AFTER PRINT PUBLICATION

An accent was left off of author Gil Benkö's last name in the version published on the Web 3/11/03 and in the March 18, 2003, issue (Vol. 42, No. 10, pp 2759–2767). The correct electronic version of the paper was published 03/12/03, and an Addition and Correction appears in the April 15, 2003 issue (Vol. 42, No. 14).

REFERENCES

1. Hom, F., Weare, J., Beukers, M. W., Horsch, S., Bairoch, A., Chen, W., Edvardsen, O., Campagne, F., and Vriend, G. (1998) *Nucleic Acids Res.* 26, 275–279.
2. Sautel, M., and Milligan, G. (2000) *Curr. Med. Chem.* 7, 889–896.
3. Ballesteros, J. A., Shi, L., and Javitch, J. A. (2001) *Mol. Pharmacol.* 60, 1–19.
4. Gether, U. (2000) *Endocr. Rev.* 21, 90–113.
5. Ballesteros, J., and Palczewski, K. (2001) *Curr. Opin. Drug Discovery Dev.* 4, 561–574.
6. Schoneberg, T., Schulz, A., and Gudermann, T. (2002) *Rev. Physiol. Biochem. Pharmacol.* 144, 143–227.
7. Seifert, R., and Wenzel-Seifert, K. (2002) *Naunyn Schmiedeberg's Arch. Pharmacol.* 366, 381–416.
8. Shenker, A. (1995) *Baillieres Clin. Endocrinol. Metab.* 9, 427–451.
9. Rattner, A., Sun, H., and Nathans, J. (1999) *Annu. Rev. Genet.* 33, 89–131.
10. Lander, E. S., Linton, L. M., and Chen, Y. J. (2001) *Nature* 409, 860–921.
11. Venter, J. C., Adams, M. D., and Zhu, X. (2001) *Science* 291, 1304–1351.
12. Joost, P., and Methner, A. (2002) *Genome Biol.* 3, 63.61–63.16.
13. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) *Science* 289, 739–745.
14. Okada, T., Ernst, O. P., Palczewski, K., and Hofmann, K. P. (2001) *Trends Biochem. Sci.* 26, 318–324.
15. Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K., and Stenkamp, R. E. (2001) *Biochemistry* 40, 7761–7772.
16. Stenkamp, R. E., Filipek, S., Driessen, C. A., Teller, D. C., and Palczewski, K. (2002) *Biochim. Biophys. Acta* 1565, 168–182.
17. Menon, S. T., Han, M., and Sakmar, T. P. (2001) *Physiol. Rev.* 81, 1659–1688.
18. Okada, T., and Palczewski, K. (2001) *Curr. Opin. Struct. Biol.* 11, 420–426.
19. Shacham, S., Topf, M., Avisar, N., Glaser, F., Marantz, Y., Bar-Haim, S., Noiman, S., Naor, Z., and Becker, O. M. (2001) *Med. Res. Rev.* 21, 472–483.
20. Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard, W. A., III (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99, 12622–12627.
21. Greasley, P. J., Fanelli, F., Rossier, O., Abuin, L., and Cotecchia, S. (2002) *Mol. Pharmacol.* 61, 1025–1032.
22. Okada, T., Takeda, K., and Kouyama, T. (1998) *Photochem. Photobiol.* 67, 495–499.

23. Okada, T., Le Trong, I., Fox, B. A., Behnke, C. A., Stenkamp, R. E., and Palczewski, K. (2000) *J. Struct. Biol.* 130, 73–80.
24. Filipek, S., Teller, D. C., Stenkamp, R. E., and Palczewski, K. (2003) *Annu. Rev. Physiol.*, in press.
25. Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M., and Shichida, Y. (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99, 5982–5987.
26. Klabunde, T., and Hessler, G. (2002) *Chembiochemistry* 3, 928–944.
27. Filipek, S., Teller, D. C., Palczewski, K., and Stenkamp, R. E. (2003) *Annu. Rev. Biophys. Biomol. Struct.*, in press.
28. Konvicka, K., Guarnieri, F., Ballesteros, J. A., and Weinstein, H. (1998) *Biophys. J.* 75, 601–611.
29. Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., and Morikawa, K. (2000) *Nature* 407, 971–977.
30. West, A. P., Jr., Llamas, L. L., Snow, P. M., Benzer, S., and Bjorkman, P. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* 98, 3744–3749.
31. Dann, C. E., Hsieh, J. C., Rattner, A., Sharma, D., Nathans, J., and Leahy, D. J. (2001) *Nature* 412, 86–90.
32. Visiers, I., Ballesteros, J. A., and Weinstein, H. (2002) *Methods Enzymol.* 343, 329–371.
33. Heck, M., Schadel, S. A., Maretzki, D., Bartl, F. J., Ritter, E., Palczewski, K., and Hofmann, K. P. (2002) *J. Biol. Chem.*
34. Gudermann, T., Schoneberg, T., and Schultz, G. (1997) *Annu. Rev. Neurosci.* 20, 399–427.
35. Ridge, K. D., Ngo, T., Lee, S. S., and Abdulaev, N. G. (1999) *J. Biol. Chem.* 274, 21437–21442.
36. Ridge, K. D., Lee, S. S., and Yao, L. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 3204–3208.
37. Ling, K., Wang, P., Zhao, J., Wu, Y. L., Cheng, Z. J., Wu, G. X., Hu, W., Ma, L., and Pei, G. (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96, 7922–7927.
38. Moritz, O. L., Tam, B. M., Papermaster, D. S., and Nakayama, T. (2001) *J. Biol. Chem.* 276, 28242–28251.
39. Sung, C. H., Makino, C., Baylor, D., and Nathans, J. (1994) *J. Neurosci.* 14, 5818–5833.
40. Tam, B. M., Moritz, O. L., Hurd, L. B., and Papermaster, D. S. (2000) *J. Cell Biol.* 151, 1369–1380.
41. Cideciyan, A. V., Zhao, X., Nielsen, L., Khani, S. C., Jacobson, S. G., and Palczewski, K. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 328–333.
42. Palczewski, K., and Benovic, J. L. (1991) *Trends Biochem. Sci.* 16, 387–391.
43. Pitcher, J. A., Freedman, N. J., and Lefkowitz, R. J. (1998) *Annu. Rev. Biochem.* 67, 653–692.
44. Benovic, J. L., Strasser, R. H., Caron, M. G., and Lefkowitz, R. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 2797–2801.
45. Benovic, J. L., Mayor, F., Jr., Staniszewski, C., Lefkowitz, R. J., and Caron, M. G. (1987) *J. Biol. Chem.* 262, 9026–9032.
46. Benovic, J. L., Regan, J. W., Matsui, H., Mayor, F., Jr., Cotecchia, S., and Leeb Lundberg, L. M. (1987) *J. Biol. Chem.* 262, 17251–17253.
47. Benovic, J. L., Mayor, F., Jr., Somers, R. L., Caron, M. G., and Lefkowitz, R. J. (1986) *Nature* 321, 869–872.
48. Shi, L., and Javitch, J. A. (2002) *Annu. Rev. Pharmacol. Toxicol.* 42, 437–467.
49. Bockaert, J., and Pin, J. P. (1999) *EMBO J.* 18, 1723–1729.
50. Dean, M. K., Higgs, C., Smith, R. E., Bywater, R. P., Snell, C. R., Scott, P. D., Upton, G. J., Howe, T. J., and Reynolds, C. A. (2001) *J. Med. Chem.* 44, 4595–4614.
51. Overton, M. C., and Blumer, K. J. (2000) *Curr. Biol.* 10, 341–344.
52. Rios, C. D., Jordan, B. A., Gomes, I., and Devi, L. A. (2001) *Pharmacol. Ther.* 92, 71–87.
53. Gouldson, P. R., Higgs, C., Smith, R. E., Dean, M. K., Gkoutos, G. V., and Reynolds, C. A. (2000) *Neuropsychopharmacology* 23, S60–77.
54. Brady, A. E., and Limbird, L. E. (2002) *Cell Signal* 14, 297–309.
55. Angers, S., Salahpour, A., Joly, E., Hilairret, S., Chelsky, D., Dennis, M., and Bouvier, M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97, 3684–3689.
56. Angers, S., Salahpour, A., and Bouvier, M. (2001) *Life Sci.* 68, 2243–2250.
57. Bouvier, M. (2001) *Nat. Rev. Neurosci.* 2, 274–286.
58. George, S. R., O'Dowd, B. F., and Lee, S. P. (2002) *Nat. Rev. Drug Discovery* 1, 808–820.
59. Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. (2002) *Nat. Rev. Mol. Cell Biol.* 3, 639–650.
60. Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A., and Palczewski, K. (2002) *Nature* 421, 127–128.
61. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics* 14, 51–55.

BI027224+