Chapter 14

Structure Prediction of G Protein-Coupled Receptors and Their Ensemble of Functionally Important Conformations

Ravinder Abrol, Adam R. Griffith, Jenelle K. Bray, and William A. Goddard III

Abstract

G protein-coupled receptors (GPCRs) are integral membrane proteins whose "pleiotropic" nature enables transmembrane (TM) signal transduction, amplification, and diversification via G protein-coupled and β arrestin-coupled pathways. GPCRs appear to enable this by being structurally flexible and by existing in different conformational states with potentially different signaling and functional consequences. We describe a method for the prediction of the three-dimensional structures of these different conformations of GPCRs starting from their amino acid sequence. It combines a unique protocol of computational methods that first predict the TM regions of these receptors and TM helix shapes based on those regions, which is followed by a locally complete sampling of TM helix packings and their scoring that results in a few (~10–20) lowest energy conformations likely to play a role in binding to different ligands and signaling events. Prediction of the structures for multiple conformations of a GPCR is starting to enable the testing of multiple hypotheses related to GPCR activation and binding to ligands with different signaling profiles.

Key words: GPCR, Protein structure prediction, Transmembrane helix, Hydrophobicity, Helix kinks, BiHelix, GPCR conformations, Activation, Transmembrane signaling

1. Introduction

G protein-coupled receptors (GPCRs) are intrinsic membrane proteins with seven transmembrane (TM) helices. As their name suggests, these receptors were thought to be G protein-coupled, but enough evidence has accumulated in favor of their exclusive coupling to β arrestins (1, 2) that these are being referred to as 7TM proteins. They are the largest superfamily in human genome with ~800 GPCRs identified, including ~340 nonolfactory receptors organized into six families (GRAFTS) (3): Glutamate, Rhodopsin, Adhesion, Frizzled, Taste2 (Bitter), and

Nagarajan Vaidehi and Judith Klein-Seetharaman (eds.), *Membrane Protein Structure and Dynamics: Methods and Protocols*, Methods in Molecular Biology, vol. 914, DOI 10.1007/978-1-62703-023-6_14, © Springer Science+Business Media, LLC 2012

Secretin. A variety of bioactive molecules (including biogenic amines, peptides, lipids, nucleotides, and proteins) modulate GPCR activity to affect regulation of essential physiological processes (e.g., neurotransmission, cellular metabolism, secretion, cell growth, immune defense, and differentiation). Thus, many important cell recognition and communication processes involve GPCRs. Due to their mediation of numerous critical physiological functions, GPCRs are involved in all major disease areas including cardiovascular, metabolic, neurodegenerative, psychiatric, cancer, and infectious diseases.

In the last few years there has been a rapid increase in the solution of crystal structures for many GPCRs (4–15) due mainly to a technological revolution in membrane protein structure determination methods (16). The topological comparison of these crystal structures and their implications for GPCR activation has been reviewed extensively (17–21). All GPCR conformations observed so far have corresponded to the inactive state of the respective receptor, except for bovine opsin (9) and nanobody stabilized human β2 adrenergic receptor (β2AR) bound to an agonist (12). Structure determination efforts are moving towards the stabilization of GPCRs in different functional conformations (e.g., bound to agonists, G proteins or β arrestins), and structural computational biology can help by mapping the energy landscape sampled by GPCRs during their life-cycle and characterizing the critical conformations along the way to link with those that are observed in experimental structures.

It has been well established that during the process of GPCR activation a sequence of conformational changes takes place that take the receptor from an inactive state to an active state. This enables the cell to convert extracellular signals into intracellular signaling cascades through G protein-coupled and/or β arrestin-coupled pathways resulting in specific physiological responses (1, 2). Strong experimental evidence for this flexibility includes fluorescence lifetime measurements that have identified distinct agonist-induced conformational states for the β_2 Adrenergic Receptor (22). It has also been recently shown that conformational coupling between the extracellular surface (ECS) and orthosteric binding site in the TM region can stabilize different conformations of a GPCR (23). It is expected that functionally different ligands (agonists, antagonists, inverse agonists, and allosteric modulators) bind to different receptor conformations among the continuum of conformational states that lie along the activation pathway, so if there was a method to identify all low-energy conformations accessible to GPCRs, it can provide a starting point to test hypotheses related to the binding of different ligands and associating specific functions with different active state conformations some of which may only couple to G protein pathways and some to β arrestin pathways.

Protein structure prediction and modeling is playing an increasingly important role in providing detailed structural information

that is relevant to GPCR activation and holds the most promise to describe the continuum of conformational states involved in GPCR function. Membrane proteins and their environment have been the focus of structure prediction and dynamics simulations for some years now (24). The interaction of these proteins with their lipid environment is considered critical to their in vivo folding and many recent studies have attempted to quantify this interaction on an absolute thermodynamic basis (25) by providing, for example, thermodynamic costs for the insertion of amino acids (that make up the TM helices) into the lipid bilayer (26). The three-dimensional structure of these α-helical membrane proteins, to which GPCRs belong, is strongly affected by interhelical interactions (mainly H-bonds and salt-bridges) (27). An accurate GPCR structure prediction methodology needs to be able to sample and describe these interhelical interactions very thoroughly. The conformational changes that accompany GPCR activation are known to occur on the millisecond or higher time scales, suggesting that explicit all-atom molecular dynamics will not be able to describe these large conformational changes for many years and even then cannot ensure complete sampling of available conformational space.

Many methods have been used to obtain model structures for membrane proteins due to their pharmacological importance. These methods have been reviewed elsewhere (24, 28). For GPCRs the main approach has been homology modeling (using the X-ray structure of Bovine Rhodopsin as a reference till 2007, and others more recently). Because of their low homology to other GPCRs of pharmacological interest, most studies have used constraints based on mutation and binding experiments coupled to the homologous Rhodopsin or later structures to guide additional mutation experiments. These structures have not generally been sufficiently accurate for predicting binding sites of ligands. Methods are also available for predicting structures of membrane proteins in general (see e.g., (29)). Some recent computational studies have started from inactive structures (from crystals) and provided a more detailed view of ligand-stabilized GPCR conformations (30, 31) which is consistent with experimental observations. Our group has been developing de novo computational approaches (not based on pure homology) such as MembStruk and HierDock (32), for predicting the 3D structure of a GPCR, and its ligand binding sites that have been used in the past (33–38).

This review focuses on the new generation of methods developed in our group aimed at generating multiple conformations of a GPCR. It is referred to as the GEnSeMBLE (GPCR Ensemble of Structures in Membrane BiLayer Environment) method, and includes a highly efficient procedure to completely sample the available conformational space. It is described in the next section followed by practical caveats in the Notes section.

2. Methods

In order to provide the 3D structures for these various conformations needed to understand the function of GPCRs and to help design new ligands, we developed the GEnSeMBLE (*G*PCR *E*nsemble of *S*tructures in *M*embrane *Bi**L*ayer *E*nvironment) method to predict the 3D structure (without using homology to known 3D structures). GEnSeMBLE predicts the ensemble of low-energy conformational states for a GPCR. The methodology consists of the following key steps:

1. *PredicTM*: A new method to predict the TM regions for a helical membrane protein that doesn't use any fitted parameters and also accounts for TM region extensions beyond the hydrophobic region.
2. *OptHelix*: A new and unbiased approach using molecular dynamics to generate relaxed helices with their natural kinks.
3. *BiHelix/SuperBiHelix*: A novel and highly efficient sampling algorithm to sample multiple conformations of a GPCR, which divides the N-helix interaction problem into a limited number of 2-helix interactions and uses the SCREAM method for side-chain optimization (39).
4. *CombiHelix/SuperCombiHelix*: An algorithm that takes the results from BiHelix step, generates the multiple rotational combinations, optimizes the side-chains using SCREAM, evaluates an implicit membrane solvation contribution, and evaluates the total energy of the optimized multihelix bundle. This step provides an ensemble of multiple low-energy structures and docking of agonists, antagonists, and inverse agonists to all these conformations will provide a unique route to understand GPCR activation.

GEnSeMBLE provides a unique protocol to predict not only the lowest energy structure but also other low-energy conformations that will be relevant to understand the function of GPCRs in relation to their interaction with different ligands as well as to their relevance for GPCR activation. The individual steps in the GEnSeMBLE method are now described in detail.

2.1. PredicTM

Accurate ab initio prediction of helical membrane proteins like GPCRs begins with the identification of the location of the TM regions of the protein within the amino acid sequence. There are many methods already available for predicting the location of TM regions. These methods encompass simple hydrophobicity-based procedures, hidden Markov models, and neural networks. PredicTM combines hydrophobic analysis with helicity predictions to generate TM regions that encompass hydrophobic regions and

can also potentially extend beyond them. The difficulty with this approach is sufficiently reducing noise in the hydrophobicity analysis without losing data.

The steps of the PredicTM procedure are broken into six parts:

1. Retrieval of similar protein sequences from a database.
2. Multiple sequence alignment of similar sequences.
3. Hydrophobic profile generation and noise removal.
4. Initial transmembrane region predictions.
5. Helicity predictions and extension of hydrophobic regions.
6. Identification of hydrophobic centers.

Step 1: *Similarity search*. The first step of the procedure, retrieving sequences related to the target sequence, is carried out through a modified BLAST search. BLAST, the Basic Local Alignment Search Tool, is a bioinformatics program used to search databases for related sequences and is widely used (2). One of the standard implementations of BLAST, and the implementation used by PredicTM, is at the Expert Protein Analysis System (ExPASy) server at the Swiss Institute of Bioinformatics (SIB). Typically, one gives a sequence or sequence identifier as an input to a Web-based form, along with some parameters, and the BLAST program will return either all sequences satisfying a statistical cutoff threshold or up to a user-specified number of sequences, whichever is reached first. The goal of this step for PredicTM is to always retrieve all sequences satisfying the statistical cutoff. For this reason, PredicTM incrementally increases the requested number of sequences until no new sequences are returned by BLAST. Once this completeness has been achieved, the sequences are returned in FASTA format for the multiple sequence alignment step (see Note 1).

Step 2: *Multiple sequence alignment*. The second step of the PredicTM procedure is multiple sequence alignment of the sequences returned from BLAST. This is executed using the MAFFT method. It has several different algorithms available, but the one used by PredicTM is the "E-INS-i" method, which is best suited for sequences with multiple aligning segments separated by nonaligning segments, which perfectly describes the situation for GPCRs (see Note 2).

Step 3: *Hydrophobicity Profile*. The third step of PredicTM is the generation of the hydrophobicity profile from the multiple sequence alignment. Because the sequences in the alignment typically have at least some nonaligning portions, there will almost always be gaps in the alignment of the target sequence. PredicTM is only concerned with the portions of the alignment that correspond directly to the target sequence, thus the first step of the profile generation procedure is to remove all portions of the alignment that align to gaps in the target sequence. The result is called a condensed alignment.

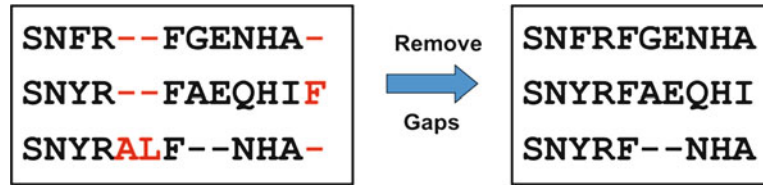

Fig. 1. Example showing the procedure to generate a condensed alignment.

Note that while there are no longer any gaps in the target sequence, there may still be gaps in the aligned sequences. This step is illustrated in Fig. 1.

The next step is to replace the amino acids in the condensed alignment with their corresponding hydrophobicity values. PredicTM uses the Wimley–White whole-residue octanol scale (25), which is a thermodynamic scale derived from transfer of residues from water into *n*-octanol. Unresolved amino acids in the alignment (B, Z, J, X) are replaced with gaps rather than potentially using an incorrect hydrophobicity value. With hydrophobicity values assigned to each residue in the alignment, a raw average hydrophobicity profile is generated. This is done by taking the average of the hydrophobicity values for all nongap residues for each position in the condensed alignment (Fig. 2).

It should be noted that by removing gaps in the alignment to create the condensed alignment and by ignoring gaps when averaging hydrophobicities, PredicTM is a gap-tolerant and gap-unbiased method. The raw average hydrophobicity is too noisy to serve for final prediction of helices. Left panel in Fig. 3 shows a sample raw average hydrophobicity profile (with a flipped sign) taken from a default application of PredicTM to the dopamine D1 receptor. Moving window averages are applied to reduce the noise in the raw average profile. Hydrophobicity averages are calculated using window sizes 7, 9, 11, …, 19, 21 and then averaged over these window sizes to obtain a final profile shown in the right panel of Fig. 3 (see Note 3).

Step 4: *Raw TM Predictions*. The initial prediction of helices based on the final profile is quite simple. Because a thermodynamic scale is being used, hydrophobicity values greater than zero are hydrophobic and thus are part of a TM region, and values less than zero are hydrophilic and are part of the loops or termini. These are referred to as raw TM regions. Typically a rule is applied that eliminates helices less than a certain length, which by default is 10 residues in PredicTM. These eliminations are noted for the user so that they can be visually inspected to ensure that they are not actually part of another helix (see Note 4).

Step 5: *TM region capping*. The raw helices predicted from the hydrophobicity profile in the previous step represent the core of

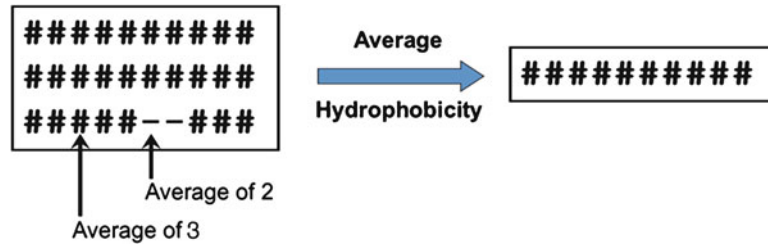

Fig. 2. Scheme to calculate average hydrophobicities along a condensed alignment.

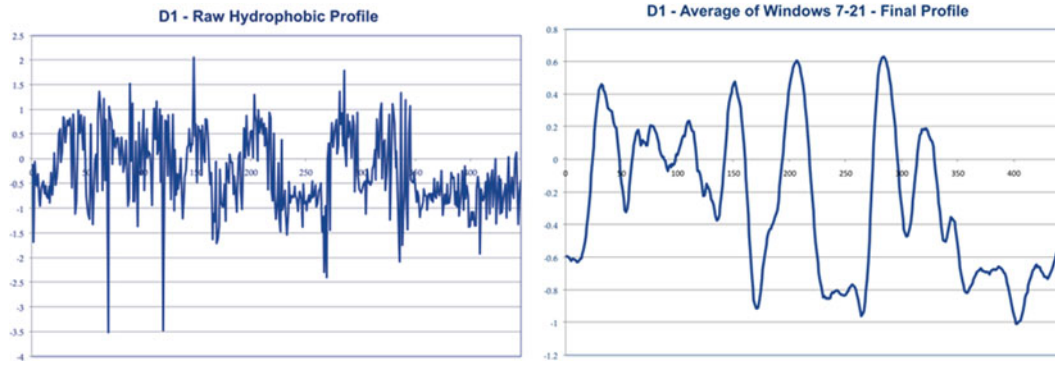

Fig. 3. Raw (*left*) and averaged (*right*) hydrophobic profile for Dopamine D1 receptor.

the TM region and do not account for their extensions beyond the membrane, though many TM helices extend beyond their hydrophobic cores. For an accurate representation of these extensions, we first use secondary structure prediction servers (APSSP2, Porter, PsiPred, Jpred, and SSPro) to obtain helicity predictions for the whole sequence. These predictions are then used to extend the raw TM regions using consensus among the five servers. An example of such predictions is shown in Fig. 4.

Step 6: *Hydrophobic Centers.* Later steps in the ab initio prediction of GPCR structures require the orientation of helices with respect to a plane representing the center of the membrane. This requires that a predefined point on the helix be placed on that plane. Two different hydrophobic centers are calculated for this reason. First is the raw midpoint center, which corresponds to the geometric center of the raw predicted helices. Second is the "area" or "centroid" center, which corresponds to the position in the raw helix where the area under the hydrophobicity profile curve is equal on both sides. This physically corresponds to the buoyant center of the TM helix in the lipid bilayer.

2.2. OptHelix

OptHelix is a program for the optimization of TM helix shapes, particularly with respect to proline kinks. Helices are individually

TMR1	PredicTMRaw	37	FSMLAAYMFLLIMLGFPINFLTLYV	61
	PredicTM2Hel	34	PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQ	64
	CrystalRhod	33	EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH	65
	CrystalOpsin	33	EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH	65
TMR2	PredicTMRaw	73	NYILLNLAVADLFMVFGGFTTTLYT	97
	PredicTM2Hel	71	PLNYILLNLAVADLFMVFGGFTTTLYTSLH	100
	CrystalRhod	70	TPLNYILLNLAVADLFMVFGGFTTTLYTSLHG	101
	CrystalOpsin	70	TPLNYILLNLAVADLFMVFGGFTTTLYTSLHG	101
TMR3	PredicTMRaw	111	NLEGFFATLGGEIALWSLVV	130
	PredicTM2Hel	107	PTGCNLEGFFATLGGEIALWSLVVLAIERYVVV	139
	CrystalRhod	105	FGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC	140
	CrystalOpsin	105	FGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCK	141
TMR4	PredicTMRaw	155	MGVAFTWVMALACAAPPLV	173
	PredicTM2Hel	150	ENHAIMGVAFTWVMALACAAPPLVGW	175
	CrystalRhod	149	GENHAIMGVAFTWVMALACAAPPLV	173
	CrystalOpsin	149	GENHAIMGVAFTWVMALACAAPPLV	173

Fig. 4. Comparison of Raw and Helicity Capped TM region predictions for Rhodopsin against Retinal-bound and Retinal-free crystal forms.

optimized through molecular dynamics (MD) in vacuum using the following steps:

(a) The method begins by identifying the location of prolines in the TM regions that were output in the previous PredicTM step. When located near a terminus, these prolines can have excessively strong bending characteristics in the dynamics simulation. To alleviate this bending and to somewhat mimic the presence of the rest of the protein, alanines are added to the terminus until the proline is 8 residues from the terminus. As an additional option, 8 alanines can be added to the termini of all helices, regardless of proline position, or no alanines can be added.

(b) A canonical helix consisting solely of alanines is produced matching the length of the sequence. Any prolines or glycines are placed in the helix using SCREAM (39). A conjugate gradient minimization, typically down to a 0.5 kcal/mol/Å RMS force threshold, is performed on this structure. At this point, any serines or threonines are placed on the helix using SCREAM. These residues have been shown to interact with the backbone hydrogen bonding network and can influence proline kinks (40). The helix is now ready for molecular dynamics.

(c) Short 10 ps (5,000 steps of 2 fs) dynamics runs are performed at 50 K, 100 K, etc., up to 250 K to warm up the system. Equilibrium dynamics is performed at 300 K following the warm-up and is

typically 2 ns in length (1,000,000 steps of 2 fs) and snapshots are taken at 10 ps intervals.

(d) Two structures are derived from the equilibrium dynamics. First, the lowest potential energy snapshot is selected from the last 1.5 ns of the dynamics. Second, the snapshot with an RMSD closest to that of the average structure during the last 1.5 ns of dynamics is selected.

(e) Finally, the original side chains are replaced using SCREAM, the structure is minimized, and any alanines that were added are removed. Application of OptHelix to the TMs of human β2 and bovine rhodopsin yields the RMSD values shown in Fig. 5. These values are taken from only one MD run. However, all RMSD values are less than the resolution of the crystal structures being compared to.

(f) Both MinRMSD and MinEnergy helix conformations are produced as output for input into the helix bundle sampling procedure.

2.3. BiHelix/ SuperBiHelix

1. *Starting Template*: To pack the seven helices from OptHelix into a bundle, requires the definition of six quantities for each

	Beta2		Bovine Rhodopsin	
	MinRMSD	MinEnergy	MinRMSD	MinEnergy
TM1	0.78	0.86	0.82	0.70
TM2	1.07	1.52	2.02	1.81
TM3	1.14	1.34	0.98	1.18
TM4	0.63	0.64	1.24	0.93
TM5	1.61	1.95	1.62	1.45
TM6	1.55	1.37	1.00	0.79
TM7	1.95	2.05	1.77	1.89
Min	0.63	0.64	0.82	0.70
Max	1.95	2.05	2.02	1.89
Avg	1.25	1.39	1.35	1.25

Fig. 5. CαRMSD of the seven TM helices predicted by OptHelix for β2 and rhodopsin relative to their crystals using the two criteria described in the text.

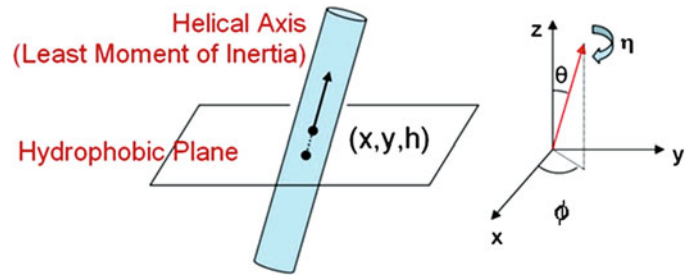

Fig. 6. Coordinates specifying the orientation of a TM helix in a membrane.

helix ($6 \times 7 = 42$ total; see Fig. 6): the x, y, h of some reference point (like hydrophobic center h from PredicTM), the tilt (θ) of each axis from the z axis, the azimuthal orientation (φ) of this tilt; and the rotation (η) of the helix about the helical axis. We choose $z = 0$ for the hydrophobic center (from PredicTM step) for each helix and the choice for the rotation angle η will be described below.

The remaining $4 \times 7 = 28$ coordinates are taken from a template structure. Currently we have multiple choices for this template structure:

(a) The 7.5 electron density map of frog rhodopsin (41).

(b) The bovine rhodopsin structure (4, 14).

(c) The structure for the CCR1 receptor with an antagonist BX 471 bound that was subjected to 10 ns of MD using an infinite membrane and full solvent (36).

(d) The structure for the DP receptor with the CDP2 agonist bound that was subjected to 2 ns of MD using an infinite membrane and full solvent (37).

(e) The structure for the MrgC11 receptor with an agonist FdMRFa bound that was subjected to 7 ns of MD using an infinite membrane and full solvent (42).

(f) Human β2 Adrenergic Receptor (5).

(g) Turkey β1 Adrenergic Receptor (7).

(h) Human A_{2A} Adenosine Receptor (6).

(i) Bovine Opsin (9).

GEnSeMBLE allows for each of these templates to be used in separate predictions, providing an ensemble of bundles among which we can select on the basis of energy or binding energy. As new structures are predicted, validated, and subjected to full MD, they will be added to the ensemble of templates.

To specify an initial rotation (η) of each helix, we use a conserved residue to match the rotation angle of its Cα projection on the $x-y$ plane, to that of the corresponding one in the template structure. The helical axis for rotation is defined as the one corresponding to the least moment of inertia axis obtained using all backbone atoms.

2. *Conformational Sampling of Helices*: Critical to defining the binding sites for ligands is the rotations and tilts of each helix about its axis (θ, ϕ, η), which are determined by the interhelical interactions between the various residues in different helices. The procedure for a complete sampling of all helical rotational combinations is called BiHelix, and even with a 30° increment for each of the seven helices, requires the sampling of $12^7 \sim 35$ million conformations, where the side chains need to be optimized for each helical rotation combination. The procedure for

a complete sampling of all helical tilt and rotation combinations is called SuperBiHelix and requires the sampling of trillions of conformations. This is computationally intractable and actually not necessary as in any given template some helices do not interact with each other. This led to the idea of the BiHelix sampling method described below.

The Sampling Method: In this procedure, first interacting pairs of helices are identified based on the template being considered. The rhodopsin template for Class A GPCRs allows for 12 interacting pairs of helices (H1–H2, H1–H7, H2–H3, H2–H4, H2–H7, H3–H4, H3–H5, H3–H6, H3–H7, H4–H5, H5–H6, H6–H7) as shown with two-way arrows in Fig. 7a. Now, in BiHelix, for each interacting pair of helices, we sample all combinations of a full 360° rotation for each helix with 30° increments leading to $12 \times 12 = 144$ combinations (see Note 5). During this sampling, the other 5 helices are not present, for example shown in Fig. 7b for helix 1–2 pair. In SuperBiHelix, helix tilt angles (θ, ϕ) are also sampled in addition to helix rotation angles η (see Note 6).

For each rotational combination, we optimize the side chains using the rotamer placement method SCREAM (39). SCREAM uses a library of residue conformations ranging from a heavy atom RMS diversity of 0.4A–1.6A in conjunction with a Monte Carlo sampling using full valence, hydrogen bond and electrostatic interactions, but special vdW potentials that reduce somewhat the penalty for contacts that are slightly too short while retaining the normal attractive interactions at full strength. With SCREAM, we find that we can now base the selections on the total energy, without separate considerations of valence, electrostatic, hydrogen bond, and vdw. Because SCREAM does not do energy minimization, it may still lead to vdW interactions slightly too large, hence we minimize each bihelical conformation for ten steps and evaluate the total energy. We take the interhelical and intrahelical components of this energy for the 144 combinations for each helix pair and use them in a pairwise addition equation scheme 1 shown above to obtain energies for

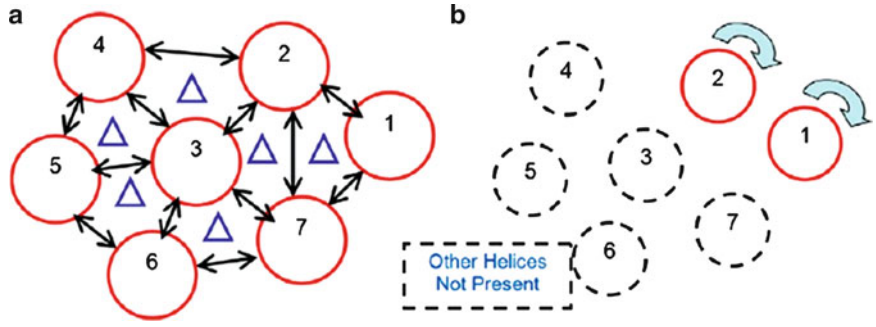

Fig. 7. BiHelix sampling scheme, where sampling is done two helices at a time for all interacting helix pairs. (**a**) All interacting helix pairs shown with a *double arrow*. (**b**) Helix1-Helix2 optimization shown in the absence of other helices.

$$e_{total}^{intra}(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6,\eta_7)=\sum_{i=1}^{7}\frac{1}{N_i}\sum_{j=J_1}^{J_{N_i}}\left[e_{ij}^{i,intra}(\eta_i,\eta_j)\right]$$

$$e_{total}^{inter}(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6,\eta_7)=\sum_{i=1}^{6}\sum_{j=J_1}^{J_{N_i}}\left[e_{ij}^{inter}(\eta_i,\eta_j)\right]$$

$$E_{total}^{bundle}(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6,\eta_7)=e_{tot}^{intra}+e_{tot}^{inter}$$

Scheme 1. Pairwise summation of BiHelix energies to obtain seven-helix bundle energies.

all possible $12^7 \sim 35$ million conformational combinations and output the top 2,000 structures with lowest total energy.

2.4. CombiHelix/ SuperCombiHelix

(a) *Starting Bundles*: The top 2,000 structures coming out of BiHelix/SuperBiHelix analysis are built explicitly using the rotations and/or tilts specified for each helix in the combination. The helical axis for rotation is the same as used in BiHelix analysis.

(b) *BiHelix/CombiHelix Bundle Optimization*: For each of the bundles built in the previous step, the side chains are optimized using SCREAM, like in the BiHelix method. The resulting bundle is minimized for ten steps and total energy reported for each bundle.

For each of the six available GPCR crystal structures, the application of this BiHelix/CombiHelix procedure resulted in the crystal packing being ranked first. These results show that the energies used are reliable enough that, upon complete sampling of the conformational space they can still identify the lowest energy conformation assumed to be the crystal conformation. Since the crystal ligand is absent during sampling (except for Bovine Opsin case which is the ligand-free form of Rhodopsin), these results indicate that the GPCR conformations are in their minimum energy state for both apo- and cocrystal forms at least in the helix rotation angle η space.

First we will look at top ten conformations for Bovine Rhodopsin after the application of BiHelix/CombiHelix procedure using 30° sampling of helix rotation angle. Top 2,000 conformations from BiHelix were taken through CombiHelix. Figure 8 shows the top ten conformations after CombiHelix, along with their energy (SCH-Energy) in kcal/mol, BiHelix rank (SBHRnk), and BiHelix energy (SBH-Energy) also in kcal/mol. Since the starting template was the crystal conformation, that conformation corresponds to all seven η angles being "zero" and according to the Fig. 8 (first row), it is ranked first after conformational sampling using BiHelix/CombiHelix. This is again a validation of the energies being used to rank/score the conformations as well as additional validation of the side-chain placement program SCREAM (39).

SCHRnk	SCH-Energy	SBHRnk	SBH-Energy	Eta1	Eta2	Eta3	Eta4	Eta5	Eta6	Eta7	CRmsd
1	250	1	211	0	0	0	0	0	0	0	0.0
2	252	3	246	0	0	0	0	0	330	0	0.7
3	268	12	267	0	0	0	0	150	0	0	1.6
4	284	4	250	0	0	0	0	60	0	0	0.9
5	285	20	275	0	0	0	0	180	0	0	1.7
6	287	11	267	0	0	0	0	300	0	0	0.9
7	292	335	338	0	0	330	0	0	0	0	0.6
8	293	2	241	0	0	0	0	90	0	0	1.2
9	302	5	253	0	0	0	0	330	0	0	0.5
10	305	7	255	0	0	0	0	0	90	0	1.7

Fig. 8. Top 10 Bovine Rhodopsin conformations after BiHelix/CombiHelix.

Even though only helix rotation angle was sampled during this procedure, the top ten conformations provide signatures for additional conformations that may be important as part of the function of Rhodopsin. First thing that is evident is the high conformational flexibility of TM5 in the helix rotational space, which is not inconsistent with the flexibility associated with TM5 for class A GPCRs. In addition, we see counterclockwise 30° rotations in TM3 and TM6 as shown in Fig. 8 by red cells, which is consistent with the helix rotations observed in the Bovine Opsin (the retinal-free form of Bovine Rhodopsin) crystal structure, −28° rotation for TM3 and −31° rotation for TM6, relative to Bovine Rhodopsin. We do not see a single conformation in the top 2,000 that shows both TM3 and TM6 with this rotation, because these rotations are also associated with changes in helix tilts which were not sampled in this case. The results for Bovine Rhodopsin show that: a. BiHelix/CombiHelix procedure can identify the crystal conformations; b. This procedure has the potential to identify multiple GPCR conformations that may have a physiological role in their function.

Figure 9 shows the ensemble of conformations that result from the application of SuperBiHelix/SuperCombiHelix procedure to adenosine A_{2A} receptor, which sampled helix tilt and rotation angles simultaneously. The sampling grid was: θ (−10, 0, +10), ϕ (−30, −15, 0, +15, +30), and η (−30, −15, 0, +15, +30) (see Note 7). Figure 9a shows the top ten conformations after SuperBiHelix and the conformation observed in the crystal (shaded cells) is ranked second out of ~10 trillion conformations. Figure 9b shows the top ten conformations after SuperCombiHelix on the top 2,000 conformations out of SuperBiHelix. The conformation observed in the crystal (shaded cells) is ranked sixth and all the top conformations are near-native. These results highlight the accuracy of the

a SuperBiHelix

θ H1	H2	H3	H4	H5	H6	H7	ϕ H1	H2	H3	H4	H5	H6	H7	η H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
0	0	0	-10	10	0	0	0	0	15	-30	0	0	0	0	0	0	-30	30	0	0	392.6	1.3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	396.9	0
0	0	0	-10	0	0	0	0	0	15	-30	0	-15	0	0	0	0	-30	-15	0	0	398.3	1
0	0	0	-10	10	0	0	0	0	15	-30	0	0	0	0	0	15	-30	30	0	0	400.6	1.3
0	0	0	-10	0	0	0	0	0	15	-30	0	0	0	0	0	0	-30	0	0	0	400.7	0.9
0	0	0	-10	0	0	0	0	0	15	-30	0	0	0	0	0	0	-30	-15	0	0	401.2	0.9
0	0	0	-10	10	0	0	0	0	15	-30	-15	0	0	0	0	0	-30	15	0	0	401.3	1.3
0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	402.4	0.4
0	0	0	-10	10	0	0	0	-15	15	-30	0	0	0	0	0	0	-30	30	0	0	402.4	1.3
0	0	0	-10	0	0	0	0	0	15	-30	-15	0	0	0	0	0	-30	0	0	0	402.8	1

b SuperComBiHelix

θ H1	H2	H3	H4	H5	H6	H7	ϕ H1	H2	H3	H4	H5	H6	H7	η H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
0	0	0	0	0	0	0	0	0	0	0	-15	0	0	0	0	0	0	0	0	0	59.1	0.4
0	0	0	0	0	0	0	0	0	-15	0	0	0	0	0	0	0	0	0	0	0	71.2	0.4
0	0	0	0	0	0	0	0	0	-15	0	0	15	0	0	0	0	0	0	0	0	74.7	0.6
0	0	0	0	0	0	0	0	0	-15	0	0	-15	0	0	0	0	0	0	0	0	76.3	0.7
0	0	0	0	0	0	0	0	0	0	-15	0	-15	0	0	0	0	0	0	0	0	77	0.6
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	0
0	0	0	0	10	0	0	0	0	0	0	0	-15	0	0	0	0	0	15	0	0	79.1	1
0	0	0	0	0	0	0	0	0	0	0	0	-15	0	0	0	0	0	0	0	0	85.8	0.5
0	0	0	0	0	0	0	0	0	0	0	0	-15	-15	0	0	0	0	0	0	0	86.2	0.6
0	0	0	0	0	0	0	0	0	-15	0	0	15	15	0	0	0	0	0	0	0	87.8	0.6

Fig. 9. Ensembles of top ten conformations for adenosine A_{2A} receptor after SuperBiHelix (**a**) and SuperCombihelix (**b**). Crystal conformation is highlighted by shaded cells.

energies used that allow the SuperBiHelix/SuperCombiHelix procedure to identify near-native conformations even out of ~10 trillion conformations.

The GEnSeMBLE procedure brings together methods for TM helix region prediction, helix shape optimization, and exhaustive but efficient TM helix bundle conformational sampling. This exhaustive sampling predicts an ensemble of low-energy GPCR conformations some of which are expected to play different functional roles in GPCR mediated signaling pathways.

3. Materials

1. The GEnSeMBLE suite of programs consists of the following modules: PredicTM, OptHelix, BiHelix, CombiHelix, SuperBiHelix, and SuperCombiHelix. Some of these modules call many helper routines along with SCREAM for optimal side-chain placement.
2. These programs run on a standard linux cluster of PCs running CentOS (a Red Hat Linux Clone). OptHelix through SuperCombiHelix modules run on multiple processors and

require number of processors ranging from 7 (OptHelix) to 24 (SuperCombiHelix).

3. No programming knowledge of scripting languages like Perl and Python is required to run the programs.

4. Notes

1. *PredicTM*: This step also allows the user to select some of the standard BLAST options. The ExPASy implementation of BLAST typically searches the Swiss-Prot and TrEMBL databases. However, TrEMBL is not curated and can be excluded from the searches, which is the default option for PredicTM. Additionally, the database can be restricted to a specific taxon, such as eukaryota, vertebrata, or mammalia. These options allow the user to restrict results to the most relevant species for a given protein. The E threshold is the statistical measure used by BLAST and can also be adjusted by the user. Larger E thresholds allow more loosely related proteins to appear in the results while smaller E thresholds restrict the results to more closely related proteins. Finally, BLAST allows the user to filter sequences for regions of low complexity. These regions can often result in unrelated proteins, skewing the results. However, because nothing is known about the protein beforehand, the low-complexity filter can also mask relevant portions of the sequence and must thus be used with caution.
2. *PredicTM*: MAFFT is used as the multiple sequence alignment program in PredicTM module. MAFFT has several different algorithms available, but the one used by PredicTM is the "E-INS-i" method, which is best suited for sequences with multiple aligning segments separated by nonaligning segments, which perfectly describes the situation for GPCRs. All GPCRs have seven conserved TM regions (the aligning portions) separated by nonconserved loops (the nonaligning portions). MAFFT correctly aligns all helices of the five dopamine receptors. Additionally, MAFFT does well for cases where non-GPCR sequences are introduced into the alignment. The dopamine D4 receptor has a long IC3 loop that is proline-dense. When performing a BLAST search with low-complexity filtering turned on, this loop is completely masked. When the low-complexity filter is not used, the results for a BLAST search against D4 have a large number of non-GPCR sequences. When aligned with MAFFT, these unrelated sequences typically do not align in the TM regions, thus allowing an accurate prediction to be made despite having a large number of unrelated sequences in the alignment.

3. *PredicTM*: The final step in generating a hydrophobicity profile is to average the moving windows. The default in PredicTM is to average windows 7 through 21, with 7 corresponding to roughly one helical turn above and below a residue and 21 corresponding to roughly the length of one TM region.
4. *PredicTM*: Prediction of TM3 is expected to be a problem for GPCRs because it is primarily an internal helix and is generally less hydrophobic in nature than the other helices. Additionally, TMs 2 and 3 and TMs 6 and 7 are often connected by very short loops, which can result in the hydrophobic peaks merging into one signal.
5. *BiHelix/CombiHelix*: This sampling mode is useful if the target GPCR is not similar to any of the starting templates. In this case, it is recommended to first perform sampling of helix rotation angles η to obtain a structure with local minimum in the η space. This can then be followed by the sampling of all helix tilts and rotations (SuperBiHelix/SuperCombiHelix).
6. *SuperBiHelix/SuperCombiHelix*: This sampling mode is useful if the target GPCR is highly homologous to one of the starting templates. In this case, sampling of helix tilts and rotations (SuperBiHelix/SuperCombiHelix) can be performed simultaneously as the starting structure can be assumed to be close to the minimum in the η rotation angle space.
7. *SuperBiHelix/SuperCombiHelix*: During the SuperBiHelix step, the following sampling grids have been found useful: θ (−10, 0, +10), ϕ (−30, −15, 0, +15, +30), and η (−30, −15, 0, +15, +30). These are derived from the known topological diversity seen among all available templates.

Acknowledgments

Our development of predictive methods for structures and function of GPCRs started ~1998 with an ARO-MURI grant oriented toward olfaction, a time when no 3D structures were available. The early advances were made by Dr. N. Vaidehi (now Professor of Immunology at City of Hope, Duarte, California) and Wely Floriano (now Associate Professor of Chemistry at Lakehead University, Ontario, Canada). Other important early contributions were made by Deqiang Zhang, Spencer Hall, Peter Freddolino, Eun Jung Choi, Georgios Zamanakos, Peter Kekenes-Huskey, Peter Freddolino, Yashar Kalani, Rene Trabanino, Victor Kam, Art Cho, John Wendel, P. Hummel, Peter Spijker; Joyce Yao-chun Peng, Youyong Li, and Jiyoung Heo. More recently the team has been led by Ravinder Abrol, with major contributions from Jenelle Bray, Soo-Kyung Kim, Bartosz Trzaskowski, Adam Griffith, Caitlin Scott, Caglar Tanrikulu, and Andrea Kirkpatrick. The codes for

PredicTM and OptHelix were written by Adam Griffith. The codes for BiHelix/SuperBiHelix were written by Ravinder Abrol and Jenelle Bray with some modules written by Bartosz Trzaskowski. The funding sources that have allowed us to continue this project came from Sanofi Aventis, Boehringer-Ingelheim, Pfizer, Schering AG, and PharmSelex. Also some funding came from NIH.

Note added in proof

Since this chapter was submitted the number of crystallized GPCR structures has more than doubled, which along with new predicted structures provide additional starting templates (introduced in Section 2.3) for the structure prediction of GPCRs. These additional templates from newly crystallized GPCRs correspond to human Histamine H1, human Dopamine D3, human chemokine CXCR4, human Muscarinic Acetylcholine M2, rat Muscarinic Acetylcholine M3, mouse Mu-Opioid, human Kappa-Opioid, mouse Delta-Opioid, human N/OFQ-Opioid, Sphingosine-1-Phosphate 1, bovine meta II rhodopsin, agonist-bound human Adenosine 2A, and Gs-bound human Beta2 Adrenergic receptors. Some of these have been compared in Abrol et al. (2011) Methods 55(4):405–414. Additional predicted templates correspond to human GLP-1, human Somatostatin type 5, human Chemokine CCR5, human Serotonin 2B, human Olfactory 1G1, human Urotensin 2, human Cannabinoid 1 and 2, human bitter taste member 38, and human Adenosine A3 receptors.

References

1. Barki-Harrington L, Rockman HA (2008) B-arrestins: Multifunctional cellular mediators. Physiology 23(1):17–22
2. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on New drug discovery. Pharmacol Rev 62(2):265–304
3. Lagerstrom MC, Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov 7(4):339–357
4. Palczewski K et al (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. Science 289(5480):739–745
5. Cherezov V et al (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 318(5854):1258–1265
6. Jaakola VP et al (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science 322(5905):1211–1217
7. Warne T et al (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. Nature 454(7203):486–491
8. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature 453(7193):363–U33
9. Park JH et al (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature 454(7201):183–187
10. Chien EYT et al (2010) Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. Science 330(6007):1091–1095
11. Wu BL et al (2010) Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. Science 330 (6007):1066–1071

12. Rasmussen SG et al (2011) Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. Nature 469(7329):175–180
13. Rosenbaum DM et al (2011) Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. Nature 469(7329):236–240
14. Okada T et al (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. J Mol Biol 342(2):571–583
15. Scheerer P et al (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature 455(7212):497–U30
16. Blois TM, Bowie JU (2009) G-protein-coupled receptor structures were not built in a day. Protein Sci 18(7):1335–1342
17. Kobilka B, Schertler GFX (2008) New G-protein-coupled receptor crystal structures: insights and limitations. Trends Pharmacol Sci 29(2):79–83
18. Mustafi D, Palczewski K (2009) Topology of class a G protein-coupled receptors: insights gained from crystal structures of rhodopsins, adrenergic and adenosine receptors. Mol Pharmacol 75(1):1–12
19. Rosenbaum DM, Rasmussen SG, Kobilka BK (2009) The structure and function of G-protein-coupled receptors. Nature 459(7245):356–363
20. Hanson MA, Stevens RC (2009) Discovery of New GPCR biology: one receptor structure at a time. Structure 17(1):8–14
21. Worth CL, Kleinau G, Krause G (2009) Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. PLoS One 4(9):e7011
22. Kobilka BK (2007) G protein coupled receptor structure and activation. Biochimica Et Biophysica Acta-Biomembranes 1768(4):794–807
23. Bokoch MP et al (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. Nature 463(7277):108–U121
24. Fleishman SJ, Unger VM, Ben-Tal N (2006) Transmembrane protein structures without X-rays. Trends Biochem Sci 31(2):106–113
25. Wimley WC, Creamer TP, White SH (1996) Experimentally determined hydrophobicity scales for membrane proteins. Biophys J 70(2):Tuam1–Tuam1
26. Hessa T et al (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450(7172):1026–U2
27. White SH (2006) How hydrogen bonds shape membrane protein structure. Peptide Solvation H-Bonds 72:157–172
28. Fanelli F, De Benedetti PG (2005) Computational Modeling approaches to structure-function analysis of G protein-coupled receptors. Chem Rev 105(9):3297–3351
29. Yarov-Yarovoy V, Schonbrun J, Baker D (2006) Multipass membrane protein structure prediction using Rosetta. Proteins-Struct Func Bioinform 62(4):1010–1025
30. Bhattacharya S, Hall SE, Vaidehi N (2008) Agonist-induced conformational changes in bovine rhodopsin: Insight into activation of G-protein-coupled receptors. J Mol Biol 382(2):539–555
31. Bhattacharya S et al (2008) Ligand-stabilized conformational states of human beta(2) adrenergic receptor: Insight into G-protein-coupled receptor activation. Biophys J 94(6): 2027–2042
32. Vaidehi N et al (2002) Prediction of structure and function of G protein-coupled receptors. Proc Natl Acad Sci USA 99(20):12622–12627
33. Kalani MYS et al (2004) The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists. Proc Natl Acad Sci USA 101(11):3815–3820
34. Freddolino PL et al (2004) Predicted 3D structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists. Proc Natl Acad Sci USA 101(9):2736–2741
35. Peng JY et al (2006) The predicted 3D structures of the human M1 muscarinic acetylcholine receptor with agonist or antagonist bound. ChemMedChem 1(8):878–890
36. Vaidehi N et al (2006) Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. J Biol Chem 281(37):27613–27620
37. Li Y et al (2007) Prediction of the 3D structure and dynamics of human DP G-protein coupled receptor bound to an agonist and an antagonist. J Am Chem Soc 129(35):10720–10731
38. Bray JK, Goddard WA 3rd (2008) The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists. J Mol Graph Model 27(1):66–81
39. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theor Comput 4(12):2160–2169
40. Deupi X et al (2004) Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. Biophys J 86(1 Pt 1):105–115
41. Schertler GF (1998) Structure of rhodopsin. Eye (Lond) 12(Pt 3b):504–510
42. Heo J et al (2007) Prediction of the 3-D structure of rat MrgA G protein-coupled receptor and identification of its binding site. J Mol Graph Model 26(4):800–812