JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Strategies for Improved Modeling of GPCR-Drug Complexes: Blind Predictions of Serotonin Receptors Bound to Ergotamine

David Rodríguez,[†,‡,§] Anirudh Ranganathan,[†,‡,§] and Jens Carlsson*[,†,‡,§]

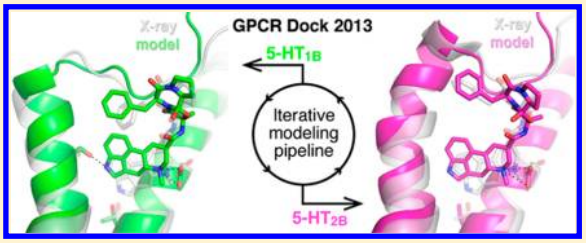
[†]Science for Life Laboratory, Stockholm University, Box 1031, SE-171 21 Solna, Sweden

[‡]Swedish e-Science Research Center (SeRC), SE-100 44 Stockholm, Sweden

[§]Department of Biochemistry and Biophysics and Center for Biomembrane Research, Stockholm University, SE-106 91 Stockholm, Sweden

Supporting Information

ABSTRACT: The recent increase in the number of atomic-resolution structures of G protein-coupled receptors (GPCRs) has contributed to a deeper understanding of ligand binding to several important drug targets. However, reliable modeling of GPCR-ligand complexes for the vast majority of receptors with unknown structure remains to be one of the most challenging goals for computer-aided drug design. The GPCR Dock 2013 assessment, in which researchers were challenged to predict the crystallographic structures of serotonin 5-HT_{1B} and 5-HT_{2B} receptors bound to ergotamine, provided an excellent opportunity to benchmark the current state of this field. Our contributions to GPCR Dock 2013 accurately predicted the binding mode of ergotamine with RMSDs below 1.8 Å for both receptors, which included the best submissions for the 5-HT_{1B} complex. Our models also had the most accurate description of the binding sites and receptor−ligand contacts. These results were obtained using a ligand-guided homology modeling approach, which combines extensive molecular docking screening with incorporation of information from multiple crystal structures and experimentally derived restraints. In this work, we retrospectively analyzed thousands of structures that were generated during the assessment to evaluate our modeling strategies. Major contributors to accuracy were found to be improved modeling of extracellular loop two in combination with the use of molecular docking to optimize the binding site for ligand recognition. Our results suggest that modeling of GPCR-drug complexes has reached a level of accuracy at which structure-based drug design could be applied to a large number of pharmaceutically relevant targets.



1. INTRODUCTION

G protein-coupled receptors (GPCRs) are responsible for a large fraction of cellular signaling processes and are the targets of nearly 30% of all marketed drugs.[1] These receptors share a common topology characterized by seven transmembrane (TM) helices interspersed with three intra- (ILs) and extracellular loops (ELs). Despite their overall topological similarity, GPCRs have evolved extracellular binding sites that recognize a diverse set of signaling molecules, ranging from neurotransmitters to hormones and peptides. Consequently, the extracellular region is extremely divergent between different receptor families (Figure 1).[2] Binding of the endogenous ligand shifts the conformational equilibrium of the receptor, which leads to structural rearrangements in the cytosolic region and activation of intracellular signaling pathways. Drugs targeting the orthosteric sites of GPCRs mediate their effects by either activating the receptor (agonist) or blocking the binding of the endogenous ligand (antagonist or inverse agonist).[3]

For a long time, molecular level understanding of drug binding to GPCRs has been hampered by the lack of experimental high-resolution structural information. Determination of the crystal structure of rhodopsin provided the first reasonable template for homology modeling of pharmaceutically relevant GPCRs.[4] However, the low sequence similarity and large differences in loop regions between rhodopsin and many GPCRs made it extremely challenging to obtain accurate models of receptor-drug complexes.[5] The next breakthrough was a high-resolution structure of the human β_2 adrenergic receptor (β_2AR) in complex with an inverse agonist.[6,7] This receptor is part of the group of biogenic amine GPCRs, which includes important drug targets such as the adrenergic, histamine, muscarinic, dopamine, and serotonin receptors.[8,9] Advances in GPCR structural biology have resulted in the determination of crystal structures for members of all five aforementioned receptor families.[7,10−15] Several of these have also been exploited in structure-based ligand discovery efforts and contributed to the identification of novel lead compounds.[16−20] In total, high-resolution GPCR structures for 24 unique members of the superfamily are now available. However, structure determination for membrane proteins via X-ray crystallography is still very challenging. Six GPCR structures were solved during 2013, and, at this rate, it would take over 60 years to solve at least one structure for all the 391

Received: April 11, 2014
Published: July 17, 2014

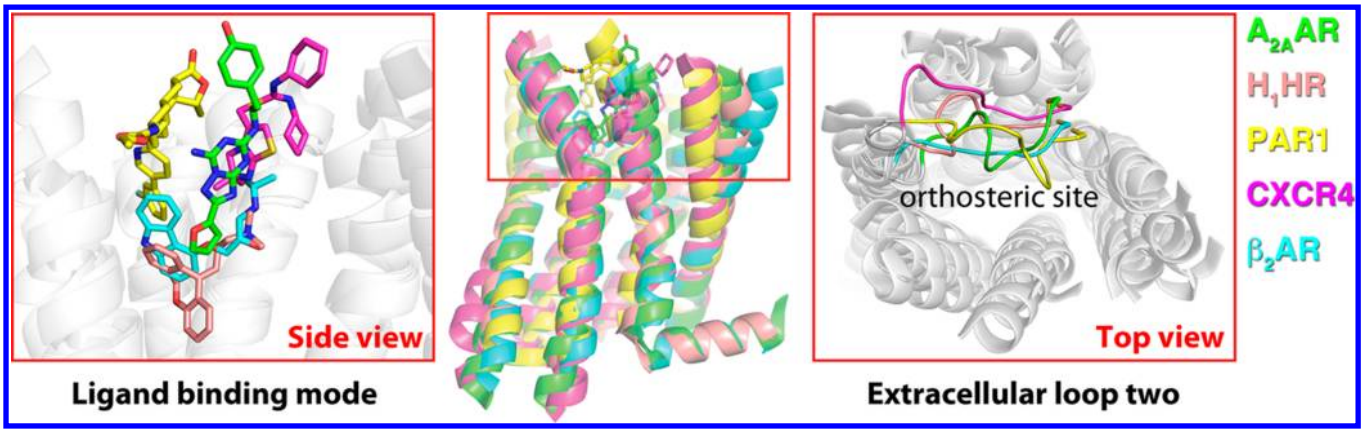


Figure 1. Structures of five class A GPCRs: the A_{2A} adenosine (A_{2A}AR, green), H_1 histamine (H_1HR, salmon), protease-activated-1 (PAR1, yellow), chemokine CXCR4 (CXCR4, magenta), and β_2 adrenergic (β_2AR, cyan) receptors. The topology of the transmembrane region is strongly conserved, but there are large variations in the (extracellular) orthosteric site, the ligand binding modes (side view), and the structure of the second extracellular loop (top view).

nonolfactory human GPCRs. In addition, it would be necessary to also determine multiple structures representing different bound ligand classes or functional states. For receptors of unknown structure, understanding of drug binding at the molecular level is restricted to the use of homology models. At a cutoff of 30% TM sequence identity, more than 50% of the class A GPCRs can be modeled based on currently available crystal structures.[21] If accurate models of GPCR-drug complexes can be generated, then structure-based methods have the potential to make major contributions to the development of novel pharmaceuticals as a tool for ligand discovery and lead optimization.

Opportunities for the molecular modeling field to make blind predictions in community-wide assessments have become valuable avenues for evaluating the state and progress of computational methods. This has been particularly evident from the CASP assessments,[22] which have had an important role in testing the accuracy of protein structure prediction methods during the last two decades. More recently, similar assessments have been arranged for predicting structures of protein−protein complexes (CAPRI),[23] protein−ligand binding modes and affinities (CSAR),[24] and small molecule solvation energies (SAMPL).[25] Since 2008, Stevens and colleagues have evaluated the state of GPCR modeling in three editions of the GPCR Dock assessment. In each of these, the scientific community has been challenged to predict structures of GPCR-ligand complexes before the release of the experimentally determined coordinates. The aim of the first GPCR Dock was blind prediction of the complex between the A_{2A} adenosine receptor (A_{2A}AR) and a potent antagonist, ZM241385.[26] The evaluation of the submitted models revealed difficulties in predicting features in structurally divergent regions, which reflected that only a few GPCR structures (with low sequence identity to the target) were available in 2008. This was particularly evident for EL2, which had an unexpected α helix turn that was in direct contact with the cocrystallized ligand.[27] The best model was generated by the Costanzi group and demonstrated the utility of experiment-guided modeling of ligand−receptor complexes.[28,29] This model captured close to 50% of the receptor−ligand contacts observed in the crystal structure, but inaccurate positioning of two key residues in EL2 led to a slight misplacement of the ligand (RMSD = 2.8 Å). The following GPCR Dock assessment was arranged in 2010 and involved complexes for two GPCRs: the D_3 dopamine receptor (D_3DR) bound to eticlopride[14] and the CXCR4 chemokine receptor in complex with either a small molecule ligand (IT1t) or a cyclic peptide (CVX15).[30] Availability of templates with high sequence identity enabled the possibility to achieve predictions that were strikingly close to the experimental D_3DR-eticlopride structure. The best prediction of the binding pose of eticlopride even had an RMSD below 1.0 Å and captured more than 50% of the ligand contacts.[31] Still, none of the top ranked groups were able to predict the key nonpolar interaction between Ile183 in EL2 and the ligand, which again emphasized the importance of accurate modeling of this loop region.[32] The peptide-binding receptor CXCR4 was considerably more challenging to model due to the lack of suitable templates. No accurate binding modes for the peptide ligand (CVX15) were found, which was likely due to its large size and conformational flexibility.[32] The best models of the CXCR4−IT1t complex were generated via careful interpretation of available mutagenesis data, leading to accurate prediction of key polar contacts for this ligand.[33] However, the predicted binding mode, with an RMSD equal to 4.9 Å, captured only 36% of the ligand−receptor contacts. The first two GPCR Dock assessments have revealed the many challenges involved in prediction of receptor-drug complexes and have been important for determining the state-of-the-art for this field. A majority of the participating research groups have not been able to predict complexes that would be considered accurate in assessments of docking methods (ligand RMSD < 2 Å).[24] For the D_3DR, which had templates above 40% sequence identity, only five groups were able to provide at least one prediction with ligand RMSD below 2 Å. No models achieved this level of accuracy for the more difficult modeling cases (A_{2A}AR and CXCR4).[27,32] These results reflect that prediction of both receptor structure and ligand binding mode is still tremendously difficult and that there is a need for development of more reliable modeling approaches.

In the most recent GPCR Dock assessment (GPCR Dock 2013), researchers were invited to predict four complexes involving receptors of different classes: the human 5-hydroxytryptamine (5-HT, serotonin) receptor subtypes 1B and 2B bound to ergotamine (class A) and the smoothened receptor bound to the antitumor agents LY-2940680 and

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

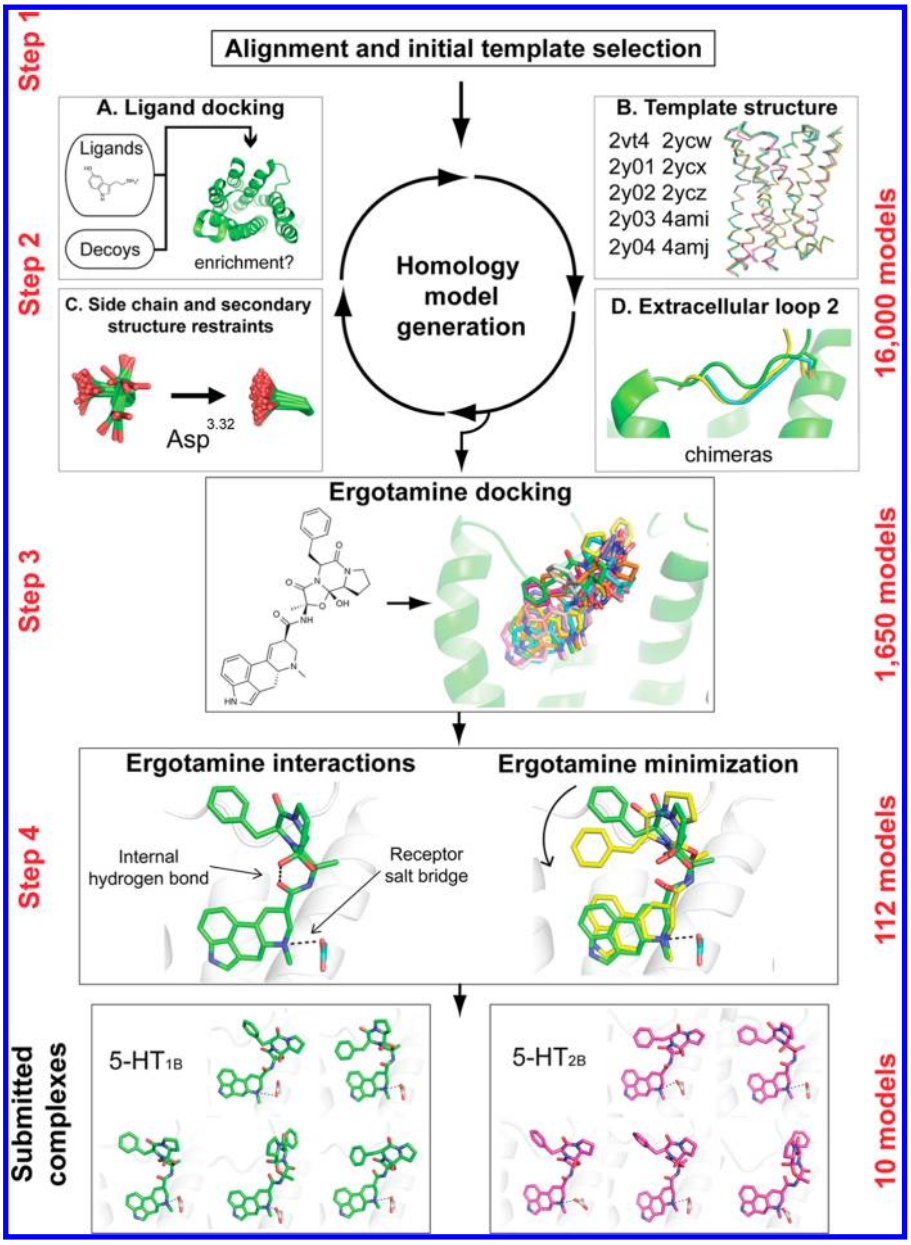


Figure 2. Summary of our modeling strategy in the GPCR Dock 2013 assessment.

SANT-1 (class F).[34] As the serotonin receptors are therapeutic targets for a wide range of central nervous system disorders,[35] the challenge to model the 5-HT_{1B} and 5-HT_{2B} receptors represented an important benchmark for computer-aided drug design. In this paper, we present our strategy for modeling GPCR-drug complexes and, in particular, its performance in GPCR Dock 2013 for the 5-HT_{1B} and 5-HT_{2B} receptors in complex with ergotamine. One of our key questions was if the recent increase in the number of available GPCR structures could be utilized to improve prediction accuracy. We combined homology modeling based on multiple templates with extensive molecular docking screening to generate the best models of the binding pockets and accurate predictions of the ergotamine complexes for both receptors. A key contribution to the quality of our predictions was improved modeling of EL2, which was accomplished by combining information from several different GPCR structures. The potential to extend our modeling strategies to other GPCRs of unknown structure is also discussed.

2. METHODS

Summary of Modeling Strategy. The strategy followed by our group (Stockholm-Carlsson in ref 34) to predict the structures of 5-HT_{1B} and 5-HT_{2B} receptors bound to ergotamine is summarized schematically in Figure 2. Our modeling approach was based on a combination of iterative generation of homology model ensembles and molecular docking screens. In each round of homology modeling, the ability of the generated structures to enrich known ligands among decoys was used as a measure of the quality of the ensemble. Modifications to the homology modeling protocol could be separated into three categories: the choice of template (Figure 2, Step 2B), tailored EL2 sampling (Figure 2, Step 2D), and introduction of additional side chain and secondary structure restraints (Figure 2, Step 2C). Every new ensemble

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

of homology models was assessed using molecular docking screens of known ligands (Figure 2, Step 2A). Ergotamine was finally docked to a few optimized homology model ensembles (Figure 2, Step 3). A set of filtering rules was used to identify complexes that captured predicted receptor−ligand interactions and internal ligand contacts, which were derived from available experimental data. Representative complexes were also subjected to an additional minimization procedure (Figure 2, Step 4). In a final step, five models of the receptor−ligand complex for each subtype were selected and submitted to the GPCR Dock 2013 assessment.

Sequence Alignment and Template Selection. The initial sequence alignment and template selection were performed with the aid of the GPCR-ModSim Web server (http://gpcr.usc.es).[36] The sequences of human 5-HT_{1B} and 5-HT_{2B} receptors provided by the organizers of the assessment were aligned to a multiple sequence alignment (MSA) profile of crystal structures of GPCRs. Members of the aminergic receptor family were predicted to be the best templates for modeling both serotonin receptors. The turkey β_1 adrenergic (β_1AR) receptor showed the highest sequence identity for the TM region (∼40%) and binding site residues (∼50%). Several structures of this receptor bound to ligands with diverse functional profiles were explored as possible templates (PDB accession codes: 2VT4, 2YCW, 2YCY, 2YCZ; 2Y00, 2Y01, 2Y03, 2Y04; 2Y02; 4AMI, 4AMJ).[37−39] The obtained sequence alignment was edited manually for loop regions and the NPxxY motif in TM7 (Supporting Information Figure S1).

Residue Numbering. The Ballesteros-Weinstein residue numbering for GPCRs[40] is indicated in superscript for TM residues throughout the text. This scheme follows the *X.YY* pattern: *X* denotes the TM helix and *YY* is a sequence-based correlative number focused on the value 50, which is assigned to the most conserved residue of each helix. For residues belonging to EL2, we follow the alternative notation that employs the *45.ZZ* pattern. In this case, the value 50 is given to the cysteine of EL2 involved in the conserved disulfide bridge with TM3.[41]

Loop Modeling Strategies. All homology models were generated with MODELLER 9.11.[42] Based on MSA of orthologous sequences, we added additional restraints to obtain an intra-EL3 cysteine bridge in both serotonin receptors. The GPCR-conserved EL2-TM3 disulfide bridge ($Cys^{45.50}$-$Cys^{3.25}$) was already present in the initial template-target sequence alignment step (Supporting Information Figure S1) and was kept in subsequent modeling rounds. We then focused on the region of EL2 between $Cys^{45.50}$ and TM5 ($EL2^{SS\text{-}TM5}$). Sequence analyses of the serotonin receptors revealed that they had longer $EL2^{SS\text{-}TM5}$ sections compared to the initial template, β_1AR: +1 and +2 residues for 5-HT_{1B} and 5-HT_{2B}, respectively. We considered two different approaches for modeling this region: *i*) introduction of different gaps in the β_1AR-target alignment and *ii*) creation of template chimeras between the β_1AR structure with EL2 conformations from other GPCR crystal structures. The combined templates were based on a β_1AR receptor structure except in the EL2 region (residues 180-HWWRDEDPQALKCYQDPGCCDFVTN-204). Stretches of the $EL2^{SS\text{-}TM5}$ section from the crystallographic structure of the D_3DR were instead used as templates for EL2 (see Table 2 for details).

Ligand-Guided Optimization of Homology Models Using Molecular Docking Screening. Model refinement was carried out using a combination of homology modeling with MODELLER 9.11[42] and molecular docking with DOCK3.6.[43] In the optimization of the homology models, different restraints on binding site residues were added to improve their ability to recognize known ligands. Up to 500 homology models were generated in each iteration, and the structures with the best DOPE scores[44] were assessed using molecular docking.

All molecular docking calculations were carried out using DOCK3.6.[45] The sampling algorithm of DOCK3.6 generates ligand orientations in the binding site by superimposing atoms of the docked molecule onto binding site matching spheres, which are labeled according to the local receptor environment for chemical matching.[46] The ligand binding site was automatically defined with 45 such spheres based on cocrystallized ligands in β_1AR structures. Several sampling protocols were explored and modified based on initial enrichment calculations for each receptor. Ligand conformations with more than one overlapping heavy atom with the receptor, i.e. a distance shorter than 2.3 and 2.6 Å for polar and nonpolar atoms, respectively, were discarded. For conformations passing this filter, the binding energy was calculated as the sum of the electrostatic and van der Waals interaction energies,[45] corrected for ligand desolvation.[47] The three-dimensional map of the electrostatic potential in the binding site was generated with Delphi.[48] The energy contributions were extracted from precalculated grids, based on parameters from a united-atom version of the AMBER force field.[49] The desolvation penalty for a ligand conformation was estimated from a precalculated grid of the transfer free energy of each docked molecule from aqueous solution to a low-dielectric medium, which also takes into account the degree of burial of the ligand in the receptor binding site.[47,50] For the best scoring pose of each ligand, 100 steps of rigid-body energy minimization were carried out.

Ligand enrichment was used to assess the quality of the generated homology models. For each receptor, known ligands were extracted from the ChEMBL15 database,[51] and property-matched decoys were subsequently generated with the DUD-E resource (http://dude.docking.org)[52] in a 1:50 ratio. Two sets of ligands, lead-like (molecular weight <350) and serotonin-like compounds (molecular weight <500), were used in retrospective docking against each receptor model. The enrichments presented in this work correspond to the results obtained for the serotonin-like sets, which contained compounds that shared a tryptamine-like substructure. Up to 600 conformations of the docked molecules were pregenerated using the software OMEGA.[53] Partial charges and desolvation energies were calculated with AMSOL,[54] and all-atom van der Waals parameters were derived from an AMBER force field.[55] These sets of ligands and decoys were screened against the homology models, and the ability of the orthosteric site to enrich ligands was quantified using the adjusted LogAUC (aLogAUC) metric.[47] The aLogAUC includes a correction for random selection (normalized to zero) and favors early ligand enrichment. The models with the best aLogAUC were inspected visually to identify side chain conformations that were able to recognize known ligands. Side chain rotamers identified as important for ligand enrichment were restrained in subsequent rounds of homology modeling (Supporting Information Table S1).

Docking of Ergotamine: Pose Generation and Filtering Based on Contacts. Ergotamine was docked to top-enriching ensembles of 5-HT_{1B} and 5-HT_{2B} receptor models using DOCK3.6.[45] During the assessment, ergotamine was only

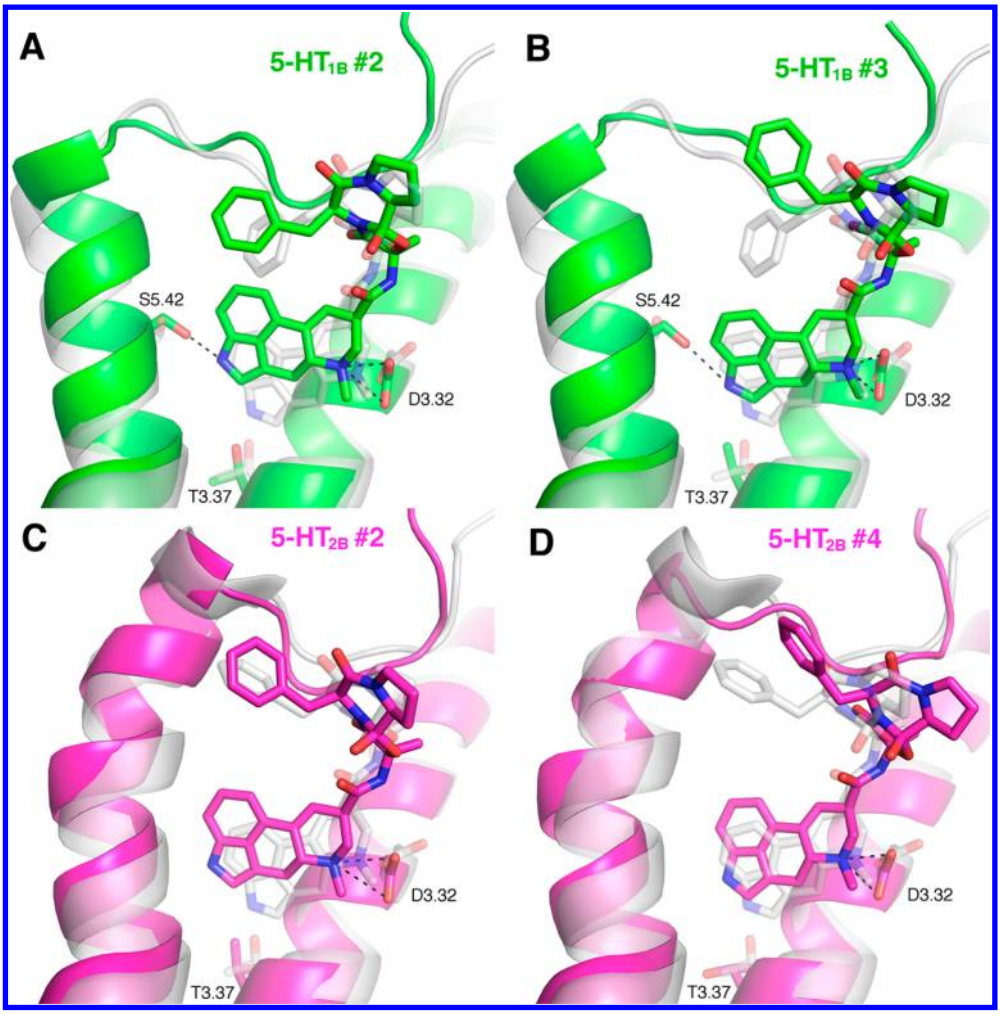


Figure 3. Our two best ranked predictions of the 5-HT_{1B} (**A, B**) and 5-HT_{2B} (**C, D**) receptor-ergotamine complexes in the GPCR Dock 2013 assessment. The carbons of the predicted complexes are colored green and magenta for 5-HT_{1B} and 5-HT_{2B} receptors, respectively. The corresponding crystallographic structures are superimposed and shown in transparent white.

docked to structures generated in the last rounds of homology modeling optimization. We also report retrospective docking of ergotamine to early ensembles of homology models in order to evaluate the performance of our modeling strategy (Table 2). Only the lowest energy pose of ergotamine was kept in all cases. A series of in-house scripts was used to detect models that captured receptor−ligand interactions that had been observed in crystal structures of other aminergic receptors.[37,56]

Energy Minimization and Selection of Final Docking Poses. Selected complexes from docking were subjected to an additional refinement step using the program PLOP[57] in which the ligand was energy minimized with the receptor held rigid. Minimization of the receptor coordinates was not considered as extensive sampling of side chain conformations was already achieved by considering multiple homology models and to maintain ligand enrichment of each model. The minimization was performed with the OPLS-AA force field[58] using the truncated Newton method as implemented in the TNPACK package.[59] Minimization was carried out for a maximum of 65 steps, and a RMS gradient of 0.001 kcal/mol/Å was set as convergence criteria. The pre- and postoptimized ligand poses were visually inspected. In addition to key polar contacts, the burial of the ergoline moiety of ergotamine in the binding site was used as a criterion for pose selection. Moreover, the presence of an internal hydrogen bond observed in the crystal structure of ergotamine (available at the Cambridge Crystallographic Data Center, http://www.ccdc.cam.ac.uk)[60,61] was also added to the selection criteria. Five diverse complexes for each receptor, which were blindly ranked from #1 to #5 by visual inspection, were submitted to the GPCR Dock 2013 assessment.

Comparison of Submitted Models to Crystal Structures. In order to assess the contribution of each modeling step to the accuracy of our predictions, a series of in-house scripts and the program PyMOL[62] were used. The atom selections of receptors regions (TMs and loops) defined by the organizers of the GPCR Dock 2013 assessment were used to align the models to the crystal structures and to perform RMSD calculations.[34] In our analysis, we also put particular focus on the last part of the second extracellular loop ($EL2^{SS\text{-}TM5}$, from $Cys^{45.50}$ to TM5). Receptors were superimposed based on residues in the TM bundle with the align algorithm available in PyMOL. After that, all RMSDs calculations were carried out with the 'rms_cur' tool. All RMSD values were corrected for ligand and side chain symmetry.

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

Table 1. Accuracy and aLogAUC Values for Our 5-HT$_{1B}$ and 5-HT$_{2B}$ Receptor Models Submitted to the GPCR Dock 2013 Assessment

		RMSD (Å)[b]					Polar contact distances (Å)[c]				
receptor	model[a]	ergotamine	pocket	TM	EL2 (full)	EL2$^{SS\text{-}TM5}$	Asp$^{3.32}$	Thr$^{3.37}$	Val/Leu$^{45.52}$	internal h-bond	aLogAUC[d]
5-HT$_{1B}$	#1	4.1	1.6	1.9	5.8	3.5	3.3	7.8	3.9	2.8	27
	#2	1.7	1.6	1.9	4.3	2.6	2.7	6.0	2.7	2.7	24
	#3	2.2	1.5	1.8	5.4	2.5	2.6	4.8	2.8	2.9	26
	#4	5.5	1.6	1.9	4.2	2.6	2.5	8.9	8.0	2.7	25
	#5	3.1	1.7	2.0	6.3	2.9	2.8	5.9	4.0	5.8	20
5-HT$_{2B}$	#1	3.5	2.3	2.0	5.3	1.9	3.0	8.2	4.5	3.1	27
	#2	1.7	2.1	2.0	6.5	1.3	3.2	7.9	2.9	2.7	30
	#3	2.9	2.4	2.0	6.2	1.9	2.6	5.9	2.9	2.8	30
	#4	2.3	2.0	2.0	4.4	1.6	2.8	6.6	2.8	3.5	20
	#5	6.2	2.4	2.0	4.3	1.7	2.5	8.3	7.2	2.6	32

[a]Blind rank assigned by our group at the time of submission. [b]RMSD of non-hydrogen (ergotamine and pocket) or backbone (Cα, C, N; TM and EL2) atoms compared to the crystallographic structure. Values are corrected for ligand or side chain symmetry where applicable. [c]Heteroatom distance of our models for the indicated polar contact found in the corresponding crystallographic complex. Experimentally observed distances for ligand−receptor interactions (Asp$^{3.32}$, Thr$^{3.37}$, and Val/Leu$^{45.52}$) and the internal ligand hydrogen bonds are 2.7, 2.9, 3.2, and 2.7 Å for 5-HT$_{1B}$ and 2.6, 3.0, 2.9, and 2.7 Å for 5-HT$_{2B}$. [d]Docking enrichment of tryptamine-like ligands from the ChEMBL database for each receptor. The corresponding values for the crystallographic structures were 20 and 19 aLogAUC units for the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors, respectively.

3. RESULTS AND DISCUSSION

Summary of Performance in GPCR Dock 2013. The GPCR Dock 2013 assessment allowed participants to submit five models for each receptor−ligand complex. The duration of the assessment was approximately one month, and, in early March 2013, we submitted five structures for each of the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors in complex with ergotamine. Thirty-nine other research groups also participated in the assessment for the 5-HT receptors. All predictions were fully prospective as the two crystal structures were not made public until 2 weeks after the submission deadline. The evaluation of the predicted complexes by Kufareva et al. was primarily focused on the binding mode of ergotamine, which was quantified based on a weighted score of ergotamine RMSD and correctly predicted ligand−receptor contacts.[32,34]

The performance of our models in the GPCR Dock 2013 assessment is summarized in Supporting Information Table S2. Our two best predictions for each receptor−ligand complex are shown in Figure 3, and the six other submitted models are depicted in Supporting Information Figure S2. Two of our solutions (5-HT$_{1B}$#3 and 5-HT$_{2B}$#4) had the best binding pocket RMSDs (1.3 and 1.9 Å) and the most accurate descriptions of the receptor−ligand contacts (contact prediction accuracy: 51% and 53%) for each subtype. Two 5-HT$_{1B}$ models were also ranked as the best and second best predictions of the 5-HT$_{1B}$-ergotamine complex out of the 182 submissions (Figures 3A and 3B). The reported RMSDs of ergotamine for these two models (5-HT$_{1B}$#2 and 5-HT$_{1B}$ #3) were equal to 1.5 and 2.1 Å, respectively. For the 5-HT$_{2B}$ receptor, our two most accurate models were ranked as the fifth and seventh best among the 171 submitted complexes and had ligand RMSD values of 1.6 and 2.5 Å, respectively (5-HT$_{1B}$#2 and 5-HT$_{2B}$#4, Figures 3C and 3D). We achieved ligand RMSDs of less than 2 Å for both receptors, which is often used as a cutoff for successful docking predictions.[24] This result is remarkable considering that the challenge in this case was to predict the binding mode of a relatively large and flexible ligand in a receptor of unknown structure.

All homology models generated during the GPCR Dock assessment (33 and 28 different modeling protocols for the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors, respectively) were stored. This resulted in over 8,000 structures for each receptor subtype, which provided a unique opportunity to retrospectively assess our GPCR modeling strategy after the crystal coordinates were released. In the following sections, we present an analysis of how our modeling strategies affected the accuracy of the predictions, focusing in particular on the binding mode of ergotamine, the receptor binding pocket, and EL2. We mainly evaluated model accuracy based on the RMSD values to the crystal structure. It should be noted that in this work we used a different method to align the TM region of the models to the crystal structures compared to the organizers of the assessment. This resulted in the same internal ranking of our submitted models, but there were some small differences in the absolute RMSD values. For comparison, our values are shown in Table 1, whereas those from the evaluation of Kufareva et al.[34] are summarized in Supporting Information Table S2. Unless noted otherwise, the values presented in the following are based on our alignment protocol.

Initial Template Selection. Multiple sequence alignment (MSA) of the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptor sequences to the GPCRs for which structures were available showed that the thermostabilized turkey β_1AR had the highest TM sequence identity (>40%) to both serotonin receptor subtypes. In early 2013, 15 crystal structures of this receptor with slight structural variations depending on the cocrystallized ligand were available. Some of these templates were also cocrystallized with ligands that share commonalities with the functional profiles of ergotamine,[37,39] but the overall conformation was representative of an inactive state. This suggested that the active structures of the closely related β_2AR, with substantial conformational changes in the cytosolic region and bound to intracellular partners,[56] would not be suitable templates in this case. Instead, we expected the overall conformation of the serotonin receptors to be more similar to an inactive state, a prediction that was later confirmed by the crystal structures.[63] The complex with the partial agonist dobutamine (PDB accession code: 2Y00)[37] was selected as the initial template as it had the highest resolution (Figure 2, Step 1). As both the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors had EL2$^{SS\text{-}TM5}$ regions longer than the β_1AR

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

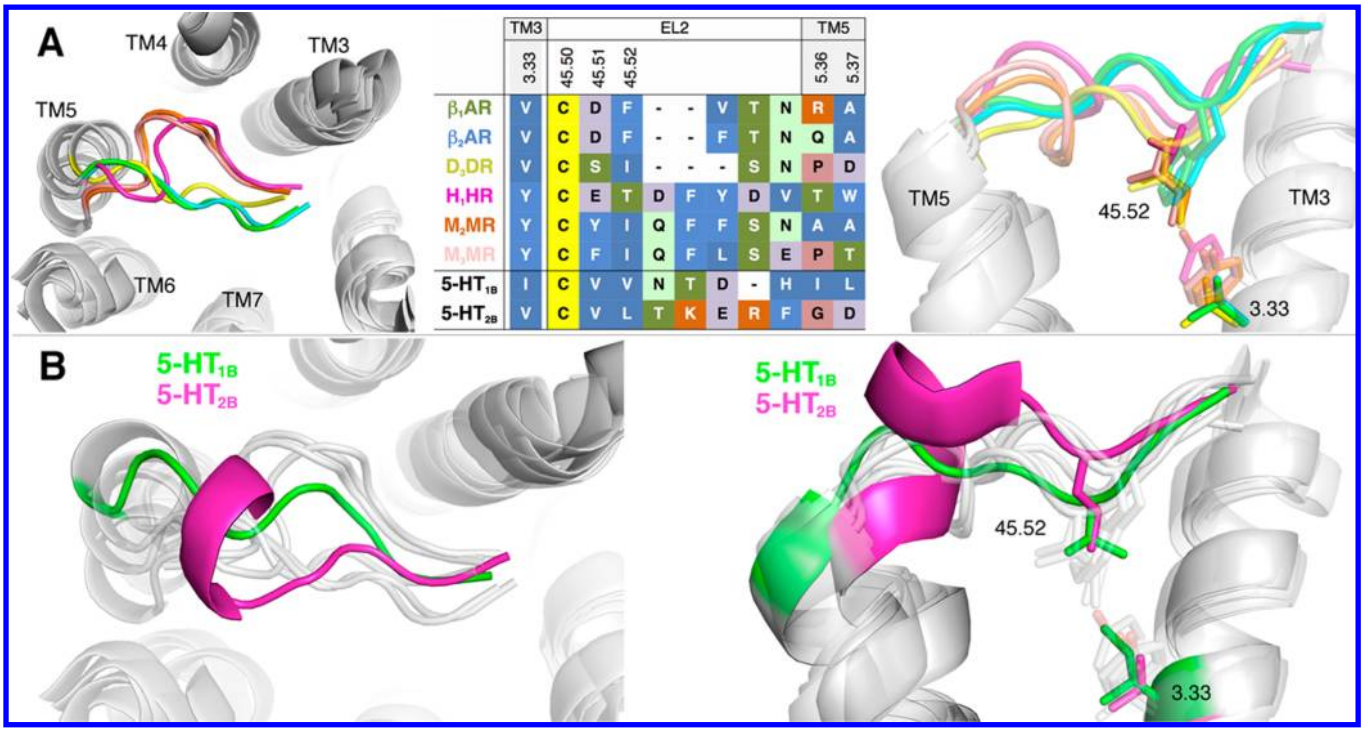


Figure 4. (**A**) Shapes of the EL2$^{SS\text{-}TM5}$ region for aminergic receptors from top (left) and side (right) views. Sequence alignment for selected residues in EL2$^{SS\text{-}TM5}$ for the β_1AR, β_2AR, D_3DR, H_1HR, M_2MR, and M_3MR (color-coded) as well as the 5-HT_{1B} and 5-HT_{2B} receptors (center). The structures correspond to PDB codes 2VT4, 2RH1, 3PBL, 3RZE, 3UON, and 4DAJ. (**B**) Comparison of the EL2$^{SS\text{-}TM5}$ folds for the 5-HT_{1B} (green, PDB code 4IAR) and 5-HT_{2B} (magenta, PDB code 4IB4) receptors to the other aminergic receptors (transparent white).

template, two possible alignments per subtype were explored to accommodate the additional residues (Supporting Information Figure S1B). We will refer to the ensembles of generated models based on the 2Y00 crystal structure and the two EL2 alignments as HM0a and HM0b. The choice of template structure was later optimized in an iterative fashion using molecular docking screening, a point to which we will return.

Modeling of EL2: Incorporation of Knowledge from Multiple GPCR Structures. GPCR crystal structures have revealed important roles of EL2 in ligand binding, receptor signaling, selectivity, and interactions with allosteric modulators.[14,26,64,65] However, the sequence diversity and length of EL2 have hampered accurate prediction of this region and has been highlighted as one of the major challenges in the modeling of GPCR-ligand complexes.[27,32] In the GPCR Dock 2013 assessment, we were able to improve the modeling of EL2 by incorporating information from multiple receptor crystal structures.

(a). Analysis of EL2 Topologies for the Aminergic Receptors. A series of structural motifs has emerged for EL2 as more GPCR crystal structures have been made available.[2,66] As orthosteric ligands have not been found to interact with the section of EL2 that stretches from TM4 to the conserved disulfide bridge, no focus was put on modeling this region. Instead, all our efforts were dedicated to the second stretch of residues from Cys$^{45.50}$ to TM5 (referred to as EL2$^{SS\text{-}TM5}$). The EL2$^{SS\text{-}TM5}$ region appeared to adjust in topology and depth to occupy similar volumes in the different aminergic receptors. Alignment of crystal structures for six aminergic receptors revealed two main folds for the EL2$^{SS\text{-}TM5}$ (Figure 4A, left panel). From the extracellular perspective, the first fold follows a straight path from the disulfide bridge to TM5. This loop conformation was present in the D_3DR, β_1AR, and β_2AR structures. The second fold had a bell-shaped trace and was found in the structures of H_1HR, M_2MR, and M_3MR. From the perspective perpendicular to the plane of the membrane, a common feature for all the EL2$^{SS\text{-}TM5}$ sequences was a hydrophobic residue located in a central position (45.52) of this stretch, which anchored the loop in the orthosteric site (Figure 4A, right panel). The EL2 backbone for receptors with small anchoring side chains (position 45.52: D_3DR/Val, H_1HR/Thr, M_2MR/Ile, and M_3MR/Ile) penetrated slightly deeper into the site compared to the two adrenergic receptors (β_1AR and β_2AR), which had a phenylalanine in the same position. There also appeared to be a correlation between the structure of EL2 and the residue in position 3.33. For the first fold (β_1AR, β_2AR, and D_3DR), there was a small amino acid (Val$^{3.33}$) in TM3, which interacted with the hydrophobic anchor in EL2. Conversely, for the second EL2 fold (represented by the M_2MR, M_3MR, and H_1HR), there were small residues anchoring the loop (Thr and Ile), which were compensated by larger amino acids in TM3 (Tyr$^{3.33}$).

(b). Prediction of EL2 Topologies for the 5-HT_{1B} and 5-HT_{2B} Receptors. Inspection of the 5-HT_{1B} and 5-HT_{2B} receptor sequences suggested that the hydrophobic anchors in EL2 were likely to be residues Val201 and Leu209, respectively (Figure 4A). The smaller amino acids in EL2 indicated that the β_1AR structure was not an optimal choice as template for the EL2$^{SS\text{-}TM5}$ of the serotonin receptor subtypes. Instead, there appeared to be more similarities to the D_3DR, H_1HR, M_2MR, and M_3MR receptor structures. As judged by the size of the side chains in position 3.33 (Ile130/5-HT_{1B} and Val136/5-HT_{2B}), the D_3DR receptor was the closest template for the target receptors. Thus, a set of homology model ensembles based on a chimera of the β_1AR and D_3DR loop were generated and will be referred to as HM1 (Table 2). As will be described in detail

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

Table 2. Summary of Representative Homology Modeling Ensembles Generated in Our Participation in the GPCR Dock 2013 Assessment

			EL2 template					average RMSD (Å)			
receptor	HM ensemble	main template	loop	TM5 extension	additional restraints (dihedral/distance)[a]	GPCR Dock 2013 models	no. of models (generated/docked)[b]	ergotamine[c]	EL2$^{SS\text{-}TM5}$[d]	pocket[c]	average aLogAUC
5-HT$_{1B}$	HM0a	β_1/2Y00	β_1/2Y00	no	0/0		150/50	9.0 (*8*)	3.7	1.8	19.9
	HM0b	β_1/2Y00	β_1/2Y00	no	0/0		150/50	8.4 (*32*)	5.2	1.8	12.4
	HM1	β_1/2Y00	D$_3$/3PBL[e]	no	0/0		150/50	8.6 (*8*)	3.2	1.6	19.1
	HM28	β_1/2Y02	D$_3$/3PBL[f]	no	15/1	#1, #3	500/250	5.9	3.0	1.6	24.8
	HM29	β_1/2Y02	D$_3$/3PBL[g]	no	15/1	#5	500/250	5.7	2.8	1.6	24.6
	HM29	β_1/2Y04	D$_3$/3PBL[g]	no	15/1	#2, #4	500/250	4.2	2.7	1.7	25.9
5-HT$_{2B}$	HM0a	β_1/2Y00	β_1/2Y00	no	0/0		150/50	9.8 (*5*)	3.9	2.4	6.1
	HM0b	β_1/2Y00	β_1/2Y00	no	0/0		150/50	9.9 (*11*)	3.5	2.7	7.7
	HM1	β_1/2Y00	D$_3$/3PBL[e]	yes	0/0		150/50	10.3	2.1	2.4	11.6
	HM10	β_1/2Y00	D$_3$/3PBL[e]	yes	4/1		150/50	9.9	2.0	2.7	17.1
	HM16	β_1/2Y00	D$_3$/3PBL[e]	yes	6/1		250/49	6.6	2.0	2.7	23.0
	HM21	β_1/2Y04	D$_3$/3PBL[e]	yes	9/2	#3, #5	500/200	4.9	1.7	2.5	27.2
	HM26	β_1/4AMJ	β_1/4AMJ	yes	6/2	#4	500/200	6.5	1.8	2.6	16.9
	HM27	β_1/2Y04	D$_3$/3PBL[f]	yes	11/2	#1	500/250	4.5	1.7	2.4	26.9
	HM28	β_1/2Y04	D$_3$/3PBL[f]	yes	12/2	#2	500/250	4.7	1.7	2.5	26.8

[a]Number of side chain dihedral and distance restraints added in MODELLER. See Supporting Information Table S1 for details. [b]Models generated in the ensemble were ranked by their DOPE scores. A subset was then further assessed using molecular docking. [c]Heavy atom RMSD compared to the crystallographic structure. Values are corrected for ligand or side chain symmetry. In italics, the number of models that could not accommodate any conformation of ergotamine without clashes. These models are not considered for the average of ergotamine RMSD, but are included in the rest of calculations. [d]Backbone atom (Cα, C, N) RMSD to the EL2$^{SS\text{-}TM5}$ region of the corresponding crystallographic structure. [e]D$_3$DR EL2 residues: 180$^{45.49}$-VCSIS-184$^{45.43}$. [f]D$_3$DR EL2 residues: 180$^{45.49}$-VCSISN-185$^{45.54}$. [g]D$_3$DR EL2 residues: 180$^{45.49}$-VCSISNA-186$^{45.55}$. Residue Pro186$^{45.55}$ was modeled as alanine.

in the following section, the modeling of EL2 was further optimized using ligand-guided homology modeling. All generated homology model ensembles for the 5-HT$_{1B}$ receptor, except HM0a and HM0b, were based on the chimeric template between the β_1AR and the D$_3$DR. The submitted models originated from two optimized modeling protocols (HM28 and HM29) using two different β_1AR templates, resulting in a total of 750 models.

For the 5-HT$_{2B}$ subtype, a chimera between the β_1AR and the D$_3$DR loop was also the most explored alternative. However, EL2$^{SS\text{-}TM5}$ was three residues longer for the 5-HT$_{2B}$ receptor compared to the D$_3$DR (Figure 4A), which led to structures that had loop residues blocking the orthosteric site in initial homology model ensembles. Analysis of EL2 regions for other GPCR crystal structures showed that peptide-binding receptors had elongated TM5 helices in the extracellular part.[30,67] Addition of modeling restraints to achieve an analogous extra helical turn for the 5-HT$_{2B}$ receptor provided better accommodation of the three consecutive charged residues 211-KER-213 that previously were considered to be part of the loop. These 5-HT$_{2B}$ receptor models were thus based on features from three different GPCR structures. Four of the 5-HT$_{2B}$ receptor structures that were submitted to the GPCR Dock 2013 assessment originated from homology model ensembles generated with the chimera template based on the β_1AR, D$_3$DR EL2$^{SS\text{-}TM5}$, and an extended TM5 helix (HM21, HM27, and HM28; a total of 700 models). As the anchoring residue for the 5-HT$_{2B}$ receptor was a leucine, which is slightly larger than the corresponding Ile183 for the D$_3$DR, we also considered that this residue might not penetrate as deep in the orthosteric site. To explore this alternative, an ensemble of models based on a chimera of the helical extension of TM5 and the β_1AR template was also generated (HM26, a total of 200 models). The parameters used for the generation of each of these homology model ensembles are summarized in Table 2.

(c). Retrospective Analysis of EL2 Predictions for the 5-HT$_{1B}$ and 5-HT$_{2B}$ Receptors. The crystal structures confirmed that residues in position 45.52, Val201 (5-HT$_{1B}$) and Leu209 (5-HT$_{2B}$), anchored EL2 in the orthosteric site. As predicted based on the D$_3$DR, these residues were closer to TM5 compared to Phe201$^{45.52}$ of the β_1AR. The D$_3$DR structure also suggested that the loop was anchored deeper compared to the β_1AR. The backbone trace of the 5-HT$_{1B}$ structure reached a similar depth as the D$_3$DR around position 45.52. However, Leu209$^{45.52}$ did not move as far down into the orthosteric site for the 5-HT$_{2B}$ subtype. This was probably because of the slightly larger size of this residue compared to Val201$^{45.52}$ of the 5-HT$_{1B}$ receptor (Figure 4B, right panel), which could contribute to the stabilization of a more elevated conformation as observed for the βAR structures. This also explained why the alternative template chimera, in which the β_1AR was combined with an extended TM5 helix (HM26), also led to good predictions for the 5-HT$_{2B}$ loop (Tables 1 and 2). The EL2$^{SS\text{-}TM5}$ fold for the 5-HT$_{1B}$ receptor appeared to be an intermediate between the two folds observed for the crystal structures that were available during the assessment (Figure 4B, left panel). This suggested that a chimeric template with, for example, the H$_1$HR would also have resulted in an accurate loop model for the 5-HT$_{1B}$ subtype given an optimal alignment. The 5-HT$_{2B}$ loop followed the predicted fold, and the extra helical turn anticipated based on peptide-binding receptor structures was also correct.

The average EL2$^{SS\text{-}TM5}$ RMSDs for the initial and final homology model ensembles are shown in Table 2. First, it should be pointed out that there was no significant difference in

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

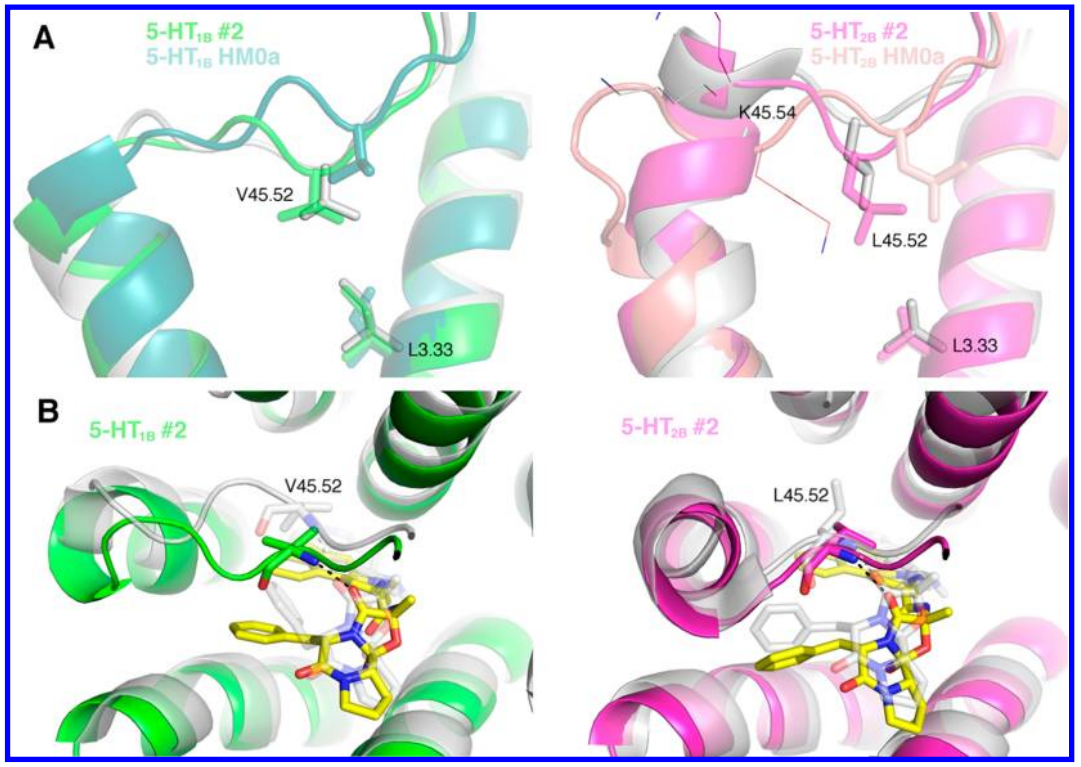


Figure 5. (**A**) Comparison of representative conformations of $EL2^{SS\text{-}TM5}$ from initial (HM0) and submitted models of the 5-HT_{1B} (left) and 5-HT_{2B} (right) receptors. Key residues for anchoring the loop in the orthosteric site (45.52, 3.33) are shown in sticks. $Lys211^{45.54}$ of the 5-HT_{2B} receptor, which was found to block the binding site in HM0a, is shown in lines. (**B**) Top view of our best complexes for each receptor. The hydrogen bond between the carbonyl of the L-alanine moiety of ergotamine and the backbone nitrogen of residue 45.52 is depicted as a black dashed line. The crystallographic structures of the 5-HT_{1B} and 5-HT_{2B} receptors are colored in white in both (**A**) and (**B**).

the accuracy of the full EL2 loop for any of the receptors (data not shown), but this was not surprising as the first part of the loop was not targeted by our modeling protocol. Instead, comparing $EL2^{SS\text{-}TM5}$ for the 5-HT_{1B} receptor homology models that were only based on the β_1AR template (HM0a and HM0b) to the final ensembles (HM29), the average backbone RMSD was reduced from 4.5 to 2.8 Å. Similarly, the average RMSD was reduced from 3.7 to 1.7 Å for the 5-HT_{2B} receptor. Encouragingly, the best models of $EL2^{SS\text{-}TM5}$ were obtained for homology modeling ensembles that combined the β_1AR, the EL2 of the D_3DR, and in the case of the 5-HT_{2B} receptor, a helical extension of TM5. Our best submission for the 5-HT_{2B} subtype had a strikingly accurate $EL2^{SS\text{-}TM5}$ region with an RMSD of only 1.3 Å (Figure 5).

(d). Extension of the Combined Template Approach for EL2 to Other GPCRs. Our approach for predicting the structure of EL2 can also be extended to other receptors. Alignment of the 41 human aminergic GPCRs suggests that chimeras of crystal structures may also be appropriate templates for modeling EL2 for other family members. For example, the β_1AR receptor crystal structures are suitable templates for modeling several of the αARs based on both high sequence identity and recognition of the same endogenous ligand. However, only the three βAR subtypes have bulky phenylalanine side chains in the position of the hydrophobic anchor (45.52) and a small residue in TM3, whereas the α_{1A}, α_{1B}, and α_{1D} subtypes instead have the same combinations of residues as the D_3DR (Supporting Information Figure S3). This suggests that the use of a chimeric template of the β_1AR TM region and the D_3DR loop would be appropriate for modeling several αAR receptors. In the case of the 5-HT_{1B} and 5-HT_{2B} receptors, the combined templates were designed after careful analysis of structures for related receptors. An extension of this approach would be to enumerate all crystal structures and possible combinations of TM regions and EL2 stretches in model building, which would currently result in 64 different templates for the aminergic receptors. For the αARs, this would be a particularly suitable approach as the full-length D_3DR crystal structure by itself is also a good template for some subtypes. The same strategy can likely also be applied to other GPCR families. Conserved features for EL2 have also been identified among peptide-binding receptors, for which the same β-hairpin fold has been observed in a large number of crystal structures.[66] This suggests that combined templates of different peptide receptors could contribute to more accurate models for the many receptors of unknown structure in this branch of the GPCR superfamily.

Ligand-Guided Optimization of Homology Models. The great interest in GPCRs as drug targets has resulted in the characterization of a large number of ligands for several receptor families. After decades of medicinal chemistry efforts, there are thousands of GPCR ligands with annotated activities in databases such as ChEMBL.[51] Quite remarkably, there are at least five annotated ligands for 145 GPCRs, with a median number of 380 compounds per receptor.[68] The ability of molecular docking methods to distinguish ligands from decoys has become a widely used approach to assess if crystal structures could be useful in prospective virtual screening.[52,69] Examples in the field of GPCRs include the work of Katritch et al., in which a 41% hit rate was achieved using a crystallographic structure of A_{2A}AR optimized to recognize known antagonists.[70] An analogous approach was used by Lane et al. to

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

identify D_3DR ligands with an impressive hit rate of 56%.[71] In another recent study, Weiss et al. used molecular docking and ligand enrichment to customize a β_2AR structure for recognition of agonists.[72] Prospective molecular docking screening against the optimized structure led to the discovery of several β_2AR ligands with the desired efficacy. The same approach can be employed to optimize homology models of GPCRs, a procedure referred to as ligand-guided, -supported, or -steered homology modeling.[73−76] This method was pioneered by Evers et al.[77] and was further developed by Abagyan and co-workers.[78] In early applications to GPCRs, the technique was used successfully in prospective virtual screening for α_1AR[79] and melanin-concentrating hormone receptor ligands.[78] Here, we present our implementation of ligand-guided receptor optimization, which was used to generate our 5-HT_{1B} and 5-HT_{2B} receptor models in the GPCR Dock 2013 competition.

(a). Optimization of the 5-HT_{1B} and 5-HT_{2B} Receptor Models. Ligand-guided optimization of the homology modeling protocols for the 5-HT_{1B} and 5-HT_{2B} receptors was carried out in an iterative fashion (Figure 2, Step 2). We systematically explored changes to the homology modeling protocol that involved the use of different template crystal structures of the β_1AR, tailored sampling of EL2, and additional restraints on residues in the orthosteric site. In each round of homology modeling, a set of known ligands together with thousands of decoys[52] were docked to the orthosteric site of the generated structures. Ligand enrichment was then quantified using the aLogAUC metric.[47] The average aLogAUC value for the structures in an ensemble was used as a measure of the quality of the modeling protocol. A majority of modifications that resulted in improved ability to recognize known ligands were taken forward into subsequent rounds of modeling. It should be noted that no particular focus was put on the docking of ergotamine in the initial optimization rounds. Instead, we presumed that generation of receptor models that were able to enrich a large number of ligands would also capture representative binding site conformations of the receptor-ergotamine complex.

Initial homology models for the 5-HT_{1B} and 5-HT_{2B} receptor structures were based on the β_1AR with two alternative alignments for EL2 (HM0a and HM0b, Supporting Information Figure S1B). For the 5-HT_{1B} receptor, both the HM0a and HM0b ensembles displayed positive enrichment of known ligands. The average enrichments for the first 5-HT_{2B} model ensembles were also better than random selection, but their aLogAUC values were significantly lower than for the 5-HT_{1B} subtype (Table 2). Subsequently, roughly 30 homology modeling protocols were explored for each receptor to optimize the orthosteric sites, following the same modeling strategy for both subtypes. Here we will focus on the ligand-guided optimization procedure for the 5-HT_{2B} receptor, which turned out to be more challenging. Several strategies for modeling the EL2 of the 5-HT_{2B} receptor were considered. The most extensively explored option was a combination of the D_3DR loop and a helical extension of TM5. In this case, introduction of the chimeric template between the β_1AR, the D_3DR loop, and the helical extension of TM5 led to a significant increase in the aLogAUC value. The subsequent rounds of optimization primarily involved restriction of side chain rotamers. These were identified either from available crystal structures of other aminergic receptors or consensus rotamer orientations from top enriching homology models. For example, the side chain of Asp135$^{3.32}$, which is conserved among the aminergic receptors, was tightly restrained to the rotamer of the template structure (Figure 2, Step 2C). The side chains of several residues in TM5 and TM6 with strong influence on the binding site volume were restrained based on analysis of predicted ligand binding modes for high enriching models (Supporting Information Table S1). After these restraints were applied, the average ligand enrichment for the homology model ensemble increased from 11.6 (HM1) to 17.1 (HM10) (Table 2). In the following rounds, a slight adjustment of the EL2 chimera and additional side chain restraints further improved the aLogAUC to 23.0 (HM16) (Figure 2, Step 2D). In the next step, 11 different crystal structures of the β_1AR, which all had the EL2 section replaced by the corresponding region of the D_3DR, were used to generate homology models (Figure 2, Step 2B). Higher enrichments compared to the initial template (PDB code: 2Y00) were found for several structures. Two of the best enriching β_1AR templates (PDB codes 2Y02 and 2Y04) had slightly different cavity shapes compared to the 2Y00 structure.[37] These structures were used as the main templates in further optimization steps, and, from this point, larger ensembles of models were considered. A few restraints added in the last steps of receptor refinement did not significantly alter enrichment but were supported by crystal structures of related receptors. For example, a distance restraint was added to favor a salt bridge between Lys211$^{45.54}$ in EL2 and Glu351$^{6.62}$ in EL3. Salt bridge interactions between these loops have been observed in several GPCR crystal structures, e.g. for the β_1AR, β_2AR, M_3MR, and A_{2A}AR.[7,10,12,80] Minor modifications were also made to the EL2 modeling protocol, and additional helical restraints were introduced in the extracellular tip of TM7, which resulted in another small increase in enrichment. Four out of the five 5-HT_{2B} structures that were submitted to the GPCR Dock 2013 assessment originated from the described optimization of the homology modeling protocol (HM21, HM27, HM28). All these ensembles had average aLogAUC values close to 27, which is 4-fold higher than the initial models (Table 2). The remaining 5-HT_{2B} model (5-HT_{2B}#4) was based on the alternative chimeric template of the β_1AR (PDB code: 4AMJ) with an extended helical structure for TM5, which was subjected to the same optimization procedure, but only reached an average aLogAUC of 16.9 (HM26). For the 5-HT_{1B} receptor subtype, the average enrichment also increased during optimization, from 19.9 to 25.5 from the best initial (HM0a) to the final (HM29) homology model ensembles, respectively.

(b). Retrospective Analysis of 5-HT_{1B} and 5-HT_{2B} Model Optimization. As our predictions were made without any knowledge of the receptor structure or ligand binding mode, the impact of our optimization protocol on the accuracy of our models was unknown. It was possible that our initial homology model ensembles, without any optimization, would have been more accurate than those submitted to the assessment. The release of the 5-HT_{1B} and 5-HT_{2B} receptor crystal structures allowed us to evaluate how ergotamine and binding site RMSDs were influenced by the ligand-guided optimization (Table 2).

The initial homology model ensembles of the 5-HT_{1B} receptor (HM0a and HM0b) differed only in the alignment of EL2 but diverged in terms of their ability to recognize known ligands (19.9 vs 12.4). Encouragingly, the alternative with the lower EL2$^{SS\text{-}TM5}$ RMSD to the crystal structure also resulted in higher ligand enrichment. For the initial 5-HT_{2B} receptor

models, the aLogAUC values were significantly lower due to steric hindrance in the binding pocket caused by the $EL2^{SS\text{-}TM5}$ region. The sequence alignments for EL2 positioned bulky residues such as $Lys211^{45.54}$ (HM0a) or $Arg213^{45.56}$ (HM0b) in the binding site of many models, which precluded accurate docking of ligands in this region (Figure 5A). The introduction of the chimeric template and the extracellular extension of TM5 (HM1), which shifted both positively charged residues out of the orthosteric site, significantly improved the accuracy of EL2 (Table 2). There was also a close correspondence between increasing ligand enrichment and better models for both receptors. For the HM0 ensembles, ergotamine could not even be docked to 40% and 16% of the 5-HT_{1B} and 5-HT_{2B} receptor models, respectively. The occurrence of these occluded binding sites was either reduced (5-HT_{1B}) or completely solved (5-HT_{2B}) in HM1, but there was still no significant improvement of average RMSD of ergotamine for the models, which remained close to 10 Å. This likely reflected that ergotamine is a relatively large ligand and that the orientations of some residues were still not accurately modeled. The first significant change for ergotamine docking to the 5-HT_{2B} receptor was found in the HM16 ensemble of structures, for which the average RMSD was 6.6 Å. At that point the average enrichment was more than 3-fold higher than the initial models. The following optimization steps, which involved switching to a different β_1AR crystal structure template, led to further reduction of the average ergotamine RMSD as well as increasing ligand enrichment. The RMSD of $EL2^{SS\text{-}TM5}$ was also improved, in part due to the salt bridge restraint between EL2 and EL3 ($Lys211^{45.54}$-$Glu351^{6.62}$), which biased the side chain conformation of $Lys211^{45.54}$ toward that observed in the 5-HT_{2B} crystal structure (Figure 5A). Encouragingly, one of the final optimized homology modeling ensembles for the 5-HT_{2B} receptor had the overall best average ergotamine RMSDs of all protocols. The average RMSD for ergotamine was 9.9 and 4.5 Å for the initial (HM0a,b) and final (HM27) homology model ensembles, respectively. The binding site RMSD was only marginally reduced from 2.6 to 2.5 Å for the same two sets of homology models. For the 5-HT_{1B} receptor subtype, the increase in enrichment also led to improved predictions for the orthosteric site and binding modes for the cocrystallized ligand. The smaller increase in ligand enrichment for the 5-HT_{1B} subtype was explained by less challenging modeling of $EL2^{SS\text{-}TM5}$ region. From the initial (HM0a,b) to final ensembles (HM29) of the 5-HT_{1B} receptor, the average ergotamine RMSD decreased from 8.7 to 4.2 Å, whereas the binding site RMSD was only slightly reduced from 1.8 to 1.7 Å (Table 2).

Figures 6 and 7 summarize the changes in ligand enrichment, binding site, and ergotamine RMSD for initial and final homology model ensembles. Encouragingly, as the ability of the homology models to enrich ligands increased, the RMSDs of ergotamine and $EL2^{SS\text{-}TM5}$ were reduced. In the first homology model ensembles (HM0a,b and HM1), the most accurate ergotamine poses had RMSD values equal to 4.4 and 3.9 Å for the 5-HT_{1B} and 5-HT_{2B} receptors, respectively. The fraction of accurate ligand poses was increased during the optimization process, and in the final homology model ensembles a significant population of solutions with RMSD values around 2 Å was obtained. The binding site RMSDs appeared to be relatively constant, and only marginal improvements were obtained after optimization. Analysis of the models showed that optimization often did not directly influence binding sites residues but instead shifted side chains that were not involved

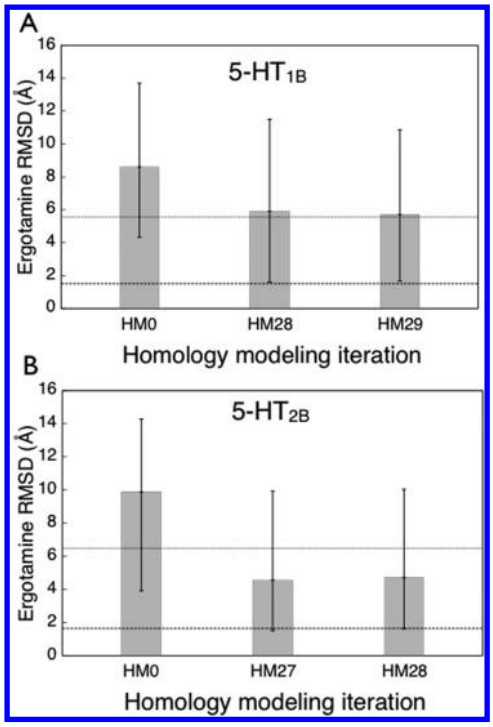


Figure 6. Average ergotamine RMSD (bars) for initial (HM0) and final homology model ensembles (HM27, HM28, or HM29) for the 5-HT_{1B} (**A**) and 5-HT_{2B} (**B**) receptors. The dashed lines represent our best and worst solutions submitted to the GPCR Dock 2013 assessment. The best and worst ergotamine RMSDs from each round of modeling are shown as error bars. The averages do not consider models that could not accommodate any conformation of ergotamine, which was the case for a considerable number of structures in HM0 for both receptors (see Table 2 for details).

in ligand binding out of the binding pocket (e.g., the helical restraints on TM5 introduced in modeling of the 5-HT_{2B} receptor). Such changes were of course crucial for improving the description of the orthosteric site but did not lead to any change in binding site RMSD. An alternative measure of the accuracy of the binding site is the number of correctly predicted receptor−ligand contacts. Comparison of the models with the best ergotamine RMSDs from the initial (HM0) and optimized (HM28 and HM29) homology model ensembles showed that ligand-guided optimization had resulted in at least a 2-fold increase of the number of correctly predicted contacts. In this context, it should also be noted that our submitted models to the GPCR Dock 2013 assessment were also the only predictions that captured more than 50% of the receptor−ligand contacts ($5HT_{1B}$#3/51% and $5HT_{2B}$#4/53%).

Final Selection of Receptor-Ergotamine Complexes: Contact Filtering and Ligand Energy Minimization. In a last step, ergotamine was docked to the final homology modeling ensembles, which resulted in a total of 750 and 900 predicted complexes for the 5-HT_{1B} and 5-HT_{2B} subtypes, respectively (Figure 2, Step 3). Close to all of these models displayed good ligand enrichment, but the docking of ergotamine still resulted in a large number of different binding modes. To facilitate identification of well-enriching models with accurate ergotamine binding poses, the docking solutions were analyzed based on a set of receptor−ligand and internal ligand contacts, which were derived from available experimental data. These hypothesized contacts were not used as strict constraints

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

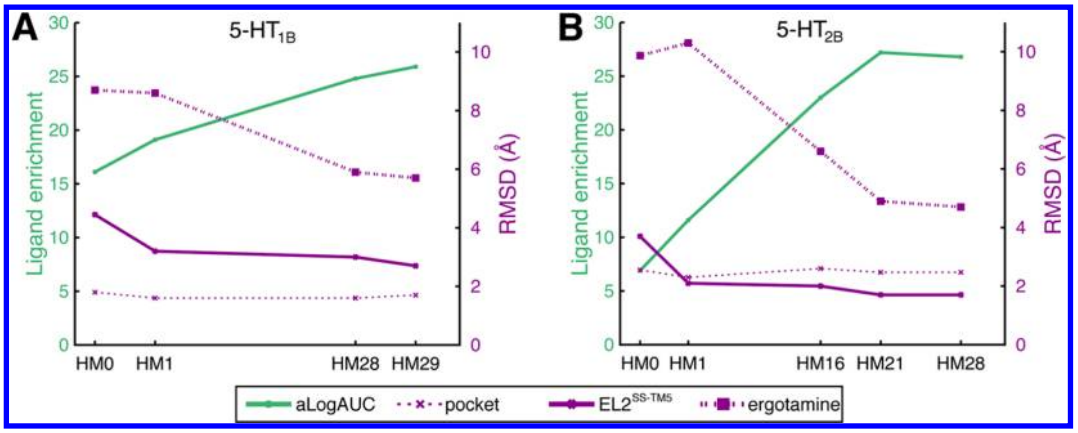


Figure 7. Summary of model accuracy (RMSD for binding pocket, ergotamine, and $EL2^{SS\text{-}TM5}$) and ligand enrichment (aLogAUC) for different modeling ensembles for the 5-HT_{1B} (**A**) and 5-HT_{2B} (**B**) receptors. The models submitted to the GPCR Dock 2013 assessment originated from HM28 and HM29 (5-HT_{1B}) and HM16, HM21, and HM28 (5-HT_{2B}). The averages do not consider models that could not accommodate any conformation of ergotamine, which was the case for some structures from ensembles HM0 and HM1 (see Table 2 for details).

Table 3. Summary of the Effect of Ligand-Residue and Internal-Ligand Contact Filters on the Accuracy of the Ergotamine Binding Mode for the Final Homology Model Ensembles of 5-HT_{1B} and 5-HT_{2B} Receptors

filter applied	5-HT_{1B}[a]				5-HT_{2B}[c]			
	models	ligand RMSD (Å)[b]			models	ligand RMSD (Å)[b]		
		average	min	max		average	min	max
none	750	5.3	1.5	11.5	900	5.1	1.5	16.4
$Asp^{3.32}$	268	2.9	1.5	6.9	731	4.6	1.5	8.2
$Ser/Gly^{5.42}$	452	4.3	1.5	7.7	255	2.8	1.5	5.9
$Asp^{3.32}$ + $Ser/Gly^{5.42}$	203	2.8	1.5	6.1	239	2.8	1.5	5.9
internal h-bond	36	5.8	2.3	9.4	159	5.8	1.5	10.1
all constraints	11	2.9	2.3	5.7	10	2.1	1.5	2.9

[a] 5-HT_{1B} receptor models from ensembles HM28, HM29/2Y02, and HM29/2Y04. See Table 2 for details. [b] Heavy atom RMSD of ergotamine compared to the corresponding crystallographic structure. [c] 5-HT_{2B} receptor models from ensembles HM21, HM26, HM27, and HM28. See Table 2 for details.

but strongly biased our model selection. Further refinement of complexes was also carried out using force field minimization.

Two hypothesized interactions between ergotamine and the 5-HT receptors were derived based on homologous GPCR crystal structures and mutagenesis data. The first was the well-established salt bridge between orthosteric ligands and $Asp^{3.32}$, which is conserved among the aminergic receptors and has been observed in all crystal structures of members of this family. The distance between the aliphatic nitrogen of ergotamine and the closest $Asp^{3.32}$ carboxylate oxygen was less than 3.5 Å for all submitted models (Table 2). The second monitored distance involved the indole nitrogen of ergotamine, but in this case there was no clear interaction partner. Hydrogen bonds with serine residues in TM5 (e.g., $Ser^{5.42}$, $Ser^{5.43}$, and $Ser^{5.46}$) have been suggested to be important for agonist binding and activation of several aminergic receptors based on mutagenesis studies[81,82] and available crystal structures.[37,56] Position 5.42 was also occupied by a serine in the 5-HT_{1B} receptor, whereas there was a glycine in the 5-HT_{2B} subtype. Side chain hydroxyl groups were found for both receptors in the 5.43 position (Ser/5-HT_{1B} and Thr/5-HT_{2B}), but as these residues pointed out of the orthosteric site in our models, they were not considered as a candidates for hydrogen bond interactions with the indole. In the 5.46 position, there was an alanine for both receptors. Based on these observations, we hypothesized that the indole nitrogen formed a hydrogen bond to the side chain hydroxyl of $Ser212^{5.42}$ for the 5-HT_{1B} receptor, but the heavy atom distance cutoff used in the analysis of the docking poses was permissive (<4.5 Å) as there was no strong experimental support for this interaction. For the 5-HT_{2B} receptor, the lack of a polar side chain in positions 5.42 and 5.46 led us to speculate that the indole might interact with the backbone carbonyl of $Gly221^{5.42}$, and the same distance cutoff was used in this case. Nine out of the 10 models that were submitted to the GPCR Dock 2013 assessment fulfilled this criterion, and in seven cases the two atoms were within hydrogen-bonding distance. The crystallographic structure of ergotamine from the Cambridge Structural Database (CSD)[60,61] was also analyzed with the hypothesis that it may provide insights into the bound conformation of the ligand. The ergotamine structure had an internal hydrogen bond, which further restricted the conformational degrees of freedom for this ligand. The presence of this internal hydrogen bond (heavy atom distance cutoff <3.5 Å) was also part of our criteria for model selection. Nine out of the 10 models that were submitted to the GPCR Dock 2013 assessment had this predicted interaction.

Based on contact filtering and visual inspection of models from the final ensembles, 56 representative complexes (20/5-HT_{1B}, 36/5-HT_{2B}) were selected for further refinement, in which an energy minimization of the ligand with the receptor held rigid was performed with the software PLOP.[57] In the final

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

step, the homology models together with the pre- and postminimized poses of ergotamine, a total of 112 complexes, were inspected visually (Figure 2, Step 4). Close to all of these fulfilled the aforementioned receptor−ligand contacts. In selection of the final complexes, we analyzed the receptor−ligand interactions in detail and in particular considered binding energy terms that were not explicitly taken into account by the DOCK3.6 scoring function, e.g. receptor desolvation and ligand internal energy. Five diverse complexes for each receptor were submitted to the GPCR Dock 2013 assessment. The impact of the filtering on ligand contacts and optimization steps on the accuracy of our predictions are presented in the following sections.

(a). Receptor−Ligand Contact Filtering. The predicted interaction between $Asp^{3.32}$ and ergotamine was confirmed by the crystal structures. Retrospective filtering based on this contact (heavy atom distance cutoff <3.5 Å) for the 1,650 optimized homology models also resulted in an average improvement of the ligand RMSD of 2.4 and 0.5 Å for the 5-HT_{1B} and 5-HT_{2B} receptors, respectively (Table 3). The smaller improvement observed for the 5-HT_{2B} receptor reflected that a majority of the docking solutions (81%) already captured this interaction, but this was clearly not sufficient to ensure an overall good binding mode. However, as the solutions that lacked the $Asp^{3.32}$ interaction had an average RMSD of 7.2 Å, this filter essentially discarded a large set of inaccurate poses. The second predicted ligand−receptor interaction (with $Ser/Gly^{5.42}$) was not correct. The indole nitrogen was instead found to interact with $Thr^{3.37}$. Filtering on the distance to $Ser/Gly^{5.42}$ (heavy atom distance cutoff <4.5 Å) resulted in improvements of 1.0 and 2.3 Å for the 5-HT_{1B} and 5-HT_{2B} receptors, respectively. This was largely due to the fact that the application of this constraint anchored the tryptamine group deeper in the orthosteric site. In contrast to the salt-bridge to $Asp^{3.32}$, the second filter influenced model selection more for the 5-HT_{2B} receptor than the 5-HT_{1B} subtype. Whereas the salt bridge was responsible for filtering out a vast majority (nearly 64%) of the models of 5-HT_{1B} receptor, it was the $Gly221^{5.42}$ distance filter for the 5-HT_{2B} subtype that reduced the number of models by nearly 72%. Encouragingly, if the two distance filters were applied together there was a consistent average improvement of ligand RMSDs of 2.5 and 2.3 Å for 5-HT_{1B} and 5-HT_{2B} receptors, respectively, leading to average values below 3 Å for both receptors (Table 3).

(b). Internal Ligand Contact Filter. The internal hydrogen bond predicted based on a small molecule crystal structure of ergotamine was confirmed by the 5-HT_{1B} and 5-HT_{2B} complexes. If the intramolecular hydrogen bond of ergotamine had been applied as a constraint by itself (heavy atom distance cutoff <3.5 Å), it would not have improved the average ergotamine RMSD of either receptor. However, if this criterion was combined with the two receptor-ergotamine contact filters, the average ligand RMSDs would have been improved by 0.7 Å for 5-HT_{2B} receptor, whereas no significant effect would have been observed for the 5-HT_{1B} subtype. This resulted in tightly distributed ensembles of structures with average RMSD values equal to 2.9 and 2.1 Å for the 5-HT_{1B} and 5-HT_{2B} receptors, respectively. Many of these models would likely rank among the best predictions in the GPCR Dock 2013 assessment. It should also be noted that although the filtering on receptor−ligand and internal ligand contacts had a significant impact on the ergotamine RMSDs of our submitted solutions, it was the combination with ligand-guided optimization that enabled accurate predictions. In the first modeling rounds, the distribution of ergotamine RMSDs was shifted toward significantly higher values compared to the final ensembles. For the HM0 and HM1 ensembles, the most accurate docking solution only had an RMSD of 3.9 Å (5-HT_{2B}, HM0b), which is close to the average RMSD of the final ensembles (4.2 Å, 5-HT_{1B}/HM29) and more than 2-fold worse than the best solution that was submitted to the assessment (1.7 Å).

(c). Refinement via Minimization. Retrospective analysis showed that the additional step of ligand refinement for 56 complexes only led to an average improvement of ergotamine RMSDs of 0.1 Å for both receptors. Ligand minimization increased accuracy for 65% and 66% of the models of the 5-HT_{1B} and 5-HT_{2B} receptors, respectively. There was no correlation between the decrease of the ligand RMSD and the initial accuracy of the complex. Thus, although eight out of our ten submitted models had been improved after PLOP optimization, there were equally or more accurate complexes available among those generated only by molecular docking to the homology models.

Blind Ranking of Models. The GPCR Dock assessment also involves ranking the submitted models by their expected accuracy. The complexes blindly ranked as the best predictions were only our fourth most accurate complexes for each receptor (Supporting Information Table S2). The two main factors reducing the accuracy for these complexes were higher placement of the ergoline moiety and a divergent conformation of the phenyl ring of ergotamine (Supporting Information Figure S2). In the two previous assessments, the research group responsible for the most accurate complex has never been able to rank this model as its best prediction.[27,32,34] Our most accurate predictions were instead blindly ranked as our second best solution for both receptors. As the final selection and ranking of models by GPCR Dock participants is typically made by manual inspection, it is difficult to draw conclusions about the accuracy of different modeling strategies from GPCR Dock results. A more detailed analysis of the several thousand structures predicted during our participation revealed strengths and limitations of our modeling strategy. On one hand, our results show that optimized homology models can enable predictions of GPCR-drug complexes that are very similar to the experimental structures (ligand RMSD < 2 Å), but it is still very challenging to rank these by accuracy.[31] However, it is also important to note that among all the generated structures, the most accurate binding mode of ergotamine still had an RMSD of 1.3 Å. This clearly shows that there is a need for development of novel modeling approaches that can account for receptor flexibility and induced fit. In practice, our results suggest that in the case of structure-based ligand optimization, it is likely necessary to work on several models in parallel, which are further refined as more structure−activity data becomes available for a lead compound.

The Role of Experimental Data in Modeling of GPCR-Drug Complexes. A large increase of the available experimental data during the last years has been crucial for improving the accuracy of structure prediction for GPCR-ligand complexes. First of all, it is the revolution in GPCR structural biology that has enabled more accurate modeling of the orthosteric site. Structure prediction for other aminergic receptors can now take advantage of templates with TM sequence identities of close to 40%. With good templates it can be assumed that the overall accuracy of the TM region will be reasonably high using homology modeling. The main challenge

instead comes from prediction of the ligand binding mode, which can be very sensitive to side chain orientations in the orthosteric site as well as the structure of the extracellular loops. For many receptors there is also a vast amount of experimental data in the form of mutagenesis experiments, ligand binding and functional assays, which could be incorporated in model generation.

(a). GPCR Crystal Structures. As in the GPCR Dock 2010 assessment for the D_3DR receptor, the adrenergic receptor crystal structures had the highest TM sequence identity to the target. However, at that time, only one crystal structure had been determined for the β_1AR. Access to multiple structures not only led to identification of models with better ligand enrichment but also enabled more accurate prediction of EL2 and the ergotamine complexes. Further extension of this approach could involve the use of crystal structures of other related receptors or ensembles of elastic network models.[83] However, our results also demonstrate that TM sequence identity is not necessarily a good measure of the difficulty of predicting a GPCR-drug complex. As illustrated by the modeling of the two 5-HT receptor subtypes, small differences in EL2 length can dramatically change the difficulty of modeling the binding site. A significant improvement in the modeling of the orthosteric site was accomplished by utilizing information from multiple GPCR crystal structures. Compared to previous assessments, it appeared as if the improved $EL2^{SS\text{-}TM5}$ predictions increased the number of correct ligand−receptor contacts in this region. First of all, the hydrogen bond between ergotamine and EL2 was captured for both receptors (Table 1 and Figure 5B). There were also interactions between the hydrophobic side chains of residues 45.52 and the ligand, which was one of the contacts that top-ranked groups of the GPCR DOCK 2010 assessment failed to predict for the D_3DR-eticlopride complex.[32] Based on our successful use of combined templates for the EL2, it would also be interesting to investigate if multiple template modeling for the TM region[84] with current crystallographic structures could further improve modeling accuracy for GPCR-drug complexes.

(b). Mutagenesis Data. Site-directed mutagenesis has been a widely used approach to study GPCR-ligand interactions. For example, in the case of the adrenergic, dopamine, and adenosine receptors, several key residues for ligand binding were identified two decades before crystal structures were determined.[8,85,86] However, despite such information being available during the GPCR Dock 2008 and 2010 assessments, few research groups could provide an accurate ligand binding mode for the A_{2A}AR or D_3DR complex. There could be several reasons to why access to extensive mutagenesis data does not necessarily lead to better predictions. First, as demonstrated by our analysis, it may not be sufficient to use one single interaction to improve predictions for a receptor−ligand complex. Two ligand−receptor constraints had to be imposed simultaneously to significantly reduce the RMSD value of ergotamine for both the 5-HT receptors. Second, if all available mutagenesis data is compiled for a specific receptor, the results are often contradictory, which was also noted by participants in the GPCR Dock 2008 assessment.[87] The only polar interaction that our models did not capture for any of the receptors was the hydrogen bond between the indole group and the side chain hydroxyl of $Thr^{3.37}$. Retrospective analysis of all the generated complexes showed that there was not a single complex that predicted the hydrogen bond to this residue. To our knowledge, there was no mutagenesis data available for this residue during the GPCR Dock 2013 assessment. However, would access to such data have improved our predictions? Wang et al. published a series of mutagenesis experiments for the $Thr^{3.37}$ together with the crystal structures. The $Thr^{3.37}$Ala mutant was found to reduce binding of serotonin significantly for the 5-HT_{1B} and 5-HT_{2B} subtypes, but the affinity of ergotamine itself was even slightly enhanced.[15] If this data had been available during the assessment, it would have suggested that ergotamine did not interact with $Thr^{3.37}$, which in the light of the crystal structures would have been misleading. Based on available structures and mutagenesis data of the adrenergic receptors,[8,37,56] we also assumed that residues in TM5 would also play a role in recognition of 5-HT receptor agonists. However, no interactions with polar side chains in TM5 were observed for the 5-HT_{1B} or 5-HT_{2B} receptors. Overall, our experience from the use of mutagenesis experiments is that such data should be interpreted and implemented very cautiously in model building. Receptor−ligand interactions observed in crystal structures of the template may not even be transferrable to closely related GPCRs and changes in ligand binding for mutants should not be considered to be sufficient evidence for a direct interaction with a specific residue.

(c). Small Molecule Crystal Structures. Experimentally determined structures of small molecules are available for a large number of drugs in the CSD and have been widely used in development of ligand-based approaches for drug design.[60] We used the structure of ergotamine to identify an internal hydrogen bond, which was used as a criterion for selecting binding modes. Similarly, a crystal structure for eticlopride was also available in the CSD during the preceding GPCR Dock 2010 challenge to predict the complex of D_3DR with this antagonist. In that case, the observed small molecule crystal structure of eticlopride, which was taken into account by several of the top-ranked research groups in the assessment, turned out to be strikingly similar to the receptor-bound conformation.[32] For drug-like molecules with more conformational degrees of freedom, differences between crystal and receptor-bound conformations are expected to be larger. However, as demonstrated by our predictions, small molecule structures may reveal important features that could be used in predictions of GPCR-drug complexes.

(d). Ligand Binding Data. One of the underlying assumptions of ligand-guided homology modeling optimization is that the experimentally determined coordinates, or at least a closely related structure, will have the ability to enrich active compounds. Retrospective docking calculations for the crystallographic structures revealed significant enrichment of ligands for both the 5-HT_{1B} and 5-HT_{2B} subtypes (aLogAUC values of 20 and 19, respectively). These values agreed well with the results obtained for the submitted models (Table 1), which were in the ranges 20−27 and 20−32 for the 5-HT_{1B} and 5-HT_{2B} receptors, respectively. Our top-ranked models (5-HT_{1B}#2 and 5-HT_{2B}#2) displayed higher enrichment than the crystal structures, but this was not surprising given the fact that the models were optimized for ligand recognition.[18,87] Our results thus support that ligand docking can be a powerful tool for evaluating the accuracy of GPCR models. This approach has also provided excellent results to previous GPCR Dock assessments by Katritch et al. for the A_{2A}AR (ranked as number 2 out of 206 models)[27,87] and Carlsson et al. for the D_3DR (ranked as number 4 out of 118 models).[32] These models were also demonstrated to be useful for structure-based ligand discovery in retrospective[87] and prospective molecular

dx.doi.org/10.1021/ci5002235 | J. Chem. Inf. Model. 2014, 54, 2004−2021

docking screens.[18] Compared to the ligand-guided homology modeling approach of Carlsson et al., which was based on a single docking screen against several thousand models, our current method was instead based on iterative optimization of smaller ensembles of models. Currently, our efforts are focused on full automation of the optimization procedure, which will enable large-scale modeling of GPCR structures. This could give access to *in silico* screens for subtype selective ligands, which is one of the major challenges in GPCR drug discovery.[64,88] However, it is not yet clear if homology models are accurate enough to capture subtle differences between closely related receptors of unknown structure.[89] It should also be noted that all these strategies are strongly dependent on access to known ligands, which hinders the application of the method to less characterized receptors. Molecular modeling of GPCR-drug complexes for receptors with low sequence identity to available templates and few known ligands remains to be very challenging.[34] For more challenging targets, it is not clear if homology models can contribute to the discovery of novel ligands, which was illustrated by the prospective virtual screen of Mysinger et al. In this case, significantly lower docking hit rate was achieved with a homology model compared to the crystal structure of the CXCR4 receptor.[90] Returning to the serotonin receptors, the strong ligand enrichment and high accuracy of the blindly predicted binding modes suggests that these models would be useful in structure-based ligand discovery.[18] Accurate models can now likely be generated for the 33 remaining aminergic receptors as well as several peptide-binding GPCRs, like the chemokine receptor family, for which valuable templates have recently become available.

4. CONCLUSIONS

The GPCR Dock assessments have challenged the community to predict GPCR-ligand complexes at atomic resolution, one of the long-standing goals in the field of structure-based drug design. The results from these assessments have not only provided guidelines for the limitations of GPCR modeling techniques and their application in structure-based drug design but have also stimulated development of new modeling strategies. Here, we have presented one of the approaches that resulted in accurate predictions for both the 5-HT_{1B} and 5-HT_{2B} orthosteric sites and the binding modes of ergotamine. Retrospective analysis of over 1000 receptor−ligand complexes generated during our participation in the GPCR Dock 2013 assessment led to the identification of the following guidelines for more accurate predictions of GPCR-drug complexes: (*i*) *Access to multiple GPCR crystal structures enables more accurate homology modeling of EL2.* Significant improvements in accuracy could be achieved by taking advantage of structural features of EL2 from up to three different receptors and slight variations in the binding pockets of β_1AR crystal structures. In particular, the use of chimeric templates is straightforward to implement in any homology modeling method and will likely be useful for modeling of a large number of other GPCRs. (*ii*) *Ligand-guided homology modeling is a powerful approach for modeling GPCR-drug complexes.* Guided by molecular docking and ligand enrichment, we were able to optimize both the loop and binding site structure for ligand recognition, which also enabled more accurate predictions of the ergotamine binding mode. In addition, this approach also ensures that the generated receptor structure is suitable for prospective virtual screening. (*iii*) *All available experimental data should be considered but assessed critically.* Our prediction of the binding mode for ergotamine involved the use of a large number of known ligands, extensive analysis of homologous receptor structures, a small molecule crystal structure of ergotamine, and mutagenesis data. Whereas access to crystal structures of several closely related receptors and ligand binding data likely made the most important contributions to the accuracy of our predictions, our analysis suggests that protocols for modeling of a GPCR-drug complex should be tailored to take into account all available experimental data. The probability for selecting an accurate binding pose was increased by taking into account mutagenesis and small molecule crystal structure data, but this was not necessarily true when they were used individually. Finally, as illustrated by the results for the smoothened receptor in the GPCR Dock 2013 assessment, prediction of receptor-drug complexes with low sequence similarity to available templates and few characterized ligands remains very challenging.

To summarize, successful prediction of GPCR-drug complexes relies on optimal utilization of available experimental data from multiple sources together with a combination of a diverse set of molecular modeling techniques. The accurate predictions achieved in the GPCR Dock 2013 assessment suggest that successful structure-based drug design is not restricted to crystal structures but can also be extended to many other pharmaceutically relevant receptors.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1, S2, and S3 and Tables S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jens.carlsson@dbb.su.se.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the groups of Professors Stevens and Abagyan for the organization of the GPCR Dock 2013 assessment. This work was supported by grants from the Knut and Alice Wallenberg foundation, the Center for Biomembrane Research, the Swedish Foundation for Strategic Research, and the Swedish e-Science Research Center to J.C. Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) and National Supercomputer Centre (NSC) in Linköping. We thank OpenEye Scientific Software (Santa Fe, NM) for the use of Omega and OEChem at no cost. The authors participate in the European COST Action CM1207 (GLISTEN).

■ ABBREVIATIONS

GPCR, G protein-coupled receptor; TM, transmembrane helix; IL, intracellular loop; EL, extracellular loop; βAR, β-adrenergic receptor; A_{2A}AR, A_{2A} adenosine receptor; αAR, α-adrenergic receptor; RMSD, root-mean-square deviation; DR, dopamine receptor; MR, muscarinic receptor; HR, histamine receptor; 5-HT, 5-hydroxytryptamine, serotonin; MSA, multiple sequence alignment; CSD, Cambridge Structural Database

■ REFERENCES

(1) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* **2006**, *5* (12), 993−996.

(2) Katritch, V.; Cherezov, V.; Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **2011**, *33* (1), 17−27.
(3) Rosenbaum, D. M.; Rasmussen, S. G.; Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **2009**, *459* (7245), 356−363.
(4) Filipek, S.; Teller, D. C.; Palczewski, K.; Stenkamp, R. The Crystallographic Model of Rhodopsin and Its Use in Studies of Other G Protein-Coupled Receptors. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 375−397.
(5) Tebben, A. J.; Schnur, D. M. Beyond rhodopsin: G protein-coupled receptor structure and modeling incorporating the beta2-adrenergic and adenosine A(2A) crystal structures. *Methods Mol. Biol.* **2011**, *672*, 359−386.
(6) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Yao, X.-J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **2007**, *318* (5854), 1266−1273.
(7) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318* (5854), 1258−1265.
(8) Strader, C. D.; Fong, T. M.; Tota, M. R.; Underwood, D.; Dixon, R. A. Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* **1994**, *63*, 101−132.
(9) Shi, L.; Javitch, J. A. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 437−467.
(10) 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. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **2008**, *454* (7203), 486−491.
(11) Shimamura, T.; Shiroishi, M.; Weyand, S.; Tsujimoto, H.; Winter, G.; Katritch, V.; Abagyan, R.; Cherezov, V.; Liu, W.; Han, G. W.; Kobayashi, T.; Stevens, R. C.; Iwata, S. Structure of the human histamine H1 receptor complex with doxepin. *Nature* **2011**, *475* (7354), 65−70.
(12) Kruse, A. C.; Hu, J.; Pan, A. C.; Arlow, D. H.; Rosenbaum, D. M.; Rosemond, E.; Green, H. F.; Liu, T.; Chae, P. S.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Wess, J.; Kobilka, B. K. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **2012**, *482* (7386), 552−556.
(13) Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; Kobayashi, T. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **2012**, *482* (7386), 547−551.
(14) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Han, G. W.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **2010**, *330* (6007), 1091−1095.
(15) Wang, C.; Jiang, Y.; Ma, J.; Wu, H.; Wacker, D.; Katritch, V.; Han, G. W.; Liu, W.; Huang, X. P.; Vardy, E.; McCorvy, J. D.; Gao, X.; Zhou, X. E.; Melcher, K.; Zhang, C.; Bai, F.; Yang, H.; Yang, L.; Jiang, H.; Roth, B. L.; Cherezov, V.; Stevens, R. C.; Xu, H. E. Structural basis for molecular recognition at serotonin receptors. *Science* **2013**, *340* (6132), 610−614.
(16) Kolb, P.; Rosenbaum, D. M.; Irwin, J. J.; Fung, J. J.; Kobilka, B. K.; Shoichet, B. K. Structure-based discovery of beta2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6843−6848.
(17) Carlsson, J.; Yoo, L.; Gao, Z.-G.; Irwin, J. J.; Shoichet, B. K.; Jacobson, K. A. Structure-based discovery of A2A adenosine receptor ligands. *J. Med. Chem.* **2010**, *53* (9), 3748−3755.
(18) Carlsson, J.; Coleman, R. G.; Setola, V.; Irwin, J. J.; Fan, H.; Schlessinger, A.; Sali, A.; Roth, B. L.; Shoichet, B. K. Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nat. Chem. Biol.* **2011**, *7* (11), 769−778.
(19) de Graaf, C.; Kooistra, A. J.; Vischer, H. F.; Katritch, V.; Kuijer, M.; Shiroishi, M.; Iwata, S.; Shimamura, T.; Stevens, R. C.; de Esch, I. J.; Leurs, R. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J. Med. Chem.* **2011**, *54* (23), 8195−8206.
(20) Kruse, A. C.; Weiss, D. R.; Rossi, M.; Hu, J.; Hu, K.; Eitel, K.; Gmeiner, P.; Wess, J.; Kobilka, B. K.; Shoichet, B. K. Muscarinic Receptors as Model Targets and Antitargets for Structure-Based Ligand Discovery. *Mol. Pharmacol.* **2013**, *84* (4), 528−540.
(21) Rodríguez, D.; Gutiérrez-de-Teran, H. Computational Approaches for Ligand Discovery and Design in Class-A G Protein-Coupled Receptors. *Curr. Pharm. Des.* **2013**, *19*, 2216−2236.
(22) Moult, J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **2005**, *15* (3), 285−289.
(23) Lensink, M. F.; Méndez, R.; Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins: Struct., Funct., Bioinf.* **2007**, *69* (4), 704−718.
(24) Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B., Jr.; Stuckey, J. A.; Carlson, H. A. CSAR benchmark exercise 2011−2012: evaluation of results from docking and relative ranking of blinded congeneric series. *J. Chem. Inf. Model.* **2013**, *53* (8), 1853−1870.
(25) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput.-Aided Mol. Des.* **2010**, *24* (4), 259−279.
(26) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322* (5905), 1211−1217.
(27) Michino, M.; Abola, E.; GPCR Dock 2008 participants; Brooks, C. L.; Dixon, J. S.; Moult, J.; Stevens, R. C. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discovery* **2009**, *8* (6), 455−463.
(28) Costanzi, S.; Siegel, J.; Tikhonova, I. G.; Jacobson, K. A. Rhodopsin and the others: a historical perspective on structural studies of G protein-coupled receptors. *Curr. Pharm. Des.* **2009**, *15* (35), 3994−4002.
(29) Costanzi, S. Homology modeling of class A G protein-coupled receptors. *Methods Mol. Biol.* **2012**, *857*, 259−279.
(30) Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Broun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, *330* (6007), 1066−1071.
(31) Obiol-Pardo, C.; Lopez, L.; Pastor, M.; Selent, J. Progress in the structural prediction of G protein-coupled receptors: D3 receptor in complex with eticlopride. *Proteins: Struct., Funct., Bioinf.* **2011**, *79* (6), 1695−1703.
(32) Kufareva, I.; Rueda, M.; Katritch, V.; GPCR Dock 2010 participants; Stevens, R. C.; Abagyan, R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **2011**, *19* (8), 1108−1126.
(33) Roumen, L.; Sanders, M. P. A.; Vroling, B.; De Esch, I. J. P.; De Vlieg, J.; Leurs, R.; Klomp, J. P. G.; Nabuurs, S. B.; De Graaf, C. In Silico Veritas: The Pitfalls and Challenges of Predicting GPCR-Ligand Interactions. *Pharmaceuticals* **2011**, *4* (9), 1196−1215.
(34) Kufareva, I.; Katritch, V.; GPCR Dock 2013 Participants; Stevens, R. C.; Abagyan, R. Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* **2014**.
(35) Berger, M.; Gray, J. A.; Roth, B. L. The expanded biology of serotonin. *Annu. Rev. Med.* **2009**, *60*, 355−366.
(36) Rodríguez, D.; Bello, X.; Gutiérrez-de-Terán, H. Molecular Modelling of G Protein-Coupled Receptors Through the Web. *Mol. Inf.* **2012**, *31* (15), 334−341 In press.
(37) Warne, T.; Moukhametzianov, R.; Baker, J. G.; Nehme, R.; Edwards, P. C.; Leslie, A. G.; Schertler, G. F.; Tate, C. G. The

structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor. *Nature* **2011**, *469* (7329), 241−244.
(38) Moukhametzianov, R.; Warne, T.; Edwards, P. C.; Serrano-Vega, M. J.; Leslie, A. G.; Tate, C. G.; Schertler, G. F. Two distinct conformations of helix 6 observed in antagonist-bound structures of a beta1-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (20), 8228−8232.
(39) Warne, T.; Edwards, P. C.; Leslie, A. G.; Tate, C. G. Crystal structures of a stabilized beta1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* **2012**, *20* (5), 841−849.
(40) Ballesteros, J. A.; Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. *Methods Neurosci.* **1995**, *25*, 366−428.
(41) de Graaf, C.; Foata, N.; Engkvist, O.; Rognan, D. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins: Struct., Funct., Bioinf.* **2008**, *71* (2), 599−620.
(42) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234* (3), 779−815.
(43) Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: a feasibility study. *J. Med. Chem.* **2009**, *52* (18), 5712−5720.
(44) Shen, M. Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, *15* (11), 2507−2524.
(45) Lorber, D. M.; Shoichet, B. K. Hierarchical docking of databases of multiple ligand conformations. *Curr. Top. Med. Chem.* **2005**, *5* (8), 739−749.
(46) Shoichet, B. K.; Kuntz, I. D. Matching chemistry and shape in molecular docking. *Protein Eng.* **1993**, *6* (7), 723−732.
(47) Mysinger, M. M.; Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **2010**, *50* (9), 1561−1573.
(48) Nicholls, A.; Honig, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson−Boltzmann equation. *J. Comput. Chem.* **1991**, *12* (4), 435−445.
(49) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765−784.
(50) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Proteins: Struct., Funct., Bioinf.* **1999**, *34* (1), 4−16.
(51) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40* (Database issue), D1100−7.
(52) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582−6594.
(53) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572−584.
(54) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First Solvation Shell Effects. *J. Phys. Chem.* **1996**, *100* (40), 16385−16398.
(55) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *J. Comput. Chem.* **1986**, *7* (2), 230−252.
(56) Rasmussen, S. G.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **2011**, *477* (7366), 549−555.
(57) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 351−367.
(58) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(59) Schlick, T.; Overton, M. A Powerful Truncated Newton Method for Potential-Energy Minimization. *J. Comput. Chem.* **1987**, *8* (7), 1025−1039.
(60) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B* **2002**, *58* (Pt 3 Pt 1), 380−388.
(61) Klepetářováa, B.; Čejkaa, J.; Kratochvíla, B.; Pakhomovaa, S.; Císařováb, I.; Cvakc, L.; Jegorov, A. Interesting Solvent Area in Crystal Structures of Two Natural Ergot Alkaloids. *Collect. Czech. Chem. Commun.* **2005**, *70* (1), 41−50.
(62) *The PyMOL Molecular Graphics System*, Version 1.4.1; Schrödinger, LLC.
(63) Wacker, D.; Wang, C.; Katritch, V.; Han, G. W.; Huang, X. P.; Vardy, E.; McCorvy, J. D.; Jiang, Y.; Chu, M.; Siu, F. Y.; Liu, W.; Xu, H. E.; Cherezov, V.; Roth, B. L.; Stevens, R. C. Structural features for functional selectivity at serotonin receptors. *Science* **2013**, *340* (6132), 615−619.
(64) Katritch, V.; Kufareva, I.; Abagyan, R. Structure based prediction of subtype-selectivity for adenosine receptor antagonists. *Neuropharmacology* **2011**, *60* (1), 108−115.
(65) Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J.; Hu, K.; Eitel, K.; Hubner, H.; Pardon, E.; Valant, C.; Sexton, P. M.; Christopoulos, A.; Felder, C. C.; Gmeiner, P.; Steyaert, J.; Weis, W. I.; Garcia, K. C.; Wess, J.; Kobilka, B. K. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **2013**, *504* (7478), 101−106.
(66) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular signatures of G-protein-coupled receptors. *Nature* **2013**, *494* (7436), 185−94.
(67) Wu, H.; Wacker, D.; Mileni, M.; Katritch, V.; Han, G. W.; Vardy, E.; Liu, W.; Thompson, A. A.; Huang, X. P.; Carroll, F. I.; Mascarella, S. W.; Westkaemper, R. B.; Mosier, P. D.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* **2012**, *485* (7398), 327−332.
(68) Lin, H.; Sassano, M. F.; Roth, B. L.; Shoichet, B. K. A pharmacological organization of G protein-coupled receptors. *Nat. Methods* **2013**, *10* (2), 140−146.
(69) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49* (20), 5912−5931.
(70) Katritch, V.; Jaakola, V. P.; Lane, J. R.; Lin, J.; Ijzerman, A. P.; Yeager, M.; Kufareva, I.; Stevens, R. C.; Abagyan, R. Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J. Med. Chem.* **2010**, *53* (4), 1799−1809.
(71) Lane, J. R.; Chubukov, P.; Liu, W.; Canals, M.; Cherezov, V.; Abagyan, R.; Stevens, R. C.; Katritch, V. Structure-based ligand discovery targeting orthosteric and allosteric pockets of dopamine receptors. *Mol. Pharmacol.* **2013**, *84* (6), 794−807.
(72) Weiss, D. R.; Ahn, S.; Sassano, M. F.; Kleist, A.; Zhu, X.; Strachan, R.; Roth, B. L.; Lefkowitz, R. J.; Shoichet, B. K. Conformation guides molecular efficacy in docking screens of activated beta-2 adrenergic G protein coupled receptor. *ACS Chem. Biol.* **2013**, *8* (5), 1018−1026.
(73) Evers, A.; Klebe, G. Ligand-supported homology modeling of g-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew. Chem., Int. Ed. Engl.* **2004**, *43* (2), 248−251.
(74) Phatak, S. S.; Gatica, E. A.; Cavasotto, C. N. Ligand-steered modeling and docking: A benchmarking study in class A G-protein-coupled receptors. *J. Chem. Inf. Model.* **2010**, *50* (12), 2119−2128.
(75) Katritch, V.; Rueda, M.; Abagyan, R. Ligand-guided receptor optimization. *Methods Mol. Biol.* **2012**, *857*, 189−205.

(76) Kolaczkowski, M.; Bucki, A.; Feder, M.; Pawlowski, M. Ligand-optimized homology models of d1 and d2 dopamine receptors: application for virtual screening. *J. Chem. Inf. Model.* **2013**, *53* (3), 638−648.
(77) Evers, A.; Gohlke, H.; Klebe, G. Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *J. Mol. Biol.* **2003**, *334* (2), 327−345.
(78) Cavasotto, C. N.; Orry, A. J.; Murgolo, N. J.; Czarniecki, M. F.; Kocsi, S. A.; Hawes, B. E.; O'Neill, K. A.; Hine, H.; Burton, M. S.; Voigt, J. H.; Abagyan, R. A.; Bayne, M. L.; Monsma, F. J., Jr. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.* **2008**, *51* (3), 581−588.
(79) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.* **2005**, *48* (17), 5448−65.
(80) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; IJzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **2012**, *337* (6091), 232−236.
(81) Strader, C. D.; Candelore, M. R.; Hill, W. S.; Sigal, I. S.; Dixon, R. A. Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor. *J. Biol. Chem.* **1989**, *264* (23), 13572−13578.
(82) Liapakis, G.; Ballesteros, J. A.; Papachristou, S.; Chan, W. C.; Chen, X.; Javitch, J. A. The forgotten serine. A critical role for Ser-2035.42 in ligand binding to and activation of the beta 2-adrenergic receptor. *J. Biol. Chem.* **2000**, *275* (48), 37779−37788.
(83) Rueda, M.; Bottegoni, G.; Abagyan, R. Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *J. Chem. Inf. Model.* **2009**, *49* (3), 716−725.
(84) Mobarec, J. C.; Sanchez, R.; Filizola, M. Modern homology modeling of G-protein coupled receptors: which structural template to use? *J. Med. Chem.* **2009**, *52* (16), 5207−5216.
(85) Woodward, R.; Coley, C.; Daniell, S.; Naylor, L. H.; Strange, P. G. Investigation of the role of conserved serine residues in the long form of the rat D2 dopamine receptor using site-directed mutagenesis. *J. Neurochem.* **1996**, *66* (1), 394−402.
(86) Martinelli, A.; Tuccinardi, T. Molecular modeling of adenosine receptors: new results and trends. *Med. Res. Rev.* **2008**, *28* (2), 247−277.
(87) Katritch, V.; Rueda, M.; Lam, P. C.; Yeager, M.; Abagyan, R. GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (1), 197−211.
(88) Congreve, M.; Langmead, C. J.; Mason, J. S.; Marshall, F. H. Progress in structure based drug design for G protein-coupled receptors. *J. Med. Chem.* **2011**, *54* (13), 4283−311.
(89) Rodríguez, D.; Brea, J.; Loza, M. I.; Carlsson, J. Structure-based Discovery of Selective Serotonin 5-HT1B Receptor Ligands. *Structure* **2014**, In press. DOI: 10.106/j.str.2014.05.017.
(90) Mysinger, M. M.; Weiss, D. R.; Ziarek, J. J.; Gravel, S.; Doak, A. K.; Karpiak, J.; Heveker, N.; Shoichet, B. K.; Volkman, B. F. Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (14), 5517−5522.