Supporting Information

Bray et al. 10.1073/pnas.1321233111

SI Text

BiHelix Energy Calculations

We use the BiHelix energy calculation method as described in ref. 1. The BiHelix sampling is done for all 12 nearest-neighbor helix pairs, resulting in (12) × (144) = 1,728 energies, which can be combined to estimate the energy of all possible $12^7 \sim 35$ million conformations in our sampling demonstration as follows. Each of the 1,728 energies corresponds to a specific helix pair *i-j*, for a specific combination of η_i and η_j, and is reported in the form of three energy components:

i) $e_{ij}^{inter}(\eta_i, \eta_j)$: 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 side chain–side chain, side chain–backbone, and backbone–backbone interactions across the two helices.

ii) $e_{ij}^{i,intra}(\eta_i, \eta_j)$: This component refers to the intrahelical energy of helix *i* while it is interacting with helix *j* in the *i-j* pair.

iii) $e_{ij}^{j,intra}(\eta_i, \eta_j)$: 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 Eqs. **S1–S3**, where N_i is the number of helix pairs involving helix *i*, and J_1 through J_{Ni} are the helix partners of helix *i* in those N_i pairs:

$$e_{total}^{intra}(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6,\eta_7) = \sum_{i=1}^{7} \frac{1}{N_i} \sum_{j=J_1}^{J_{N_i}} \left[e_{ij}^{intra}\left(\eta_i,\eta_j\right)\right], \quad \textbf{[S1]}$$

$$e_{total}^{inter}(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6,\eta_7) = \sum_{i=1}^{6} \sum_{j=J_1}^{J_{N_i}} \left[e_{ij}^{inter}\left(\eta_i,\eta_j\right)\right], \quad \textbf{[S2]}$$

$$E(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6,\eta_7) = e_{total}^{intra} + e_{total}^{inter}. \quad \textbf{[S3]}$$

Binding Site Prediction Methodology

For the ligand-docking step, we use techniques developed as part of the GenDock and DarwinDock procedure. First, each structure was prepared for docking by mutating all of the bulky, nonpolar residues (phenylalanine, isoleucine, leucine, methionine, tyrosine, valine, and tryptophan) to alanine, a procedure we call alanization. Then a sphere set representing the binding region was formed by taking a 2.0-Å radius around the coordinates of the ligand bound to the template. The ligands were assigned Mulliken charges from quantum mechanics (B3LYP flavor of DFT using the 6–31G** basis set, calculated with Jaguar). Then GenDock was used to dock the ligand with the DarwinDock method.

DarwinDock generates a large number of poses (no energy calls) using Dock6, and clusters them into families by seeding the families with pose pairs closest to each other in heavy-atom rmsd and expanding the families or merging them as long as each family member pose is within 2.0-Å rmsd of all other family members. The program then adds 5,000 new ligand poses (again using Dock6). If the fraction of new families is less than 1/20, then we consider that completeness is achieved. Generally, this leads to 30,000–50,000 poses partitioned into 2,000–5,000 families. The family head pose is defined as the centroid pose in each family (based on smallest average rmsd to all other poses in the family); for rare families of just two poses, the family head is picked randomly. DarwinDock then scores these family heads and chooses the best 10% family heads and their family members for further analysis. Then, these remaining families are scored completely. Finally, the top 100 structures are passed on to the next step.

For each of these 100 ligand configurations, the bulky nonpolar residues (that went alanization) are restored, and their side chains along with others in a 4-Å unified binding site (defined as the union of all residues within a 4-Å radius of the docked ligand in all of the complexes) are optimized with SCREAM (2). This procedure leads to a different set of optimum binding-site residue side-chain positions for each of the 100 poses. The complexes are ranked by total energy, and the best 50% are kept for neutralization.

During the neutralization step, which uses the method described in ref. 3, we neutralize all charged residues of the system and the ligand by transferring the hydrogen of each salt bridge from the acceptor back to the donor and by adding a proton to each exposed Asp or Glu and removing one from each Lys and Arg. We use a modified Dreiding force field that included special hydrogen-bonding parameters chosen to reproduce the binding for dimers of analogous residues found from quantum mechanics. The neutral residue scheme is an improvement over the charged residue scheme for the binding-energy calculations because it decreases the large variations between complexes caused by exaggerated long-range coulombic interactions between charged groups. These exaggerated interactions are due to the fact that the charges are fixed, so that the charge screening present in the experimental system is not present to damp the long-range interactions. The neutralization procedure is carried out for the binding-energy calculations and not the docking procedure because the large coulombic interactions are important to ensure that binding modes with a salt bridge are selected.

After the neutralization, the complexes are ranked by total energy, and the best 50% are kept for minimization. The 4.0-Å unified binding site is then minimized for 50 steps. The complexes are again ranked by total energy, and the best 50% are kept for the final step, resulting in 13 final structures. Then, for each these 13 structures, we minimize the full complex for 500 steps (or to the RMS force is 0.25 kcal/mol·Å. Finally, these 13 structures are put through a quench-anneal cycle (50 K to 600 K and back over 11.5 ps) using the nonneutralized model, selecting the configuration with the lowest potential energy structure. Then the structures are reneutralized and reminimized.

The 13 final docked structures for each structure were all compared with the ligand-bound crystal structure. The similarity to the ligand-bound crystal structure was measured with the contact rmsd. This is calculated by first determining the closest contacts for the ligand on each residue for the crystal structure, and finding their contact distances. Then the same contact distances are determined for the predicted ligand-bound complex. Finally, the rmsd is calculated between the two sets of distances.

The Effect of SuperBiHelix on Binding Site Predictions

For carazolol docked into the ligand-free β_2-adrenergic crystal structure, the contact rmsd is 2.4 Å. For the β_2-adrenergic helices in the A_{2A} adenosine template before SuperBiHelix, the lowest contact rmsd in the final 13 docked structures is 4.5 Å, and after SuperComBiHelix it is 4.4 Å. These bound structures are seen in Fig. S1. Although there is very little improvement in the contact rmsd from SuperBiHelix, inspection of the bound structures shows marked improvement. In the crystal structure, carazolol forms strong hydrogen bonds with D113(TM3) and N312(TM7). In the structure before SuperBiHelix, carazolol only has a hydrogen bond with D113(TM3). However, after SuperComBiHelix, carazolol makes strong hydrogen bonds with both D113(TM3) and N312(TM7). So, SuperBiHelix does make the binding site more similar to the crystal structure.

Predicting the Inactive Form of Rhodopsin Starting with the Active Form

For a second case to validate how well SuperBiHelix can recognize active and inactive forms of the same receptor, we placed the helices from rhodopsin [Protein Data Bank (PDB) ID 1u19] in the opsin (PDB ID 3cap) template, which we expect to have a low-energy conformation that is active when opsin helices are used, but which would have a low-energy state that is inactive when rhodopsin helices are used. The reasoning for this is as follows. Opsin is the ligand-free active-like conformation of Rhodopsin and it was crystallized without the G protein. The helix shapes have adapted to the active-like state of Opsin in the absence of any other cocrystal component. Because we do not perform helix shape optimization during SuperBiHelix, for these helix shapes the Opsin conformation is expected to have a lower energy than the Rhodopsin conformation, due to better interhelical energies.

We carried out SuperBiHelix, sampling θ with values of −10°, 0°, and 10°; ϕ with values of −45°, −30°, −15°, 0°, 15°, 30°, and 45°; and η with values of −30°, −15°, 0°, 15°, and 30°. Indeed, the lowest energy conformation from SuperCombiHelix is within 0.6-Å rmsd of the X-ray rhodopsin structure. In fact, the only degree of freedom off from the rhodopsin template is the sweep angle of helix 4, which is off by 30°. The results are shown in Fig. S2.

CB1 Binding Experimental Methods

The level of [^{35}S]GTPγS binding to CB1 was measured in the absence of ligand for the wild-type, T3.46I, T3.46A, T3.46/R2.37A, and T3.46/R2.37Q receptors. The level of [^{35}S]GTPγS binding for the mock-transfected membrane sample is shown for comparison. Data are presented as specific binding of GTPγS to the membrane preparation. Nonspecific binding was determined in the presence of 10 μM unlabeled GTPγS. Each data point represents the mean ± SEM of at least three independent experiments performed in duplicate. The dashed line indicates the level of non-CB1–mediated GTPγS binding obtained from [^{35}S]GTPγS binding to the mock-transfected membrane sample.

1. Abrol R, Bray JK, Goddard WA, 3rd (2011) Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. *Proteins* 80(2):505–518.
2. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. *J Chem Theory Comput* 4(12):2160–2169.
3. 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.

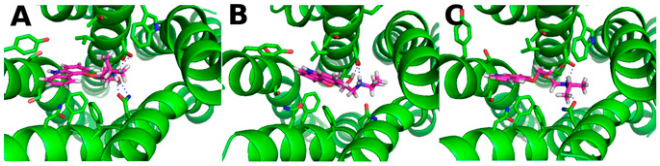

Fig. S1. (*A*) The carazolol-bound β_2-adrenergic crystal structure. (*B*) Carazolol docked into the β_2-adrenergic helices in the A_{2A} adenosine template before SuperBiHelix. (*C*) Carazolol docked into the β_2-adrenergic helices in the A_{2A} adenosine template after SuperBiHelix.

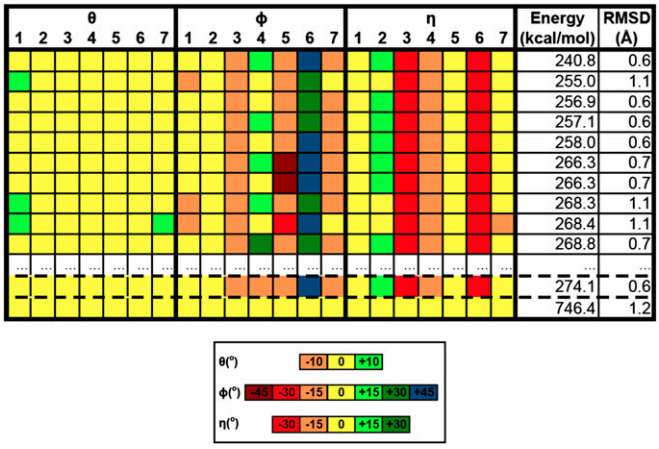

Fig. S2. SuperComBiHelix results for rhodopsin helices in the opsin template. The rmsd is the backbone rmsd to the rhodopsin crystal structure. The structure outlined by dashed lines is the structure closest to the crystal structure, given the angles sampled. The structure in all yellow (representing all 0° angles) is the original structure in the incorrect template.

Table S1. The ranking of the crystal structure conformation for each helix after the QuadHelix protocol

TM	1	2	3	4	5	6	7
β_2	1	1	1	2	3	12	1
A_{2A}	1	1	4	7	4	2	1

Results are for both the A_{2A} adenosine and the β_2-adrenergic receptors.

Table S2. The differences between the A_{2A} adenosine receptor and β_2-adrenergic receptor templates

TM	*x*, Å	*y*, Å	θ, °	ϕ, °	η, °
1	2.0	1.7	6.3	17.9	4.8
2	0.2	0.0	7.7	2.4	−18.8
3	0.0	0.0	4.9	2.0	−13.2
4	−1.5	−0.4	−4.1	−3.0	−9.2
5	0.1	−0.6	−1.2	14.9	0.0
6	0.8	0.0	10.8	−24.8	0.9
7	−0.1	1.1	4.1	3.1	2.6

The system of coordinates is described in Fig. 2 in the manuscript.