ORIGINAL PAPER

# Modeling activated states of GPCRs: the rhodopsin template

**Masha Y. Niv · Lucy Skrabanek · Marta Filizola ·
Harel Weinstein**

**Abstract** Activation of G Protein-Coupled Receptors (GPCRs) is an allosteric mechanism triggered by ligand binding and resulting in conformational changes transduced by the transmembrane domain. Models of the activated forms of GPCRs have become increasingly necessary for the development of a clear understanding of signal propagation into the cell. Experimental evidence points to a multiplicity of conformations related to the activation of the receptor, rendered important physiologically by the suggestion that different conformations may be responsible for coupling to different signaling pathways. In contrast to the inactive state of rhodopsin (RHO) for which several high quality X-ray structures are available, the structure-related information for the active states of rhodopsin and all other GPCRs is indirect. We have collected and stored such information in a repository we maintain for activation-specific structural data available for rhodopsin-like GPCRs, http://www.physiology.med.cornell.edu/GPCRactivation/gpcrindex.html. Using these data as structural constraints, we have applied Simulated Annealing Molecular Dynamics to construct a number of different active state models of RHO starting from the known inactive structure. The common features of the models indicate that TM3 and TM5 play an important role in activation, in addition to the well-established rearrangement of TM6. Some of the structural changes observed in these models occur in regions that were not involved in the constraints, and have not been previously tested experimentally; they emerge as interesting candidates for further experimental exploration of the conformational space of activated GPCRs. We show that none of the normal modes calculated from the inactive structure has a dominant contribution along the path of conformational rearrangement from inactive to the active forms of RHO in the models. This result may differentiate rhodopsin from other GPCRs, and the reasons for this difference are discussed in the context of the structural properties and the physiological function of the protein.

M. Y. Niv · L. Skrabanek · M. Filizola · H. Weinstein (✉)
Department of Physiology and Biophysics, Weill Medical College of Cornell University, 1300 York Ave., New York, NY 10021, USA
e-mail: haw2002@med.cornell.edu

L. Skrabanek · H. Weinstein
HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University, 1300 York Ave., New York, NY 10021, USA

## Introduction

G protein-coupled receptors (GPCRs) are members of the largest known family of cell-surface receptors, which transduce a large variety of extracellular stimuli such as hormones, neurotransmitters, odorant molecules and light to the intracellular signaling pathways. These proteins constitute the most prominent family of validated drug targets within biomedical research, since 40–60% of approved drugs elicit their therapeutic effects by selectively addressing members of the GPCR family [1].

To date the only GPCR for which a three-dimensional (3D) crystallographic structure has been solved is the prototypic rhodopsin in its inactive state [2]. Since the structure of an active GPCR is essential for a mechanistically correct understanding of agonist binding and downstream signaling by these proteins, several labs have attempted to construct 3D models of activated forms of GPCRs by computational means. For example, to provide insights into the initial photodynamics process in rhodopsin, Molecular Dynamics simulations were performed in the presence of isomerized retinal [3–5]. However, the timescales on which such calculations were performed (tens of nanoseconds) are not long enough to reach the MII active state of rhodopsin from the inactive structure. In fact, experimentally the MII is achieved on a millisecond timescale via several intermediate states and proton transfer processes [6, 7]. Accelerated weighted masses Molecular Dynamics [8] was applied to obtain an activated rhodopsin model which then served as a homology template for 5HT-2A [9].

Drug-design oriented approaches towards modeling of activated states of GPCRs utilized annealing in the presence of known agonists of rhodopsin-based homology models [9, 10] and of ab-initio models [11, 12] of GPCRs. Such models are very successful in virtual screening of ligand binding, achieving significant enrichment factors [10, 12], but are not geared towards modeling regions distal to the binding site. Yet, conformational changes occurring upon activation in regions distal to the ligand-binding site, were demonstrated for rhodopsin and other GPCRs using various experimental techniques, including Site-Directed Spin Labeling studies [13–15], disulfide bond cross-linking [16], and engineered zinc binding sites [51]. To address computationally such distal changes, several groups have incorporated distance constraints available from experiments into 3D models of rhodopsin [17–20], $\beta$2AR [19, 21], 5-HT2A [22, 23] and $\mu$-opioid receptors [24] to obtain active state models of these receptors.

It is becoming clearer that pharmacological diversity is measurable when the receptors adopt more than one active conformation [25–28]. The current view is that the various active states elicited by distinct agonists have physically distinguishable conformations that enable different interactions with downstream proteins in the signaling cascade to produce different signaling patterns [29–33]. We reasoned that the variety of data from experimental explorations of structural changes in rhodopsin related to activation may contain information about the possible activated forms of GPCRs. Consequently, we undertook the construction

of models of the activated forms of rhodopsin by incorporating different compilations of structural information as constraints. These compilations were extracted from a database we have collected of currently known structural boundaries. The entire set of boundaries, as well as partial sets were implemented as constraints in separate simulated annealing-based simulation protocols to construct the six different activated models of rhodopsin presented here. We analyzed these modeling results in terms of compatibility between different constraints, since some of them may be present in a specific activated state but not another. In addition, the comparative analysis of these models presented below revealed common features such as decreases in specific pairwise distances that had not been previously examined experimentally. Such specific observations from the models are being presented here as experimentally testable predictions about the properties of the activated states of rhodopsin, and possibly other cognate GPCRs.

An even more formidable task than obtaining models of the active conformations of the GPCRs, is gaining insight into the dynamic pathways leading to them from the inactive structure. Several techniques were recently developed to tackle the great challenge of long-timescale conformational changes, as reviewed recently by Elber [34]. One of the very successful and computationally simplest approximations is the Normal Mode Analysis (NMA). In NMA, molecular motions are decomposed into vibrational modes. This approach extends the timescale accessible to theoretical work and is useful for studying collective motions and mechanical properties of biological systems, reviewed and discussed in [35–37]. In the Elastic Network Model (ENM) version of the NMA, a harmonic potential with a single force constant accounts for pairwise interactions between all $C_\alpha$ atoms that are within a cutoff distance [38]. The ENM was shown to capture key features of many biological systems (as compared to atomistic potential NMA [39, 40] and to experimental data such as crystallographic $\beta$-factors [38], and H/D exchange rates [41]). The ENM energy function is a minimum for any chosen configuration of the system, thus eliminating the need for minimization prior to normal mode analysis. This is an important advantage for low-resolution structures, such as the models of rhodopsin active conformations obtained here.

In the present work, we have performed ENM-based analysis of the inactive form of rhodopsin and evaluated the structural relation of the inactive normal modes to the transition vectors towards the active conformations of the models we constructed. It has

been shown that in some cases, transition from one state of a protein to another corresponds to the changes indicated by a few low frequency normal modes of the original structure [37, 39]. In the case of rhodopsin activation, several low frequency normal modes can be identified as contributing to transition towards the active conformations, but we find that none of the normal modes has a dominant contribution. This result may differentiate rhodopsin from other GPCRs and is discussed here in view of the molecular properties associated with the physiological functions of rhodopsin as a detector of single photons.

## Methods

Experimental data on GPCRs activation were retrieved from a considerable number of articles following careful and thorough literature search. The experimental findings (summarized in our online repository http://www.physiology.med.cornell.edu/GPCRactivation/gpcrindex.html) were interpreted geometrically to derive structural constraints that can be used to define distance relations in the activated form(s). The various categories of such data are described briefly below.

### SDSL data

In the SDSL (Site-Directed Spin Labeling) approach, spin labels are introduced into recombinantly expressed proteins by site-directed mutagenesis. Magnetic dipole–dipole interactions between spins on double spin labeled mutants were analyzed to provide interspin distance distributions in the light-induced structural form of rhodopsin [13–15]. The SDSL measurements do not provide exact $C_\alpha$–$C_\alpha$ distances, but indicate a trend of change in the distance between the labeled residues. We defined $C_\alpha$–$C_\alpha$ distance constraint to be within the range of dark state distances derived from the rhodopsin available crystal structures in the PDB databank (A and B chains in 1F88, 1L9H, 1HZX, 1U19 and 1GZM) when SDSL indicates lack of significant change in the active state relative to the dark state. We impose $distance > max\_dist\_dark$, and $distance < min\_dist\_dark$ for increased and decreased distances, respectively.

### Disulfide bonds data

When engineered disulfide bonds are permissive (i.e., when introduced they do not compromise activation of the receptor), the $C_\alpha$–$C_\alpha$ distance constraint in the MII is set to $distance < 8$ Å [42]. For cross-links that block

activation, $C_\alpha$–$C_\alpha$ $distance > 8$ Å is imposed for the active conformation. When cross-links were found to block activation and additional independent data (such as increased accessibility of the residues [16], SDSL data [15] and mutational data [43–46]) indicate increased distance, the constraint was set to $distance > 18$ Å.

Promiscuous disulfide cross-linking was found for the cytoplasmic face of $M_3$ muscarinic receptors, indicating high flexibility of the cytoplasmic region [47, 48]. In this case we do not impose the $distance < 8$ Å constraint for each pair for which cross-linking was found in the presence of agonist, but rather impose a distance decrease constraint for those pairs for which the ratio of agonist-dependent increase in cross-linking is high relative to cross-linking in inactive receptor [48]. Thus, we imposed distance decrease constraints for the pairs 227/251, 227/252, 227/253. (These correspond to positions 5.62/6.34, 5.62/6.35 and 5.62/6.36 in our generic numbering scheme for GPCRs [49], in which residues are numbered within each helix ($n$) relative to the most conserved residues in that TM, that is given the arbitrary number 50; for example, the conserved Pro215 in TM5 is identified as P5.50).

Activation of transducin was not established for cross-linked rhodopsin between 3.55 and Helix 8 residues, and therefore we do not include a constraint for the 3.55/H8 distance [50].

### Fluorescence quenching

A distance decrease constraint was imposed for the pair 5.63/6.27 based on fluorescence quenching experiments that have shown distance decrease between these residues upon activation of $\beta$2AR [32].

### Zinc binding sites

Fowler et al. [51] define Zinc binding site constraints as follows: 6.5–8.5 Å distances between $C_\beta$–$C_\beta$ atoms and 3.2–4.0 Å distances between O, N and S atoms of the residues participating in the formation of $Zn^{2+}$ binding site. We use only the $C_\beta$–$C_\beta$ constraint because the data were obtained either for GPCRs other than rhodopsin [52–55] or for engineered sites in rhodopsin [56]. We use $C_\beta$–$C_\beta$ $distance > 8.5$ Å if the Zinc binding site blocks activation and 6.5 Å < $distance$ $C_\beta$–$C_\beta$ < 8.5 Å if it permits activation.

### NMR

The ability of the dipolar assisted rotational resonance (DARR) protocol to identify proximity of specific

tyrosine, serine and threonine residues to the retinal was established for dark state rhodopsin. DARR was further applied to the MII state to obtain distance constraints for interactions between the retinal and the active protein [57]. Cross peaks are generated in the 2D spectrum when 13C labels in the retinal are within 5.5 Å of the 13C-labeled amino acids. Therefore, cross peaks that disappear in the MII spectrum compared to the dark state generate the constraint *distance* >5.5 Å between the labeled atoms, whereas cross peaks that appear in MII become *distance* <5.5 Å constraints.

## Helical constraints

Helical constraints for the distance between O (residue *i*) and N (residue *i+4*) are introduced as constraints between 2.7 and 3.5 Å (based on dark state differences), and a force constant of 14 kcal/(mol Å$^2$). Such constraints were imposed for the following residues: from 35 to 48, 50 to 59 (i.e., 1.30 to 1.43, 1.45 to 1.54); from 73 to 88, 90 to 95 (2.40 to 2.55, 2.57 to 2.62); from 109 to 135 (3.24 to 3.50); form 150 to 165 (4.39 to 4.54); from 199 to 209, 212 to 221 (5.34 to 5.44, 5.47 to 5.56); from 247 to 262 (6.30 to 6.45, force constant 4) 267 to 274 (6.50 to 6. 57, force constant 24); from 288 to 293, 301 to 303 (7.35 to 7.46, 7.48 to 7.50). No constraints were imposed for prolines and glycines.

## 3D modeling of activated states

To start the procedure we chose chain A of the 1GZM.pdb entry for the rhodopsin structure, because in the trigonal crystal form rhodopsin retained an amphipathic molecular environment characteristic of the native membrane. This allows the hydrophilic segments to display conformations present in a membrane environment unhindered by packing interactions [58]. Based on a validated computational approach for the calculation of p$K_a$ values [59] we assigned the following protonation states to the structure: His65 (1.60), His152 (4.41), His195 (5.30) and His278 (6.61) with hydrogen on N$\epsilon$, His100 (2.67) and His211 (5.46) with hydrogen on N$\delta$ (M. Ceruso, personal communication). Glu122 (3.37), Glu181 (EL2) and Asp83 (2.50) are protonated [7, 59]. The isomerization of the retinal is performed by switching the dihedral potential of the C$_{11}$–C$_{12}$ bond to the *trans* form [4].

After photoisomerization, the dihedral potential is switched back to the ground state and Glu113 (3.28) [60] and Glu134 (3.49) [61] are protonated, while the Schiff base is deprotonated [6]. The protonated Schiff

base parameters were adapted from Roux and coworkers [62, 63].

In the following stage, the constraints derived from the experiments are implemented using the "NOE" option in CHARMM [64]. The NOE potential in CHARMM is zero for allowed distances of $R$, i.e., $R_{min} < R < R_{max}$; has a harmonic form outside the allowed region, i.e., $E(R) = 0.5 \cdot K_{min} \cdot (R - R_{min})^2$ for $R < R_{min}$ or $R_{max} < R < R_{lim}$; and it takes the form $E(R) = F_{max} \cdot (R - 0.5 \cdot (R_{lim} + R_{max}))$, $R > R_{lim}$; $R_{lim} = R_{max} + F_{max} \div K_{max}$ for distances exceeding a specified limit. $F_{max}$ was set to 2 kcal/(mol*Å$^2$). The force constant $K_{min}$ (and/or $K_{max}$) was initially set to 0.1kcal/(mol*Å$^2$).

The system was cooled from 360 K to 300 K during 2 ps for 100 cycles, with the force constant being raised by 0.1 kcal/(mol*Å$^2$) per cycle, and for 100 additional cycles for which constraints are gradually reduced to zero. Steepest descent minimization was performed for 500 steps each time the new set of force constants was introduced. The procedure was repeated four times for each set of constraints designated **R0** through **R5,** and for each set of constraints the structure with the lowest potential energy was chosen as the representative model.

*Normal mode analysis* of the structures and calculation of the contribution of individual modes to the conformational change was performed using NO-MAD-Ref website (http://www.lorentz.immstr.pasteur.fr/nomad-ref.php) [65]. NOMAD-Ref uses the Elastic Network Model [38] where the potential energy function is replaced by Eq. 1:

$$E(\vec{r}_a, \vec{r}_b) = \begin{cases} \frac{k}{2}\left(||\vec{r}_a - \vec{r}_b|| - ||\vec{r}_a^0 - \vec{r}_b^0||\right)^2 & \text{for } ||\vec{r}_a^0 - \vec{r}_b^0|| \leq R_C \\ 0 & \text{for } ||\vec{r}_a^0 - \vec{r}_b^0|| > R_C \end{cases}$$

where $\vec{r}_a - \vec{r}_b$ denotes the vector connecting C$_\alpha$-atoms *a* and *b*, the zero superscript indicates the initial configuration of the C$_\alpha$-atoms, $R_C$ is the spatial cutoff for interconnections between C$_\alpha$ atoms (set to the default value of 10 Å), and *k* is the force constant arbitrarily set to 100 kcal/(mol*Å$^2$). The normal modes are the eigenvectors obtained from diagonalization of the Hessian matrix due to Eq. 1. The overlap between the conformational change vector $\Delta r$ and the normal modes is the cosine of these vectors. The cumulative overlap (which should reach "1" when contributions from enough normal modes are added) is the sum of squared cosines.

*Proline kink* calculation was performed with Prokink [66] as implemented in simulaid (http://www.inka.-mssm.edu/~mezei/simulaid/)

*Iterative fitting of the models and pairwise RMSD calculations* were performed using VMD (http://www.ks.uiuc.edu/Research/vmd/) plugins RMSDTT and iTrajComp developed by Dr. Luis Gracia (http://www.physiology.med.cornell.edu/faculty/hweinstein/vmdplugins/, and personal communication). The iterative fitting implemented in the RMSDTT plugin performs by-residue weighted pairwise fitting of all against all models, such that after each iteration these residues with lower average RMSD get higher weights in the next iteration (Luis Gracia, personal communication). A similar idea has independently been described recently by Damm and Carlson [67]. Pairwise RMSD after iterative fitting were then analyzed using the iTrajComp plugin.

## Results and discussion

The experimental constraints we obtained from the literature are available in Supplementary Table 1 and via http://www.physiology.med.cornell.edu/GPCRactivation/gpcrindex.html. The information contained in this website is up to date, and includes the activation-dependent set of NMR constraints published recently by Smith et al. [57, 68, 69]. To the best of our knowledge, the web site contains the largest set of constraints used to build activated forms of rhodopsin or any other GPCR.

We grouped the total of 83 constraints (53 protein/protein and 30 retinal/protein constraints, detailed in Supplementary Table 1) into classes, depending on the receptors for which they were derived, the TM domains involved, and the increase/decrease of distance.

Since the data for the constraints were obtained in separate experiments, and in some cases for different Class A GPCRs, there is no guarantee that all of the constraints must be satisfied simultaneously by a particular receptor. More importantly, different constraints may potentially correlate with different active states of a GPCR. To explore whether the constraints are compatible with each other, and whether partial sets of constraints may result in models with similar features, we constructed six different sets of constraints, **R0**–**R5** as shown in Table 1.

Model **R0** was built by including all of the constraints. **R1** does not include the newest constraint (derived for Ste2p, the G protein-coupled receptor (GPCR) for the tridecapeptide pheromone alpha-factor of *Saccharomyces cerevisiae* [70]) and the constraint that involves a sulfide bound residue in the second extracellular loop, 97/110 (2.64/3.25). **R2** includes all constraints reported for rhodopsin, and none of the constraints that were obtained for other GPCRs. **R3** includes only constraints reported for rhodopsin, except those for the retinal and the 121/265 (3.36/6.48), 207/167 (5.42/4.56), 211/167 (5.46/4.56) pairs [57, 68, 69]. Because the increase in the distance between cytoplasmic parts of TM3 and TM6 has been the hallmark of GPCR activation, and is well established for various GPCRs [23], it is the only constraint used in model **R4** to establish a "generic" model of activation, possibly applicable across the board. Finally, **R5** is another extreme of a model that includes almost all possible constraints except for the retinal constraints that are clearly specific to rhodopsin only, and the 121/265 (3.36/6.48), 207/167 (5.42/4.56), 211/167 (5.46/4.56) pairs.

**Table 1** Distance constraints for the various active state models. Constraints that were used to derive each model are shaded in the respective column

| derived for | residues involved | change | R0 | R1 | R2 | R3 | R4 | R5 |
|---|---|---|---|---|---|---|---|---|
| RHO | TM1/TM7-H8 (1.60/7.53) | increase | | | | | | |
| RHO | TM1/TM7-H8 (1.60/7.57) | increase | | | | | | |
| RHO | TM1/TM7-H8 (1.56-1.60/7.59) | ~ | | | | | | |
| MC4R | TM2/TM3 (2.60/3.28) | 6.5<>8.5 | | | | | | |
| MC4R | TM2/TM3 (2.64/3.25) | 6.5<>8.5 | | | | | | |
| RHO | TM3/TM5 (3.51/5.57) | increase | | | | | | |
| RHO | TM3/TM5 (3.51/5.60) | **decrease** | | | | | | |
| RHO | TM3/TM5 (3.53/5.57) | increase | | | | | | |
| RHO | TM3/TM5 (3.53/6.30) | **decrease** | | | | | | |
| RHO | TM3/TM6 (3.49-52/6.30-35) | increase | | | | | | |
| NK1, β2ar | TM3/TM7 (3.32/7.39) | 6.5<>8.5 | | | | | | |
| RHO | TM5/TM6 (5.33-5.39/6.59) | **decrease** | | | | | | |
| Mm3 mACHR | TM5/TM6 (5.62/6.34-6.36) | **decrease** | | | | | | |
| β2ar | TM5/TM6 (5.63/6.27) | **decrease** | | | | | | |
| Ste2p | TM5/TM6 (5.30/6.61) | **<10** | | | | | | |
| RHO | various | **various** | | | | | | |

The models incorporating the **R0**-to-**R5** constraints were obtained using the protocol described in Methods. For each of the sets of constraints, the resulting lowest energy model was analyzed for the extent and identity of constraints conservation. We find that all of the constraints that were imposed in the construction of the model are conserved within 1 Å in that model, as detailed in Supplementary Table 1.

Comparison of activated state model conformations

Proline kinks [66] in transmembrane helices are known to play a crucial role in signaling mechanisms of membrane proteins [71]. TM5, TM6 and TM7 of GPCRs have sequence-conserved prolines. To evaluate the contribution of such prolines to the active state models, we calculated the proline bend angles using the default parameters in simulaid (http://www.inka.mssm.edu/~mezei/simulaid/) for the prolines in those three TMs. These are listed in Table 2.

Inspection of Table 2 shows that the bend of TM5 increases in all of the models compared to the inactive state. TM6 is more straight in model **R2**, and more bent in model **R4**, compared to the inactive form, suggesting that the degree of TM6 bending may vary in different active states. Proline kink angles in TM7 vary to a lesser extent than in TM5 and TM6, except for the rather straight conformation in **R3**.

A milestone in the development of a structure-based insight about the intramolecular mechanisms of ligand-dependent GPCR activation was parsing the receptor structure into specific regions identified as structural motifs (SM) acting as (often conserved) functional microdomains (FM) (SM/FMs) [23]. These motifs were shown to be sufficiently conserved in structure and function to merit a specific designation, and therefore we described them as SM/FMs in several different GPCRs, as reviewed recently [28]. These SM/FMs include, for example, the "ionic lock of the arginine cage". Information on the disruption of the ionic lock

during activation was used here to support the increase in TM3–TM6 distance in all of the models.

The cluster of aromatic residues in TM6 that surround the conserved Trp6.48, and straddle the conserved Pro6.50 was shown to be a key SM/FM motif that triggers the regulation of the "ionic lock" through a series of specific structural rearrangements in the upper (more extracellular) end of the GPCR molecule [23]. Notably, this motif can vary somewhat in composition in different receptors, depending on the nature of the ligand. But these variations preserve the steric properties that can trigger the rearrangement of the other aromatic residues in the SM/FM in the manner of a "toggle switch" [23] that connects the ligand-binding event to the rearrangements in the receptor structure, leading to activation.

Earlier simulations modeling the manner in which ligands are recognized in the GPCR binding pocket [23] suggested that the ability to trigger this toggle switch determines the efficacy of a ligand [30]. Thus, the position of a ligand in the binding site affects its interaction with the aromatic cluster, so that the ligands that differ in orientation will differ in the extent to which they can affect the toggle switch. For example, in the 5-HT$_2$AR model, compounds with bulky substitutions of the cationic amine moiety that correlated with the appearance of hallucinogenic properties, adopt a very different position in the binding pocket compared with non-hallucinogenic congeners in the same family [30].

It had previously been shown that upon activation the orientation of Trp6.48 sidechain changes from being "perpendicular" to the membrane plane to "parallel" to membrane plane [28] in accord with experimental studies [72]. Notably, we observe the same phenomenon, both in models that include explicit constraints on the 6.48 sidechain orientation from NMR experiments [68] (**R0**, **R1** and **R2**) and in models that do not include any constraints on 6.48 sidechain orientation (**R3**, **R4** and **R5**). Thus, although the Trp6.48 orientation may differ among the models, the change towards "parallel" orientation is a common feature of activated state conformation they represent.

To reveal the differences in their overall structural properties compared to each other and to the inactive structure, the models were overlapped using the RMSDTT plugin (Dr. Luis Gracia, http://www.physiology.med.cornell.edu/faculty/hweinstein/vmdplugins/) to the VMD program. This newly developed procedure iteratively redefines weights for regions in the protein to be used in the structural alignment, such that the most conserved regions have highest weights, while the

**Table 2** Proline kink angle values for TM5, TM6 and TM7 in the different models

|  | P215 bend | P267 bend | P303 bend |
|---|---|---|---|
| GZM | 4 | 36 | 23 |
| **R0** | 12 | 37 | 17 |
| **R1** | 16 | 36 | 17 |
| **R2** | 12 | 18 | 27 |
| **R3** | 20 | 33 | 7 |
| **R4** | 26 | 52 | 30 |
| **R5** | 15 | 30 | 26 |

least conserved ones have the lowest. The resulting overlapped structures, the RMSD vs. Cα and the pairwise RMSD (calculated with iterCOMP plugin, http://www.physiology.med.cornell.edu/faculty/hweinstein/vmdplugins/) are shown in Fig. 1.

Inspection of these results reveals that the active models differ from the inactive form more than from each other, and TM2, followed by TM1, are conserved regions of the activated conformations in our study. Most of the conformational differences involve the intracellular loops $IC_2$, $IC_3$ and the cytoplasmic ends of TMs 5,6 and 7 (Fig. 1C).

The **R0** and **R1** models, built with similar sets of constraints, are similar (2.6 Å backbone RMSD). **R0** and **R2** are much less similar to each other (3.6 Å). **R2** is

constructed solely with constraints that were derived for rhodopsin directly and, indeed, most of the constraints that were not specifically introduced in **R2** were not satisfied by this model. The significant difference between **R0** and **R2** implies that different GPCRs might vary in their active forms. The **R3** model was constructed with the same constraints as **R2**, but without the protein/ retinal and 121/265 (3.36/6.48), 207/167 (5.42/4.56) and 211/167 (5.46/4.56) constraints derived from NMR. The difference between the two models is 3 Å, and most of the difference is in TM5 and $IC_3$. The intracellular end of TM3 and extracellular ends of TM1, TM4 and TM7 also exhibit deviations from **R2**. **R4** is a model built from the minimal set of constraints—only the hallmark TM3/ TM6 increase in distance, which was shown for many
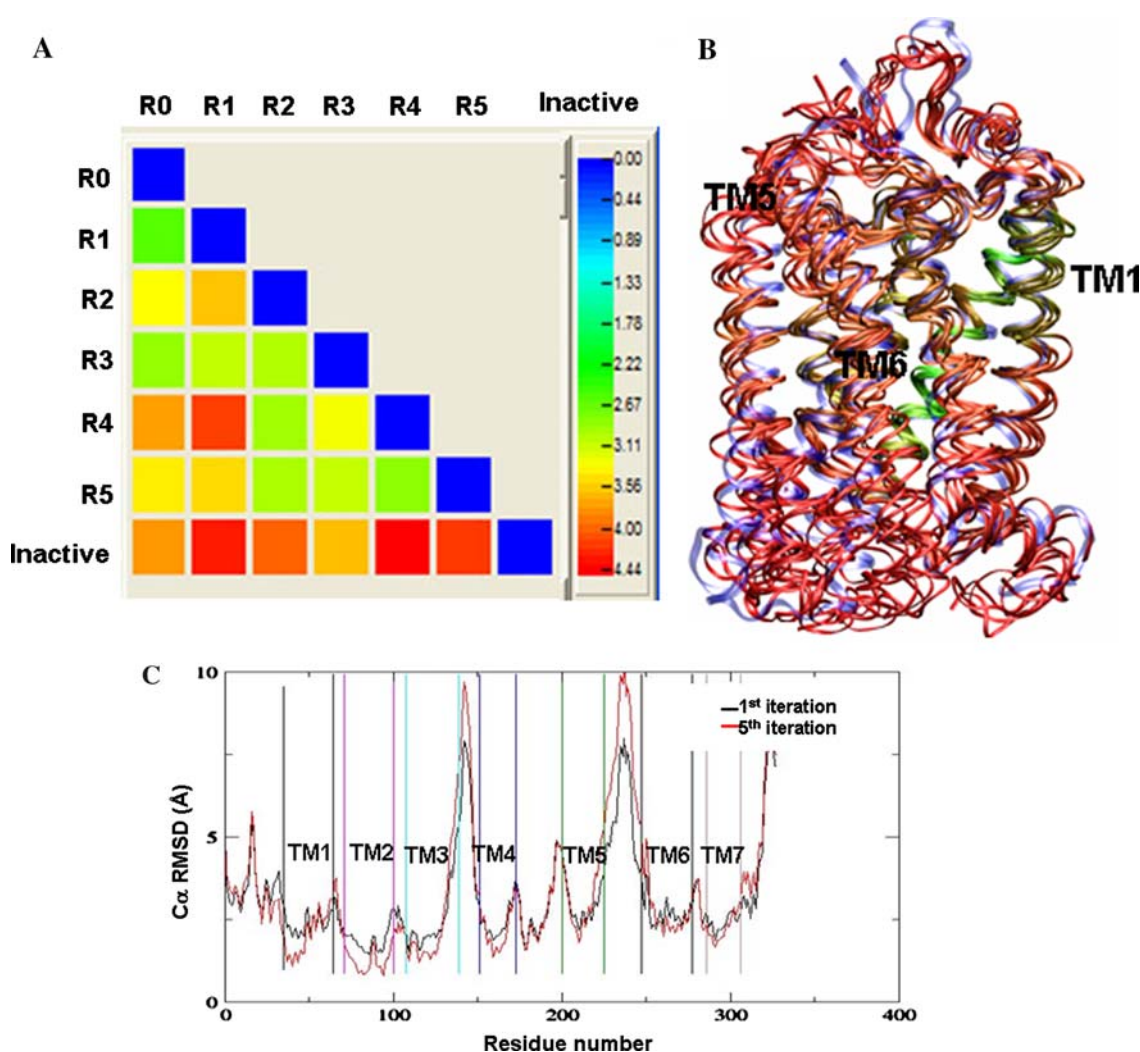


**Fig. 1** Average RMSD per residue, based on iterative alignment of the dark state and five active models. (**A**) Pairwise backbone RMSD for the models and the inactive form is represented in a 2D matrix. (**B**) Ribbon representation of the inactive form (1GZM, chain A, colored blue) and the models, colored by RMSD calculated after five iterations of structural alignment. (**C**) RMSD values calculated after first and fifth iterations are plotted versus $C_\alpha$ number

different GPCRs. After five iterations of weighted superimposition of **R0** and **R4**, the RMSD is 3.4 Å. The differences between the models are spread throughout the protein, with the most conserved parts being the extracellular parts of TM1 and TM2, as well as helix 8.

The **R0** and **R5** models were built based on the same (maximal) set of constraints, except that **R5** is lacking all the retinal set and the 121/265 (3.36/6.48), 207/167 (5.42/4.56), 211/167 (5.46/4.56) constraints. Similarly to what we have found by comparing **R2** and **R3** models, we observe that the most pronounced differences between **R0** and **R5** are in TM5 and IC$_3$.

To define an ''activation pattern'', we identified in each model the residue pairs that exhibited the largest (top 1%) distance difference between the inactive and the active state. Most of these residues are expected to be the ones on which constraints were originally imposed to build the active state, but a few additional pairs emerged, listed in Table 3.

The decrease in distance between intracellular ends of TM2 and TM5 occurs in all of the models, but is most pronounced in **R0**, **R1** and **R5**. There is also a decrease in distance between the intracellular end of TM6 and the beginning of the amphipatic helix 8 in all models, most pronounced in models **R0**, **R1** and **R3**. The intracellular TM3/TM4 distance also decreases in all models except **R5**, most significantly in **R4**. *These distance changes are in the ranges accessible to SDSL experiments, and our prediction based on the models is that positions 2.39/5.61, as well as 6.26/7.59 should have increased cross-linking in the active states. Thus far unnoticed, these residue-pair distances are interesting candidates for experimental testing of activated state properties.*

Faced with the added complexity of dimers/oligomers rather than monomers in the current signaling models of GPCRs, we are currently building models of activated GPCR complexes using a similar strategy. Given the information about GPCR dimerization interfaces available from rhodopsin models [73] and data for the dopamine D2 receptor [74], it is very interesting to note the movement of TM5, IC$_2$, and IC$_3$ resulting from the application of all sets of distance constraints used in this study. This observation suggests

a perturbation of the interface of dimerization of rhodopsin upon activation that is consistent with our proposed rearrangement of the interface of dimerization of dopamine D2 receptor upon activation [74], and the transformation from an inactive model of the rhodopsin dimer based on inferences from atomic force microscopy [73].

## Concerning the path of conformational rearrangement from inactive to active forms of rhodopsin—Normal Mode Analysis (NMA)

The realization that experimentally observed functional motions of proteins can be predicted by coarse-grained normal mode analysis has renewed interest in applications of this technique to structural biology [36]. One notable application of NMA is the prediction of biologically relevant motions of proteins and supramolecular structures driven by their structure-encoded collective dynamics. It has been shown that structural transitions from one state of a protein to another are often represented by few (low frequency) normal modes of the original structure. For example, Tama and Sanejouand [39] have analyzed 20 cases where 2 conformations of a protein are known. They showed that the maximal overlap of a single normal mode with the conformational change towards the other state of the protein ranges from 33 to 85%. Krebs and coworkers [75] have analyzed the protein motions in a database framework, showing that motions often lie along the direction of a combination of two modes, and the most commonly observed weight of a single mode is 0.5–0.6. Tobi and Bahar have emphasized the pre-existing equilibrium/conformational selection as a mechanism for protein–protein interaction and the concept that in their native conformation, proteins are predisposed to undergo conformational fluctuations that are relevant to, or required for, their biological functions [76].

To probe whether this relation holds in the more complex case of rhodopsin where the transition from the inactive to activated forms involves the deposition of the significant energy of a photon, we calculated the lowest normal modes of the ground state of inactive

**Table 3** Distances that undergo a significant change upon transition to the active state, but were not included in the constraints

| Regions involved | Residue pair | Inactive (Å) | **R0** (Å) | **R1** (Å) | **R2** (Å) | **R3** (Å) | **R4** (Å) | **R5** (Å) |
|---|---|---|---|---|---|---|---|---|
| TM2/TM5 | 72/226 (2.39/5.61) | 16.7 | 10 | 8 | 12 | 12 | 11 | 11 |
| TM6/Helix 8 | 243/312 (6.26/7.59) | 14.2 | 7 | 9 | 10 | 8 | 9 | 13 |
| TM3/TM4 | 140/152 (3.55/4.41) | 17.2 | 14 | 13 | 12 | 13 | 7 | 17 |

**Fig. 2** Overlap of the transformation change towards the different models with the normal modes of the inactive form. The numbers identify the most prevalent modes
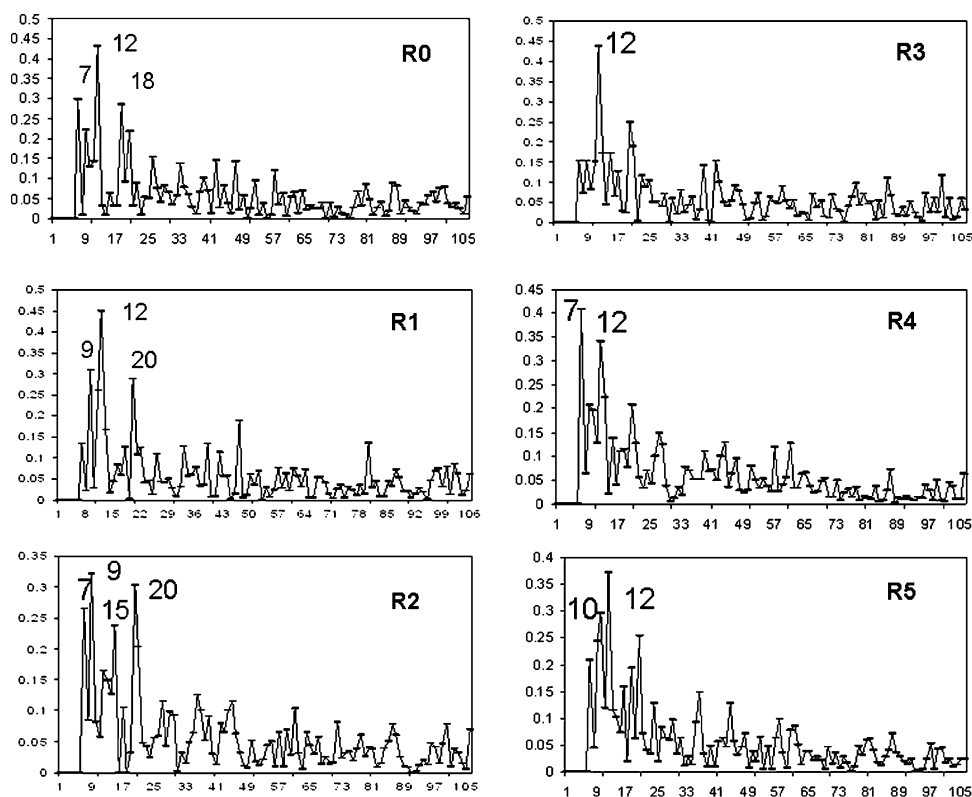
rhodopsin (using 1GZM.pdb chain A) and their contribution towards transition to each one of the activated models using NOMAD-Ref server [65] (Fig. 2).

For any of the active models we described here, the difference vector between the model and the inactive state has a maximum overlap value of 30–45% (shown in Fig. 2) with the normal modes of the inactive state. Thus, unlike many of the examples cited above, there is no single mode or pair of modes of the inactive structure of rhodopsin that can represent the conformational change that would accompany activation. The cumulative sum of the squares of such individual contributions (which should reach the value of ''1'' if the conformational change was fully described) reaches 50% only after summing the contributions from the 20–30 lowest frequency modes, as shown in Fig. 3.

Nevertheless, it is worth noting the modes that have the most significant contributions to the change. In particular, modes 7, 9, 12 and 20 are the most significant contributors to transition towards more than one active model. The displacements of the $C_\alpha$ atoms in these modes are depicted in Fig. 4. The largest displacements in mode 7 are for residues 220 through 247: the intracellular tip of TM5 and the intracellular loop 3, which are often implicated in interaction with downstream proteins [77–80]. Mode 7 contributes most

to transitions to the active states of models **R0**, **R2**, **R4** and **R5**. Mode 9 consists of correlated motions of $IC_3$ and $IC_2$, another well established scaffolding point for intracellular partners of GPCRs [28, 81–84], and is involved in transitions to models **R1**, **R2** and **R5**.

## Conclusions

It is now established that GPCRs are characterized by several conformationally distinct active states [25, 26, 29, 33]. In this study we obtained models of putative
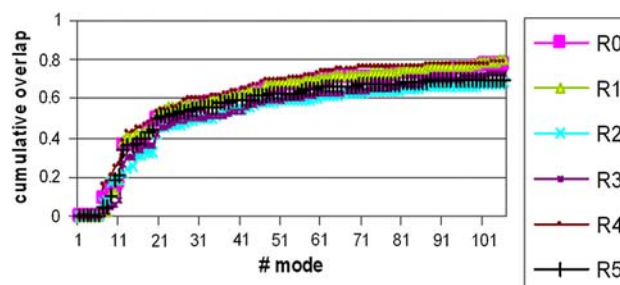


**Fig. 3** Cumulative overlap (calculated as sum of squares) between transformation to various models and the normal modes of the inactive form. The six lowest frequency modes are translation and rotation

**Fig. 4** $C_\alpha$ displacements in normal modes 7, 9, 12 and 20. Panels on the left show the values as a function of sequence. Structures on the right represent the motions for the corresponding modes by superposition of conformations perturbed along the normal modes on the equilibrium structure of RHO
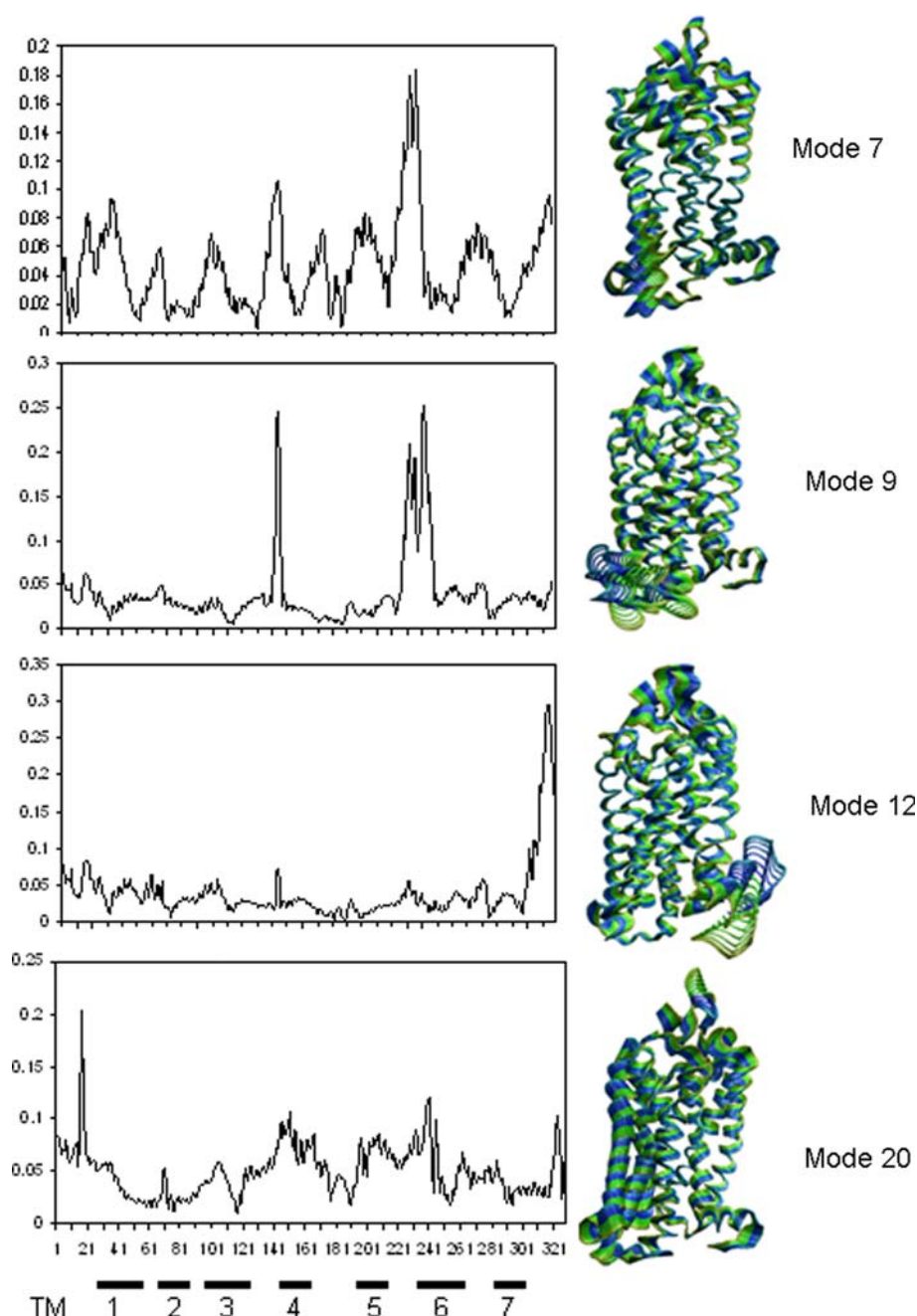


active conformations, using different subsets of constraints representing experimentally determined information about the conformational changes that occur upon activation. Some of the structural changes produced by imposing the experiment-derived constraints occur in regions that have not been previously tested experimentally, and thus emerge as interesting candidates for further experimental exploration of the conformational space of activated GPCRs. Specifically, we predict from the models a decrease in distances between intracellular ends of TM2/TM5, TM6/helix 8,

and TM3/TM4, which may be common features of the active states. For example, we have identified distances that decrease upon activation: $C_\alpha$–$C_\alpha$ distances 2.39/5.61, 6.26/7.59 are 16.7 and 14.2 Å, respectively in the inactive form, but are below 12 Å in models **R0** through **R4**. Therefore, disulfide cross-linking of these residues would be more compatible with the active, than with the inactive forms. These distance changes, as well as the 3.55/4.41 distance decrease from 17.2 to 13 Å and below are testable experimentally also by spin labeling the positions of interest and

monitoring the change in the EPR spectra upon illumination.

To the best of our knowledge, this work is the first attempt to evaluate the normal mode motions of the inactive form of rhodopsin in the context of the changes involved in transitions to specific structures of activated form of the protein. We find that unlike many other cases of structural changes related to protein functions [75], the conformational transition related to rhodopsin activation is not dominated by a small number of low-frequency modes. This finding does not come as a surprise. Given the nature of its physiological function, rhodopsin is a tightly controlled protein, for which the level of constitutive activity is extremely low [85] compared to other GPCRs, and naturally occurring constitutive active mutations lead to serious diseases [86–88]. The activation is therefore unlikely to be a spontaneous or low barrier process, but rather is induced by absorption of a high-energy photon. As expected, therefore, the transition of inactive rhodopsin to the active state cannot be described as a simple motion of a single low frequency normal mode.

Notably, we find that the pattern of contributions of normal modes to transformation towards different models is similar, and the models cannot be distinguished on the basis of the number of contributing normal modes. Perhaps as a result of the physiological considerations described above, this finding stands in some contrast to a recent suggestion that correct putative active-state models of smooth muscle myosin ATPase can be distinguished from less correct models by the ability to conform those models to the final inhibited state along a small number of low-frequency normal modes [89].

In order to identify structural changes that may be important for the connection of the activated state to the downstream components of the signaling cascade, we note that in addition to the hallmark TM6 movement, TM3 and TM5 also appear to have a prominent role in the activation process. It is tempting to speculate that the differences observed in the TM5 and $IC_3$ regions in the various models may be responsible for the variety of changes in the GPCR interfaces, that would lead to differential signaling pathways via changes in monomer-monomer communication among GPCRs. This pathway is an alternative, or perhaps complementary to the changes involved directly in the interactions with downstream soluble proteins such as G proteins, kinases, arrestins, PDZ domains etc. (GPCR interaction molecules reviewed in [90]). In a study reported elsewhere in this same issue, Filizola et al. present an equilibrated model of a rhodopsin dimer simulated in an explicit lipid–water environment. Starting from this equilibrated structure, and using the activation protocol described above, we are currently building models of activated GPCR dimers in order to test the emerging hypothesis that different activated forms of GPCRs might differentially influence the dimer interfaces of the proteins, leading to selective coupling to different signaling pathways.

# References

1. Muller G (2000) Curr Med Chem 7(9):861
2. Li J, Edwards PC, Burghammer M Villa C, Schertler GF (2004) J Mol Biol 343(5):1409
3. Rohrig UF, Guidoni L, Rothlisberger U (2002) Biochemistry 41(35):10799
4. Saam J, Tajkhorshid E, Hayashi S, Schulten K (2002) Biophys J 83(6):3097
5. Lemaitre V, Yeagle P, Watts A (2005) Biochemistry 44(38):12667
6. Kuwata O, Yuan C, Misra S, Govindjee R, Ebrey TG (2001) Biochemistry (Mosc) 66(11):1283
7. Yan EC, Kazmi MA, Ganim Z, Hou JM, Pan D, Chang BS, Sakmar TP, Mathies RA (2003) Proc Natl Acad Sci USA 100(16):9262
8. Elamrani S, Berry MB, Phillips GN Jr, McCammon JA (1996) Proteins 25(1):79
9. Chambers JJ, Nichols DE (2002) J Comput Aided Mol Des 16(7):511
10. Bissantz C, Bernard P, Hibert M, Rognan D (2003) Proteins 50(1):5
11. Shacham S, Marantz Y, Bar-Haim S, Kalid O, Warshaviak D, Avisar N, Inbal B, Heifetz A, Fichman M, Topf M, Naor Z, Noiman S, Becker OM (2004) Proteins 57(1):51
12. Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O, Bar-Haim S, Warshaviak D, Fichman M, Noiman S (2004) Proc Natl Acad Sci USA 101(31):11304
13. Altenbach C, Klein-Seetharaman J, Cai K, Khorana HG, Hubbell WL (2001) Biochemistry 40(51):15493
14. Altenbach C, Cai K, Klein-Seetharaman J, Khorana HG, Hubbell WL (2001) Biochemistry 40(51):15483
15. Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG (1996) Science 274(5288):768
16. Dunham TD, Farrens DL (1999) J Biol Chem 274(3):1683
17. Choi G, Landin J, Galan JF, Birge RR, Albert AD, Yeagle PL (2002) Biochemistry 41(23):7318
18. Nikiforovich GV, Marshall GR (2003) Biochemistry 42(30):9110
19. Gouldson PR, Kidley NJ, Bywater RP, Psaroudakis G, Brooks HD, Diaz C, Shire D, Reynolds CA (2004) Proteins 56(1):67
20. Slusarz R, Slusarz MJ, Lammek B, Ciarkowski J (2006) QSAR Comb Sci 25:105
21. Elling CE, Frimurer TM, Gerlach LO, Jorgensen R, Holst B, Schwartz TW (2006) J Biol Chem 281(25):17337

22. Filizola M, Visiers I, Skarabanek L, Campagne F, Weinstein H (2003) Molecular neuropharmacology: strategies and methods. Humana Press. Inc., Totowa, NJ
23. Visiers I, Ballesteros JA, Weinstein H (2002) Methods Enzymol 343:329
24. Fowler CB, Pogozheva ID, Lomize AL, LeVine H 3rd, Mosberg HI (2004) Biochemistry 43(50):15796
25. Kenakin T (2004) Trends Pharmacol Sci 25(4):186
26. Vauquelin G, Van Liefde I (2005) Fundam Clin Pharmacol 19(1):45
27. Perez DM, Karnik SS (2005) Pharmacol Rev 57(2):147
28. Weinstein H (2006) AAPS J 7(4):E871
29. Ghanouni P, Gryczynski Z, Steenhuis JJ, Lee TW, Farrens DL, Lakowicz JR, Kobilka BK (2001) J Biol Chem 276(27):24433
30. Ebersole BJ, Visiers I, Weinstein H, Sealfon SC (2003) Mol Pharmacol 63(1):36
31. Gether U, Lin S, Ghanouni P, Ballesteros JA, Weinstein H, Kobilka BK (1997) EMBO J 16(22):6737
32. Ghanouni P, Steenhuis JJ, Farrens DL, Kobilka BK (2001) Proc Natl Acad Sci USA 98(11):5997
33. Swaminath G, Deupi X, Lee TW, Zhu W, Thian FS, Kobilka TS, Kobilka B (2005) J Biol Chem 280(23):22165
34. Elber R (2005) Curr Opin Struct Biol 15(2):151
35. Tama F, Brooks CL 3rd (2006) Ann Rev Biophys Biomol Struct 35:115
36. Bahar I, Rader AJ (2005) Curr Opin Struct Biol 15(5):586
37. Ma J (2005) Structure 13(3):373
38. Tirion MM (1996) Phys Rev Lett 77(9):1905
39. Tama F, Sanejouand YH (2001) Protein Eng 14(1):1
40. van Vlijmen HW, Karplus M (2005) J Mol Biol 350(3):528
41. Bahar I, Wallqvist A, Covell DG, Jernigan RL (1998) Biochemistry 37(4):1067
42. O'Connor BD, Yeates TO (2004) Nucleic Acids Res 32(Web Server issue): W360
43. Huang P, Visiers I, Weinstein H, Liu-Chen LY (2002) Biochemistry 41(40):11972
44. Huang P, Li J, Chen C, Visiers I, Weinstein H, Liu-Chen LY (2001) Biochemistry 40(45):13501
45. Scheer A, Fanelli F, Costa T, De Benedetti PG, Cotecchia S (1996) EMBO J 15(14):3566
46. Ballesteros J, Kitanovic S, Guarnieri F, Davies P, Fromme BJ, Konvicka K, Chi L, Millar RP, Davidson JS, Weinstein H, Sealfon SC (1998) J Biol Chem 273(17):10445
47. Zeng FY, Hopp A, Soldner A, Wess J (1999) J Biol Chem 274(23):16629
48. Ward SD, Hamdan FF, Bloodworth LM, Wess J (2002) J Biol Chem 277(3):2247
49. Ballesteros JA, Weinstein H (1995) Methods Neurosci 25:366
50. Yu H, Kono M, Oprian DD (1999) Biochemistry 38(37):12028
51. Fowler CB, Pogozheva ID, LeVine H 3rd, Mosberg HI (2004) Biochemistry, 43(27):8700
52. Holst B, Elling CE, Schwartz TW (2000) Mol Pharmacol 58(2):263
53. Elling CE, Thirstrup K, Holst B, Schwartz TW (1999) Proc Natl Acad Sci USA 96(22):12322
54. Sheikh SP, Vilardarga JP, Baranski TJ, Lichtarge O, Iiri T, Meng EC, Nissenson RA, Bourne HR (1999) J Biol Chem 274(24):17033
55. Lagerstrom MC, Klovins J, Fredriksson R, Fridmanis D, Haitina T, Ling MK, Berglund MM, Schioth HB (2003) J Biol Chem 278(51):51521
56. Sheikh SP, Zvyaga TA, Lichtarge O, Sakmar TP, Bourne HR (1996) Nature 383(6598):347
57. Patel AB, Crocker E, Eilers M, Hirshfeld A, Sheves M, Smith SO (2004) Proc Natl Acad Sci USA 101(27):10048
58. Li J, Edwards PC, Burghammer M, Villa C, Schertler GF (2004) J Mol Biol 343(5):1409
59. Periole X, Ceruso MA, Mehler EL (2004) Biochemistry 43(22):6858
60. Jager F, Fahmy K, Sakmar TP, Siebert F (1994) Biochemistry 33(36):10878
61. Fahmy K, Sakmar TP, Siebert F (2000) Biochemistry 39(34):10607
62. Nina M, Smith JC, Roux B (1993) J Mol Struct (Theochem) 105:231
63. Nina M, Roux B, Smith JC (1995) Biophys J 68(1):25
64. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187
65. Lindahl E, Azuara C, Koehl P, Delarue M. (2006) Nucleic Acids Res, 34(Web server issue): 38
66. Visiers I, Braunheim BB, Weinstein H (2000) Protein Eng 13(9):603
67. Damm KL, Carlson HA (2006) Biophys J 90(12):4558
68. Patel AB, Crocker E, Reeves PJ, Getmanova EV, Eilers M, Khorana HG, Smith SO (2005) J Mol Biol 347(4):803
69. Crocker E, Patel AB, Eilers M, Jayaraman S, Getmanova E, Reeves PJ, Ziliox M, Khorana HG, Sheves M, Smith SO (2004) J Biomol NMR 29(1):11
70. Lee YH, Naider F, Becker JM (2006) J Biol Chem 281(4):2263
71. Sansom MS, Weinstein H (2000) Trends Pharmacol Sci 21(11):445
72. Lin SW, Sakmar TP (1996) Biochemistry 35(34):11149
73. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A (2003) J Biol Chem 278(24):21655
74. Guo W, Shi L, Filizola M, Weinstein H, Javitch JA (2005) Proc Natl Acad Sci USA 102(48):17495
75. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M (2002) Proteins 48(4):682
76. Tobi D, Bahar I (2005) Proc Natl Acad Sci USA 102(52):18908
77. Acharya S, Saad Y, Karnik SS (1997) J Biol Chem 272(10):6519
78. Lee NH, Geoghagen NS, Cheng E, Cline RT, Fraser CM (1996) Mol Pharmacol 50(1):140
79. Varrault A, Le Nguyen D, McClue S, Harris B, Jouin P, Bockaert J (1994) J Biol Chem 269(24):16720
80. Ulfers AL, McMurry JL, Kendall DA, Mierke DF (2002) Biochemistry 41(38):11344
81. Gether U (2000) Endocr Rev 21(1):90
82. Visiers I, Hassan SA, Weinstein H (2001) Protein Eng 14(6):409
83. Miura S, Karnik SS (2002) J Biol Chem 277(27):24299
84. Visiers I, Hassan SA, Weinstein H (2001) Protein Eng 14(6):409
85. Milligan G (2003) Mol Pharmacol 64(6):1271
86. Jin S, Cornwall MC, Oprian DD (2003) Nat Neurosci 6(7):731
87. Iakhine R, Chorna-Ornan I, Zars T, Elia N, Cheng Y, Selinger Z, Minke B, Hyde DR (2004) J Neurosci 24(10):2516
88. Rao VR, Oprian DD (1996) Ann Rev Biophys Biomol Struct 25:287
89. Tama F, Feig M, Liu J, Brooks CL 3rd, Taylor KA (2005) J Mol Biol 345(4):837
90. Tilakaratne N, Sexton PM (2005) Clin Exp Pharmacol Physiol 32(11):979