



Published in final edited form as:

*Proteins*. 2012 February ; 80(2): 505–518. doi:10.1002/prot.23216.

## BiHelix: Towards *de novo* Structure Prediction of an Ensemble of G-Protein Coupled Receptor Conformations

Ravinder Abrol<sup>1,\*</sup>, Jenelle K. Bray<sup>2</sup>, and William A. Goddard III<sup>1,\*</sup>

<sup>1</sup>Materials and Process Simulation Center (MC 139-74), California Institute of Technology, Pasadena, CA 91125

<sup>2</sup>Simbios, NIH Center for Biomedical Computation, Stanford University, Stanford, CA 94305

### Abstract

G-Protein Coupled Receptors (GPCRs) play a critical role in cellular signal transduction pathways and are prominent therapeutic targets. Recently there has been major progress in obtaining experimental structures for a few GPCRs. Each GPCR, however, exhibits multiple conformations that play a role in their function and we have been developing methods aimed at predicting structures for all these conformations. Analysis of available structures shows that these conformations differ in relative helix tilts and rotations. The essential issue is, determining how to orient each of the 7 helices about its axis since this determines how it interacts with the other 6 helices. Considering all possible helix rotations to ensure that no important packings are overlooked, and using rotation angle increments of 30° about the helical axis would still lead to 12<sup>7</sup> or 35 million possible conformations each with optimal residue positions. We show in this paper how to accomplish this. The fundamental idea is to optimize the interactions between each pair of contacting helices while ignoring the other 5 and then to estimate the energies of all 35 million combinations using these pair-wise interactions. This BiHelix approach dramatically reduces the effort to examine the complete set of conformations and correctly identifies the crystal packing for the experimental structures plus other near-native packings we believe may play an important role in activation. This approach also enables a detailed structural analysis of functionally distinct conformations using helix-helix interaction energy landscapes and should be useful for other helical transmembrane proteins as well.

### Keywords

GPCRs; Helical membrane proteins; Protein structure prediction; Conformational ensemble; Helix-Helix interactions

### 1. Introduction

The G protein-coupled receptors (GPCRs) form the largest superfamily of integral membrane proteins and comprise of seven transmembrane (TM)  $\alpha$ -helices. The GPCRs are activated by a variety of bioactive molecules, like biogenic amines, peptides, lipids,

\*Corresponding authors: Materials and Process Simulation Center (MC 139-74) California Institute of Technology Pasadena, CA 91125 Ph: 1-626-395-8143 (RA); 1-626-395-2731 (WAG) Fax: 1-626-585-0918 abrol@wag.caltech.edu; wag@wag.caltech.edu.

nucleotides, proteins, tastants, odorants and non-molecular sensory signals like light, touch etc, making them essential for a range of physiological processes (e.g. neurotransmission, cellular metabolism, secretion, cell growth, immune defense, and differentiation).<sup>1,2</sup> Besides G proteins, these receptors couple to  $\beta$ -arrestins for signaling.<sup>3</sup> The receptors have also been shown to couple exclusively to  $\beta$ -arrestins<sup>4-6</sup> (in a G protein independent manner) and hence are also referred to as 7TM proteins.

It is well established that the process of GPCR activation<sup>7-11</sup> involves a sequence of conformational changes that enable the cell to convert extracellular signals into physiological responses and is at the heart of their pleiotropic nature as nicely reviewed recently by Kenakin and Miller.<sup>12</sup> GPCR activation also occurs spontaneously in the absence of any signal (called constitutive activity) due to inherent conformational flexibility of these receptors.<sup>13</sup> Kenakin has provided a conceptual framework to think about these signal-induced and signal-independent conformational changes in terms of shifting energy landscapes and transforming ensemble of conformations.<sup>14,15</sup> Strong experimental evidence for this flexibility includes fluorescence lifetime measurements that have identified distinct ligand-induced conformational states for the  $\beta_2$ -Adrenergic Receptor ( $\beta_2$ AR) along the activation pathways<sup>10</sup>, as well as the recently crystallized agonist-bound conformation of  $\beta_2$ AR stabilized by a G protein mimicking nanobody.<sup>16</sup> Single point mutations have also been shown to induce constitutively more-active conformations.<sup>17,18</sup> Thus, to fully gain structural insight into the pleiotropic function of GPCRs will require the functional characterization of their multiple conformations to highlight how different ligands or mutations modulate (stabilize and/or induce) conformations to cause their effects.<sup>12,19-22</sup>

In the last few years there has been a rapid increase in the solution of crystal structures for many GPCRs<sup>16,23-31</sup> due mainly to a technological revolution in membrane protein structure determination methods.<sup>32</sup> The topological comparison of these crystal structures and their implications for GPCR activation has been reviewed extensively<sup>33-37</sup>. All GPCR conformations observed so far have corresponded to the inactive state of the respective receptor, except for bovine opsin<sup>28</sup> and nanobody stabilized human  $\beta_2$ AR bound to an agonist<sup>16</sup>. Structure determination efforts are moving towards the stabilization of GPCRs in different functional conformations (e.g., bound to agonists, G proteins or  $\beta$  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.

Homology based computational methods have had successes with soluble proteins where there may be experimental structures for a variety of similar proteins, particularly where the homology is above 30%,<sup>38</sup> although it is not clear whether such structures are sufficiently accurate for drug design, though they may be accurate enough to be helpful in virtual ligand screening (VLS) efforts<sup>39,40</sup>, in addition to those obtained from de novo methods, which have been successful in VLS efforts as well as in correctly predicting structure-activity relationships<sup>41,42</sup>. The homology methods do not appear to be a viable alternative (yet) for predicting multiple GPCR conformations due to the very limited conformational sampling (if any) that is done to predict structures and will require significant modifications to be able to predict multiple diverse conformations for a GPCR. Homology models may be good for

obtaining structures of very closely related receptors (like dopamine D2 from dopamine D3) and that too only for the same functional form (inactive form or one of many active GPCR forms) as that of the template, but homology modeling doesn't provide a recipe to obtain multiple low-energy conformations, which the BiHelix approach is targeting and which is essential to describe GPCR activation and function.

The conformational flexibility associated with GPCRs can, however, be targeted by structure prediction methods that sample and identify all low energy (physiologically important) conformations for a GPCR. The lipid bilayer environment of the plasma membrane forces specific boundary conditions on these structures and how they fold. In contrast to the almost immeasurable ways of sampling tertiary conformations of soluble proteins, the tertiary structure of GPCRs with seven almost-parallel TM  $\alpha$ -helices allows one to quantify the conformational space available to these TM bundles as will be shown in the next section. Molecular events like the binding of ligands or G proteins or  $\beta$  arrestins to GPCRs as well as point mutations in GPCRs will cause a different sampling of this available space and stabilize different sets of conformations. Methods based mainly on molecular dynamics (MD) are ill-positioned to sample all relevant conformations due to the large dimensionality of the energy landscape and high energetic barriers to access a diverse set of conformations. MD simulations are ideal to test the stability of GPCR structures (predicted or crystal structures) in their local minima, though it takes a long time and lot of effort to establish convergence<sup>43</sup>.

Knowledge-based methods have been developed to predict GPCR structures<sup>44-46</sup> and their ability to predict structures for GPCRs distant from the templates remains to be seen. The Membstruk<sup>47,48</sup> and PREDICT<sup>49</sup> methods tried to sample TM helix bundles as well but do very coarse sampling. The Membstruk method sampled helix rotation angles on a  $5^\circ$  grid but didn't sample anyway near the complete conformational space of the BiHelix approach presented here, because when a helix underwent rotation from  $0^\circ$  to  $360^\circ$ , the other 6 helices were held at  $0^\circ$ . So, the conformational space sampled in Membstruk was  $72 \times 7 \sim 504$  conformations compared to the  $\sim 35$  million or  $\sim 4.5$  billion conformations sampled by the BiHelix approach in this work. The PREDICT method sampled and refined structures obtained from templates/decoys generated by placing the 7 helices as 7 discs on a 2-D plane in different possible topologies. The decoys sampled a highly unrealistic helix packing space as evidenced by the fact that their most stable helix packing for dopamine D2 receptor (Fig. 3 in ref.<sup>49</sup>) shows TM7 in the middle which is at odds with all the GPCR crystal structures observed so far (including the very closely related one for dopamine D3 receptor<sup>29</sup>) all of which show TM3 in the middle.

Monte carlo (MC) type methods have the best chance to sample the conformational space allowed to GPCRs by the lipid-bilayer of the membranes as long as they can ensure a complete sampling of this available conformational space because MC methods cannot guarantee the complete sampling of this space. If we take, as an example, the space of helix rotations in the TM bundles of GPCRs, our previous approaches to predict the structures of GPCRs indicated that one needs to sample rotations of each helix at increments no larger than  $30^\circ$  to identify important helical packings. Since we want to make sure that we sample all possible conformations in this space, this suggests that we should consider  $12^7 \sim$

35,000,000 possible packings (using a 30° grid in the full rotational space from 0° to 360° for all seven helices). Packing such a huge number of conformations into 7-helix bundles, optimizing the side chains for each, and evaluating the energy after removing bad contacts would be an enormous undertaking. This paper introduces the BiHelix method that enables a complete sampling of this huge conformational space at a very modest computational cost. We illustrate this approach with a demonstration of the efficiency of a complete sampling of the helical rotational degree of freedom (DOF) using six of the available GPCR crystal structures. It will also be shown how this sampling can begin to identify other functionally important GPCR conformations and also enable an analysis of the functionally distinct conformations through helix-helix energy potential landscapes.

The efficiency of the BiHelix method will become even more critical when we sample helix tilts (see next section for definitions) in addition to helix rotations, which increases the available conformational space to > 10 trillion conformations. Even though this method has been demonstrated for GPCRs, it is directly applicable to helical membrane proteins in general and also has the potential to assist with some of the challenges associated with the TM helix cores that lie ahead for the prediction of membrane protein structures.<sup>50</sup>

## 2. Materials and Methods

### 2.A Coordinate System

To characterize the conformational space available to GPCRs, we need to define a reference frame (relative to the plasma membrane) and a set of coordinates that uniquely define the orientation of each of the seven  $\alpha$ -helices in the TM bundle. As crystal structures don't provide absolute membrane orientation of GPCRs, we use their orientation as predicted by the OPM (Orientation of Proteins in Membrane) database,<sup>51</sup> which aligns each newly deposited membrane protein structure to an implicit membrane. During this alignment, it optimizes the membrane width as well as tilt of the whole protein relative to the membrane middle to maximize the free energy of protein's insertion into the membrane. The middle of the membrane corresponds to the  $z=0$  plane and will be called the hydrophobic plane.

Each GPCR structure can then be characterized by the six orientation parameters of the seven  $\alpha$ -helices relative to this plane. Figure 1 shows how the helix position and tilt are defined. Helix position ( $R$ ) on the hydrophobic plane is given by  $x$  and  $y$ . The  $h$  value for a helix refers to the hydrophobic center residue of the helix and specifies the helix residue that intersects with the  $z=0$  or hydrophobic plane described above. Two angles,  $\theta$  and  $\phi$  specify the tilt angles of the as yet undefined helical axis and the angle  $\eta$  corresponds to the helix rotation angle about this axis. The two tilt angles ( $\theta, \phi$ ) and the rotation angle ( $\eta$ ) require a unique definition of the helical axis which also account for the reality of bent helices caused mainly by Proline residues in the middle of TM helices. We use a helical axis that corresponds to the least moment of inertia vector for the helix, obtained by diagonalizing the moment of inertia matrix for the helix using only heavy backbone atoms.

In this coordinate system, each GPCR conformation will be represented by 42 numbers (6 orientation parameters defined above for each of the seven helices). In a GPCR structure prediction problem, two scenarios typically arise:

- a. Structure for one conformation is available for a GPCR, usually in the inactive form from solving the crystal structure. This will be an ideal starting template for sampling to characterize the conformational changes possible in this GPCR as part of its function and ligand binding properties. During this sampling, it is likely that only a small subset of orientation parameters (out of 42) may need to be sampled locally around the starting template.
- b. Structure is available for one conformation of a GPCR closely related to the target GPCR. That template conformation, mutated to reflect the residues in the target GPCR, becomes the starting template for sampling to characterize the potential conformational changes in the target GPCR. During this sampling, probably a larger subset of orientation parameters (compared to the previous case) may need to be sampled over a bigger range locally around the starting template.

## 2.B Starting TM Bundle

The BiHelix method, described in the next section, predicts the lowest energy structures in the conformational space being sampled. As mentioned in the previous section, the starting structure for sampling can be obtained either from a known conformation (e.g., from crystal structure) or from a homology structure based on a related protein structure. To demonstrate the BiHelix method, we use 6 of the available GPCR structures: bovine rhodopsin (PDB ID: 1u19)<sup>52</sup>, h $\beta$ 2-AR (2rh1)<sup>24</sup>, hA2AR (3eml)<sup>25</sup>, turkey  $\beta$ 1-AR (2vt4)<sup>26</sup>, bovine opsin (3cap)<sup>28</sup>, and squid rhodopsin (2z73)<sup>27</sup>. The PDB structures aligned to the membrane from the OPM database<sup>51</sup>, are minimized using the Dreiding force-field (FF)<sup>53</sup> and their loop regions are removed leaving the GPCR bundles with only full length TM helices. These TM bundles serve as starting structures for the BiHelix sampling that is to be described next.

## 2.C BiHelix Sampling

The BiHelix sampling method deals with the optimization of the TM helix bundles and implicitly assumes that this optimization is not affected by the loops that connect the TM regions, although the effect of this assumption remains to be seen. This is a reasonable assumption as the loop regions are flexible compared to the TM regions and can potentially adapt to any TM bundle conformation that has strong TM bundle interactions. We anticipate that in the full sampling of all helix orientation angles ( $\theta, \phi, \eta$ ) for the seven helices (conformational space  $\sim 10$  trillion) loop closure condition can potentially eliminate the disallowed TM bundle packings by restricting the sampling space and make it computationally tractable.

For each of the starting templates mentioned in the previous section, we will need to determine the optimum orientation parameters ( $\mathbf{x}, \mathbf{y}, \mathbf{h}, \theta, \phi, \eta$ ) for the packing of each of the 7 helices into a 7-helix bundle. Complete sampling of these 42 parameters is neither practical nor the most efficient. The availability of many starting templates provides a good initial sampling of the  $\mathbf{x}, \mathbf{y}$  space. The parameter  $\mathbf{h}$  or the hydrophobic center can be obtained from the starting template or from the estimation of the buoyant center of a TM helix from a hydrophobic profile (part of a new TM helix prediction algorithm to be covered in a forthcoming publication). So, one would need to sample helix orientation angles ( $\theta, \phi, \eta$ ) for


each of the 7 helices. The BiHelix sampling methodology will now be described and demonstrated using the helix rotation angle ( $\eta$ ) as the DOF being sampled. The sampling method is independent of the number of DOFs being sampled.

Consider a  $30^\circ$  sampling of the  $\eta$  angle (keeping  $\theta$  and  $\phi$  fixed) for each of the 7 helices over full  $360^\circ$ , which leads to  $(12)^7 \sim 35$  million conformations, for each of which the side chains must be optimized. This number reaches trillions of conformations if all three angles ( $\theta$ ,  $\phi$ ,  $\eta$ ) were being sampled. To make such a huge sampling computationally tractable, we devised the BiHelix sampling procedure. A typical class A GPCR template is shown in Figure 2A. It has 12 nearest-neighbor helix pairs shown by double arrows in Figure 2B. Since helix pairs 1-3, 1-4, 1-5, 1-6, 2-5, 2-6, etc don't interact with each other, such pairs of helices don't need to be sampled. In the BiHelix procedure, the seven-helix bundle is split into 12 two-helix bundles comprising of nearest-neighbor helices shown explicitly in Figure 2C, whose conformational space can be sampled completely, during which other five helices are not present.



Using the sampling of rotational angle  $\eta$  as an example, each two-helix bundle will sample  $(12)^2 = 144$  conformations (using a  $30^\circ$  grid over full  $360^\circ$ ). Each of these conformations is built for which the side chains need to be optimized now. We use the SCREAM method<sup>54</sup> that uses an updated DREIDING force field (FF)<sup>53</sup> (referred to as DREIDING-III or D3FF). The energy functions based on D3FF don't include the effects of the nonpolar environment because we have found (data not shown) that accounting of this environment only reorders the protein conformations locally and doesn't affect the overall conformational preference of the structures.

SCREAM uses a library of residue conformations ranging from a CRMS diversity of  $0.4\text{\AA}$  to  $1.6\text{\AA}$  in conjunction with a Monte Carlo sampling using full valence, hydrogen bond and electrostatic interactions from D3FF, but with a special flat bottom van der Waals (vdW) potentials that reduces the penalty for contacts that are slightly too short while retaining the normal attractive interactions at full strength. To benefit from this penalty, the two-helix bundle with SCREAM optimized side-chains is minimized for 10 steps using a conjugate gradient (CG) method. The resulting D3FF energy is accurate enough to correctly select the best cases (without separate considerations of valence, electrostatic, hydrogen bond and vdW terms) as will be seen in the results.

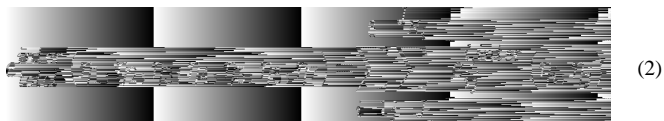
This sampling is done for all 12 nearest-neighbor helix pairs mentioned earlier resulting in  $(12) \times (144) = 1728$  energies, which can be combined to estimate the energy of all possible  $\sim 35$  million  $(12)^7$  conformations in our sampling demonstration as follows. Each of the 1728 energies corresponds to a specific helix pair  $i$ - $j$ , for a specific combination of  $\eta_i$  and  $\eta_j$  and is reported in the form of three energy components:

- a.   $(\eta_i, \eta_j)$ : This component corresponds to interhelical energy of the helix pair that describes the total interaction energy between the helices. It is calculated by subtracting the internal energy of individual helices from the total energy of the two interacting helices and captures sidechain-sidechain, sidechain-backbone, backbone-backbone interactions across the two helices.



- b.   $(\eta_i, \eta_j)$ : This component refers to the intrahelical energy of helix  $i$  while it is interacting with helix  $j$  in the  $i$ - $j$  pair.
- c.   $(\eta_i, \eta_j)$ : This component refers to the intrahelical energy of helix  $j$  while it is interacting with helix  $i$  in the  $i$ - $j$  pair.

The interhelical component of the energy is additive and can be summed to give the total interhelical energy of a seven-helix bundle. The intrahelical component of helix 3, for example, in the 2-3 helix pair will in general be different from that in the 3-4 helix pair. To accommodate this, the intrahelical energy of a helix in the seven-helix bundle is approximated as the average of that energy from all helix-pairs involving that helix. Using this “mean-field” approximation, the energy of the ~35 million conformations for the seven-helix bundle can be estimated. The corresponding equations are shown in **Equations** (1) through (3), where  $N_i$  is the number of helix pairs involving helix  $i$ , and  $J_i$  through  $J_{N_i}$  are the helix partners of helix  $i$  in those  $N_i$  pairs. The ordering of the energies of all ~35 million conformations can be used to select top M (typically 1000 or 2000) seven-helix conformations to be built explicitly and reordered using more accurate energies in the next step (as will be described), because until this point, only two-helix bundles have been optimized in the absence of other five helices.



## 2.D CombiHelix Rotational Sampling

In the CombiHelix step, the top M structures from the BiHelix analysis are built explicitly using the rotations specified for each helix in the combination. The helical axis for rotation is identical to the one used in BiHelix. For each bundle, the side chains are optimized using SCREAM<sup>54</sup> as described in the previous section. This is followed by 10 steps of CG minimization.

The energy function in this CombiHelix stage refers to the energy of the full seven-helix bundle whose residue side-chains were optimized and the resulting structure minimized. The energy function in the BiHelix stage estimated the energy of the seven-helix bundle by combining the energies of the nearest-neighbor two-helix bundles (with optimized side-chains) using **Equations** (1) through (3). So, the energy of a seven-helix bundle at the

CombiHelix stage is expected to be more accurate than at the BiHelix stage. The resulting energies are reordered to provide a new thermodynamic ranking of the M structures taken from BiHelix and the resulting lowest energy conformations can be analyzed in the context of the conformational space that was sampled, (e.g.,  $\eta$  for all seven helices).

In the next section, this BiHelix/CombiHelix procedure will be applied to six of the GPCR crystal templates, to show that the complete sampling scheme works and that the energies are reliable enough to produce near-native GPCR conformations as ones with the lowest energy. Rhodopsin has been crystallized in ligand-free form (Opsin)<sup>28</sup> which has a different TM helix packing. These two systems will be used in the next section to analyze the results of the BiHelix methodology in terms of understanding their functional differences on the basis of structure. The methodology also lends itself to analyzing helix-helix interactions using potential energy landscapes that can provide additional insights into the function of these receptors.

### 3. RESULTS and DISCUSSION

#### 3.A Validation of BiHelix/CombiHelix Sampling and Energy Function

For validation, we have applied this procedure to six of the available GPCR crystal structures: bovine rhodopsin (PDB ID: 1u19)<sup>52</sup>, h $\beta$ 2-AR (2rh1)<sup>24</sup>, hA2AR (3eml)<sup>25</sup>, turkey  $\beta$ 1-AR (2vt4)<sup>26</sup>, bovine opsin (3cap)<sup>28</sup>, squid rhodopsin (2z73)<sup>27</sup>. The TM seven-helix bundles for the application of the computational procedure were obtained as described in section 2.B.

In the first demonstration of BiHelix, we compared energies obtained for three-helix bundles extracted from h $\beta$ 2-AR and hA2AR using the BiHelix approximation and the corresponding exact energies from a full sampling of three-helix (trihelix) conformations. These energies were obtained for all possible rotations of 3 helices at 30° increment ( $12^3=1728$  conformations) for the following six trihelices: H1-H2-H7, H2-H3-H4, H2-H3-H7, H3-H4-H5, H3-H5-H6, and H3-H6-H7. The trihelix energies are compared to those obtained by a BiHelix mean-field approximation utilizing three BiHelix pairs and the results are shown in three supplementary Figures S1, S2, and S3. Figure S1 has results for the h $\beta$ 2-AR system and Figure S2 has results for the hA2AR system. The correlation between bihelix mean-field and trihelix energies holds for most of the trihelix energy range. Big deviations are seen between the two energies only in the very high energy (> 500kcal/mol) conformational space, seen especially in Figure S1.E for h $\beta$ 2-AR and Figures S2.B and S2.D for hA2AR. Figure S3 shows these three cases with a focus on the lowest 1000 kcal/mol range, which shows that the correlation between these two energies holds in the relevant low energy range. Next we will look at the BiHelix energies at the seven-helix bundle level.

In the next demonstration of BiHelix, each of the seven-helix bundles was taken through a complete sampling of the conformational space using helix rotation angle  $\eta$  as a sampling variable. The angle was sampled over full 360° using two grids, a 30° grid and a 15° grid. The 30° grid samples ~35 million conformations and the 15° grid samples  $\sim 4.5 \times 10^9$  conformations. After BiHelix sampling, the top 2000 conformations were taken through CombiHelix to produce a final ranking of lowest energy conformations. The results for all



six systems are summarized in Table 1, which lists the rank of the crystal conformation (XTAL Rank) at the BiHelix level and at the CombiHelix level. It also lists the worst BiHelix ranked structure that shows up in the top 10 conformations after CombiHelix.

As can be seen from Table 1, the crystal conformation is ranked “first” for all cases in 30° sampling and all but one case in 15° sampling. The exception in the 15° sampling mode is for the turkey  $\beta_1$  case, where the crystal conformation is ranked second. 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 predicted low-energy GPCR conformations in the absence of ligands are identical to or close to those observed experimentally at least in the helix rotation angle  $\eta$  space.

Table 1 also shows the worst BiHelix Rank for the top 10 conformations after the CombiHelix step, which is a function of the sampled conformational space as evident by higher worst ranks for the 15 degree sampling. Since CombiHelix energies are more accurate than those at the BiHelix level, this worst BiHelix Rank also provides a qualitative estimate of the accuracy of BiHelix energies for top 10 conformations after CombiHelix. Next, we will look at results for some GPCR systems to understand their implications and specific correlations between BiHelix/CombiHelix energies and C $\alpha$ -RMSD of conformations to the crystal conformation.

### 3.B Conformational and Correlation Analysis

First we will look at the top 10 conformations for Bovine Rhodopsin after the application of BiHelix/CombiHelix procedure using 30° sampling of helix rotation angle. The top 2000 conformations from BiHelix were taken through CombiHelix. Table 2 shows the top 20 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  $\eta$  angles being “zero” and according to the Table 2 (green cells), it is ranked at the top after conformational sampling using BiHelix/CombiHelix. This is a validation of the energies being used to rank/score the conformations as well as additional validation of the side-chain placement program SCREAM.<sup>54</sup> Even though only helix rotation angle was sampled during this procedure, the top 20 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 flexibility associated with TM5 for class A GPCRs. In addition, we see counterclockwise 30° rotations in TM3 and TM6 as shown in Table 2 by red cells, which is consistent with the helix rotations observed in the Bovine Opsin (the retinal-free form of Bovine Rhodopsin) crystal structure, as characterized by the comparison of their templates shown in Table S1 in the supplementary information, which shows the Opsin TM3 and TM6 rotated by -28° and -31° respectively relative to Rhodopsin.

We don't see a single conformation in the top 2000 that shows both TM3 and TM6 with this rotation, because these rotations are also associated with changes in helix tilts (  $\phi$  column in

Table S1) which were not sampled. The results for Bovine Rhodopsin show that: a. BiHelix/CombiHelix procedure can identify the crystal conformation to be the lowest energy conformation; b. The procedure has the potential to identify multiple GPCR conformations that may have a physiological role in their function. These results for other 5 systems are included in the supplementary information (Tables S2 through S6).

Next we will use human  $\beta_2$  adrenergic receptor system to highlight different correlations that are present in the BiHelix/CombiHelix procedure. Figure 3A shows the correlation between BiHelix and CombiHelix energies as raw (red) and smoothed (blue) data (obtained by using a traveling window average of three points:  $i-1$ ,  $i$ , and  $i+1$ , in the y-axis). This correlation is good for conformations ranked up to  $\sim 200$ , as expected due to less noise in the low-energy regions for both BiHelix and CombiHelix energies. Beyond that range of conformations, BiHelix energies flatten out, whereas CombiHelix energies continue to rise.

This is also clear from Figure 3D, which shows BiHelix/CombiHelix energies as a function of conformation rank, with BiHelix energies flattening at higher energies (presumably due to lack of steric clash in BiHelix) and CombiHelix energies rising steeply at high energies. Figure 3B shows the rank correlation between BiHelix and CombiHelix ranks. The raw data (in red) fluctuates a lot but the underlying behavior can be extracted by smoothing the raw data by a traveling window average using a window size of three points ( $i-1$ ,  $i$ , and  $i+1$ ) shown in blue. The rank correlation of Figure 3B qualitatively mimics the energy based correlation in Figure 3A. The correlation between C $\alpha$ -RMSD of predicted conformations (relative to the crystal conformation) versus CombiHelix energy is shown in Figure 3C, as this energy determines the final rank of the conformations. It shows a good correlation and all high energy conformations have a high C $\alpha$ -RMSD as well. This correlation should breakdown at high C $\alpha$ -RMSD values as these can have potentially similar energies. The corresponding results for other 5 GPCR systems are included in the supplementary information (Figures S4 through S8). For predicting multiple GPCR conformations the energy and rank correlations of Figures 3A and 3B provide good signatures to assign confidence to the predictions.

### 3.C Helix-Helix interaction potential energy landscapes from BiHelix

The bihelix energies presented in the previous sections can also be used to look at the potential energy landscapes that describe helix-helix interactions. This is especially useful for the following two related applications: **a.** It can provide snapshots of energy landscapes that these receptors will encounter during the protein folding process, which for all membrane proteins is still not very well understood, except for evidence for a “two-stage” model for  $\alpha$ -helical membrane protein folding<sup>55</sup> which suggests that helix formation takes place before helix-helix associations can take place in the membrane; **b.** It can also provide vital information about the flexibility in helix-helix packings (in terms of the degree of freedom being sampled and accessible conformational space) which when compared across related systems with different functional signatures can begin to provide useful insight into the structural implications of those functional differences. These implications will now be demonstrated using Bovine Rhodopsin and Bovine Opsin as examples bearing in mind that the ultimate fate of these observations is determined by the sum of all BiHelix energies.

Figures 4 and 5 display the potential energy surfaces for all nearest neighbor helix-helix interactions as a function of the rotation angles of the corresponding helices for Bovine Rhodopsin. Panels A through F in Figure 4 show these surfaces for helix pairs 1-2, 1-7, 2-7, 2-3, 2-4, and 3-4. Panels A through F in Figure 5 show these surfaces for helix pairs 3-5, 3-6, 3-7, 4-5, 5-6, and 6-7. These figures only display data for BiHelix conformations that are within 100 kcal/mol of the lowest energy conformation in each case. The empty conformational space in the figures then refers to high energy conformations, which will be assumed to be inaccessible for the current discussions.

The darkest blue regions correspond to the most stable/favorable regions for the involved helix pairs. The center point in each panel refers to the (0°, 0°) conformation corresponding to the crystal structure. For majority of the helix pairs, the native (crystal structure) conformation and near-native conformations fall in the most stable (dark blue) regions. The exceptions appear to be helix pairs 3-5 and 3-6 (Figure 5 panels A and B). Helix pairs 4-5 and 5-6 (Figure 5 panels D and E) have the largest blue area in terms of conformational space in helix rotation angles suggesting a flatter energy landscape as compared to all other helix pairs.

Helix pairs 1-2, 2-3, and 3-4 (Figure 4 panels A, D, F) have the most inaccessible conformational space, which might be due to the necessity to restrict available conformations during the protein-folding process given the fact that these helix pairs involve the first four helices of the 7-helix bundle. This, however, remains to be verified. This observation about large inaccessible conformational space for these helix pairs is also seen in the corresponding helix pairs in most of the other GPCR systems studied in this work, suggesting something unique or unusual about these helix pairs requiring future investigation.

Figures 6 and 7 show the corresponding potential energy surfaces for Bovine Opsin, considered to be the near-active GPCR conformation. Figure 6 contains these surfaces for helix pairs 1-2, 1-7, 2-7, 2-3, 2-4, and 3-4. Figure 7 contains these surfaces for helix pairs 3-5, 3-6, 3-7, 4-5, 5-6, and 6-7.

Here also, for majority of the helix pairs, the native (crystal structure) conformation and near-native conformations fall in the most stable (dark blue) regions. The exceptions appear to be helix pairs 3-7 and 5-6 (Figure 7 panels C and E). Helix pairs 4-5 and 3-6 (Figure 7 panels D and B) have the largest blue area in terms of conformational space in helix rotation angles suggesting a flatter energy landscape as compared to all other helix pairs.

The critical helix-helix contact in Rhodopsin crystal structure appears to be the TM3-TM6 salt bridge which gets broken upon activation and Opsin (one of the near-active conformations downstream of meta Rhodopsin II) displays a TM5-TM6 salt bridge. Energy landscapes corresponding to these helix pairs for Rhodopsin and Opsin can provide some useful insight. It is curious that crystal conformations for TM5-TM6 Opsin BiHelix pair (Figure 7 panel E) and TM3-TM6 Rhodopsin BiHelix pair (Figure 5 panel B) are not in their lowest energy regions (dark blue), so interactions with other helices play a dominant role in making the crystal conformation more stable and attain lowest energy. Since the BiHelix

method implicitly takes into account the effect of non-nearest-neighbor helices when we combine all the two-helix energies to get seven-helix bundle energies, the BiHelix/CombiHelix energies are able to correct this curious observation at the seven-helix bundle level. This is a testament to the strength of the BiHelix/CombiHelix method. In addition, the TM3-TM6 BiHelix pair exhibits a flat energy landscape in Opsin (Figure 7 panel B), consistent with its crystal structure which as described above shows that the TM3-TM6 Rhodopsin contact is broken in Opsin. It is the TM5-TM6 BiHelix pair that exhibits this flat energy landscape for Rhodopsin (Figure 5 panel E) consistent with its crystal structure showing a TM3-TM6 salt-bridge and a broken TM5-TM6 contact. This same BiHelix pair (TM5-TM6) exhibits a restricted landscape for Opsin (Figure 7 panel E), not inconsistent with its crystal structure showing stronger coupling between helices 5 and 6. This doesn't appear to be the case for the analogous case in Rhodopsin (TM3-TM6, Figure 5 panel B), suggesting a stronger involvement of other helices (compared to Opsin) in restricting TM3-TM6 to the crystal conformation. These analyses suggest that connecting the energy landscapes with functional information about different receptors can begin to provide valuable mechanistic insight into the function of these receptors.

Opsin also displays Helix pairs 1-2, 2-3, and 3-4 (Figure 7 panels A, D, F) with the most inaccessible conformational space like Rhodopsin. The supplementary information contains the corresponding potential energy surfaces for other 4 systems studied:  $\beta 2$ ,  $\beta 1$ , A2AR, SquidRhod (see Figures S9 through S16), which also share this observation. There is a wealth of information in all these landscapes which is currently being investigated. So, in addition to structure prediction, the BiHelix methodology can help rationalize functional behavior with structural insights and may even be helpful in protein folding studies of GPCRs.

In terms of computational resources consumed by BiHelix and CombiHelix, the BiHelix sampling of 1728 two-helix conformations (on the  $\eta$  grid of  $30^\circ$ ) is fast and takes about 2 hours on 12 Xeon 2.33GHz processors, whereas the CombiHelix sampling of top 2000 conformations (from the BiHelix step) takes about 16 hours on 12 processors, as this step now involves the optimization of residue side-chains on all seven helices along with minimization.

## 4. CONCLUSIONS

We have presented the BiHelix/CombiHelix methodology aimed towards *de novo* structure prediction of multiple conformations of helical membrane proteins using GPCRs as the example applications. The methodology efficiently performs a *complete sampling* of the available conformational space by breaking the full helix bundle into multiple pairs of interacting helices, whose energies are pooled together to get the energies of all conformations of the full helix bundle. Application of this approach to six of the available GPCRs is able to identify native and near-native conformations as the top conformations by energy. These results validate the sampling method as well as the scoring energy for assessing the best bundle conformations. The demonstrated sampling of helix rotation angles for Bovine Rhodopsin showed that the methodology can identify other physiologically important conformations for a GPCR. We expect that several of the best 10 or 20 predicted

structures from BiHelix/CombiHelix conformational sampling might play a role in differential binding of agonists, antagonists and inverse agonists or in activation pathways. The BiHelix energy landscapes are also capable of providing structural insights into functionally different receptors as demonstrated for Rhodopsin vs Opsin comparisons.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work has been supported in part by NIH grants R01NS071112, R21MH073910, and by gifts to the Materials and Process Simulation Center (MSC) at the California Institute of Technology. We would like to acknowledge fruitful discussions with Soo-Kyung Kim, Bartosz Trzaskowski, and Adam Griffith. Bartosz also assisted with some of the analysis of the CombiHelix output.

## F. REFERENCES

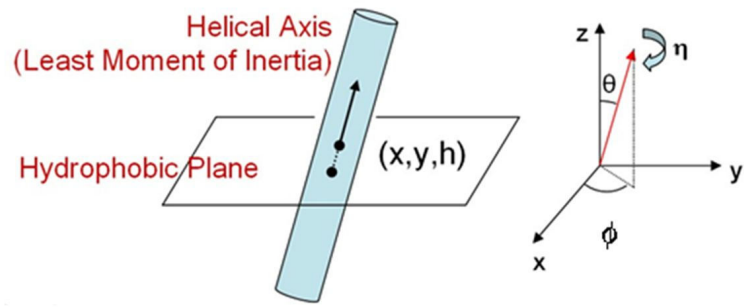
1. Lagerstrom MC, Schiöth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov.* 2008; 7(4):339–357. [PubMed: 18382464]
2. Ji TH, Grossmann M, Ji IH. G protein-coupled receptors I. Diversity of receptor-ligand interactions. *J Biol Chem.* 1998; 273(28):17299–17302.
3. Lefkowitz RJ. G protein-coupled receptors III. New roles for receptor kinases and beta-arrestins in receptor signaling and desensitization. *J Biol Chem.* 1998; 273(30):18677–18680. [PubMed: 9668034]
4. Rajagopal K, Lefkowitz RJ, Rockman HA. When 7 transmembrane receptors are not G protein-coupled receptors. *Journal of Clinical Investigation.* 2005; 115(11):2971–2974. [PubMed: 16276410]
5. Gesty-Palmer D, Chen MY, Reiter E, Ahn S, Nelson CD, Wang ST, Eckhardt AE, Cowan CL, Spurney RF, Luttrell LM, Lefkowitz RJ. Distinct beta-arrestin- and G protein-dependent pathways for parathyroid hormone receptor-stimulated ERK1/2 activation. *J Biol Chem.* 2006; 281(16):10856–10864. [PubMed: 16492667]
6. Kim J, Ahn S, Rajagopal K, Lefkowitz RJ. Independent beta-Arrestin2 and G(q)/Protein Kinase C zeta Pathways for ERK Stimulated by Angiotensin Type 1A Receptors in Vascular Smooth Muscle Cells Converge on Transactivation of the Epidermal Growth Factor Receptor. *J Biol Chem.* 2009; 284(18):11953–11962. [PubMed: 19254952]
7. Gether U, Kobilka BK. G protein-coupled receptors - II. Mechanism of agonist activation. *J Biol Chem.* 1998; 273(29):17979–17982. [PubMed: 9660746]
8. Weis WI, Kobilka BK. Structural insights into G-protein-coupled receptor activation. *Curr Opin Struc Biol.* 2008; 18(6):734–740.
9. Park PSH, Lodowski DT, Palczewski K. Activation of G protein-coupled receptors: Beyond two-state models and tertiary conformational changes. *Annual review of pharmacology and toxicology.* 2008; 48:107–141.
10. Kobilka BK. G protein coupled receptor structure and activation. *Bba-Biomembranes.* 2007; 1768(4):794–807. [PubMed: 17188232]
11. Ahuja S, Smith SO. Multiple Switches in G Protein-Coupled Receptor Activation. *Trends Pharmacol Sci.* 2009; 30(9):494–502. [PubMed: 19732972]
12. Kenakin T, Miller LJ. Seven Transmembrane Receptors as Shapeshifting Proteins: The Impact of Allosteric Modulation and Functional Selectivity on New Drug Discovery. *Pharmacol Rev.* 2010; 62(2):265–304. [PubMed: 20392808]
13. Gether U, Ballesteros JA, Seifert R, SandersBush E, Weinstein H, Kobilka BK. Structural instability of a constitutively active G protein-coupled receptor - Agonist-independent activation due to conformational flexibility. *J Biol Chem.* 1997; 272(5):2587–2590. [PubMed: 9006889]

14. Kenakin T. Ligand-selective receptor conformations revisited: the promise and the problem. *Trends Pharmacol Sci.* 2003; 24(7):346–354. [PubMed: 12871667]
15. Kenakin T. Drug efficacy at G protein-coupled receptors. *Annual review of pharmacology and toxicology.* 2002; 42:349–379.
16. Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature.* 2011; 469(7329):175–180. [PubMed: 21228869]
17. Perez DM, Hwa J, Gaivin R, Graham RM. Constitutive Activation of a Single Effector Pathway - Evidence for Induced Conformational Pleiotropy of G-Protein-Coupled Receptors. *Faseb Journal.* 1995; 9(4):A849–A849.
18. D'Antona AM, Ahn KH, Kendall DA. Mutations of CB1 T210 produce active and inactive receptor forms: correlations with ligand affinity, receptor stability, and cellular localization. *Biochemistry-Us.* 2006; 45(17):5606–5617.
19. Strange PG. G-protein coupled receptors - Conformations and states. *Biochem Pharmacol.* 1999; 58(7):1081–1088. [PubMed: 10484065]
20. Vauquelin G, Van Liefde I. G protein-coupled receptors: a count of 1001 conformations. *Fund Clin Pharmacol.* 2005; 19(1):45–56.
21. Kobilka BK, Deupi X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol Sci.* 2007; 28(8):397–406. [PubMed: 17629961]
22. Hoffmann C, Zurn A, Bunemann M, Lohse MJ. Conformational changes in G-protein-coupled receptors - the quest for functionally selective conformations is open. *Brit J Pharmacol.* 2008; 153:S358–S366. [PubMed: 18059316]
23. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science.* 2000; 289(5480):739–745. [PubMed: 10926528]
24. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science.* 2007; 318(5854):1258–1265. [PubMed: 17962520]
25. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science.* 2008; 322(5905):1211–1217. [PubMed: 18832607]
26. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature.* 2008; 454(7203):486–491. [PubMed: 18594507]
27. Murakami M, Kouyama T. Crystal structure of squid rhodopsin. *Nature.* 2008; 453(7193):363–U333. [PubMed: 18480818]
28. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature.* 2008; 454(7201):183–187. [PubMed: 18563085]
29. Chien EYT, Liu W, Zhao QA, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science.* 2010; 330(6007):1091–1095. [PubMed: 21097933]
30. Wu BL, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science.* 2010; 330(6007):1066–1071. [PubMed: 20929726]
31. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SG, Choi HJ, Devree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK. Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature.* 2011; 469(7329):236–240. [PubMed: 21228876]
32. Blois TM, Bowie JU. G-protein-coupled receptor structures were not built in a day. *Protein Sci.* 2009; 18(7):1335–1342. [PubMed: 19536805]

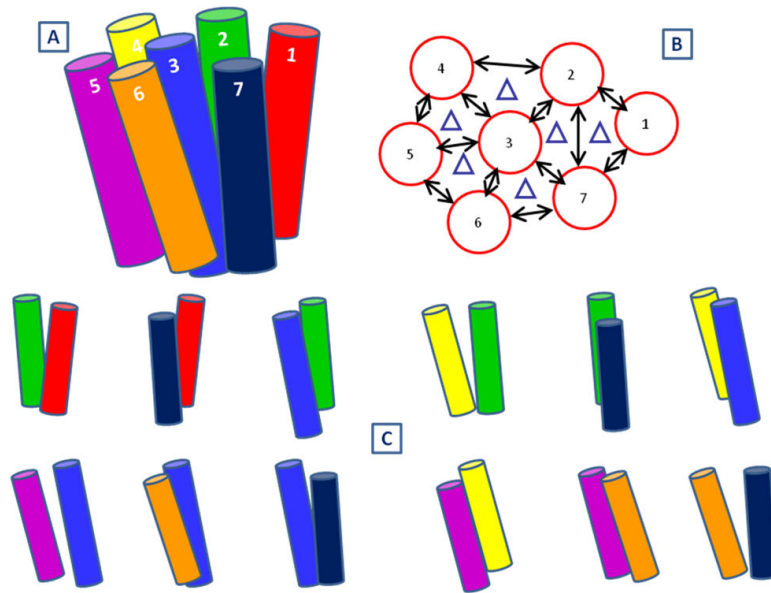


33. Kobilka B, Schertler GFX. New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol Sci.* 2008; 29(2):79–83. [PubMed: 18194818]
34. Mustafi D, Palczewski K. Topology of Class A G Protein-Coupled Receptors: Insights Gained from Crystal Structures of Rhodopsins, Adrenergic and Adenosine Receptors. *Mol Pharmacol.* 2009; 75(1):1–12. [PubMed: 18945819]
35. Rosenbaum DM, Rasmussen SG, Kobilka BK. The structure and function of G-protein-coupled receptors. *Nature.* 2009; 459(7245):356–363. [PubMed: 19458711]
36. Hanson MA, Stevens RC. Discovery of New GPCR Biology: One Receptor Structure at a Time. *Structure.* 2009; 17(1):8–14. [PubMed: 19141277]
37. Worth CL, Kleinau G, Krause G. Comparative Sequence and Structural Analyses of G-Protein-Coupled Receptor Crystal Structures and Implications for Molecular Models. *Plos One.* 2009; 4(9)
38. Schwede T, Sali A, Honig B, Levitt M, Berman HM, Jones D, Brenner SE, Burley SK, Das R, Dokholyan NV, Dunbrack RL, Fidelis K, Fiser A, Godzik A, Huang YJ, Humblet C, Jacobson MP, Joachimiak A, Krystek SR, Kortemme T, Kryshchukovych A, Montelione GT, Moulton J, Murray D, Sanchez R, Sosnick TR, Standley DM, Stouch T, Vajda S, Vasquez M, Westbrook JD, Wilson IA. Outcome of a Workshop on Applications of Protein Models in Biomedical Research. *Structure.* 2009; 17(2):151–159. [PubMed: 19217386]
39. Klebe G, Evers A. Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J Med Chem.* 2004; 47(22):5381–5392. [PubMed: 15481976]
40. Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the Alpha1A adrenergic receptor. *J Med Chem.* 2005; 48(4):1088–1097. [PubMed: 15715476]
41. Li Y, Zhu F, Vaidehi N, Goddard WA 3rd, Sheinerman F, Reiling S, Morize I, Mu L, Harris K, Ardani A, Laoui A. 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.* 2007; 129(35):10720–10731. [PubMed: 17691773]
42. Vaidehi N, Schlyer S, Trabanino RJ, Floriano WB, Abrol R, Sharma S, Kochanny M, Koovakat S, Dunning L, Liang M, Fox JM, de Mendonca FL, Pease JE, Goddard WA 3rd, Horuk R. Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. *J Biol Chem.* 2006; 281(37):27613–27620. [PubMed: 16837468]
43. Grossfield A, Feller SE, Pitman MC. Convergence of molecular dynamics simulations of membrane proteins. *Proteins-Structure Function and Bioinformatics.* 2007; 67(1):31–40.
44. Barth P, Schonbrun J, Baker D. Toward high-resolution prediction and design of transmembrane helical protein structures. *P Natl Acad Sci USA.* 2007; 104(40):15682–15687.
45. Michino M, Chen JH, Stevens RC, Brooks CL. FoldGPCR: Structure prediction protocol for the transmembrane domain of G protein-coupled receptors from class A. *Proteins-Structure Function and Bioinformatics.* 2010; 78(10):2189–2201.
46. Katritch V, Rueda M, Lam PCH, Yeager M, Abagyan R. GPCR 3D homology models for ligand screening: Lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins-Structure Function and Bioinformatics.* 2010; 78(1):197–211.
47. Vaidehi N, Floriano WB, Trabanino R, Hall SE, Freddolino P, Choi EJ, Zamanakos G, Goddard WA 3rd. Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci U S A.* 2002; 99(20):12622–12627. [PubMed: 12351677]
48. Trabanino RJ, Hall SE, Vaidehi N, Floriano WB, Kam VWT, Goddard WA. First principles predictions of the structure and function of G-protein-coupled receptors: Validation for bovine rhodopsin. *Biophys J.* 2004; 86(4):1904–1921. [PubMed: 15041637]
49. 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. PREDICT modeling and in-silico screening for G-protein coupled receptors. *Proteins-Structure Function and Bioinformatics.* 2004; 57(1):51–86.
50. Elofsson A, von Heijne G. Membrane protein structure: Prediction versus reality. *Annu Rev Biochem.* 2007; 76:125–140. [PubMed: 17579561]

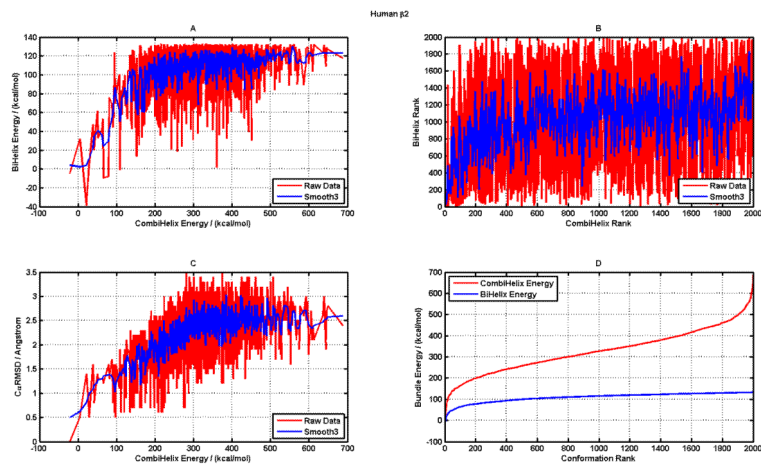
51. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: Orientations of proteins in membranes database. *Bioinformatics*. 2006; 22(5):623–625. [PubMed: 16397007]
52. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *Journal of molecular biology*. 2004; 342(2):571–583. [PubMed: 15327956]
53. Mayo SL, Olafson BD, Goddard WA. Dreiding - a Generic Force-Field for Molecular Simulations. *J Phys Chem-Us*. 1990; 94(26):8897–8909.
54. Kam VWT, Goddard WA. Flat-Bottom Strategy for Improved Accuracy in Protein Side-Chain Placements. *J Chem Theory Comput*. 2008; 4(12):2160–2169.
55. Popot JL, Engelman DM. Membrane-Protein Folding and Oligomerization - the 2-Stage Model. *Biochemistry-Us*. 1990; 29(17):4031–4037.



**Figure 1.**  
Coordinates specifying the orientation of a TM helix in a membrane.

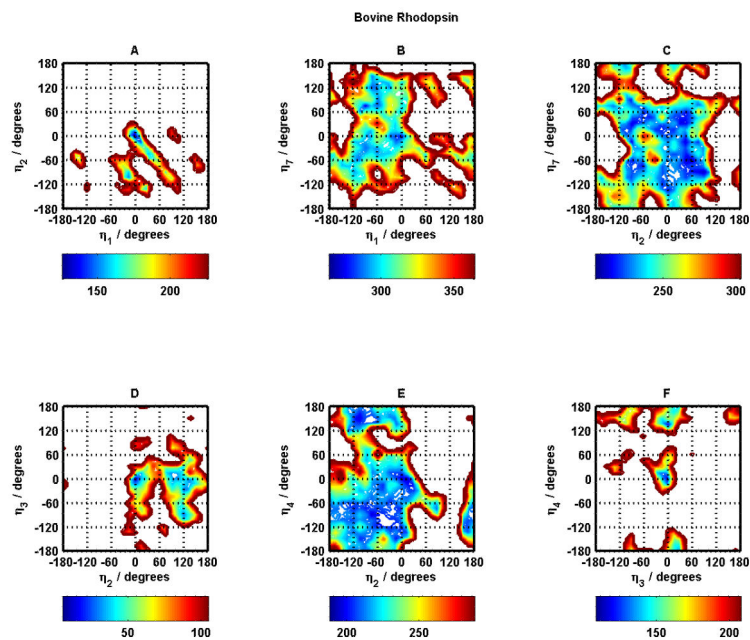


**Figure 2.**  
 A. Typical seven-helix bundle. B. Nearest-neighbor helix pairs highlighted by double arrows; C. The 12 helix pairs shown explicitly.



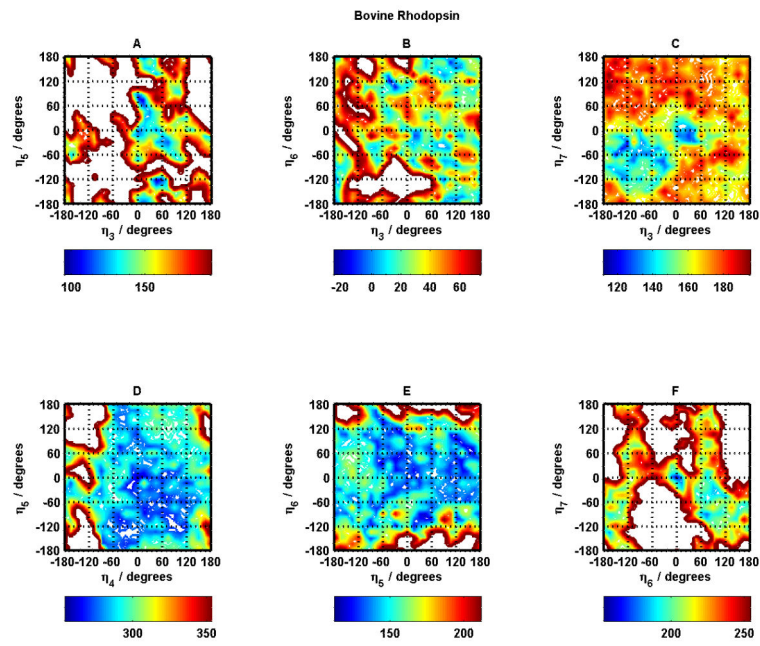
**Figure 3.**

A. Correlation between CombiHelix energies and BiHelix energies of the predicted helix bundle conformations, raw data (in red) and smoothed data using  $\pm 1$  data points (in blue); B. Correlation between CombiHelix Rank and BiHelix Rank of the predicted conformations (colors same as in panel A); C. Correlation between CombiHelix energies and the corresponding Ca-RMSD of those conformations relative to the conformation observed in the crystal structure (PDBid: 2rh1) (colors same as in panel A); D. BiHelix (Blue) and CombiHelix (Red) energies as a function of Conformation Rank.

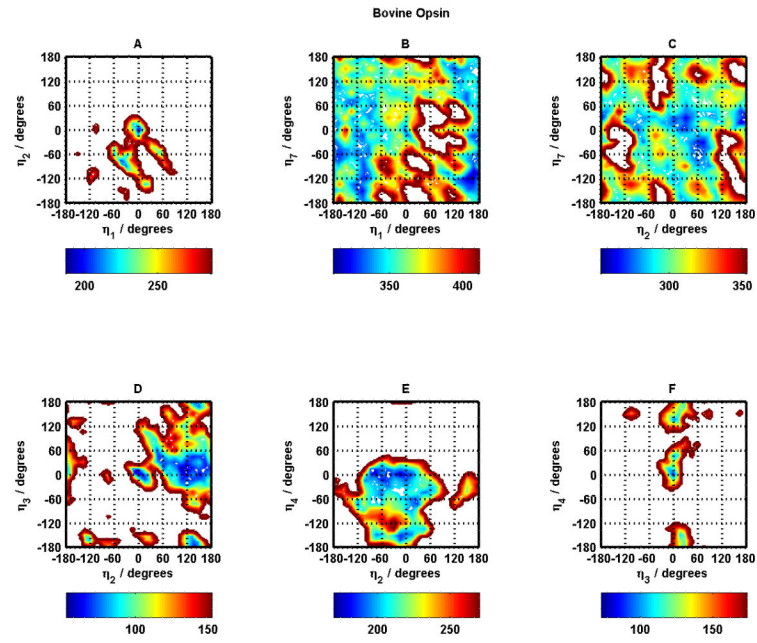


**Figure 4.** Bovine Rhodopsin potential energy surfaces for helix pairs TM1-TM2 (A), TM1-TM7 (B), TM2-TM7 (C), TM2-TM3 (D), TM2-TM4 (E), and TM3-TM4 (F) as a function of helix rotation angles for corresponding helices. The energies are in kcal/mol, where the colorbar under each panel shows the lowest 100 kcal/mol energy range.

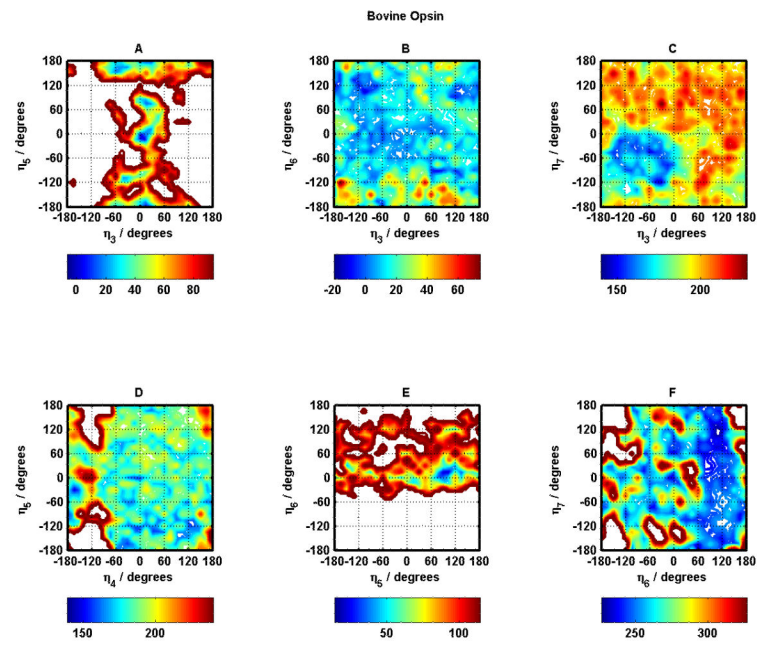




**Figure 5.** Same as Figure 4 for helix pairs TM3-TM5 (A), TM3-TM6 (B), TM3-TM7 (C), TM4-TM5 (D), TM5-TM6 (E), and TM6-TM7 (F).



**Figure 6.**  
Same as Figure 4 for Bovine Opsin.



**Figure 7.**  
Same as Figure 5 for Bovine Opsin.

Table 1

Summary of BiHelix/CombiHelix Results for all tested systems.

GPCR Combi 2000	360° Sampling of $\eta$ Angle on 30° Grid			360° Sampling of $\eta$ Angle on 15° Grid		
	XTAL Rank		Worst BiHelix Rank in Top 10 CombiHelix	XTAL Rank		Worst BiHelix Rank in Top 10 CombiHelix
	BiHelix	CombiHelix		BiHelix	CombiHelix	
A2A	1	1	88	1	1	789
Beta2	4	1	96	19	1	636
Beta1	2	1	82	6	2	1065
BovRhod	1	1	335	1	1	1278
BovOpsin	1	1	131	1	1	533
SquRhod	1	1	79	1	1	363

Table 2

Top 20 Bovine Rhodopsin conformations after BiHelix/CombiHelix. The conformation observed in the bovine rhodopsin crystal structure (1u19) is shown in green cells. The helix rotation angles for TM3 and TM6 as observed in opsin crystal structure (2z73) are shown in pink cells.

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
11	306	22	275	0	0	0	0	0	300	0	1.2
12	318	71	302	0	0	0	0	180	330	0	1.8
13	320	720	358	0	0	330	30	0	0	0	0.7
14	320	8	262	0	0	0	0	0	270	0	1.7
15	324	37	286	0	0	0	0	210	0	0	1.6
16	325	134	314	0	0	0	0	180	90	0	2.4
17	329	10	264	0	0	0	0	0	0	330	0.5
18	332	13	267	0	0	0	0	0	0	30	0.5
19	332	16	272	0	0	0	0	330	330	0	0.8
20	333	30	282	0	0	0	0	90	90	0	2.0