Journal of Structural Biology 170 (2010) 10–20

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure

William A. Goddard III *, Soo-Kyung Kim [1], Youyong Li [1], Bartosz Trzaskowski, Adam R. Griffith, Ravinder Abrol

Materials and Process Simulation Center (MC139-74), California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, United States

ARTICLE INFO

Article history:
Received 10 August 2009
Received in revised form 1 January 2010
Accepted 4 January 2010
Available online 15 January 2010

Keywords:
GPCRs
Adenosine
Structure prediction
Membrane protein structure
Ligand binding
Docking
Modeling
Receptor

ABSTRACT

G protein-coupled receptors (GPCRs) are therapeutic targets for many diseases, but progress in developing active and selective therapeutics has been severely hampered by the difficulty in obtaining accurate structures. We have been developing methods for predicting the structures for GPCR ligand complexes, but validation has been hampered by a lack of experimental structures with which to compare our predictions. We report here the predicted structures of the human adenosine GPCR subtypes (A_1, A_{2A}, A_{2B}, and A_3) and the binding sites for adenosine agonist and eight antagonists to this predicted structure, making no use of structural data, and compare with recent experimental crystal structure for ZM241385 bound human A_{2A} receptor. The predicted structure correctly identifies 9 of the 12 crystal binding site residues. Moreover, the predicted binding energies of eight antagonists to the predicted structure of A_{2A} correlate quite well with experiment. These excellent predictions resulted when we used Monte Carlo techniques to optimize the loop structures, particularly the cysteine linkages. Ignoring these linkages led to a much worse predicted binding site (identifying only 3 of the 12 important residues).

These results indicate that computational methods can predict the three-dimensional structure of GPCR membrane proteins sufficiently accurately for use in designing subtype selective ligands for important GPCR therapeutics targets.

1. Introduction

G protein-coupled receptors (GPCRs) modulate regulation of many essential physiological processes involved in cardiovascular, metabolic, neurodegenerative, psychiatric, cancer and infectious diseases (Lundstrom, 2006; Tang and Insel, 2005). They represent 30–50% of the current drug targets (Hopkins and Groom, 2002; Lundstrom, 2006), but a major impediment to developing active *and* selective therapeutics is the lack of structural data. Thus, of ~800 human GPCRs, experimental crystal structures are available only for two [β_2 Adrenergic Receptor ($h\beta_2AR$) (Cherezov et al., 2007) and adenosine A_{2A} receptor ($hAA_{2A}R$) (Jaakola et al., 2008)]. Moreover, these experimental structures have a bound inverse agonist or antagonist, providing little information about the mechanism of activation.

Development of active subtype selective ligands would be greatly aided if *in silico* computational modeling could provide sufficiently accurate structures and binding constants *for use in the development of new drugs*. We have been developing *in silico* methods for predicting the 3D structures of GPCRs and the binding sites for agonists and antagonists (Floriano et al., 2000; Vaidehi et al., 2002; Freddolino et al., 2004; Kalani et al., 2004; Trabanino et al., 2004; Peng et al., 2006), but with very little opportunity to obtain direct confirmation from experiment of predictions made in advance of the experiment (Vaidehi et al., 2006; Heo et al., 2007).

Stimulated by the Critical Assessment of GPCR Structure Modeling and Docking (CAGSMD) challenge (Michino et al., 2009), we applied our methods to predict the structure of $hAA_{2A}R$ with bound antagonist ZM241385 prior to publication of the crystal structure results (Jaakola et al., 2008). We report here our methods and results, showing how we assessed the candidate structures for submission to CAGSMD. From comparison to experiment, we concluded that it is essential to predict accurate extracellular loops (EL) to obtain an accurate ligand binding site. $hAA_{2A}R$ has eight cysteine residues distributed between the EL1, EL2, and EL3 extracellular loops and all are oxidized in the crystal structures. We find that optimizing these loops using our general Monte Carlo methods but not using any crystal structure information leads to a dramatic effect on the binding position of the ZM241385, reducing the error in the predicted ligand position from 5.6 Å to 2.8 Å RMSD relative to the crystal structure. The predicted structure correctly identifies 9 out of 12 crystal binding site residues (Table 1).

* Corresponding author. Fax: +1 626 585 0918.
E-mail address: wag@wag.caltech.edu (W.A. Goddard).
[1] These authors contributed equally to this work.

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.01.001

Table 1
Contribution from each residue to the energy change upon binding (kcal/mol) of ZM241385 to human A_{2A} adenosine receptor. The contributions are ordered by the contributions from the Exper* X-ray structure with the side-chains for H250 and M270 optimized by SCREAM. (The results for the original structure are listed under Exper.) The theory and Exper* agree on three of the four most strongly interacting residues (>3 kcal/mol) and on 9 of the 12 residues binding more strongly than 1 kcal/mol. Color coding for contributions of each residue to binding of the ZM241385 ligand: dark blue: >3 kcal/mol, blue: 1–3 kcal/mol, light blue: 0.5–1.0 kcal/mol, yellow: very repulsive by >3 kcal/mol, green: slightly repulsive by 0.01–3 kcal/mol, white: attractive by 0.01–0.5 kcal/mol. Note that the Exper and Exper* data includes explicit water, whereas the predicted structure does not, resulting in a more stable predicted cavity energy. (For interpretation of the references to color in this table caption, the reader is referred to the web version of this article.)

RES	#	Exper		Exper*		Predicted	
		Non-bond	Contact	Non-bond	Contact	Non-bond	Contact
ASN	253	−6.83	3.00	−8.57	2.76	−3.24	3.00
GLU	169	−4.79	3.42	−6.78	2.81	−3.63	3.42
PHE	168	−3.61	3.23	−6.76	3.51	−8.03	3.48
LEU	249	−3.55	3.46	−4.23	3.56	−0.58	4.01
WAT	772	3.44	2.47	−2.97	2.80	WAT	WAT
WAT	773	−4.27	3.08	−2.59	2.82	WAT	WAT
MET	177	−148	3.02	−2.30	3.17	−0.01	11.20
ILE	274	−2.26	3.94	−2.26	3.89	−0.77	3.55
LEU	267	−1.81	3.85	−2.04	3.61	−1.57	3.41
WAT	775	−3.33	3.37	−1.97	2.78	WAT	WAT
TYR	271	−1.54	4.63	−1.80	3.99	−2.34	3.41
HIS	250	2.48	3.40	−1.38	3.50	−1.43	3.84
LEU	58	−1.54	3.71	−1.31	3.42	−0.53	5.93
MET	270	263.34	3.12	−1.25	3.81	−4.86	3.55
HIS	264	0.55	3.34	−1. 09	3.72	−2.05	3.71
TRP	246	0.29	3.36	−0.99	3.54	−0.32	6.19
ASN	181	−0.86	4.40	−0.91	4.21	−0.05	10.56
LEU	167	−0.63	5.53	−0.83	4.91	−3.31	3.26
VAL	84	−0.70	5.04	−0.70	4.93	−0.40	12.75
WAT	777	1.38	3.06	−0.56	3.20	WAT	WAT
ILE	66	−0.61	5.83	−0.55	5.93	−2.10	3.99
WAT	776	−0.44	3.79	−0.40	3.79	WAT	WAT
ILE	252	−0.38	6.30	−0.38	5.97	−1.49	6.56
MET	174	−0.33	5.02	−0.35	4.99	−0.03	13.07
THR	88	−0.29	5.41	−0.34	5.23	−0.19	7.43
PHE	182	−0.20	6.75	−0.24	6.56	−0.68	4.88
SER	67	−0.18	7.07	−0.23	6.87	−0.23	6.73
VAL	186	−0.07	4.79	−0.23	5.09	−0.23	7.32
THR	256	−0.16	6.51	−0.20	6.09	−0.45	6.73
ALA	265	−0.29	6.76	−0.09	6.61	−0.18	7.00
WAT	771	−0.50	3.68	−0.07	3.86	WAT	WAT
WAT	774	0.28	3.65	−0.01	3.70	WAT	WAT
SUM		231.12		−54.35		−38.70	

We also report the predicted binding site and energies for eight antagonists (structures shown in Table 2), finding relative affinities that correlate well with experiment. In addition, we predicted structures for the human A_1, A_{2B}, and A_3 adenosine receptors (ARs) and used these structures to predict subtype selectivity of the ZM241385 antagonist to all four adenosine receptors.

2. Results

The methods used for obtaining structures (TM regions and loops) submitted in the CAGSMD challenge are described in detail in Section 4. The only change from our original procedure is that we now assume that all eight Cys in the extracellular loops (EL) are oxidized (as found in the crystal structure) (Jaakola et al., 2008), rather than reduced as in our original predictions. No other information was used from experiment. Here we discuss the details only for the best (lowest total energy) predicted protein structure (including oxidized Cys in the loops) and the best (lowest total energy) predicted ligand docked structure [using HierDock (Floriano et al., 2000; Vaidehi et al., 2002)].

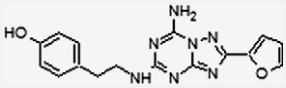
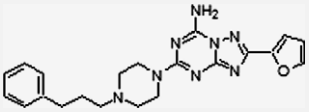
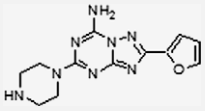
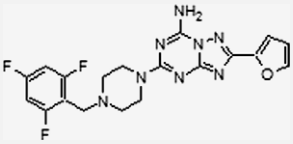
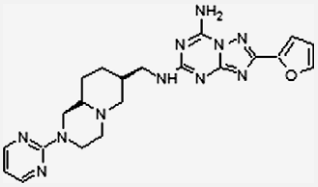
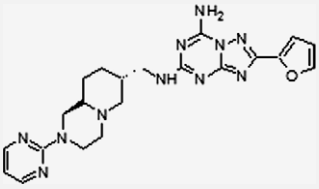
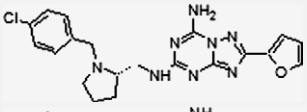
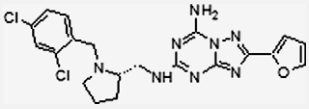
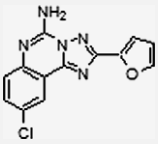
For the predictions of subtype selectivity, we matched the predicted best binding pose of the ligand in $hAA_{2A}R$ structure to our predicted apo-protein structures for the other three subtypes, then we used SCREAM (Kam and Goddard, 2008) to predict the optimum side-chain position of residues in the binding pocket, and then we minimized the energy to obtain the final ligand/protein complexes.

2.1. Antagonist ZM241385 bound to $hAA_{2A}R$

To analyze the predicted binding site for ZM241385/$hAA_{2A}R$, we calculated the interaction energy between the atoms of each residue with all atoms of the ligand (called the cavity analysis) as shown in Table 1 and compared with experiment.

The experimental crystal structure (Jaakola et al., 2008) (denoted Exper) has the C_γ heavy atom of M270 only 3.1 Å from the C6 atom of the phenoxy ring in ligand, leading to a very repulsive van der Waals (vdW) interaction and a negative (repulsive) contribution to the binding energy of 263 kcal/mol. Using our SCREAM method (Kam and Goddard, 2008), we found a better side-chain conformation for M270 with a closest distance of 3.8 Å and an attractive binding of 1.25 kcal/mol. In addition, the Exper structure has the Cε1 heavy atom of H250 3.4 Å from the C24 atom of the furan ring in the ligand, leading to a repulsive vdW interaction, with a negative (repulsive) binding contribution of 2.48 kcal/mol. SCREAM led to a better side-chain conformation of H250 with a closest distance of 3.5 Å and an attractive binding of 1.38 kcal/mol. Using these two modified side-chains (after adding hydrogens) and minimizing the structure led to the Exper* structure, which has a cavity binding energy of 54.4 kcal/mol (compared to being very repulsive by 231 kcal/mol for Exper). The heavy atom

Table 2
Compounds for Structure–Activity Relationships (SAR) studies, including experimental binding constants (nM) and predicted binding energy (kcal/mol).

Compound	Structure		Binding affinity (Ki:nM)	Binding energy(kcal mol)	Reference
1		ZM241335	0.8 (0.7–1.0)	−38.78	Ongini et al. (1999)
2		Piperazine derivative	1.300	−33.21	Vu et al. (2004)
3		Piperazine derivative	>5.000	−20.96	Vu et al. (2004)
4		Piperazine derivative	3	−36.50	Vu et al. (2004)
5		Cis Bicydtc Piperazine derivative	0 3	−39.40	Peng et al. (2004)
6		Trans Bicycltc Piperazine derivative	230	−29.02	Peng et al. (2004)
7		Triazolotriazine derivative	39	−31.89	Vu et al. (2005)
8		Triazolotriazine derivative	11	−32.56	Vu et al. (2005)
9		CGS15943	04	−32.56	Ongini et al. (1999)

root-mean-squared deviation (RMSD) changes of the ligand position in Exper* compared to Exper is only 0.302 Å.

The predicted lowest energy binding mode for ZM241385/hAA$_{2A}$R is shown in Fig. 1a and compared to the X-ray structure (Exper*) in Fig. 1b. The aligned structures (Fig. 1c) lead to a 2.78 Å RMSD in the ligand. The predicted pharmacophore of ZM241385 in hAA$_{2A}$R is consistent with experiment, identifying all important interactions to critical residues: with a hydrogen bond (HB) to N253$^{6.55}$ [N15(ZM)–O(N253) = 2.7 Å and O25(ZM)–N(N253) = 3.1 Å]; HB to E169$^{5.30}$ (EL2) [N15(ZM)–O(E169) = 2.7 Å]; hydrophobic interactions with F168 (3.7 Å), H250 (3.7 Å), W246 (6.2 Å), M270 (3.8 Å), and H264 (4.1 Å). These results indicate that including full optimization of the EL with our standard computational methods identifies an accurate binding site.

2.2. Comparison of various ligands bound to hAA$_{2A}$R

To be useful in drug design it is essential that the predicted structures properly order the binding by various ligands (Structure–Activity Relationships, SAR). Starting from ZM241385/hAA$_{2A}$R, we predicted the binding site for the seven related antagonists in Table 2, which exhibit experimental binding constants from 0.3 to 5000 nM. The consistency between experimental binding affinities and predicted binding energies (with no adjustable constants) is shown in Fig. 2a. Except for compound **2** (from Table 2) our binding energies correlate quite well with experimental binding constants, leading to R^2 = 0.979. In each case the right part of the ligands is in the pocket bounded by TM2, 6, and 7, putting the left part in the cavity between EL2, TM2, and TM7. Our predicted binding for compound **2** is too strong by ~5 kcal/mol, probably due to the much greater flexibility of **2**, which would lead to a much larger decrease in entropy upon binding than the other seven ligands. Including all eight compounds, the fit is R^2 = 0.765 (solid line in Fig. 2a).

In particular, compounds **5** and **6** are identical except for chirality at one carbon, leading to quite different binding affinities: experimental K_i = 0.3 nM (#**5**, *cis*) vs. 230 nM (#**6**, *trans*) compared with computational BE = −39.4 kcal/mol (#**5**, *cis*) and −29.0 kcal/mol (#**6**, *trans*). The origin of this difference is clear in Fig. 2b,

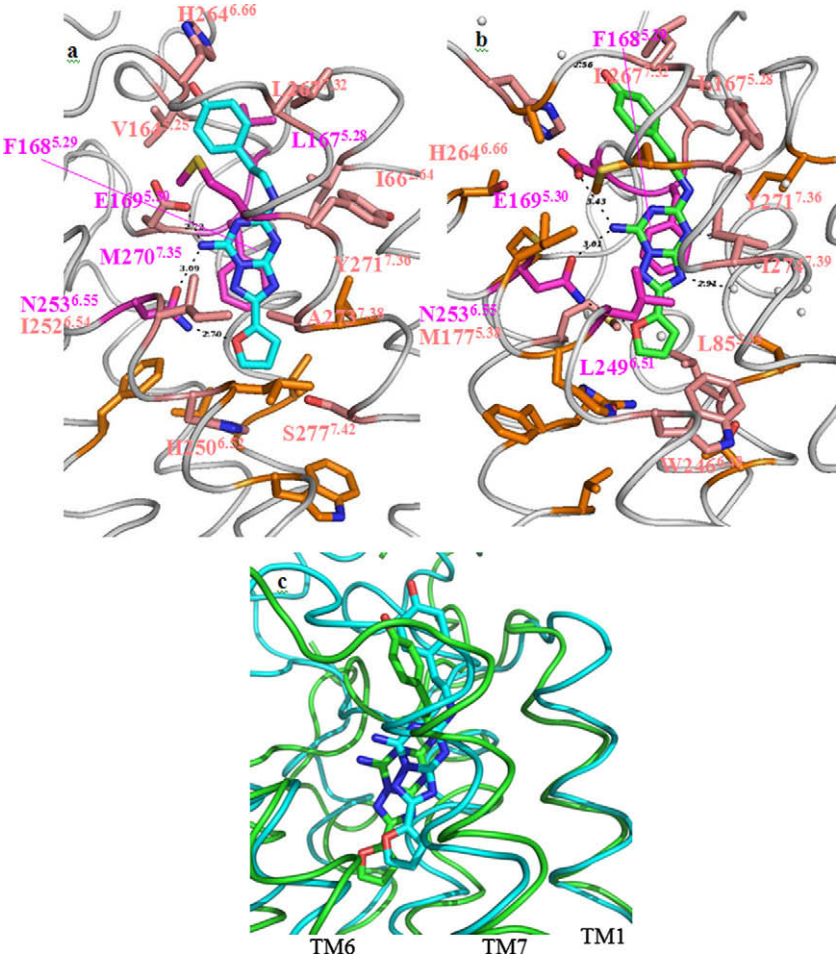


Fig. 1. Predicted structure of ZM241385/human A_{2A} adenosine receptor compared with X-ray structure. (a). Predicted binding mode for ZM241385/human A_{2A} adenosine receptor. (b). Binding mode of X-ray structure. (c). Superposition of crystal structure (green) with the predicted structure (light blue) (matching the backbond atoms of the protein). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

showing that the left part of the *cis* form is in a good cavity between EL2, TM2, and TM7 while the *trans* form puts the left part of the ligand among TM1, TM2, and TM7 with unfavorable interactions to the Y271 (TM7) and Y9 (TM1) side-chains.

2.3. Subtype selectivity

The experimental binding constants of ZM241385 (compound **1** in Table 2) to the four adenosine receptor subtypes show dramatic subtype selectivity: K_i = 0.8 nM (A_{2A}), 50 nM (A_{2B}), 255 nM (A_1) and >10,000 nM (A_3) (Ongini et al., 1999). We matched our predicted binding site for A_{2A} to the apo-protein structures we predicted for the other three subtypes (using the same techniques), leading to exactly the same trend, as shown in Fig. 3. Our cavity analysis for binding of ZM241385 to all four subtypes indicates that E169 and M270 in A_{2A} are the most important for subtype selectivity. Thus E169 of A_{2A} (−4.1 kcal/mol interaction with ligand) corresponds to E172 in A_1 (−3.3 kcal/mol), E174 in A_{2B}

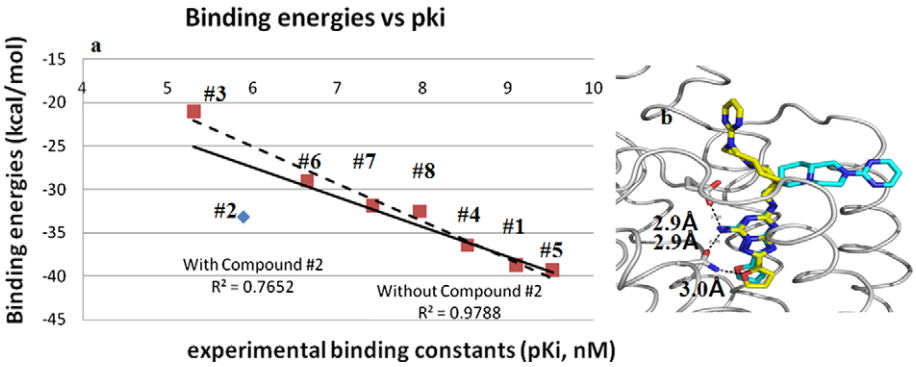


Fig. 2. Structure–Activity Relationships (SAR) based on the predicted structure of $hAA_{2A}R$. (a) The predicted binding energies (kcal/mol) to the human A_{2A} adenosine receptor of the eight antagonists listed in Table 2 compared with the experimental binding constants (pK_i). The quantity labeled as binding energy is the change in energy due to the ligand binding, so that more negative is more strongly binding. The dotted line shows the fit without compound **2** (which is much more flexible than the others). (b) Alignment of compound **5** (Cis Bicyclic Piperazine, strong binder, yellow) and compound **6** (Trans Bicyclic Piperazine, weak binder, cyan) in the binding site of hA_{2A} adenosine receptor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

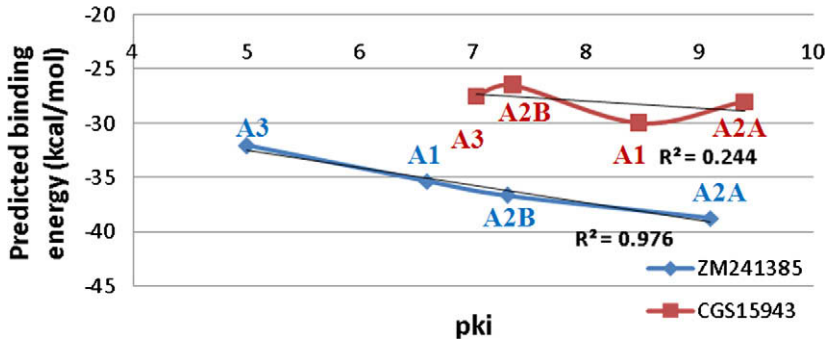


Fig. 3. Subtype selectivity study and agonist binding based on the predicted structure. The correlation of the experimental binding constants (pK_i) (Ongini et al., 1999) with the predicted binding energies (kcal/mol) of human A_{2A} selective compound **1**, ZM241385 (blue curve) and non-selective compound **9**, CGS15943 (red curve) to A_1, A_{2A}, A_{2B}, and A_3 human adenosine receptors. The computed energies show the same trends as the experiments. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

(−2.65 kcal/mol), but V169 in A_3 (−0.9 kcal/mol). The cavity analysis for M270 of A_{2A} (−1.6 kcal/mol) corresponds to T270 in A_1 (−0.66 kcal/mol), M272 in A_{2B} (−3.1 kcal/mol), and L264 in A_3 (−2.3 kcal/mol).

In contrast, the CGS15943 antagonist (compound **9** in Table 2) shows little subtype specificity with: K_i = 0.4 nM (A_{2A}), 3.5 nM (A_1), 44 nM (A_{2B}), and 95 nM (A_3). Indeed matching our predicted binding site for A_{2A} to the other three subtypes, leads correctly to only small differences in binding energies in good agreement with experiment (Fig. 3).

These results suggest that the predicted protein structures can be used to design subtype selective ligands that can lead to therapeutics with decreased side effects.

2.4. Binding site for adenosine agonist

We used the same docking techniques to predict the binding site of adenosine agonist to the predicted (lowest energy) structure of $hAA_{2A}R$. The result (Fig. 4) indicates that the most important residues (cavity analysis) are (including the interaction energy with ligand in kcal/mol in parentheses): V84 (−2.44), T88 (−0.58), E169 (−1.09), N253 (−0.61), F182 (−1.73), H250 (−2.86), S277 (−0.99), W246 (−0.70), H278 (−0.35). All these residues have been implicated in prior mutagenesis experiments of adenosine-based agonists. The polar residues S277, H278, and H250 show a weaker interaction with the ZM241385 antagonist. These might play important roles in the activation of $hAA_{2A}R$. Our previous experience is that substantial changes occur in the agonist/GPCR binding mode during dynamics as the receptor changes to its active form (Li et al., 2007). We matched the agonist to the other three ARs, finding a similar binding site and cavity analysis.

We outline here the difference between agonist-bound and antagonist-bound structure for $hAA_{2A}R$. The major difference involves two hydroxyl groups from the agonist that act as a bi-dentate hook on the ligand interacting with H250 of TM6 and S277 of TM7. Eventually, after MD simulations in the full solvent, the agonist-bound protein structure may display additional structural differences from the antagonist-bound structure. In our previous work on the human DP receptor (Li et al., 2007), we obtained detailed structural differences between the agonist-bound structure and the antagonist-bound structure, and the conformational changes that accompany receptor activation.

2.5. Loop predictions for $hAA_{2A}R$

Our first predictions of the 3D structures of GPCRs, focused on the packing of the 7-helix bundle. Then we predicted the three extracellular loops (EL1, EL2, EL3), the three intracellular loops

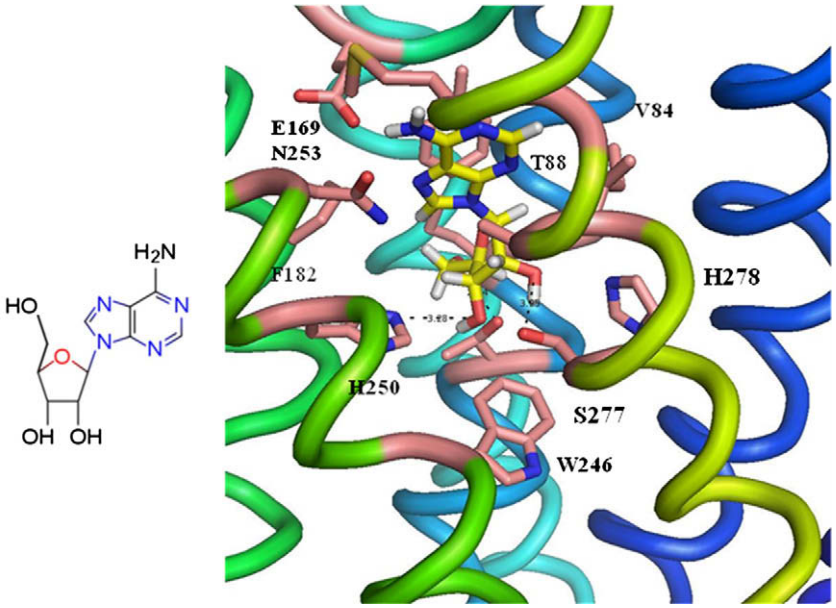


Fig. 4. Predicted binding mode (lowest energy) of the adenosine agonist to the human A_{2A} adenosine receptor. Key residues are indicated.

(IL1, IL2, IL3), the amino terminus (NT) and the carboxyl terminus (CT) using the Continuous Configuration Boltzmann Biased (CCBB) Monte Carlo method (Li and Goddard, 2006; Sadanobu and Goddard, 1997), including loop termination optimization (Debe et al., 1999). Many GPCRs have a conserved cysteine at the top of TM3 (C77 for $hAA_{2A}R$) and a conserved cysteine in EL2 (C166 for $hAA_{2A}R$) that are oxidized (coupled). Thus, our standard procedure is to separately build EL2 in two sections, accounting for this coupling. To allow maximum flexibility in positioning the loops, we alanize the loops (except for Gly and Pro) and grow ~20 loops (including the C77-C166 linkage constraint). These structures are minimized and then the side-chains added using SCREAM. The loops were built in the sequence EL1, EL3, EL2, IL1, IL2, IL3, but rather than using CCBB we determined the N-terminus and the C-terminus by optimizing an extended conformation. In each case, the lowest total energy was used to select the loop configuration. Then the six loops and two termini were minimized for 500 steps using the DREIDING 2 FF (Mayo et al., 1990), with all but the last residue on each TM constrained. In these structures, all Cys in the loops except C166 were assumed to be reduced. *No information on loops from crystal structures was used.*

The above general procedure was used in our CAGSMD challenge submission, but upon examining the crystal structure, we found that all six non-conserved cysteines in $hAA_{2A}R$ (in addition to C77 and C166 discussed above) distributed among the three EL loops (C71 and C74 in EL1, C146, and C159 in EL2, C259, and C262 in EL3), are oxidized to form three cysteine pairs. Based on this information, we extended our general method to include optimization of the loops with all cysteines oxidized, but *without using any structural or topological information from the crystal structure*. Thus, we examined all 15 ways to couple the six non-conserved cysteines.

As shown in Fig. 5a, our original loops (with reduced Cys) have the Cys on EL1 far from the Cys on EL3 (Cα–Cα > 24.6 Å), making it implausible to form disulfide bonds between EL1 and EL3. In addition, Fig. 5b shows that C71 (EL1) is 22.2 Å (Cα–Cα) from C146 (EL2) on the edge of TM4, 24.6 Å from C259 (EL3); and 22.5 Å C262 (EL3). Thus, disulfide bonds between C71 and these three cysteines are not plausible. This analysis indicates that we need consider only two of the 15 possibilities to form three disulfide bonds among six non-conserved cysteines.

- ELLA: C71 (EL1)–C159 (EL2), C74 (EL1)–C146 (EL2), C259 (EL3)–C262 (EL3)
- ELLB: C71 (EL1)–C74 (EL1), C146 (EL2)–C159 (EL2), C259 (EL3)–C262 (EL3)

For both ELLA and ELLB we first grew the alanized loops (using the CCBB process described above) with Cys–Cys constraints. This led to 20 minimized structures for both ELLA and ELLB. We then selected the six with the lowest total energy [five from ELLA and one from ELLB (5th in energy)]. The lowest energy loop conformation had the ELLA configuration of disulfide bonds as shown in Fig. 6, which was selected for subsequent docking. Indeed it has the same disulfide bond configuration as in the crystal structure, validating this procedure. The CRMSDs (of Cα atoms) of EL1, EL2, and EL3 relative to the crystal structure are 3.0 Å, 5.5 Å, and 3.9 Å, respectively, which are much higher CRMSDs than for the helix region (with a CRMSD of 2.0 Å). Meanwhile, the predicted EL2 has the lowest quality (CRMSD is 5.5 Å) due to its long length. We consider these CRMSDs to be reasonable for the conformationally flexible loops.

3. Discussion

We find that our new methods for predicting the protein structure of $hAA_{2A}R$ and of the A_{2B}, A_1, and A_3 subtypes including optimizing the loops with all Cys oxidized lead to:

- A predicted binding site for ZM241385 to $hAA_{2A}R$ in good agreement with the X-ray experiments, with RMSD error of 2.8 Å and a correct prediction of 9 of the 12 important residues in the binding site.
- Predicted binding energies for a series of eight related antagonists to $hAA_{2A}R$ in excellent agreement with experimental relative binding constants ($R^2 = 0.765$ for all eight and $R^2 = 0.979$ excluding the one very flexible ligand).
- Excellent agreement of the subtype specificities for binding of the subtype selective antagonist ZM241385 across the four adenosine receptors.
- Excellent agreement in the lack of subtype specificity for the adenosine agonist and for the non-subtype selective antagonist CGS15943, as observed experimentally.

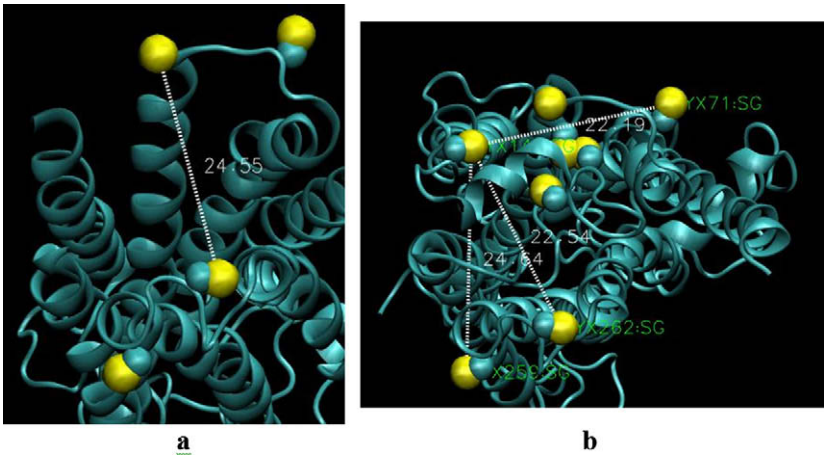


Fig. 5. (a). The best predicted structure of $hAA_{2A}R$ adenosine receptor assuming that the non-conserved disulfide bonds in the extracellular loop (EL) are reduced. Here we show two cysteines C71 and C74 in EL1 (The two yellow spheres near the top) and two cysteines C259 and C262 in EC3 (The two yellow spheres at the bottom). The closest Cα distance (C74–C262) between the cysteines on EL1 and the cysteines on EL3 is 24.6 Å, making it implausible to have disulfide bonds between EL1 and EL3. (b). The best predicted structure of A2A adenosine receptor without consideration of non-conserved disulfide bonds. Here we show all eight cysteines in yellow spheres. C146 on extracellular loop (EC2) 2 is on the edge of TM 4 and is far away from C71 on EC1 (Sγ distance: 22.5 Å); C259 on EC3 (24.6 Å); C262 on EC3 (22.5 Å). It is not plausible to make disulfide bonds between C71 and those three cysteines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

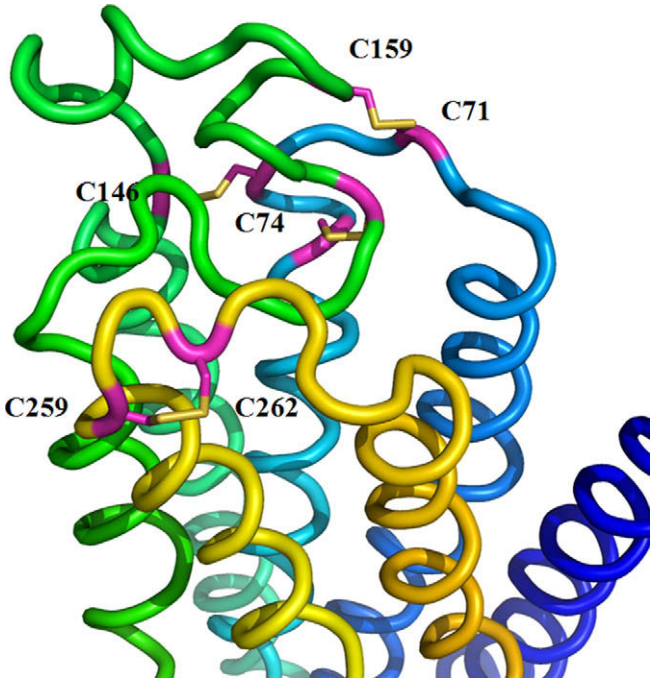


Fig. 6. The lowest energy loops predicted by the continuous Configuration Boltzmann Biased (CCBB) method.

- Excellent agreement of the binding site predicted for the adenine agonist bound to $hAA_{2A}R$ with available mutation data.

This suggests that these computational methods can provide predictions sufficiently accurate to design subtype selective ligands for the many important GPCR therapeutics targets for which no experimental structural information is available.

These studies also show how important it was to have a crystal structure for a related system. This allowed us to discover the need to optimize fully the Cys coupling in the EL loops critical in predicting accurate binding sites. This in turn enabled the theory to predict the structures of the other subtypes and the structures of agonists and antagonists bound to all subtypes, information not yet available from experiment.

Our analysis of the experimental crystal structure shows that 16% of the ligand binding energy results from its interaction with the strongly bound waters, indicating that increased accuracy would require the theory to predict the location of such strongly bound waters. Docking of the ligand to the experimental crystal structure with and without waters, shows that the presence of waters improve the accuracy of the predicted binding pose for the ligand from 1.06 Å to 0.80 Å, as shown in Supplementary data section.

4. Methods

Our structure prediction methods are focused on GPCRs, which we consider the best candidate for direct prediction methods. This is because GPCRs have a well-defined three-dimensional topology in which a single protein must thread through the membrane seven times, forming α-helices that pack tightly with each other while exposing hydrophobic regions to the surrounding lipid. Moreover, the important ligands for many GPCRs bind selectively to the extracellular facing ends of the seven helical TM domains.

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 (GPCR Ensemble of Structures in Membrane BiLayer Environment) method (Abrol et al., in preparation), an improved version of MembStruk method (Vaidehi et al., 2002) to predict the 3D structure (without using homology to known 3D structures). GEnSeMBLE predicts the ensemble of low-energy conformational states for a GPCR. We consider that it is important to have an ensemble of 10 or more of the lowest packings because different ligands might make different choices among them. Indeed cases are known in which different conformations of a GPCR are stabilized by different ligands (Kobilka, 2004). In particular several of the structures in the low energy ensemble might play a role in activation. Moreover, for mutation binding experiments to validate the predictions, it could be that even single mutations might change the optimum packing, so that we should evaluate the effect of a mutation on all low lying packings. This ensemble approach has been validated for Bovine Rhodopsin, Human β_2 and Turkey β_1 Adrenergic Receptors. The methodology consists of the following key steps:

4.1. Prediction of the 7-helix bundle

(1) PredicTM: Our method to predict the transmembrane (TM) regions for a membrane protein that does not use any fitted parameters.
(2) OptHelix: Our new approach for using molecular dynamics to generate relaxed helices with their natural kinks.
(3) BiHelix: This highly efficient sampling algorithm samples rapidly $12^7 \sim 35{,}000{,}000$ packings of the seven helices of the GPCR. BiHelix partitions the 7-helix interaction problem into 12 sets of 2-helix interactions, using SCREAM to optimize the side-chains for each case (Kam and Goddard, 2008).
(4) CombiHelix: This algorithm takes the best 1000 results from BiHelix step, generates the multiple rotational combinations, and optimizes the side-chains using SCREAM in conjunction with an implicit membrane solvation contribution to evaluate the total energy of the optimized multi-helix bundle. This provides an ensemble of ~10 low-energy structures each of which is used in docking studies of agonists, antagonists, and inverse agonists to provide the structures likely to be involved in GPCR activation.

4.2. PredicTM

Accurate ab initio prediction of helical membrane proteins begins with the identification of the location of the transmembrane (TM) regions of the protein within the amino acid sequence. Many such methods have been developed based on analysis of the hydrophobicity profile. Our current approach (PredicTM) consists of six steps:

1. Retrieval of similar protein sequences from a database: here we select ~1000 sequences with structural identities down to ~5%.
2. Multiple sequence alignment of similar sequences: here we now use the MAFFT (Katoh et al., 2005) multiple sequence alignment program, using the "E-INS-i" method, which we find best suited for sequences with multiple aligning segments separated by non-aligning segments, which perfectly describes the situation for GPCRs.
3. Hydrophobic profile generation and noise removal: we now use the Wimley–White whole-residue octanol scale, a thermodynamic scale derived from transfer of residues from water into *n*-octanol (Wimley et al., 1996). Unresolved amino acids in the alignment (B, Z, J, X) are replaced with gaps. Here we eliminate noise (as shown in Fig. 7 for Dopamine D1 receptor) by averaging windows of seven amino acids through 21 amino

acids, where seven corresponds roughly to one helical turn above and below a residue and 21 corresponds roughly to the length of one TM region.
4. Initial transmembrane region predictions: the initial TM domain predictions are taken as the regions with hydrophobicity values greater than zero, leading to "raw helices".
5. Application of capping rules: the raw helices are extended (or capped) on both N- and C-termini until a "helix breaker" residue is found.
6. Identification of hydrophobic centers: to place all helices on the same reference plane we select the "hydrophobic center" for each helix as the position in the raw helix where the area of the hydrophobicity profile is equal on both sides, providing the "buoyant" center of the helix in the lipid environment.

4.3. OptHelix

We have been refining our method of optimization of transmembrane helices, to account for the kinks often induced by prolines. We now eliminate noise due to longer side-chains and due to the helix termini by replacing these groups with alanines (except for Ser and Thr near the Pro) and optimizing with 2 ns of MD, followed by restoring the correct residues using SCREAM and minimizing the structure.

4.4. BiHelix

4.4.1. Rotational sampling of helices

Most critical to our current methods is to sample all ways of packing the seven TM domains together. We have found that with the newer methods of identifying the TM domains, of optimizing the helices, and applying SCREAM, it is sufficient to sample the orientations of helix with a stride of 30°. However, even with a 30° increment for each of the seven helices, a complete sampling of all helical rotational combinations would require the sampling of $12^7 \sim 35$ million conformations, for each of which the side-chains would need to be optimized. This is computationally intractable. Instead, we use BiHelix sampling as described below.

4.4.2. BiHelix sampling

As indicated by two-way arrows in Fig. 8 (left panel), there are 12 nearest neighbor pair-wise interactions between 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. For each such 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. During this sampling, the other five helices are not present, as indicated in Fig. 8 (right panel) for helix 1–2 pair. For each rotational combination, we optimize the side-chains using our rotamer placement method SCREAM.

SCREAM uses a library of residue conformations ranging from a CRMS diversity of 0.4–1.6 Å 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 E_{scream}, without separate considerations of valence, electrostatic, hydrogen bond and van der Walls terms. We combine the E_{scream} energies for 144 combinations for each helix pair (using a total of 144×12 helix pairs = 1728 energies) to obtain an estimated energy for all possible $12^7 \sim 35$ million conformational combinations. We have found that the best structures are always within the top 1000 structures ranked by increasing energy.

As a test, we applied this procedure to Bovine Rhodopsin experimental structure (PDB ID: 1u19), where we removed all the loops along with the ligand retinal and only kept the experimentally determined TM helix regions. Using the BiHelix method we estimated the energies for all $(12)^7 \sim 35$ million combinations. This includes the experimental structure, which corresponds to the combination 0_0_0_0_0_0_0 (first zero corresponds to helix 1, second zero corresponds to helix 2, and so on). Fig. 9a shows the top structures out of $(12)^7 \sim 35$ million, obtained using the mean field energies from the BiHelix procedure. The experimental structure is 7th best. We then used CombiHelix (described in Section 4.5) to construct the full 7-helix bundle for the best 1000 structures from the BiHelix analysis and reoptimized the side-chains using SCREAM, followed by 10 steps of minimization. The best structures ordered by energy are listed in Fig. 9b, where we see that the experimental structure is predicted to have the lowest total energy, which gives us confidence in our energy functions and in the BiHelix ordering. Among these best packings, Helices 1 through 4 all show the crystal conformation. Helix 6 shows an alternate preference for a 30° anticlockwise rotation (330° clockwise rotation), which is consistent with the change previously suggested for the active state of Rhodopsin [where TM6 undergoes anticlockwise rigid-body rotation while looking at the protein from the extracellular side (Farrens et al., 1996)]. This is also consistent with the recently published ligand-free Bovine Opsin structure (Park et al., 2008) that shows helix 6 rotated 30° anticlockwise relative to the *cis*-retinal bound Bovine Rhodopsin. Helix 5 shows a surprising flexibility, probably because the covalently attached retinal is absent during our BiHelix procedure. Indeed this is also consistent with the putative 180° rotation of Helix 5 in activated Bovine Rhodopsin (Meng and Bourne, 2001).

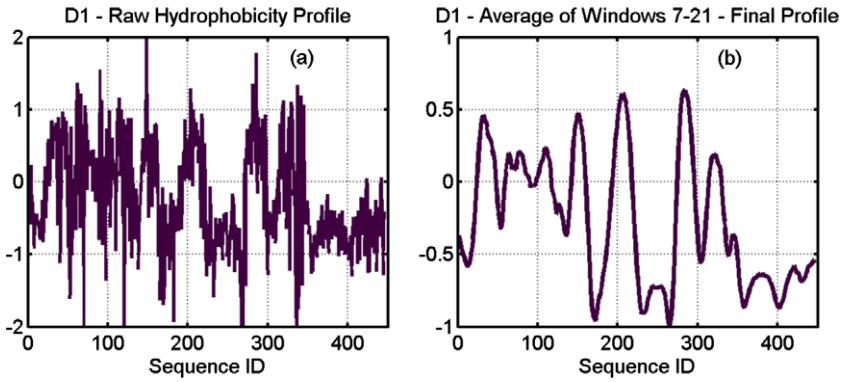


Fig. 7. (a). Raw hydrophobicity profile for human Dopamine D1 receptor. (b). Average hydrophobicity profile of multiple window averages ranging from 7 to 21 residues obtained from the raw hydrophobicity profile (see text for details).

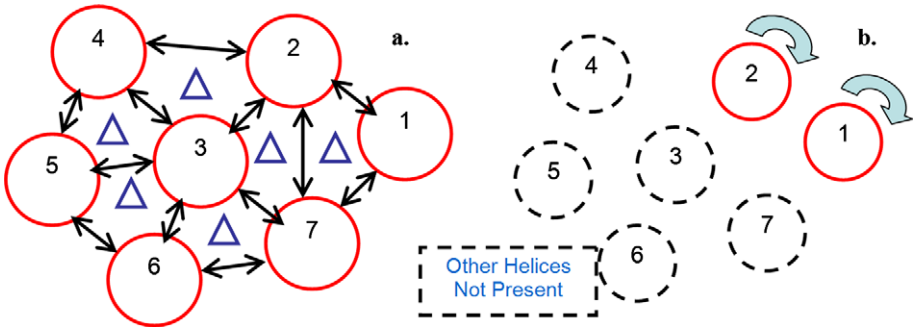


Fig. 8. a. Double arrows connecting nearest neighbor helix pairs that are sampled independently in the BiHelix procedure. b. The BiHelix procedure is highlighted using helices 1 and 2 to show that when the conformations for this helix pair are sampled, other helices are not present.

a.

Structure	TotalE	
0_0_0_0_0_330_0	-237	
0_0_0_0_300_330_0	-234	
0_0_0_0_180_330_0	-231	
0_0_0_0_90_330_0	-220	
0_0_0_0_210_330_0	-215	
0_0_0_0_60_330_0	-213	
0_0_0_0_0_0_0	-210	XTAL
0_0_0_0_330_330_0	-203	
0_0_0_0_90_0_0	-195	
0_0_0_0_180_0_0	-189	
0_0_0_0_60_0_0	-175	

b.

H1	H2	H3	H4	H5	H6	H7	Total	
0	0	0	0	0	0	0	-55.1	XTAL
0	0	0	0	210	0	0	-20.6	
0	0	0	0	90	0	0	-16.7	
0	0	0	0	0	0	330	-4.1	
0	0	0	0	180	0	0	3.1	
0	0	0	0	330	0	0	23	
0	0	0	0	210	0	330	30.5	
0	0	0	0	210	330	0	33.2	
0	0	0	0	300	330	0	36.7	

Fig. 9. (a). Top structures (out of ~35 million) from BiHelix analysis for Bovine Rhodopsin. (b). Top structures after building top 1000 bundles from BiHelix.

Template Selection: To pack the seven helices from OptHelix into a bundle, requires the definition of 6 quantities for each helix ($6 \times 7 = 42$ total): the x, y, z of the hydrophobic center 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. Our procedures focus on z and η as described above. The initial values for the remaining four coordinates for each helix (x, y, θ, ϕ) are taken from a template structure. We use a library of template structures that we consider as validated:

(1) X-ray structure for Bovine Rhodopsin (Palczewski et al., 2000).
(2) X-ray structure for Human β2 Adrenergic Receptor (Cherezov et al., 2007).
(3) X-ray structure for Turkey β1 Adrenergic Receptor (Warne et al., 2008).
(4) X-ray structure for Human Adenosine A_{2A} Receptor (Jaakola et al., 2008).
(5) X-ray structure for ligand-free Bovine Opsin (Park et al., 2008).
(6) Computational 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 (Vaidehi et al., 2006).
(7) Computational 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 (Li et al., 2007).
(8) Computational 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 (Heo et al., 2007).

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 bundle energy or ligand binding energy. As new structures are solved and predicted (validated and subjected to full MD in lipid bilayer environment), they will be added to the ensemble of templates.

To specify the reference rotation angle (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.

4.5. CombiHelix

4.5.1. Starting bundles

The top 1000 structures coming out of BiHelix analysis are built explicitly using the rotations specified for each helix in the combination. The helical axis for rotation is the same as used in BiHelix analysis.

4.5.2. CombiHelix bundle optimization

For each of the bundles built in the previous step, the sidechains are optimized using SCREAM as in the BiHelix method. The SCREAM energy is reported for each bundle. Each bundle is also immersed in an implicit membrane to compute membrane solvation effects that should disfavor helix rotations that expose charged residues to lipids. This membrane solvation is described in the next step.

4.5.3. Membrane solvation

This step evaluates protein–lipid interactions in an implicit fashion. For efficient and accurate estimation of the lipid–helix interactions, we implemented in the Delphi Poisson–Boltzmann (PB) Solver a multi-dielectric model (Rocchia et al., 2001) of lipid bilayers to account properly for the penalty of exposing polar and charged residues to lipid molecules. As shown in Fig. 10, we replace the middle 20 Å of the lipid with a slab of $\varepsilon = 2$, while the 5 Å slabs above and below the middle slab have $\varepsilon = 7$, and $\varepsilon = 80$ outside the membrane.

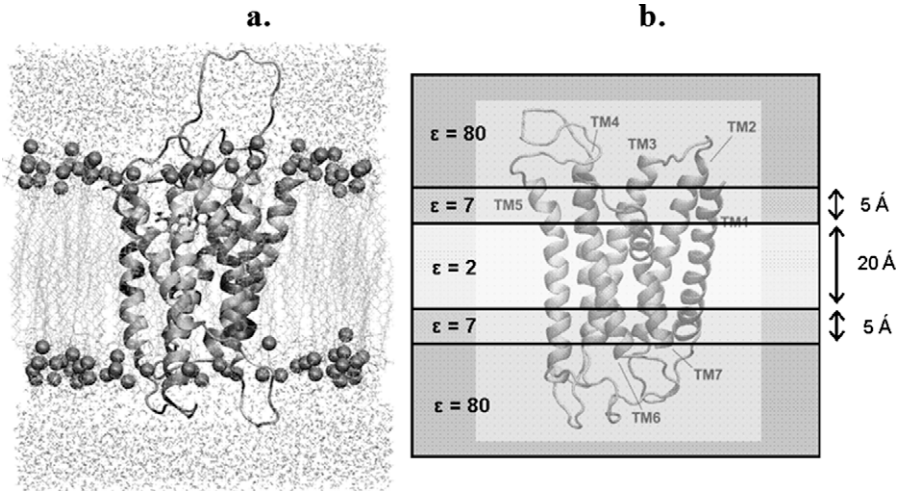


Fig. 10. (a). An example system showing a GPCR embedded in an explicit lipid bilayer environment. (b). Multi-dielectric description of the lipid bilayer environment used for the treatment of implicit solvation for membrane embedded proteins.

The energy of the bundle interacting with the lipid is denoted as Esolv. In addition to the PB term, we include a cavity term based on the White hydrophobic scale. Our earlier MembStruk methods used either implicit solvent methods or a cluster of ~40 lipid molecules to mimic the membrane. For each combinatorial conformation from BiHelix method, we use the side-chain-optimized structure to evaluate the membrane solvation energy, E_{solv}, add it to the SCREAM energy E_{scream} from previous step. This process generally leads to 2–5 final structures stable enough to play a role in differential binding of agonists, antagonists and inverse agonists to suggest possible activation pathways. Appling this to the top 1000 combinations for bovine rhodopsin, β1, and β2 adrenergic receptors correctly predicts the crystal structure to be the lowest energy structure. For bovine rhodopsin, the top structures are shown in Fig. 9b and were discussed in Section 4.4.

For $hAA_{2A}R$, in addition to the above procedure, we constructed a TM bundle based on homology to the TM region of hβ2AR (PDB ID: 2rh1) and tβ1AR (PDB ID: 2vt4) using the sequence alignment based on structurally conserved regions (SCR). Then, we used BiHelix sampling to optimize the rotations of TM5 and TM6 in the helix bundle (using a 10° grid over ±90° from the template).

We docked the ligand ZM241385 to the best five bundles from these sets. The protein structure with the lowest binding energy for the ligand was the BiHelix-optimized homology model based on the hβ2AR template with TM5 rotated clockwise 10°. This predicted structure is the one discussed in this paper.

4.6. Procedure for docking ligands to predicted protein structures

The procedures for predicting protein structure indicated above were applied to $hAA_{2A}R$ for six templates, as summarized above. We found that the best by energy was the BiHelix-optimized homology model to hβ2AR [with optimum rotations of (0, 0, 0, 0, 10°, 0, 0) relative to the template] combined with the optimized ELLA loop. In order to assess the accuracy of predicting binding to subtypes, we used the same procedure to predict structures for the AA_1R, $AA_{2B}R$, and AA_3R subtypes.

Each of these protein structures was used to dock ZM241385. The structure and charges of ZM241385 were calculated using quantum mechanics (B3LYP with the 6-311G** basis set). Only the lowest energy conformer was used for docking. We used the HierDock general procedure for docking.

The HierDock method uses a hierarchical strategy for selecting ligand-binding conformations and calculating their binding energies. The whole protein was partitioned into 35 regions (each with sides of 10 Å) and scanned to find the putative binding regions (with the six hydrophobic residues, I, L, V, F, Y, and W alanized). This ScanBindSite procedure used DOCK4.0 (Kuntz et al., 1982) to generate 1000 conformations in each of these putative regions, selecting the optimum regions based on a combination of burial score and binding energy. These optimum regions were combined and 10,000 poses were generated using DOCK4.0, which were scored using the DREIDING 2 FF (Mayo et al., 1990). The top 1000 (by energy) were de-alanized, SCREAMed, and then minimized using the DREIDING 3 FF. Then we selected the top 1% (10) for further minimization of the binding site complex (using the unified binding site including all residues within 4 Å of any of the 10 binding poses). The protein and ligand were then neutralized by transferring protons appropriately in salt bridges and protonating or deprotonating exposed side-chains (this leading to more reliable energy comparisons [Bray and Goddard, 2008]). Then the final docked structure with the best binding energy was selected.

Acknowledgments

Dr. Victor Kam and John Wendel contributed significantly to developing the methods used here but were not involved directly in the project. We thank Dr. Jenelle Bray for helpful discussions. Funding for this project was provided by gifts to the MSC at Caltech.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2010.01.001.

References

Bray, J.K., Goddard, W.A., 2008. The structure of human serotonin 2c G protein-coupled receptor bound to agonists and antagonists. J. Mol. Graph. Model. 27, 66–81.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K., Stevens, R.C., 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 318, 1258–1265.

Debe, D., Carlson, M.J., Sadanobu, J., Chan, S.I., Goddard III, W.A., 1999. Protein fold determination from sparse distance restraints: the restrained generic protein direct Monte Carlo method. J. Phys. Chem. B 103, 3001–3008.

Farrens, D.L., Altenbach, C., Yang, K., Hubbell, W.L., Khorana, H.G., 1996. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science 274, 768–770.

Floriano, W.B., Vaidehi, N., Goddard III, W.A., Singer, M.S., Shepherd, G.M., 2000. Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. Proc. Natl. Acad. Sci. USA 97, 10712–10716.

Freddolino, P.L., Kalani, M.Y.S., Vaidehi, N., Floriano, W.B., Trabanino, R.J., Kam, V.W.T., Goddard III, W.A., 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, 2736–2741.

Heo, J., Han, S.K., Vaidehi, N., Wendel, J., Kekenes-Huskey, P., Goddard III, W.A., 2007. Prediction of the 3D structure of FMRF-amide neuropeptides bound to the mouse MrgC11 GPCR and experimental validation. Chem. Biol. Chem. 8, 1527–1539.

Hopkins, A.L., Groom, C.R., 2002. The druggable genome. Nat. Rev. Drug Discov. 1, 727–730.

Jaakola, V.P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E.Y.T., Lane, J.R., IJzerman, A.P., Stevens, R.C., 2008. The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science 322, 1211–1217.

Kalani, M.Y.S., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Floriano, W.B., Kam, V.W.T., Goddard III, W.A., 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, 3815–3820.

Kam, V.W.T., Goddard III, W.A., 2008. Flat-bottom strategy for improved accuracy in protein side-chain placements. J. Chem. Theor. Comput. 4, 2160–2169.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acid Res. 33, 511–518.

Kobilka, B., 2004. Agonist binding: a multistep process. Mol. Pharmacol. 65, 1060–1062.

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E., 1982. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161, 269–288.

Li, Y.Y., Goddard III, W.A., 2006. Continuous self-avoiding walk with application to the description of polymer chains. J. Phys. Chem. B 110, 18134–18137.

Li, Y.Y., Zhu, F.Q., Vaidehi, N., Goddard III, W.A., Sheinerman, F., Reiling, S., Morize, I., Mu, L., Harris, K., Ardati, A., Laoui, A., 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, 10720–10731.

Lundstrom, K., 2006. Latest development in drug discovery on G protein-coupled receptors. Curr. Prot. Peptide Sci. 7, 465–470.

Mayo, S.L., Olafson, B.D., Goddard III, W.A., 1990. DREIDING – a generic force-field for molecular simulations. J. Phys. Chem. 94, 8897–8909.

Meng, E.C., Bourne, H.R., 2001. Receptor activation: what does the rhodopsin structure tell us? Trends Pharmacol. Sci. 22, 587–593.

Michino, M., Abola, E., GPCR Dock 2008 participants, Brooks III, C.L., Dixon, J.S., Moult, J., Stevens, R.C., 2009. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. Nat. Rev. Drug Discov. 8, 455–463.

Ongini, E., Dionisotti, S., Gessi, S., Irenius, E., Fredholm, B.B., 1999. Comparison of CGS15943, ZM241385 and SCH58261 as antagonists at human adenosine receptors. Naunyn-Schmiedeberg's Arch. Pharmacol. 359, 7–10.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M., 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 289, 739–745.

Park, J.H., Scheerer, P., Hofmann, K.P., Choe, H.W., Ernst, O.P., 2008. Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature 454, 183. U33.

Peng, J.Y.C., Vaidehi, N., Hall, S.E., Goddard III, W.A., 2006. The predicted 3D structures of the human M1 muscarinic acetylcholine receptor with agonist or antagonist bound. Chem. Med. Chem. 1, 878–890.

Peng, H.R., Kumaravel, G., Yao, G., Sha, L., Wang, J., Van Vlijmen, H., Bohnert, T., Huang, C., Vu, C.B., Ensinger, C.L., Chang, H.X., Engber, T.M., Whalley, E.T., Petter, R.C., 2004. Novel bicyclic piperazine derivatives of triazolotriazine and triazolopyrimidines as highly potent and selective adenosine A(2A) receptor antagonists. J. Med. Chem. 47, 6218–6229.

Rocchia, W., Alexov, E., Honig, B., 2001. Extending the applicability of the nonlinear Poisson–Boltzmann equation: multiple dielectric constants and multivalent ions. J. Phys. Chem. B 105, 6507–6514.

Sadanobu, J., Goddard III, W.A., 1997. The continuous configurational Boltzmann biased direct Monte Carlo method for free energy properties of polymer chains. J. Chem. Phys. 106, 6722–6729.

Tang, C., Insel, P., 2005. Genetic variation in G-protein-coupled receptors – consequences for G-protein-coupled receptors as drug targets. Expert Opin. Therap. Targets 9, 1247–1265.

Trabanino, R.J., Hall, S.E., Vaidehi, N., Floriano, W.B., Kam, V.W.T., Goddard III, W.A., 2004. First principles predictions of the structure and function of G-protein-coupled receptors: validation for bovine rhodopsin. Biophys. J. 86, 1904–1921.

Vaidehi, N., Floriano, W.B., Trabanino, R., Hall, S.E., Freddolino, P., Choi, E.J., Zamanakos, G., Goddard III, W.A., 2002. Prediction of structure and function of G protein-coupled receptors. Proc. Natl. Acad. Sci. USA 99, 12622–12627.

Vaidehi, N., Schlyer, S., Trabanino, R.J., Floriano, W.B., Abrol, R., Sharma, S., Kochanny, M., Koovakat, S., Dunning, L., Liang, M., Fox, J.M., de Mendonca, F.L., Pease, J.E., Goddard III, W.A., Horuk, R., 2006. Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. J. Biol. Chem. 281, 27613–27620.

Vu, C.B., Peng, B., Kumaravel, G., Smits, G., Jin, X.W., Phadke, D., Engber, T., Huang, C., Reilly, J., Tam, S., Grant, D., Hetu, G., Chen, L.Q., Zhang, J.B., Petter, R.C., 2004. Piperazine derivatives of [1,2,4]triazolo[1,5-alpha][1,3,5]triazine as potent and selective adenosine A(2a) receptor antagonists. J. Med. Chem. 47, 4291–4299.

Vu, C.B., Pan, D., Peng, B., Kumaravel, G., Smits, G., Jin, X.W., Phadke, D., Engber, T., Huang, C., Reilly, J., Tam, S., Grant, D., Hetu, G., Petter, R.C., 2005. Novel diamino derivatives of [1,2,4]triazolo[1,5-α][1,3,5]triazine as potent and selective adenosine A(2a) receptor antagonists. J. Med. Chem. 48, 2009–2018.

Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G.W., Tate, C.G., Schertler, G.F.X., 2008. Structure of a beta(1)-adrenergic G-protein-coupled receptor. Nature 454, 486. U2.

Wimley, W.C., Creamer, T.P., White, S.H., 1996. Solvation energies of amino acid side chains and backbone in a family of host–guest pentapeptides. Biochemistry 35, 5109–5124.