



Published in final edited form as:

Proteins. 2012 June ; 80(6): 1503–1521. doi:10.1002/prot.24035.

Do crystal structures obviate the need for theoretical models of GPCRs for structure based virtual screening

Hao Tang^{1,2}, Xiang Simon Wang^{3,*}, Jui-Hua Hsieh¹, and Alexander Tropsha^{1,*}

¹Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products and Carolina Exploratory Center for Cheminformatics Research, School of Pharmacy

²Biophysics Program, School of Medicine; University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-736

³Laboratory of Cheminformatics and Drug Design, Department of Pharmaceutical Sciences, College of Pharmacy, Howard University, Washington, District of Columbia 20059

Abstract

Recent highly expected structural characterizations of agonist-bound and antagonist-bound beta-2 adrenoreceptor (β 2AR) by X-ray crystallography have been widely regarded as critical advances to enable more effective structure-based discovery of GPCRs ligands. It appears that this very important development may have undermined many previous efforts to develop 3D theoretical models of GPCRs. To address this question directly we have compared several historical β 2AR models *versus* the inactive state and nanobody-stabilized active state of β 2AR crystal structures in terms of their structural similarity and effectiveness of use in virtual screening for β 2AR specific agonists and antagonists. Theoretical models, including both homology and *de novo* types, were collected from five different groups who have published extensively in the field of GPCRs modeling; all models were built before X-ray structures became available. In general, β 2AR theoretical models differ significantly from the crystal structure in terms of TMH definition and the global packing. Nevertheless, surprisingly, several models afforded hit rates resulting from virtual screening of large chemical library enriched by known β 2AR ligands that exceeded those using X-ray structures; the hit rates were particularly higher for agonists. Furthermore, the screening performance of models is associated with local structural quality such as the RMSDs for binding pocket residues and the ability to capture accurately most if not all critical protein/ligand interactions. These results suggest that carefully built models of GPCRs could capture critical chemical and structural features of the binding pocket thus may be even more useful for practical structure-based drug discovery than X-ray structures.

Keywords

GPCRs modeling; crystallography; beta-2 adrenoreceptor; agonist-bound; antagonist-bound; molecular docking; enrichment factor

*Correspondence to: Xiang Simon Wang, 2300 4th St. NW, Washington, DC 20059. xiang.wang@howard.edu; Alexander Tropsha, CB #7360, Beard Hall, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7360. alex_tropsha@unc.edu.

Introduction

Structure based drug discovery (SBDD) has become a major strategy in identifying novel leads for important biological targets. SBDD enabled well documented discovery of several approved drugs, e.g. dorzolamide and imatinib. Typically, the 3D structures of biomolecules obtained by the means of X-ray crystallography and NMR spectroscopy are needed for the purpose of virtual screening (VS), whose performance is strongly affected by the quality of biomolecular structure, especially with respect to binding site description. When no experimentally determined structures are available, theoretical models based on either homology(comparative) or *de novo* modeling approaches are employed instead^{17, 21, 34}. However, there were some debates about the accuracy and applicability of theoretical models^{10, 11, 20, 28, 30, 40} in SBDD. In principle, the success of theoretical modeling is typically measured by how close the models could reproduce experimental structures, which implies that the latter are regarded as inherently more appropriate choice for SBDD applications.

G-protein coupled receptors (GPCRs) is a protein family where theoretical models have been used widely because of GPCRs' importance as targets for many drugs, on one hand, and the lack of experimental structures, until recently, on the other hand. The experimental effort to characterize the 3D structure of GPCRs had been seriously hindered by membrane related issues⁸. A large number of theoretical models have been employed in the past decades^{22, 24, 35, 47, 54, 59, 64} for VS often yielding reasonable results^{7, 12, 18, 19, 33}. However, it was fully expected that the availability of the experimental structure of any GPCRs would substantially enhance the efficacy of SBDD efforts. Thus, the recent characterizations of the crystal structures of human beta-2 adrenergic GPCR (β 2AR) in both agonist-bound^{49, 52} and antagonist-bound states^{13, 50, 51}, cleared the way for the validation of previous theoretical models, as well as provided critical data for building homology based models of other GPCRs as the most obvious structural template.

It has been shown that human β 2AR features a structurally conserved rhodopsin-like 7TM core, but there exist novel structural features that had not been identified previously. It remained unclear as to whether these structural divergences would affect the outcome of VS studies. Dissimilarity of theoretical models relative to the crystal structure should lead to relatively poorer performance of the former in docking and scoring of known ligands; however, this general expectation should not necessarily be regarded as a law. One should take into account that some of the theoretical models are actually manually refined with known medicinal chemistry data and therefore, there is at least a possibility that theoretical models may be even more suitable for drug discovery by VS than the crystal structure.

In this study, we have addressed this, both scientifically and pragmatically, important question directly. We have compared the X-ray structure of β 2AR vs. several previously built theoretical models in terms of their respective ability to recover known β 2AR ligands (both agonists and antagonists) from a large external compound library in VS experiments. None of these models were generated in our group to ensure objective and unbiased comparisons. Furthermore, although our group has developed both scoring functions⁶³ and virtual screening protocols⁴⁴, for the same reasons we restricted ourselves to using several

popular commercial docking and scoring tools developed elsewhere. Thus, by design, this study lacked any user biases concerning preferred theoretical models or most familiar computational tools to emulate the situation that is most commonly faced by the majority of molecular modeling practitioners both in academia and industry.

There have been previous studies on comparing homology models among themselves or even with the crystal structure but to the best of our knowledge nobody approached the question poised herein as retrospectively and broadly as we did. For instance, Bissantz *et al* employed three human GPCRs models derived from rhodopsin for virtual screening using multiple docking programs and scoring functions¹⁰. Their work proved that homology models are suitable for VS but there was no comparison to the crystal structure. After the crystal structure of human β 2AR was published, Costanzi reported two studies where carazolol was docked both into two rhodopsin-based homology models of human β 2AR as well as into its X-ray structure of inactive state^{16, 57}. The models afforded high accuracy of the docking poses and ideal enrichment for both antagonists and agonist, especially after incorporating the biochemical data to adjust the orientation of the binding pocket residues. However, these studies were limited to their in-house models and lacked of global structural comparison. More recently, Fan *et al* reported that for 27 out of the 38 protein targets, the consensus enrichment for multiple homology models was better than or comparable to that of both the holo- and the apo- X-ray structures²⁰. However, that study was focused on soluble protein targets and applied a single homology model building tool that employed X-ray characterized structural templates. In contrast, all models included herein were built before the β 2AR crystal structures became available as possible templates. Michino et al. and Kufareva et al. recently reported two large-scale assessments of current GPCR modeling efforts and molecular docking capacities^{37, 41}. While their studies provided important insights on how to improve current structure prediction and docking techniques to reproduce the X-ray structures, our studies focus primarily on the pragmatic question as to whether computational models of GPCRs how they could be used successfully for structure based drug discovery.

We have carried out a systematic, retrospective study on a large collection of published human β 2AR theoretical models and evaluated their structural accuracies and virtual screening performances in comparison with three crystal structures, i.e., 2RH1 (released by the RCSB Protein Data Bank (PDB) on Oct. 30, 2007¹³), 3D4S (released by RCSB PDB on Jun. 17, 2008²⁶), and the latest agonist-bound structure 3P0G (released by RCSB PDB on Jan. 19, 2011⁴⁹). Two other structures, 2R4R and 2R4S⁵⁰, were not employed because they came from the same source as 2RH1 but with lower resolutions. Another agonist-bound β 2 adrenoceptor structure, 3PDS with an irreversible disulfide bond between ligand and complex, was not used because its conformation is more close to the inactive state⁵². As shown in Table I, we collected eight independently published theoretical models of human β 2AR including both apo and holo structures. Both agonist and antagonist bound models were included to account for any structural features associated with functional activity. Furthermore, we incorporated both homology and *de novo* models to cover these two major types of GPCRs modeling.

Surprisingly, we found that some of the theoretical models displayed better or comparable VS performances than the crystallographic structures. This study by no means undermines the extreme significance of the X-ray structures of β 2AR as well as other GPCRs^{13, 29, 45, 46, 60, 62} in understanding the intricate details of GPCRs structure in relation to its function nor in the significance of X-ray structures for SBDD. Nevertheless, it most certainly testifies to the importance of intelligent computational modeling approaches especially those incorporating comprehensive medicinal chemistry knowledge of receptor/ligand complex for structure based virtual screening.

Materials and Methods

Structural similarity analysis

In addition to crystal structures of inactive state (PDB ID: 2RH1) and active state (PDB ID: 3P0G) of human β 2AR, the structure of bovine rhodopsin (PDB ID: 1U19) in dark state was also included in the analysis because it had been used as the major template for all β 2AR homology models. The structural similarity was assessed in three aspects, i.e. the accuracy of the boundary definitions for each transmembrane (TM) helices, the backbone root-mean-square-deviation (RMSD) for TM regions, and the C α RMSD of the binding pocket residues. The numbering of amino acids followed the conventions set by Weinstein et al⁵⁸. The highly conserved residues embedded in each TM region were employed as anchors for the alignment. Each theoretical model was structurally aligned against 2RH1, 3P0G or 1U19 by individual TM helix as well as the whole TM bundle. The RMSDs were calculated using the entire lengths of the corresponding segments of 2RH1.

Screening libraries

Active seeds—To include most known binders, we collected 57 known antagonists of human β 2AR reported in DrugBank⁶¹ and GLIDA⁴³ databases as active seeds. All antagonists have sub-micromolar potency and can be found in additional external databases, such as PDSP Ki⁵³, PubChem⁵ and KEGG³¹. A few of them are β 2AR specific (e.g. butoxamine and aminoflisopolol) while others can act on both β 1AR and β 2AR. In the meantime, we compiled the agonist set of thirteen full agonists for human β 2AR from the DrugBank database.

Decoys set I: PDSP binding decoys plus WDI drug-like decoys—As our initial effort to establish a reasonable decoy dataset for virtual screening, we employed the World Drug Index (WDI) database version 2004¹ since most of its compounds are drug-like. The original collection of 59,000 molecules was first cleaned by removing metals, salts and fragments, then filtered to eliminate unqualified compounds according to Lipinski's rule of five and later extensions of this rule²⁵. The remaining collection of ca. 38,000 compounds was further reduced to a diverse subset of 374 compounds using MOE2007.09. In addition, 12 binding decoys with similar chemical scaffolds but poor binding affinity ($K_i > 10\mu\text{M}$) were selected from the PDSP Ki database and merged into the WDI diverse subset. In the end, 57 human β 2AR antagonists and 13 agonists were seeded amongst 386 decoys separately to constitute two different screening libraries.

Decoys set II: ChEMBL binding decoys plus Schrodinger drug-like decoys—In an effort to avoid the bias brought by one single screening library, we built another decoys set which includes 1000 drug-like decoys²³ from Schrodinger web portal and 916 known nonbinders from ChEMBL database (<https://www.ebi.ac.uk/chembl/>). The size of this decoys set is fairly large and the ratio of actives to decoys is ($r=33.6$ for β 2AR antagonists) close to the criteria set by the DUD database ($r=36.0$). The ChEMBL binding decoys were selected based on the standards similar to the ones in the decoys set I for PDSP database but of large number.

Decoys set III: Subset of II containing N+ only—Like other amine GPCRs families, β 2AR's ligands are special in that they normally contain a positively charged nitrogen at the side chain portion, which is a key ingredient of the binding event. To ensure that the power of current scoring function is based upon factors other than this simple structural feature, we created the 3rd decoys set by applying this filter (N+ present) to the decoys set II. In the end, there were 652 compounds left with positively charged nitrogen.

Molecular docking methods

We employed three popular docking programs, i.e. Glide4.01, AutoDock4.0 and eHiTS6.2, to evaluate systematically the screening performance of structural models.

Glide4.01—The calculations by Glide version 4.01²³ was carried out using Schrodinger Suite 2007. The targeted protein and theoretical models were prepared through *Protein Preparation* module with the default setting and assigned with the OPLS 2001 force field atom types and partial charges. The screening databases were prepared within the *LigPrep* module and the ionization states of each molecule were calculated as to be compatible with the pH value of 7.0 ± 2.0 . All molecules were subjected to energy minimization with MMFFs force field before the docking computation. For X-ray structure and holo models, the center of the grid box was selected as the center of bound ligands. For apo models, their binding pockets were first aligned to that of 2RH1 and the center of co-crystallized carazolol was chosen. The proper size of the enclosing box was not set to be fixed but determined by the extent of the bound ligand. The Glide SP scoring function was used to rank the docking poses and the top-ranked poses for each database molecule were saved for post-docking analysis.

AutoDock4.0—We prepared the targeted protein and docking parameters for AutoDock version 4.0^{27, 42} using the AutoDockTools graphic interface. Explicit hydrogen atoms were added to the receptor structures while atom types and partial charges were assigned to generate the pdbqt receptor files. The database molecules were prepared using the 'prepare_ligand4.py' script to merge non-polar hydrogen atoms and define flexible torsions. The center and dimension of the enclosing boxes were defined to include the whole binding pockets, similar to those in the Glide docking. The genetic algorithm were employed during the docking with a start population size of 150 individuals and 20 runs combined with a maximum number of 12,500,000 energy evaluation for each molecule. Other parameters for genetic algorithm were kept by the default value.

eHiTS6.2—The eHiTS version 6.2⁶⁵ was used through the CheVi user interface. Protein preparations, such as protonation state determination for residues, hydrogen atoms addition and partial charge assignment, were actually not needed since eHiTS' docking and scoring are based on the prior training data of its knowledge base of 97 protein families. The grid box was assigned automatically using the bound ligand's SDF file as the CLIP file. We employed the default settings for eHiTS docking and ranked the database molecule based on its lowest eHiTS score.

In addition, we added eight more scoring functions of different types by rescoring the top-ranked poses generated by Glide4.01, to ensure an unbiased and complete comparison. The multiple scoring functions in Sybyl8.0⁴ CScore module and OpenEye² FRED 2.2.4 were applied, including Chemscore, D_score, Gold_score, PMF, Chemgauss3, PLP, Screenscore and Shapegauss. The consensus scores were also used for the above scoring functions through the rank by rank strategy.

Assessment of virtual screening performance

To measure the efficiency of virtual screening we used the following conventional parameters: the enrichment factor and the receiver operating characteristic (ROC) curve that characterizes the ability if a method to recover known ligands among the top-scored screening molecules. The enrichment factor follows the most popular definition as to how many more seed compounds (*i.e.*, known ligands) were found within a defined “early recognition” fraction of the ranked list relative to a random distribution:

$$EF = H_{scr} / H_{tot} \times D_{tot} / D_{scr} \quad (1)$$

where H_{scr} is the number of target-specific seeds recovered at a specific % level of the database; H_{tot} is the total number of seeds for the target; D_{scr} is the number of compounds screened at a specific % level of the database; D_{tot} is the total number of compounds of the database. The ROC curve is generated by plotting the sensitivity (Se) vs. $(1 - \text{specificity } (Sp))$ for a binary classifier system as its discrimination threshold is varied. In the case of virtual screening for recovering the i th known active from the inactive decoys, the Se_i and Sp_i are defined as follows:

$$Se = H_{scr} / H_{tot} \quad (2)$$

$$Sp = \frac{(D_{tot} - H_{tot}) - (D_{scr} - H_{scr})}{D_{tot} - H_{tot}} \quad (3)$$

The area under the ROC curve (AUC) is the metric that is widely accepted for assessing the likelihood that a screening method assigns a higher rank to known actives than to inactive compounds. The AUC values at a specific percentage of the ranked database are calculated from the following equation:

$$AUC = \sum_{i=1}^n [Se_i (Sp_{i+1} - Sp_i)] \quad (4)$$

Here n is the total number of known actives in the screening database. One additional parameter, the yield, is also employed as the percentage of true hits retrieved by the virtual screening method:

$$Yield = H_{scr} / H_{tot} \times 100 \quad (5)$$

Cluster analysis of binding profiles

To closely evaluate the key receptor/ligand interaction patterns, we employed the LigX module¹⁴ in MOE2007.09 to analyze the crystal structure of β 2AR/carazolol complex and the docking poses generated by Glide4.01. For each antagonist, the top-ranked docking pose with the highest score was selected. Two major types of interactions that contribute to protein/ligand binding affinities were considered, i.e. hydrogen bonds (donor or acceptor) and non-bonded weak interactions. The score to assess the hydrogen bond is based on a scale of 0 to 100% that indicates the probability of being a geometrically perfect hydrogen bond while the score for non-bonded weak interaction is the pairwise distance between residue and ligand atoms. In our studies, we took the default parameters in which 4.5 angstrom is the cutoff for weak interactions and 2.5 angstrom is the closest distance between any residue/ligand atom pairs. The original score was normalized; thus, the values of the modified scores were between 0 and 1, which is proportional to the interaction intensities. To better visualize the binding patterns of docked poses for each theoretical models and crystal structures, the LigX scores were transformed into heat maps and clustered using the R statistical package⁴⁸. We applied the hierarchical clustering with the Ward linkage algorithm; thus, the patterns of interaction between 57 human β 2AR antagonists and residues in the active sites of three different structural models would be expected to be similar if the respective clusters are similar.

In addition, we have exploited the Protein Ligand Interaction Fingerprints (PLIF, also available in MOE2007.09) for the same purpose. PLIF can identify and score major protein/ligand interactions, including hydrogen acceptor from sidechain, hydrogen donor to sidechain, hydrogen acceptor from backbone, hydrogen donor to backbone, ionic attractions and surface contacts. For each docking pose, the PLIF fingerprints ranging from 30 to 50 bits were generated. The relative frequencies of each identified fingerprint can be then used to produce fingerprint significance chart, which is based on the hypothesis that 'if the bit is set, then the compound is active'.

Results and Discussion

Structural comparison of theoretical models and X-ray structures of β 2AR

Prior to VS experiments, we analyzed the similarity between theoretical models and two X-ray structures, i.e. 2RH1 and 1U19 deposited to the Protein Data Bank⁹. All theoretical models used in this study are listed in Table I; each model was aligned against the X-ray structures to evaluate relative definitions of transmembrane (TM) helices, their conformations, and relative orientation. The X-ray structure of bovine rhodopsin (1U19) was also included in addition to β 2AR because the former had been used as a common template for GPCRs homology modeling. In doing so, we were interested to explore if the failure to

predict the structural conservation and/or divergence from the template structure may cause poor VS results for the theoretical models.

Our initial efforts focused on evaluating the accuracy of TM helical boundaries as defined by models vs. X-ray structures. For this purpose we have employed MOE 2007.09 software³ to annotate the secondary structural elements in the 3D structures. We found that the MOE module assigned boundaries either at exact positions or only one amino acid apart to over 90% of the TM segments of 2RH1 and 1U19 in comparison with those in the PDB header. Thus, the MOE software was deemed reliable in identifying the helical boundaries; the results of applying MOE to the six theoretical models and two crystal structures (2RH1 and 1U19) are summarized in Figure 1a and Table II. From the alignment of the eight structures, it can be seen that the apo models (AM1-AM3) perform better than the holo models (CM1-CM3) in terms of accuracy of TM assignment. The location and length of the TM helices for all three AM models are consistently close to those in the rhodopsin structure (1U19), with the only exception that TM6 and TM7 in the AM1 model are shorter than the corresponding helices in the crystal structures. This observation can be easily rationalized since AM1 and AM2 models are solely based on homology modeling whereas AM3 is a hybrid model developed with a combination of both threading and *ab initio* methods. In all cases, the crystal structures of bovine rhodopsin were used as a template for model building. Furthermore, we have concluded that individual TM helices were very close in terms of helical length and relative orientation when compared to crystal structures of bovine rhodopsin and human β 2AR¹³. Thus, not surprisingly, given the methods used for model building, the secondary structural elements for all three AM models were found to be assigned very accurately as compared to their homologous experimental structures.

The accuracies of TM helix boundaries assignments for CM models were less satisfactory. In general, seven TM helices in all three CM models were shorter than expected with the largest disagreements located at TM1, TM4 and TM5 (cf. Figure 1a and Table II). The CM3 model gave the largest deviation in terms of the percentage of correctly defined TM helical boundaries. Moreover, it had the shortest lengths for individual TM helices. For instance, it was eight residues shorter for TM1, fourteen for TM3 and eleven for both TM5 and TM6 in comparison with the β 2AR crystal structure. One possible explanation is that all three CMs are *de novo* models, generated without any template structure. Both the Lybrand (CM1, 2) and the Goddard (CM3, 4) groups employed the standard alpha helix as a starting point and calculated the intrinsic tilt/kink and relative orientation of the TM helical bundle purely based on the physical considerations. If the rhodopsin structure is not employed as a reference, the secondary structure assignments could be affected by many factors, such as the type of phospholipid used in the MD simulations employed as part of model refinement in studies by the Goddard group^{22, 54}.

It should be noted that the engineered modification of the wild type protein using a segment of T4 lysozyme to replace most residues of IL3 introduced an artifact in the crystal structure (2RH1) of human β 2AR. This modification led to altering the boundaries of IL3, thus affecting the correct locations of both the TM5 terminus and the start of the TM6. The accurate definition of these two boundaries is less important in the comparison of structural similarities between theoretical models and the crystal structures. However, the accuracy of

predicting the TM4 terminus and the beginning of TM5 is critical considering the functional roles of EL2 in both rhodopsin and human β 2AR^{6, 55}. As can be seen from Figure 1a and Table II among the three AM models, AM3 has the highest accuracy (one residue error) for the segment between TM4 and TM5 (EL2) followed by AM2 and AM1. CM3 model is comparable to AM2 while CM1 and CM2 had much larger errors with respect to TM helix assignments.

The seven TM helices of each theoretical model were superimposed onto respective helices of β 2AR as well as the rhodopsin structure, and the backbone pairwise RMSD of individual respective TM helices was calculated (Figures 1b, 1c and Table II). As expected, the homology models (AMs) are generally more similar to the rhodopsin structure than to the β 2AR structure. The RMSDs of most helices in AM1-3 range from 0.30 Å to 1.00 Å as compared to bovine rhodopsin where the RMSDs are as big as 1.60 Å to 3.80 Å when aligned against the human β 2AR. For the whole TMs bundle, the RMSDs are 1.15 Å to 1.88 Å with respect to rhodopsin and 2.25 Å to 3.19 Å with respect to β 2AR. In comparison, the *de novo* models (CMs) deviate more significantly from both crystal structures. The RMSDs of most helices in CM1/CM2 are in the same range of 2.00 Å - 5.50 Å when aligned against both rhodopsin and β 2AR. For CM3, the RMSD increases to 2.60 Å and 7.00 Å, respectively. Similarly, the RMSDs of TMs bundle for CM1 are 3.20 Å when aligned against the rhodopsin and 3.59 Å for β 2AR. For CM3, the corresponding RMSDs are 3.83 Å and 4.41 Å (cf. Table II).

Consistently, the AMs models were observed to be aligned well to the rhodopsin structure (cf. AM2 in Figure 1b as an illustrative example). They deviated from the β 2AR structure with the noticeable shifts for TM1, TM3 and TM5, although the secondary structure assignment for these regions was relatively accurate. The large RMSD differences for these three TMHs were obviously due to the differences between rhodopsin and β 2AR crystal structures (7TMs backbone RMSD of 2RH1 vs. 1U19 is 1.85 Å). It was indeed reported in the original publication on β 2AR crystal structure¹³ that there is a noticeable shift in TM1 of β 2AR relative to bovine rhodopsin, primarily at the extracellular portion which tilts away from the TM bundle compared to bovine rhodopsin^{13, 51}. The long N-terminal fragment could not be observed in both crystal structures, but it could cause large flexibility/variation in the assignment of TM1 boundaries, especially in the upper helical region. TM5 has a proline-induced kink at conserved positions along the transmembrane segments, which is believed to be responsible for the structural rearrangements required for the GPCRs activation^{32, 36, 39, 58}. The subtle difference in the activation status of the current β 2AR structure (2RH1, bound to an inverse agonist carazolol) may lead to the structural diversification at the kink region, in terms of the amplitude of motion and rotation degree. Notably TM3 and TM5 constitute half of the binding pockets for the co-crystallized carazolol¹³ (Figure 1b).

Unlike AM models, three CMs deviate from both bovine rhodopsin and β 2AR in a similar way (cf. CM1 in Figure 1c as the representative case). A large discrepancy can be found at TM1, 4, 6 and 7 for CM1/CM2 and TM1, 3, 5 and 6 for CM3. Here the similar reasoning used in the analysis of AMs can be applied to TM1 because the N-terminal fragment was not considered as part of model building and optimization. Interestingly, the RMSD of CM1's

TM4 is as large as 5.00 Å with respect to rhodopsin and 5.21 Å when aligned against β2AR. It is surprising since TM4 seems the easiest one to model. Among all seven TMs, TM4 is the shortest and the most orthogonal to the plane of the phospholipid membrane. This observation indicates the limitation of computational protocols employed in developing the CM models in their ability to reproduce the conformation of this ‘anchor’ helix in the TM bundle. For TM5, 6 and 7, one of the common shared features is the proline-induced kink. The comparisons between β2AR models and the crystal structures highlight the difficulty associated with the accurate modeling of this unique structural feature of GPCRs.

Comparison of VS performance for theoretical models and X-ray structures of β2AR

Figure 2 shows the enrichment factor plot (a,c,e) and ROC curves (b,d,f) of 57 known β2AR antagonists against decoys by three docking methods. The yield plot is shown since it essentially delivers the same information as the ROC curves. The detailed statistical parameters characterizing the VS performance, such as the maximum EF (EF_{max}), ROC AUC and the recently proposed Boltzmann-Enhanced Discrimination of ROC (BEDROC)^{15, 56}, are summarized in Table III. All four holo models as well as the β2AR structure were used in the VS study. Because of the poor performance of the apo models in general during virtual screening, we only showed the data for the relatively better ones (AM1 and AM3) to represent the AMs group. In many cases, such as AM3 in Figures 2d and 2f, the ROC curve is close to the random expectation (the diagonal line). It is understandable because the side chain rotamers of binding pocket residues in the AM models had not been optimized in the way it was done for holo models.

Among all four CMs models, those from the Lybrand group (CM1, CM2) achieved better enrichment than the models from the Goddard group (CM3, CM4). In most cases, the CM2 model yielded comparable results to the β2AR crystal structure. In the screening by Glide4.01, CM2 model gave higher EF at the very early phase (0% - 2% of ranked database, cf. Figure 2a and Table III). After that, it remained up to one unit lower than the crystal structure until converging with the latter at the 22% of ranked database. In terms of ROC AUC, CM2 reached the value of 0.86, close to AUC of 0.88 for the crystal structure. Based on the BEDROC metric, CM2 was similar to 2RH1 when α was small ($\alpha=20$) while showing better performance when α was large ($\alpha=53.6$ or 100). We note that Glide4.01 gave better VS results in this study compared to AutoDock4.0 and eHiTS6.2. Thus, we placed more emphasis on the screening data/docking poses generated by Glide4.01. Nevertheless, the results obtained with both AutoDock4.0 and eHiTS6.2 also highlighted the impressive performance of CM2. As demonstrated by both types of plots in Figures 2c-f and most criteria in Table III, the CM2 afforded VS results superior to the crystal structure when using these two docking programs. Therefore, practically speaking, these results suggest that the use of crystal structure is not advantageous in terms of VS performance when the scoring function is not highly accurate. On the other hand, it reflects on the extreme sensitivity of the docking approaches to small structural variations. As mentioned above, CM3 and CM4 had poorer performance than CM1 and CM2 but were comparable to AM1 and AM3 models in this case (cf. Table III).

The crystal structure of β 2AR represents an inactive state of the receptor because it is bound to the inverse agonist, carazolol^{13, 38}. Thus, it may be considered unfair to compare the crystal structure of β 2AR with theoretical models as applied to agonist screening, especially when the models were created to capture known data on agonists. However, for comparison purposes, we did explore the possible utility of 2RH1 for screening for agonists. The screening results are summarized in Figures 3 and Table IV. As expected, the CM2 model showed the best performance to enrich for thirteen β 2AR agonists. With Glide4.01 method, the CM2 model could recover 100% of seed agonists at the 15% of ranked databases and its maximum EF could be as high as 36.09. Thus, it excelled over 2RH1 greatly in terms of these two parameters of VS performance. Taking into account the data for the antagonist virtual screening, we shall conclude that CM2 model demonstrated remarkable performance as a model of choice for virtual screening for both agonists and antagonists.

Figure 4, Table V and Table VI showed the results of virtual screening for antagonists and agonists with decoys sets II and III. We observed that the relative performance between crystal structures and theoretical models with new sets is similar to that observed in our previous experiments, although the absolute performance of each protein model varied by the choice of decoy sets. Notably the CM2 model was consistently ranked comparable to three experimental structures to enrich fifty-seven β 2AR antagonists (cf. Figure 4 and Table V). And remarkably it outperforms structures 2RH1 and 3D4S for thirteen β 2AR agonists on most aspects of screening, and is comparable to the structure of 3P0G at the late phase of screening (> 25% db, cf. Figure 4 and Table VI). Thus we believe that the conclusions we drew from the decoys set I are valid and expandable to other screening libraries, thus appear to be general.

The possible explanation of the better performance of CM1/CM2 models is that Lybrand *et al* exploited many site-directed mutagenesis data during the model optimization^{24, 35}. The important receptor/ligand interactions had been turned into distance restraints that were applied explicitly to specific atoms of both the receptor and its ligands during molecular dynamics simulations²⁴. In comparison, CM3 and CM4 models from Goddard *et al* did not employ such information^{22, 54}; their models were built by optimizing the target/ligand interaction using physical force field. Obviously, the differences in the type of data utilized for theoretical model building and optimization can largely affect the accuracy of binding pocket modeling, and consequently, the model performance in virtual screening experiments. In order to evaluate the similarity between binding pockets of individual models, we superimposed C α atoms of key residues inside the pocket with their counterparts in 2RH1. The binding pocket was defined by residues found within 4 Å of the co-crystallized carazolol. Carazolol was merged into the binding sites of all models as defined by the alignment. As shown in Figure 5 (A-H), the CM2 binding pocket (Figure 5B) is most similar to that of 2RH1 with respect to both the ligand pose and the position of residues interacting with the ligand. The RMSD for its C α atoms was 2.40 Å while the one for CM1 was 2.33 Å (cf. Table II). These two models also reproduced the contacts of carazolol with residues Ser203^{5,42}, Asn312^{7,39} and Phe193. For three AMs models, the RMSDs ranged from 3.39 Å to 3.71 Å. CM3 and CM4 models had the largest deviation (RMSD = 5.64 Å), as can also be seen in Figures 5C and 5D.

Furthermore, the close inspection of the top-ranked docking poses of all seed antagonists showed that the interactions between the antagonists and the binding site of the CM2 model were largely in agreement with the site-directed mutagenesis data. The protonated nitrogen in most β 2AR antagonists formed salt bridges with ASP113^{3.32} and Asn312^{7.39}; the amide hydroxyl group formed hydrogen bonds with Ser203^{5.42}, Ser204^{5.43} or Ser207^{5.46}. Another important interaction was formed between antagonists and Phe193 of EL2, i.e., the residue that was also found to interact with carazolol within the crystallographic structure of β 2AR^{13, 24, 51}. It should be pointed out that CM2/CM1 models include both extracellular and intracellular loops, whereas CM3/CM4 and AM1 models did not incorporate these regions^{22, 51}.

To elucidate the molecular basis for dissimilar virtual screening performance of different CM2 and CM3 models in comparison with 2RH1 we have conducted the cluster analysis of the binding profiles of all 57 antagonists docked to the respective binding sites. Binding profiles reflected the strengths of interaction between antagonists and active site residues. Importantly, there were significant differences in the weak interaction patterns of CM2 (Figure 6b) and CM3 (Figure 6c) in comparison with 2RH1 (Figure 6a). The major clusters formed by CM2 in the region of conserved residues matched well to those found in 2RH1, suggesting a critical role of weak interactions between binding site and antagonists. In comparison, clusters formed by CM3 were scarce and many key interactions were missing, especially for residues Trp286^{6.48}, Phe289^{6.51}, Phe290^{6.52} and Asn293^{6.55} of the binding pocket. This analysis indicates that the binding pocket of CM2 was well-organized and similar to that of 2RH1 whereas the CM3 binding pocket was formed by somewhat different residues, with the key residues found in the binding site of the X-ray structure were inaccessible to the bound antagonists. The cluster profile of hydrogen bonding pattern was less informative as some prominent patterns at 2RH1 such as the ones with Asn312^{7.39} were absent at both CM2 and CM3. Notably, Ser204^{5.43} and Ser207^{5.46} of CM2 were found to be hydrogen bonded to *ca.* 15 antagonists, but the same pattern was not observed with either CM3 or 2RH1.

During the course of our studies, Kobilka *et al.* experimentally characterized two new structures of β 2AR, one in a nanobody-stabilized active state and another in complex with an irreversible agonist. In comparison with the structures of inactive state of β 2AR, the agonist-binding pockets showed fairly subtle changes, with the differences at the hydrogen bonding contacts with Ser203^{5.42} and Ser207^{5.46} residues^{49, 52}. To determine whether these minor changes will increase the receptor's selectivity toward agonists, we carried out structure-based virtual screening studies using the nanobody-stabilized structure (PDB ID: 3P0G), and compared its performances with the inactive state structure as well as our collection of theoretical models. As expected, the structure of the active state showed better performances than its counterparts of the inactive state in enriching for thirteen agonists but less effective in enriching for fifty-seven antagonists (Figures 2a-b, 3a-b). With Glide4.01 docking protocol, the active state structure could recover 100% of seed agonists at the 25% of ranked databases with the maximum EF as high as 30.69. On the other hand, the active state model of CM2 showed comparable performances to the 3P0G structure in terms of EF and AUC metrics (cf. Table III and Table IV). As mentioned above, the *de novo* model of

CM2 had captured the critical agonist-protein interaction at Ser207^{5,46} (cf. Figure 5G). This piece of evidence provides another illustration that the properly optimized theoretical models can provide a sensible and accurate description of the binding pocket of the active state structure, and therefore they can be employed reliably for structure-based virtual screening of β 2AR agonists.

To summarize our observations, we established that theoretical models of GPCRs built by knowledge-based approaches can achieve similar if not better VS performance as experimental structures from X-ray crystallographic studies. This somewhat surprising observation is reassuring with respect to using carefully developed theoretical models of protein structures for SBDD.

Conclusions

In this study we have addressed the long-standing debate about the structural accuracy and applicability of theoretical models vs. X-ray structures of GPCRs for SBDD. We have carried out a systematic study on a large collection of historical human β 2AR theoretical models and evaluated their structural accuracies and screening performances in comparison with recent agonist-bound and antagonist-bound crystal structures. We have shown that there exists a discrepancy between global structural accuracies of β 2AR theoretical models and their screening performances. In general, β 2AR theoretical models differ largely from the crystal structure in terms of TMHs definition and global packing while many can achieve the same performance in virtual screening and as demonstrated elsewhere¹⁶, pose predictions. Our analysis indicates that the binding pockets of models showing the best performance are well-organized and they also align well to active sites in the crystal structures. The key interactions of residues in the active site with the bound antagonists were found to be preserved in models that were built and refined taking into account the site-directed mutagenesis and other experimental data (i.e. the CM1 and CM2 models). Our results emphasize that knowledge-based approaches result in structural models that can achieve the same or even better performance in virtual screening as those built with X-ray crystallographic data. Also, we must stress that our studies address very specific (i.e. β 2AR or GPCRs) and pragmatic question (i.e. virtual screening) concerning the use of protein models vs. experimental structures for virtual screening. Our observations by no means undermine the critical importance of experimental structures for understanding protein structure-function relationships as well as the role that crystal structures serve as a critical reference for evaluating the accuracy of predicted protein/ligand interactions.

Acknowledgments

We would like to thank Drs. Gert Vriend, Andrej Sali, Jeffrey Skolnick, Terry P. Lybrand, and Nagarajan Vaidehi for sharing the coordinates of their respective theoretical models of beta-2 adrenergic receptor. We are also grateful to Dr. Brian K. Kobilka for helpful discussions. Finally, we acknowledge the access to the computing facilities at the ITS Research Computing Division of the University of North Carolina at Chapel Hill.

Grant sponsor: National Institutes of Health, USA; Grant numbers: GM066940, P20-HG003898, G12 RR003048;

Grant sponsor: UNC-CH University Research Council; Grant number: A3-12988

References

1. World Drug Index (WDI). 2004
2. OpenEye scientific software. 2006
3. MOE. Chemical Computing Group; 2007.
4. sybyl8.0. 2007
5. Austin CP, Brady LS, Insel TR, Collins FS. NIH Molecular Libraries Initiative. *Science*. 2004; 306:1138–1139. [PubMed: 15542455]
6. Avlani VA, Gregory KJ, Morton CJ, Parker MW, Sexton PM, Christopoulos A. Critical role for the second extracellular loop in the binding of both orthosteric and allosteric G protein-coupled receptor ligands. *J Biol Chem*. 2007; 282:25677–25686. [PubMed: 17591774]
7. Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O, Bar-Haim S, Warshaviak D, Fichman M, Noiman S. G protein-coupled receptors: in silico drug discovery in 3D. *Proc Natl Acad Sci U S A*. 2004; 101:11304–11309. [PubMed: 15277683]
8. Becker OM, Shacham S, Marantz Y, Noiman S. Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Curr Opin Drug Discov Devel*. 2003; 6:353–361.
9. Berman HM, J W, Z F, G G, TN B, H W, IN S, PE B. The protein data bank. *Nucleic acids research*. 2000; 28:235. [PubMed: 10592235]
10. Bissanz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins*. 2003; 50:5–25. [PubMed: 12471595]
11. Bordogna A, Pandini A, Bonati L. Predicting the accuracy of protein-ligand docking on homology models. *J Comput Chem*. 2011; 32:81–98. [PubMed: 20607693]
12. Chen JZ, Wang J, Xie XQ. GPCR structure-based virtual screening approach for CB2 antagonist search. *J Chem Inf Model*. 2007; 47:1626–1637. [PubMed: 17580929]
13. 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; 318:1258–1265. [PubMed: 17962520]
14. Clark AM, Labute P. 2D depiction of protein-ligand complexes. *J Chem Inf Model*. 2007; 47:1933–1944. [PubMed: 17715911]
15. Clark RD, Webster-Clark DJ. Managing bias in ROC curves. *J Comput Aided Mol Des*. 2008; 22:141–146. [PubMed: 18256892]
16. Costanzi S. On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. *J Med Chem*. 2008; 51:2907–2914. [PubMed: 18442228]
17. Ellison KE, Gandhi G. Optimising the use of beta-adrenoceptor antagonists in coronary artery disease. *Drugs*. 2005; 65:787–797. [PubMed: 15819591]
18. 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; 48:1088–1097. [PubMed: 15715476]
19. Evers A, Klebe G. Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J Med Chem*. 2004; 47:5381–5392. [PubMed: 15481976]
20. Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A. *J Chem Inf Model*. 2009
21. Fanelli F, De Benedetti PG. Computational modeling approaches to structure-function analysis of G protein-coupled receptors. *Chem Rev*. 2005; 105:3297–3351. [PubMed: 16159154]
22. Freddolino PL, Kalani MY, Vaidehi N, Floriano WB, Hall SE, Trabanino RJ, Kam VW, Goddard WA III. Predicted 3D structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists. *Proc Natl Acad Sci U S A*. 2004; 101:2736–2741. [PubMed: 14981238]
23. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate

- docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem.* 2004; 47:1739–1749. [PubMed: 15027865]
24. Furse KE, Lybrand TP. Three-dimensional models for beta-adrenergic receptor complexes with agonists and antagonists. *J Med Chem.* 2003; 46:4450–4462. [PubMed: 14521408]
 25. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem.* 1999; 1:55–68. [PubMed: 10746014]
 26. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EY, Velasquez J, Kuhn P, Stevens RC. A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure.* 2008; 16:897–905. [PubMed: 18547522]
 27. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem.* 2007; 28:1145–1152. [PubMed: 17274016]
 28. Ivanov AA, Barak D, Jacobson KA. Evaluation of homology modeling of G-protein-coupled receptors in light of the A(2A) adenosine receptor crystallographic structure. *J Med Chem.* 2009; 52:3284–3292. [PubMed: 19402631]
 29. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC. The 2.6 Å crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science.* 2008; 322:1211–1217. [PubMed: 18832607]
 30. Kairys V, Fernandes MX, Gilson MK. Screening drug-like compounds by docking to homology models: a systematic study. *J Chem Inf Model.* 2006; 46:365–379. [PubMed: 16426071]
 31. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28:27–30. [PubMed: 10592173]
 32. Katritch V, Reynolds KA, Cherezov V, Hanson MA, Roth CB, Yeager M, Abagyan R. Analysis of full and partial agonists binding to beta2-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J Mol Recognit.* 2009; 22:307–318. [PubMed: 19353579]
 33. Katritch V, Rueda M, Lam PC, Yeager M, Abagyan R. GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins.* 2010; 78:197–211. [PubMed: 20063437]
 34. Klabunde T, Hessler G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem.* 2002; 3:928–944. [PubMed: 12362358]
 35. Kontoyianni M, DeWeese C, Penzotti JE, Lybrand TP. Three-dimensional models for agonist and antagonist complexes with beta 2 adrenergic receptor. *J Med Chem.* 1996; 39:4406–4420. [PubMed: 8893835]
 36. Kota P, Reeves PJ, Rajbhandary UL, Khorana HG. Opsin is present as dimers in COS1 cells: identification of amino acids at the dimeric interface. *Proc Natl Acad Sci U S A.* 2006; 103:3054–3059. [PubMed: 16492774]
 37. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure.* 2011; 19:1108–1126. [PubMed: 21827947]
 38. Lefkowitz RJ, Sun JP, Shukla AK. A crystal clear view of the beta2-adrenergic receptor. *Nat Biotechnol.* 2008; 26:189–191. [PubMed: 18259173]
 39. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem.* 2003; 278:21655–21662. [PubMed: 12663652]
 40. McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J Med Chem.* 2003; 46:2895–2907. [PubMed: 12825931]
 41. Michino M, Abola E, Brooks CL III, Dixon JS, Moulton J, Stevens RC. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat Rev Drug Discov.* 2009; 8:455–463. [PubMed: 19461661]
 42. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry.* 1998; 19:1639–1662.

43. Okuno Y, Yang J, Taneishi K, Yabuuchi H, Tsujimoto G. GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Res.* 2006; 34:D673–D677. [PubMed: 16381956]
44. Oloff S, Zhang S, Sukumar N, Breneman C, Tropsha A. Chemometric analysis of ligand receptor complementarity: identifying Complementary Ligands Based on Receptor Information (CoLiBRI). *J Chem Inf Model.* 2006; 46:844–851. [PubMed: 16563016]
45. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le T I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science.* 2000; 289:739–745. [PubMed: 10926528]
46. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature.* 2008; 454:183–187. [PubMed: 18563085]
47. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, Shen MY, Kelly L, Melo F, Sali A. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* 2006; 34:D291–D295. [PubMed: 16381869]
48. R Development Core Team. R: A language and environment for statistical computing. Vienna: 2008.
49. 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; 469:175–180. [PubMed: 21228869]
50. Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, Kobilka BK. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature.* 2007; 450:383–387. [PubMed: 17952055]
51. Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science.* 2007; 318:1266–1273. [PubMed: 17962519]
52. 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; 469:236–240. [PubMed: 21228876]
53. Roth BL, Lopez E, Patel ES, Kroeze WK. The Multiplicity of Serotonin Receptors: Uselessly diverse molecules or an embarrassment of riches? *Neuroscientist.* 2006; 6:252–262.
54. Spijker P, Vaidehi N, Freddolino PL, Hilbers PA, Goddard WA III. Dynamic behavior of fully solvated beta2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist. *Proc Natl Acad Sci U S A.* 2006; 103:4882–4887. [PubMed: 16551744]
55. Stern PS, Yu L, Choi MY, Jurenka RA, Becker L, Rafaeli A. Molecular modeling of the binding of pheromone biosynthesis activating neuropeptide to its receptor. *J Insect Physiol.* 2007; 53:803–818. [PubMed: 17512003]
56. Truchon JF, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem. *J Chem Inf Model.* 2007; 47:488–508. [PubMed: 17288412]
57. Vilar S, Ferino G, Phatak SS, Berk B, Cavasotto CN, Costanzi S. Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models. *J Mol Graph Model.* 2011; 29:614–623. [PubMed: 21146435]
58. Visiers I, Ballesteros JA, Weinstein H. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Methods Enzymol.* 2002; 343:329–371. [PubMed: 11665578]
59. Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph.* 1990; 8:52–6. 29. [PubMed: 2268628]
60. 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; 454:486–491. [PubMed: 18594507]

61. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006; 34:D668–D672. [PubMed: 16381955]
62. Wu B, Chien EY, 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; 330:1066–1071. [PubMed: 20929726]
63. Zhang S, Golbraikh A, Tropsha A. Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein-ligand interfaces. *J Med Chem.* 2006; 49:2713–2724. [PubMed: 16640331]
64. Zhang Y, Devries ME, Skolnick J. Structure modeling of all identified G protein-coupled receptors in the human genome. *PLoS Comput Biol.* 2006; 2:e13. [PubMed: 16485037]
65. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. eHiTS: a new fast, exhaustive flexible ligand docking system. *J Mol Graph Model.* 2007; 26:198–212. [PubMed: 16860582]

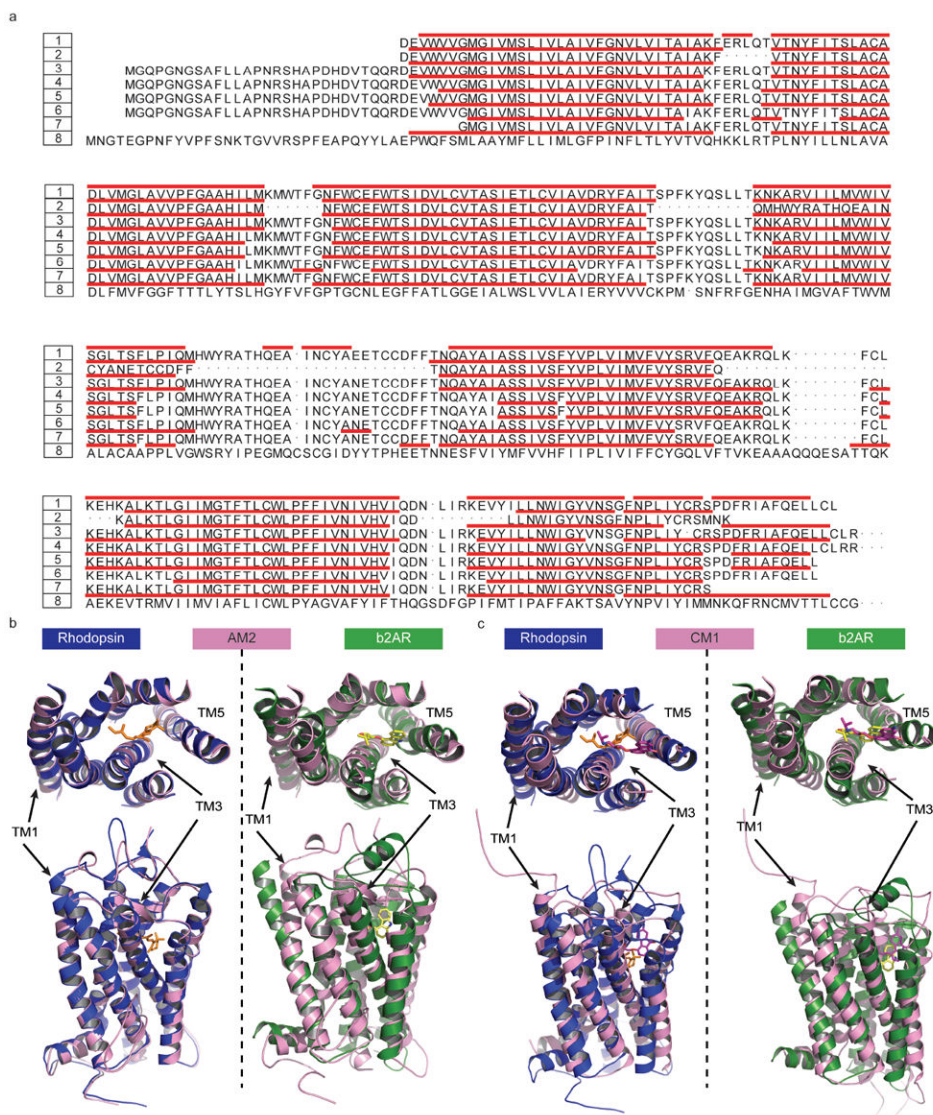


Figure 1. The structural similarity analysis of theoretical models in comparison with crystal structures. **(a)** The secondary structure assignment for TM segments of six theoretical models and two crystal structures (2RH1 and 1U19). The numbers and their corresponding structures are coded by 1:2RH1, 2:AM1, 3:AM2, 4:AM3, 5:CM1, 6:CM2, 7:CM3, 8:1U19. The remaining two models, i.e., AM4 and CM4, share similar backbone structures to CM3 with pairwise RMSD of TMs less than 0.4 Å. Therefore only CM3 is included in the sequence alignment plot. The red bars indicate the helical structure elements identified by MOE. **(b)** The structural superposition of the theoretical models AM2 (rendered in pink) to 2RH1 (rendered in dark green) and 1U19 (rendered in blue). Note that the most structurally divergent TM regions are indicated. **(c)** The structural superposition of the theoretical models CM1 (rendered in pink) to 2RH1 (rendered in dark green) and 1U19 (rendered in blue). Note that the most structurally divergent TM regions are indicated as well.

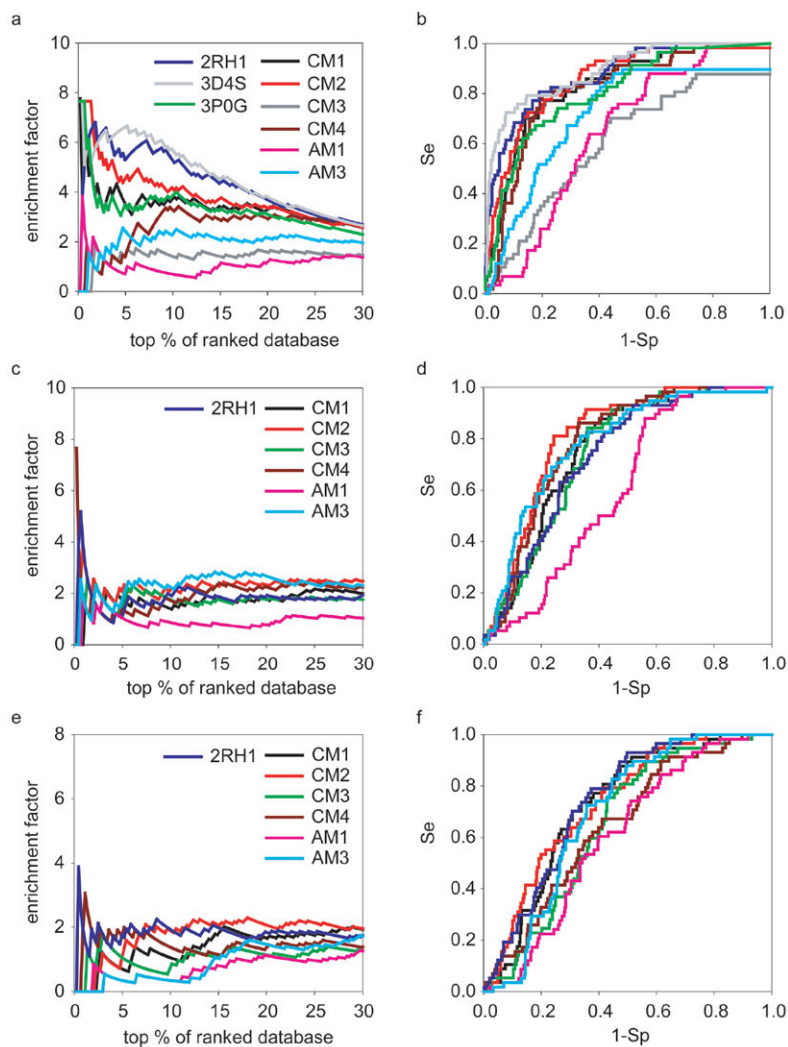


Figure 2. The docking performance of six theoretical models in comparison to three crystal structures 2RH1, 3D4S and 3P0G. The enrichment (a, c, e) and ROC curves (b, d, f) of fifty-seven known β 2AR antagonists with the decoys set I by three docking methods. The annotations are (a, b) Glide4.01 (c, d) AutoDock4.0 and (e, f) eHiTS6.2.

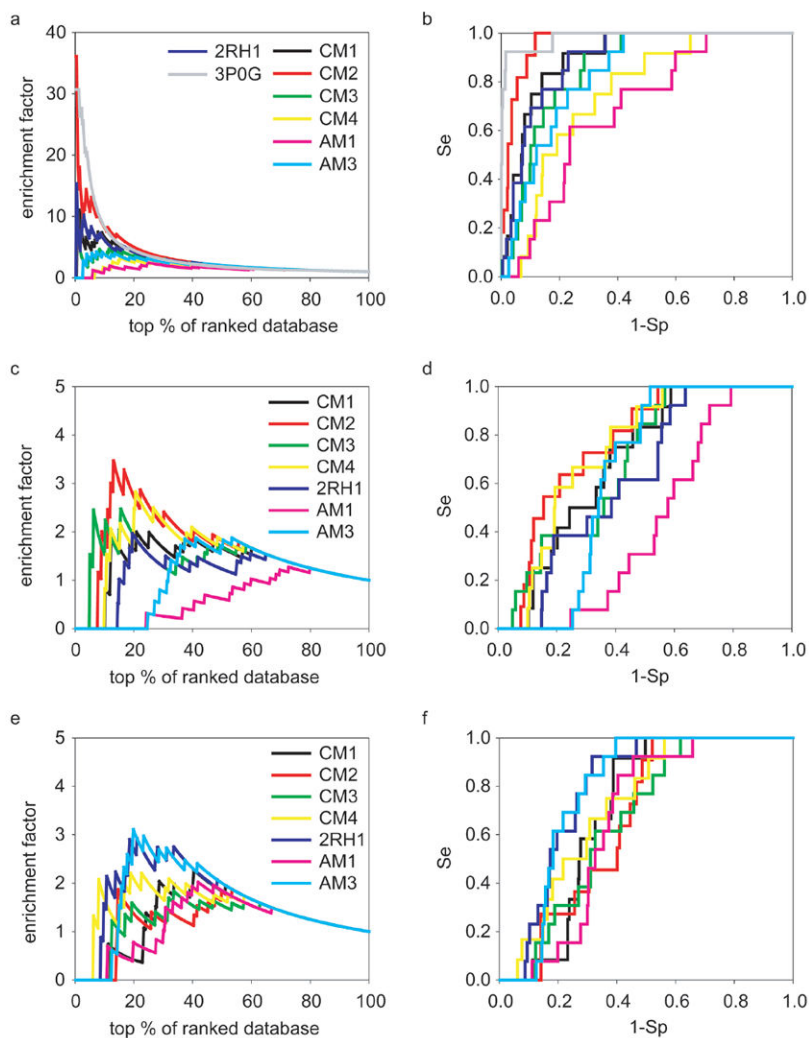


Figure 3. The enrichments and ROC curves of thirteen known β 2AR agonists with the decoys set I during the screening against six theoretical models and two crystal structures (2RH1 and 3P0G). The annotations are (a, b) Glide4.01 (c, d) AutoDock4.0 and (e, f) eHiTS6.2.

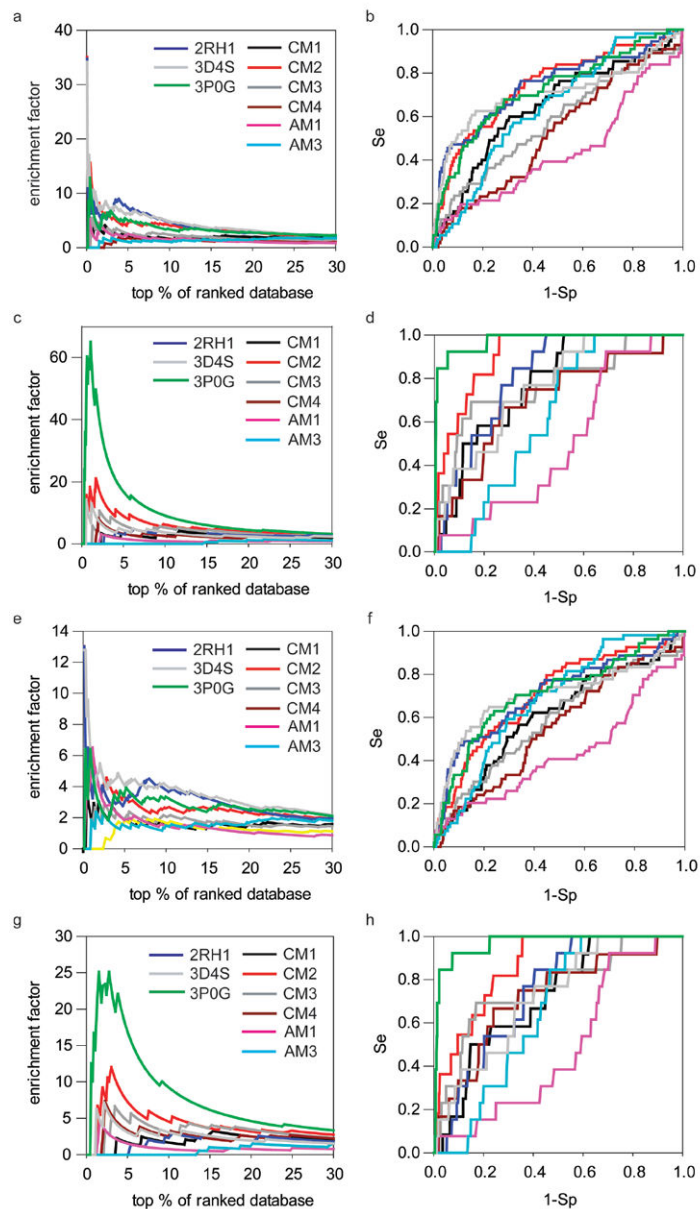


Figure 4. The docking performance of six theoretical models in comparison to three crystal structures 2RH1, 3D4S and 3POG with the decoys sets II and III. The enrichment (a) and ROC curves (b) of fifty-seven known β 2AR antagonists and the enrichment (c) and ROC curves (d) of thirteen known β 2AR agonists with the decoys set II. The enrichment (e) and ROC curves (f) of fifty-seven known β 2AR antagonists and the enrichment (g) and ROC curves (h) of thirteen known β 2AR agonists with the decoys set III.

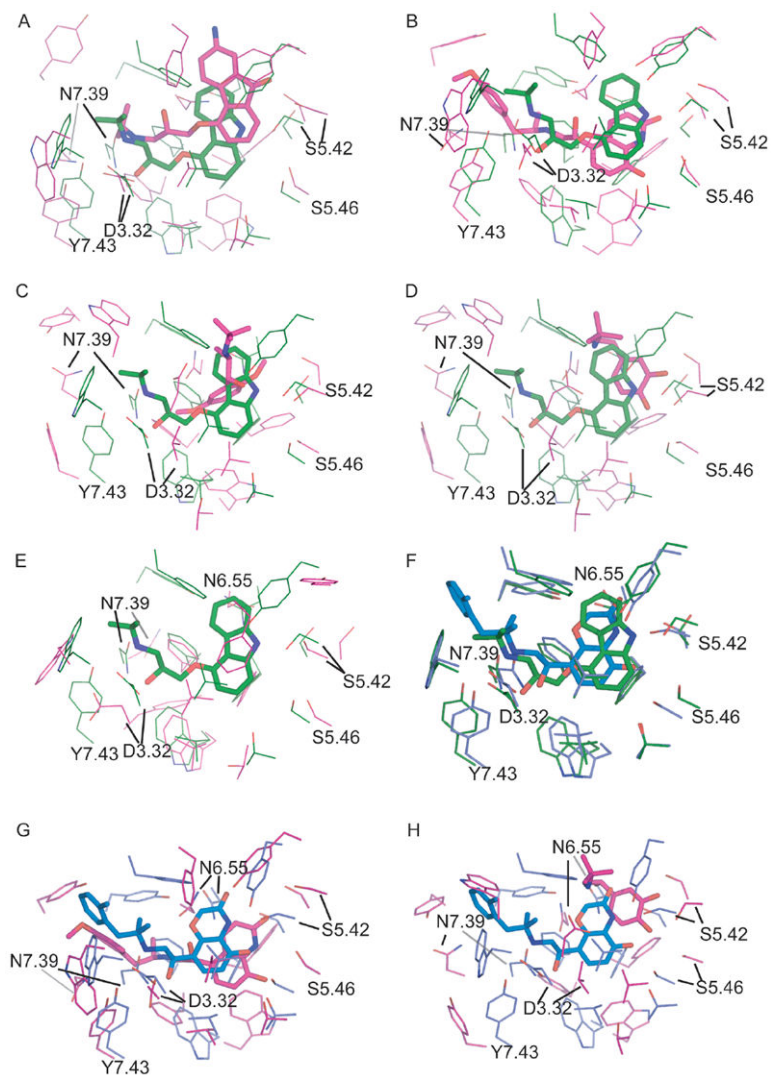


Figure 5.

The comparison of the binding pockets of theoretical models versus the β 2AR crystal structures of inactive state (2RH1; **A:CM1**, **B:CM2**, **C:CM3**, **D:CM4**, **E:AM1**) and active state (3P0G; **F:2RH1**, **G:CM2**, **H:CM4**). The active sites were superimposed by Ca atoms of key binding site residues of 2RH1 (W3.28, D3.32, V3.33, V3.36, T3.38, F5.32, Y5.38, S5.42, S5.43, S5.46, W6.48, F6.51, F6.52, N6.55, Y7.35, N7.39, Y7.43). The crystal structures 2RH1 and 3P0G are colored in green and blue respectively, while the model structures are colored in pink.

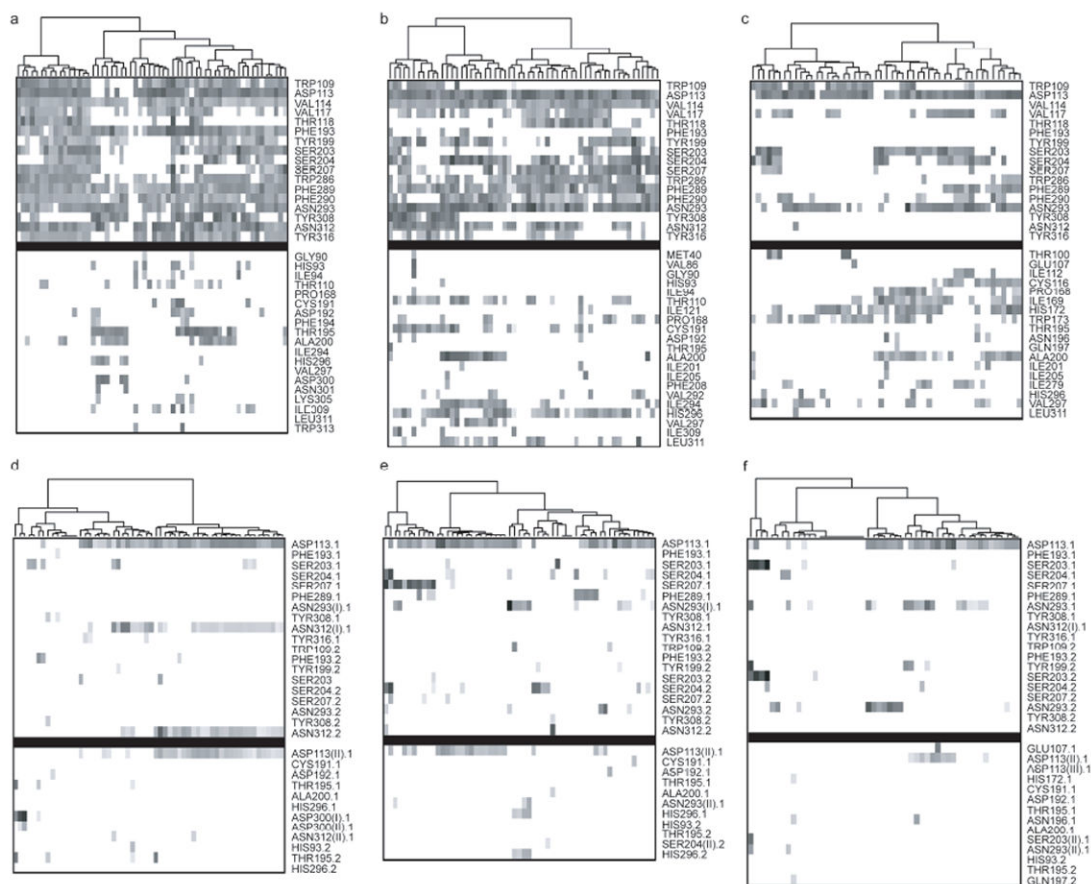


Figure 6.

The cluster analysis of the antagonists binding profile. (**Upper panel, a-c**) The comparison of weak interaction profiles of fifty-seven antagonists of human β 2AR with 2RH1 (**a**), CM2 (**b**) and CM3 (**c**). (**Lower panel, d-f**) The comparison of hydrogen bonding (HBond) profiles of fifty-seven antagonists of human β 2AR with 2RH1 (**d**), CM2 (**e**) and CM3 (**f**). The weak and HBond interactions were identified/scored by LigX module in MOE2007.09 and marked as X.1 for HBond donors and X.2 for HBond acceptors. In the case that one residue forms two HBonds to the ligand, the interactions were labeled as X(I).X or X(II).X, in which the X(I).X had the better score. The upper block in each map contains the binding pocket residues of 2RH1 while the lower block contains other interacting residues. Each point in the maps represents the identified interactions and was shaded by their respective scores, wherever darker indicates higher score and thus greater interaction strength. All the points in the map had been reorganized using hierarchical clustering by interaction scores.

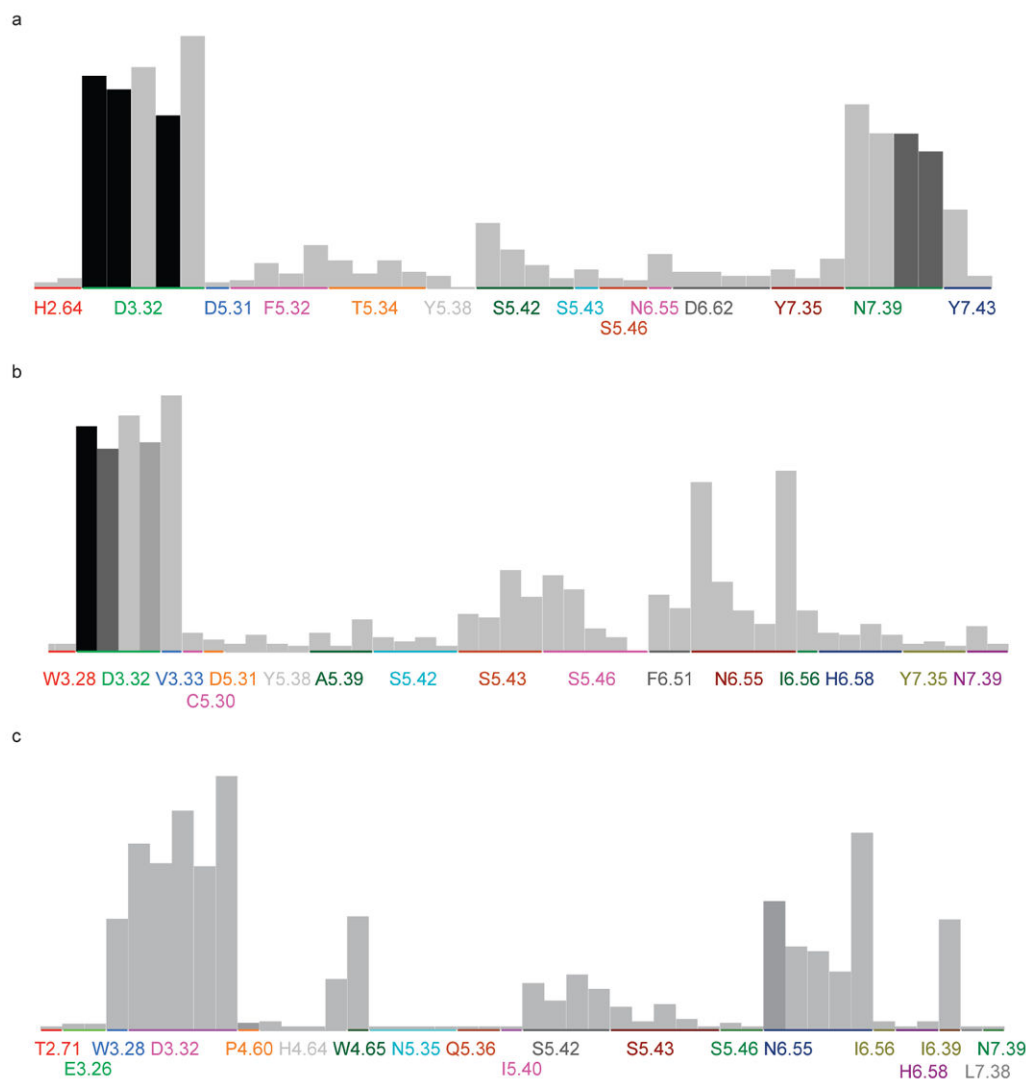


Figure 7.

The significance chart for PLIF fingerprints generated from docking poses against (a) β 2AR crystal structure (2RH1), (b) CM2 model and (c) CM3 model. The higher the bar (individual fingerprint bit) is, more frequently this type of interaction occurs in the dataset and of higher probability it contributes to the activities. The shade of the color indicates the significance of the particular bit to the actives, which is based on the hypothesis that ‘if the bit is set, then the compound is active’. The residues are randomly colored and several bars of the same residue indicate that they have different types of contacts.

Table I

The synonym of eight human β 2AR theoretical models employed in this study.

Apo model	Source	Holo model	Source
AM1	By G. Vriend , a homology model(7) based on the crystal structure of bovine rhodopsin	CM1	By T. Lybrand , a <i>de novo</i> model(1, 5) bound by aminoflisopolol (β 2AR antagonist)
AM2	By A. Sali , a homology model(6) based on the crystal structure of bovine rhodopsin	CM2	By T. Lybrand , a <i>de novo</i> model(1, 5) bound by TA2005 (β 2AR agonist)
AM3	By J. Skolnick , a hybrid model(2) combined threading and <i>ab initio</i> methods	CM3	By W. Goddard , a <i>de novo</i> model(3, 4) bound by butoxamine (β 2AR antagonist)
AM4	By W. Goddard , a <i>de novo</i> model(3) based on first principles methods	CM4	By W. Goddard , a <i>de novo</i> model(3, 4) bound by salbutamol (β 2AR agonist)

Table II

Comparisons for six β 2AR theoretical models versus two crystal structures in terms of TM helical boundaries and RMSD. The bordering residues for each TM region are identified by MOE and annotated using a specialized numbering rule. The RMSDs for each model's TM were calculated using the true length of the corresponding segments of 2RH1 and 1U19 (in italic).

Segment	2RH1		AMI		AM2		AM3		CMI		CM2		CM3	
	start/end	RMSD	start/end	RMSD	start/end	RMSD	start/end	RMSD	start/end	RMSD	start/end	RMSD	start/end	RMSD
TMI	1.30/1.59	1.29/1.60	2.77Å	1.29/1.59	3.02Å	1.28/1.59	3.01Å	1.31/1.59	3.18Å	1.32/1.58	3.19Å	1.35/1.56	5.12Å	5.12Å
	<i>1.29/1.59</i>	<i>0.30Å</i>	<i>0.30Å</i>	<i>0.30Å</i>	<i>0.30Å</i>	<i>0.83Å</i>	<i>0.83Å</i>	<i>0.83Å</i>	<i>2.93Å</i>	<i>3.03Å</i>	<i>3.03Å</i>	<i>3.03Å</i>	<i>6.90Å</i>	<i>6.90Å</i>
TMII	2.38/2.67	2.38/2.67	1.76Å	2.38/2.67	1.59Å	2.38/2.67	1.90Å	2.37/2.65	3.19Å	2.37/2.65	3.09Å	2.45/2.64	4.38Å	4.38Å
	<i>2.38/2.67</i>	<i>0.32Å</i>	<i>0.32Å</i>	<i>0.37Å</i>	<i>0.37Å</i>	<i>0.68Å</i>	<i>0.68Å</i>	<i>2.40Å</i>	<i>2.40Å</i>	<i>2.36Å</i>	<i>2.36Å</i>	<i>4.31Å</i>	<i>4.31Å</i>	
TMIII	3.21/3.55	3.22/3.55	4.32Å	3.22/3.54	2.09Å	3.23/3.54	1.67Å	3.22/3.55	2.57Å	3.23/3.55	2.56Å	3.27/3.47	6.99Å	6.99Å
	<i>3.21/3.54</i>	<i>0.34Å</i>	<i>0.34Å</i>	<i>0.37Å</i>	<i>1.08Å</i>	<i>1.08Å</i>	<i>1.08Å</i>	<i>2.21Å</i>	<i>2.21Å</i>	<i>2.25Å</i>	<i>2.25Å</i>	<i>5.00Å</i>	<i>5.00Å</i>	
TMIV	4.39/4.62	N/A	1.95Å	4.39/4.61	1.94Å	4.39/4.62	1.67Å	4.40/4.57	5.21Å	4.41/4.57	5.43Å	4.38/4.68	3.47Å	3.47Å
	<i>4.39/4.62</i>	<i>0.32Å</i>	<i>0.32Å</i>	<i>0.49Å</i>	<i>1.06Å</i>	<i>1.06Å</i>	<i>1.06Å</i>	<i>5.00Å</i>	<i>5.00Å</i>	<i>5.02Å</i>	<i>5.02Å</i>	<i>4.60Å</i>	<i>4.60Å</i>	
TMV	5.36/5.68	5.34/5.62	2.37Å	5.35/5.62	3.48Å	5.35/5.70	2.76Å	5.41/5.67	2.17Å	5.41/5.67	2.13Å	5.37/5.58	6.85Å	6.85Å
	<i>5.31/5.62</i>	<i>0.43Å</i>	<i>0.43Å</i>	<i>0.38Å</i>	<i>1.00Å</i>	<i>1.00Å</i>	<i>2.04Å</i>	<i>2.04Å</i>	<i>2.06Å</i>	<i>2.06Å</i>	<i>2.06Å</i>	<i>3.59Å</i>	<i>3.59Å</i>	
TMVI	6.29/6.60	6.33/6.60	1.67Å	6.33/6.59	3.80Å	6.27/6.60	1.96Å	6.28/6.59	4.24Å	6.28/6.59	4.21Å	6.38/6.59	8.12Å	8.12Å
	<i>6.25/6.59</i>	<i>0.32Å</i>	<i>0.32Å</i>	<i>9.11Å</i>	<i>0.80Å</i>	<i>0.80Å</i>	<i>3.63Å</i>	<i>3.63Å</i>	<i>3.69Å</i>	<i>3.69Å</i>	<i>3.69Å</i>	<i>12.54Å</i>	<i>12.54Å</i>	
TMVII	7.32/7.55	7.38/7.56	3.82Å	7.32/7.51	1.65Å	7.33/7.55	2.35Å	7.32/7.55	4.08Å	7.32/7.55	4.20Å	7.34/7.55	3.93Å	3.93Å
	<i>7.32/7.55</i>	<i>0.38Å</i>	<i>0.38Å</i>	<i>1.01Å</i>	<i>1.37Å</i>	<i>1.37Å</i>	<i>3.34Å</i>	<i>3.34Å</i>	<i>3.42Å</i>	<i>3.42Å</i>	<i>3.42Å</i>	<i>2.64Å</i>	<i>2.64Å</i>	
TMI-VII			2.47Å	3.19Å	3.19Å	2.25Å	2.25Å	3.59Å	3.59Å	N/A	N/A	4.41Å	4.41Å	
			<i>1.15Å</i>	<i>1.88Å</i>	<i>1.88Å</i>	<i>1.38Å</i>	<i>1.38Å</i>	<i>3.20Å</i>	<i>3.20Å</i>	<i>3.20Å</i>	<i>3.20Å</i>	<i>3.83Å</i>	<i>3.83Å</i>	
ECLII			N/A	13.43Å	13.19Å	13.19Å	7.00Å	7.00Å	7.18Å	7.18Å	7.18Å	19.00Å	19.00Å	
			<i>N/A</i>	<i>1.32Å</i>	<i>1.81Å</i>	<i>1.81Å</i>	<i>13.72Å</i>	<i>13.72Å</i>	<i>13.96Å</i>	<i>13.96Å</i>	<i>13.96Å</i>	<i>19.70Å</i>	<i>19.70Å</i>	
Binding pocket			3.71Å	3.39Å	3.39Å	3.58Å	2.33Å	2.33Å	2.40Å	2.40Å	2.40Å	5.64Å	5.64Å	

Table III

The enrichment factors (EF) and yields of fifty-seven known β 2AR antagonists by theoretical models versus two crystal structures with decoys set I (PDSP binding decoys plus WDI drug-like decoys).

Structure/Model	Docking method	EF at 1% db	EF at 5% db	EF at 25% db	% db to find 25% seeds	Yield 25% db	EF _{max}	% db EF _{max} occurred	ROC AUC	AUC at 25% db	$\alpha=20$	BEDROC $\alpha=53.6$	$\alpha=100$
2RH1	Glite4.01	5.83	5.30	3.08	4.74	77.19	6.80	1.81	0.88	0.11	0.72	0.75	0.74
	AutoDock4.0	3.89	1.06	1.75	11.74	43.86	0.68	5.18	0.73	0.05			
	eHITS6.2	1.94	1.77	1.82	13.54	45.61	3.89	0.45	0.74	0.05			
3D4S	Glite4.01	6.12	6.66	3.17	4.05	79.31	7.65	0.22	0.89	0.11	0.77	0.80	0.79
	Glite4.01	5.74	3.33	2.69	7.21	67.24	7.66	0.68	0.81	0.08	0.50	0.52	0.62
CM1	Glite4.01	5.83	3.53	2.87	6.77	71.93	7.77	1.35	0.83	0.08	0.50	0.52	0.58
	AutoDock4.0	0.00	1.77	2.17	13.77	54.39	2.59	1.35	0.76	0.04			
	eHITS6.2	0.00	0.71	1.75	14.67	43.86	2.03	15.58	0.72	0.04			
CM2	Glite4.01	7.66	4.52	2.90	5.63	72.41	7.66	1.35	0.86	0.09	0.61	0.72	0.83
	AutoDock4.0	3.83	2.09	2.48	11.49	62.07	3.83	0.90	0.81	0.06			
	eHITS6.2	0.00	1.04	2.14	12.16	53.45	2.30	18.02	0.73	0.05			
CM3	Glite4.01	0.00	1.41	1.54	17.61	38.60	1.94	0.90	0.61	0.04	0.20	0.15	0.11
	AutoDock4.0	1.94	1.41	1.82	14.90	45.61	2.59	1.35	0.75	0.05			
	eHITS6.2	0.00	1.06	1.26	23.25	31.58	1.79	2.93	0.65	0.03			
CM4	Glite4.01	1.91	1.39	2.90	2.70	72.41	3.42	10.59	0.82	0.07	0.34	0.19	0.13
	AutoDock4.0	1.91	1.39	2.34	12.61	58.62	7.66	0.23	0.78	0.05			
	eHITS6.2	1.91	1.74	1.52	16.89	37.93	3.06	1.13	0.65	0.04			
AM1	Glite4.01	1.91	0.70	1.31	20.27	32.76	3.83	0.45	0.65	0.03	0.14	0.17	0.22
	AutoDock4.0	1.91	1.04	1.03	22.75	25.86	1.91	0.90	0.61	0.03			
	eHITS6.2	0.00	0.35	0.97	25.90	24.14	1.42	37.61	0.60	0.03			
AM3	Glite4.01	0.00	2.44	2.07	10.36	51.72	2.55	4.73	0.72	0.06	0.28	0.20	0.14
	AutoDock4.0	1.91	1.74	2.34	10.81	58.62	2.83	14.64	0.78	0.07			
	eHITS6.2	0.00	0.35	1.45	17.57	36.21	1.81	32.43	0.69	0.03			

The enrichment factors (EF) and yields of thirteen known β 2AR agonists by six theoretical models versus the crystal structure 2RH1 with decoys set I (PDSP binding decoys plus WDI drug-like decoys).

Table IV

Structure/Model	Docking method	% db to find 25% agonists	EF at 25% db	Yield at 25% db	Max.EF	% db EF _{max} occurred	AUC	AUC at 25% db
2RH1	Glide4.01	3.01	3.38	84.62	15.35	0.50	0.90	0.15
	AutoDock4.0	18.55	1.53	38.46	1.97	19.55	0.63	0.03
	eHTS6.2	13.78	2.46	61.54	2.92	21.05	0.79	0.06
CM1	Glide4.01	3.77	3.65	91.67	11.06	0.75	0.91	0.15
	AutoDock4.0	12.31	1.99	50.00	2.03	12.31	0.70	0.04
	eHTS6.2	23.37	1.33	33.33	2.27	40.45	0.70	0.02
CM2	Glide4.01	1.51	3.97	100.00	36.09	0.25	0.96	0.20
	AutoDock4.0	11.08	2.53	63.64	3.47	13.10	0.77	0.08
	eHTS6.2	14.61	1.08	27.27	1.87	53.40	0.66	0.03
CM3	Glide4.01	7.77	3.07	76.92	15.35	0.50	0.86	0.11
	AutoDock4.0	15.04	1.53	38.46	2.48	15.54	0.69	0.06
	eHTS6.2	19.30	1.23	30.77	1.83	33.58	0.66	0.03
CM4	Glide4.01	12.31	2.32	58.33	3.26	15.33	0.75	0.07
	AutoDock4.0	26.13	2.32	58.33	2.83	20.60	0.74	0.05
	eHTS6.2	15.33	1.99	50.00	2.21	18.84	0.72	0.05
AM1	Glide4.01	17.04	2.46	61.54	2.48	24.81	0.69	0.05
	AutoDock4.0	44.11	0.31	7.69	1.27	72.68	0.44	0.00
	eHTS6.2	30.08	0.61	15.38	2.02	41.85	0.66	0.02
AM3	Glide4.01	7.27	3.07	76.92	4.23	7.27	0.83	0.10
	AutoDock4.0	31.33	0.31	7.69	1.87	41.10	0.64	0.00
	eHTS6.2	15.79	2.76	69.23	3.11	19.80	0.78	0.05

Table V

The EFs and yields of 57 β 2AR antagonists by theoretical models versus three crystal structures with decoys set II (ChEMBL binding decoys plus Schrodinger drug-like decoys). The data from decoys set III (subset of II containing N+ only) were marked in italic.

Structure/Model	EF at 1% db	EF at 5% db	EF at 25% db	% db to find 25% seeds	Yield at 25% db	EF _{max}	% db EF _{max} occurred	ROC AUC	AUC at 25% db	$\alpha=20$	BEDROC $\alpha=53.6$	$\alpha=100$	Seeds Num.	Decoys Num.
2RHI	6.95	7.02	2.47	3.35	61.82	34.75	0.05	0.74	0.08	0.32	0.23	0.21	55	1856
	3.72	3.25	2.09	6.96	52.83	<i>13.02</i>	<i>0.14</i>	<i>0.71</i>	<i>0.07</i>	<i>0.31</i>	<i>0.28</i>	<i>0.33</i>	53	637
	10.42	6.27	2.49	3.29	62.50	34.14	0.05	0.71	0.09	0.33	0.26	0.26	56	1856
3D4S	5.48	4.02	2.54	5.79	64.81	12.80	0.14	0.70	0.08	0.37	0.37	0.44	54	637
	5.15	4.90	2.41	4.63	60.71	12.86	0.42	0.72	0.07	0.24	0.18	0.17	56	1865
3P0G	5.51	3.93	2.35	7.78	59.26	6.43	0.29	0.71	0.06	0.26	0.25	0.28	54	640
CM1	3.51	2.17	1.95	12.47	49.09	7.03	0.26	0.64	0.04	0.14	0.10	0.10	55	1878
	1.88	1.88	1.57	16.21	39.62	4.38	0.43	0.61	0.03	0.17	0.17	0.18	53	644
	6.68	4.60	2.28	5.25	57.14	35.05	0.05	0.74	0.08	0.28	0.22	0.23	56	1907
CM2	4.89	3.26	2.12	9.80	53.70	13.04	0.14	0.71	0.06	0.28	0.30	0.35	54	650
	3.57	2.50	1.45	11.42	36.36	5.49	0.66	0.58	0.04	0.15	0.11	0.10	55	1906
CM3	1.66	2.15	1.50	14.94	37.74	2.95	2.56	0.59	0.04	0.18	0.15	0.13	53	650
	0.00	1.77	0.98	22.67	25.00	1.84	3.87	0.52	0.03	0.08	0.03	0.01	56	1907
CM4	0.00	1.81	1.10	21.31	27.78	1.92	8.66	0.55	0.03	0.12	0.06	0.02	54	650
	1.75	2.48	0.84	27.46	21.43	5.84	0.31	0.44	0.03	0.12	0.09	0.08	56	1907
AM1	6.52	1.96	0.89	28.84	22.22	6.52	1.14	0.43	0.03	0.18	0.23	0.28	54	650
	0.00	1.05	1.84	17.01	46.43	1.85	24.15	0.64	0.03	0.08	0.04	0.02	56	1907
AM3	1.63	1.41	1.92	16.05	48.15	2.37	1.56	0.68	0.03	0.15	0.12	0.10	54	650

Table VI

The EFs and yields of 13 β 2AR agonists by six theoretical models versus three crystal structures with decoys set II (ChEMBL binding decoys plus Schrodinger drug-like decoys). The data from decoys set III (subset of II containing N+ only) were marked in italic.

Structure/Model	EF at 1% db	EF at 2% db	% db to find 25% agonists	EF at 25% db	Yield at 25% db	EF _{max}	% db EF _{max} occurred	AUC at 25% db	$\alpha=20$	BEDROC $\alpha=53.6$	$\alpha=100$	Seeds Num.	Decoys Num.
2RHI	0.00	0.00	6.26	2.46	61.54	4.92	6.26	0.81	0.06	0.16	0.01	13	1856
	<i>0.00</i>	<i>0.00</i>	<i>10.46</i>	<i>2.15</i>	<i>53.85</i>	<i>2.94</i>	<i>10.46</i>	<i>0.74</i>	<i>0.04</i>	<i>0.09</i>	<i>0.00</i>	<i>13</i>	<i>637</i>
	7.57	7.57	6.10	1.84	46.15	14.38	0.54	0.76	0.06	0.22	0.10	13	1856
3D4S	0.00	3.85	8.92	1.84	46.15	7.14	2.15	0.72	0.05	0.19	0.08	13	637
	60.83	41.82	0.59	3.99	100.00	65.01	1.06	0.98	0.18	0.79	0.55	13	1865
	<i>14.35</i>	<i>23.44</i>	<i>1.38</i>	<i>3.98</i>	<i>100.00</i>	<i>25.12</i>	<i>1.53</i>	<i>0.97</i>	<i>0.17</i>	<i>0.72</i>	<i>0.54</i>	<i>13</i>	<i>640</i>
CM1	0.00	0.00	8.94	2.33	58.33	4.28	11.69	0.77	0.06	0.14	0.01	12	1878
	<i>0.00</i>	<i>0.00</i>	<i>12.04</i>	<i>2.29</i>	<i>58.33</i>	<i>3.25</i>	<i>15.40</i>	<i>0.72</i>	<i>0.05</i>	<i>0.10</i>	<i>0.01</i>	<i>12</i>	<i>644</i>
	17.44	17.88	1.56	3.60	90.91	21.13	1.72	0.91	0.12	0.40	0.16	11	1907
CM2	0.00	8.58	2.57	3.26	81.82	12.02	3.03	0.86	0.09	0.32	0.11	11	650
	0.00	7.38	3.49	2.76	69.23	10.80	2.14	0.77	0.09	0.25	0.06	13	1906
	<i>0.00</i>	<i>0.00</i>	<i>5.43</i>	<i>2.77</i>	<i>69.23</i>	<i>6.65</i>	<i>3.47</i>	<i>0.76</i>	<i>0.08</i>	<i>0.22</i>	<i>0.04</i>	<i>13</i>	<i>650</i>
CM3	0.00	8.20	7.30	2.28	58.33	8.42	1.98	0.70	0.05	0.16	0.03	12	1907
	<i>0.00</i>	<i>3.94</i>	<i>6.50</i>	<i>2.66</i>	<i>66.67</i>	<i>7.36</i>	<i>2.27</i>	<i>0.72</i>	<i>0.05</i>	<i>0.18</i>	<i>0.04</i>	<i>12</i>	<i>650</i>
	0.00	0.00	41.77	0.92	23.08	3.28	2.34	0.50	0.02	0.06	0.01	13	1907
AM1	0.00	3.64	42.84	0.92	23.08	4.64	1.66	0.48	0.02	0.07	0.04	13	650
	0.00	0.00	21.88	1.22	30.77	1.69	49.95	0.62	0.01	0.01	0.00	13	1907
	<i>0.00</i>	<i>0.00</i>	<i>20.21</i>	<i>1.21</i>	<i>30.77</i>	<i>1.78</i>	<i>47.51</i>	<i>0.65</i>	<i>0.01</i>	<i>0.02</i>	<i>0.00</i>	<i>13</i>	<i>650</i>