

# Snooker: A Structure-Based Pharmacophore Generation Tool Applied to Class A GPCRs

Marijn P. A. Sanders,<sup>†</sup> Stefan Verhoeven,<sup>‡</sup> Chris de Graaf,<sup>§</sup> Luc Roumen,<sup>§</sup> Bas Vroeling,<sup>||</sup> Sander B. Nabuurs,<sup>†</sup> Jacob de Vlieg,<sup>†,‡</sup> and Jan P. G. Klomp<sup>\*,‡</sup>

<sup>†</sup>Computational Drug Discovery Group, CMBI, Radboud University Nijmegen, Nijmegen, The Netherlands

<sup>‡</sup>Department of Molecular Design and Informatics, MRL, MSD, Oss, The Netherlands

<sup>§</sup>Division of Medicinal Chemistry, LACDR, VU University Amsterdam, Amsterdam, The Netherlands

<sup>||</sup>Modeling and Data Mining Group, CMBI, Radboud University Nijmegen, Nijmegen, The Netherlands

**S** Supporting Information

## ABSTRACT:



G-protein coupled receptors (GPCRs) are important drug targets for various diseases and of major interest to pharmaceutical companies. The function of individual members of this protein family can be modulated by the binding of small molecules at the extracellular side of the structurally conserved transmembrane (TM) domain. Here, we present Snooker, a structure-based approach to generate pharmacophore hypotheses for compounds binding to this extracellular side of the TM domain. Snooker does not require knowledge of ligands, is therefore suitable for apo-proteins, and can be applied to all receptors of the GPCR protein family. The method comprises the construction of a homology model of the TM domains and prioritization of residues on the probability of being ligand binding. Subsequently, protein properties are converted to ligand space, and pharmacophore features are generated at positions where protein ligand interactions are likely. Using this semiautomated knowledge-driven bioinformatics approach we have created pharmacophore hypotheses for 15 different GPCRs from several different subfamilies. For the beta-2-adrenergic receptor we show that ligand poses predicted by Snooker pharmacophore hypotheses reproduce literature supported binding modes for ~75% of compounds fulfilling pharmacophore constraints. All 15 pharmacophore hypotheses represent interactions with essential residues for ligand binding as observed in mutagenesis experiments and compound selections based on these hypotheses are shown to be target specific. For 8 out of 15 targets enrichment factors above 10-fold are observed in the top 0.5% ranked compounds in a virtual screen. Additionally, prospectively predicted ligand binding poses in the human dopamine D3 receptor based on Snooker pharmacophores were ranked among the best models in the community wide GPCR dock 2010.

## INTRODUCTION

G-protein coupled receptors (GPCRs) represent a large superfamily of membrane proteins responsible for the signal transduction from the extracellular to intracellular side of the cell membrane in many different physiological pathways. Therefore, they are effective drug targets for various diseases and of major interest to pharmaceutical companies.<sup>1,2</sup>

The GPCR family is characterized by seven conserved alpha-helices, which span the cell membrane. To date the crystal structure of only six different GPCRs have been elucidated

because proteins from this family are, like many other membrane proteins, difficult to crystallize. The lack of high resolution structural data complicates the process of rational drug design especially for large and structurally diverse families such as the GPCRs. In modern drug discovery, computer-aided techniques are often used to speed up the design process. These techniques are typically divided into ligand-based and structure-based

Received: February 21, 2011

Published: August 25, 2011

approaches. Ligand-based drug design techniques rely on the availability of known active compounds and have proven to be very successful in the design of new compounds.<sup>3–8</sup>

Commonly used ligand-based techniques that are frequently combined with structure-based approaches include the use of a spatial arrangement of key chemical features, a so-called pharmacophore, to discriminate active from inactive compounds. A disadvantage of ligand derived pharmacophore hypotheses is the assumption that active compounds bind in a similar binding mode, and, consequently, the designed compounds usually have less novelty.<sup>9</sup>

Structure-based drug design, on the other hand, does require a three-dimensional structure of the protein, acquired by means of several techniques, including electron microscopy, atomic force microscopy, X-ray crystallography, and NMR spectroscopy, or by computational methods e.g. homology modeling. Depending on the accuracy of a three-dimensional model, structure-based searching strategies, such as docking, have proven to be very successful in the design of new active compounds<sup>10–13</sup> also for GPCRs.<sup>14–16</sup> However, working with less accurate structures, as typically obtained by homology modeling, remains a major challenge. New developments, such as induced fit docking, have increased the accuracy of results,<sup>17</sup> but are computationally expensive and remain dependent to some extent on prior knowledge.

A protein-based approach that depends less on protein structure resolution is the association of sequence motifs with ligand (interaction) features. In the thematic analysis method,<sup>18</sup> structure activity relationships (SAR) of class A and B GPCRs were used to generate a pairing of sequence patterns/themes and ligand structural motifs. Next, focused libraries can be designed by inclusion of compounds with structural motifs which occur in ligands for receptors which share similar sequence patterns. This method contains limited three-dimensional structural information and is limited to the structural motifs observed in known ligands.

Another low resolution approach in structure-based drug design is the use of pharmacophore models derived from protein binding sites. These structure based pharmacophore models can be derived from homology models and have been successfully applied for characterizing ligand binding pockets and virtual screening for ligands for various protein targets, including GPCRs.<sup>14,19–22</sup> Kratochwil et al.<sup>23</sup> developed a method to characterize a GPCR pocket with 35 pharmacophore features representing the 35 residues aligning the ligand binding pocket.

Klabunde et al. presented an approach to construct structure-based pharmacophore hypotheses for class A GPCRs based on chemoprints of the pharmacophore hypotheses derived from 10 different homology models and 3 X-ray structures of GPCRs.<sup>24</sup>

Integrating ligand- and structure-based approaches allows for optimal use of all available data and has resulted in a number of successful virtual screens in which new compounds have been identified.<sup>25,26</sup>

Recently, the structures of several new GPCRs together with an increasing amount of annotated ligand data have been available in databases such as PDB, CheMBL,<sup>27</sup> DrugBank,<sup>28</sup> BindingDB,<sup>29–31</sup> PDDBBind,<sup>32,33</sup> MOAD,<sup>34,35</sup> WOMBAT,<sup>36</sup> and Glida-DB.<sup>37</sup> All these data repositories allow data mining to construct for example structure activity relationships (SAR), including Glida-DB and GPCR SARfari (based on CheMBL), which are dedicated to GPCRs. However, they generally do not provide structural understanding of receptor–ligand interactions

and offer only limited tools for the discovery of new chemical entities. There is an urging need for tools to connect structure and ligand data especially in the field of GPCRs where only limited crystal structures are available.

We present here Snooker, a low resolution approach whereby structure-based pharmacophore hypotheses are constructed with no prior knowledge of ligand structure or interactions. The pharmacophore hypotheses are obtained from homology models constructed on-the-fly, based on an in-house sequence alignment of the seven transmembrane domains<sup>38</sup> and a crystal structure template. Residues important for ligand binding are identified by analysis of Shannon entropies of structurally conserved positions in a multiple sequence alignment and chemical features representing protein–ligand interactions are positioned in the binding pocket. This abstraction of protein–ligand interaction properties into pharmacophores allows for the discovery of new active compounds and connects structural knowledge to ligand data.

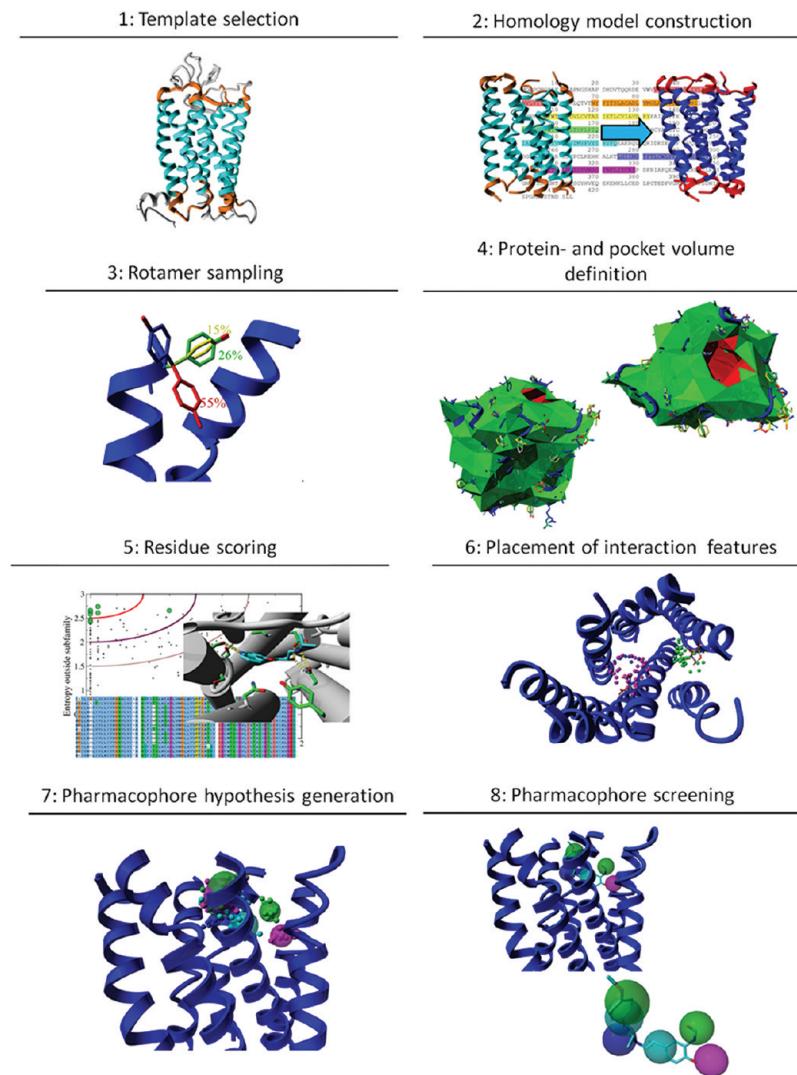
The validity of our pharmacophores is tested in three different experiments. First of all, poses of known active compounds are reproduced in a retrospective ligand binding mode prediction experiment in the beta-2 adrenergic receptor (ADRB2). Second, the eticlopride binding mode in the dopamine D3 receptor (DRD3) is successfully predicted in the community wide GPCR DOCK 2010 assessment prior to the release of the DRD3 cocrystal structure. Finally, structure-based pharmacophore models of 15 different GPCRs (matching key ligand binding residues) are successfully used to identify target specific ligand sets and discriminate active from inactive compounds in a virtual cross-screen.

## ■ OUTLINE OF THE APPROACH

The Snooker approach consists of several stages outlined and discussed in sequential order below. Figure 1 shows the construction of a pharmacophore model and an example of a subsequent pharmacophore search. For clarity and reference, Figure 1 shows the results obtained at each different stage of pharmacophore construction for the human beta2 adrenoceptor based on the rhodopsin crystal structure cocrystallized with retinal (Protein Data Bank (PDB) entry 1GZM<sup>39</sup>) and the subsequent pharmacophore search of R-R-formoterol using the obtained pharmacophore.

**Template Selection.** The initial stage of the Snooker protocol starts with the selection of one or multiple template structures. When this research was conducted only the structures of the rhodopsin, adenosine A2A, and the beta 1 and 2 adrenergic receptor were available. However future crystal structures and custom-made homology models or combinations thereof can be used as template structures as well. Because loop modeling for GPCRs remains very challenging, the templates are stripped down until only the 7 transmembrane helices remain (see Appendix 1, Supporting Information for the remaining residue numbers of each available template structure and their respective Ballesteros and Weinstein numbers<sup>40</sup>). Inclusion of loops or even single residues is still possible but requires manual adjustments of the provided templates and is only recommended for highly similar loop regions or where prior knowledge on the target is available. The orange and cyan colored residues of the template shown in panel 1 of Figure 1 are used as the starting point for the construction of the homology model.

**Homology Model Construction.** The 7 transmembrane (TM) domain sequences of the desired GPCR are extracted



**Figure 1.** Visual outline of the Snooker approach, illustrated using the pharmacophore hypothesis generation for the human beta 2 adrenergic receptor based on a bovine rhodopsin crystal structure (pdb code: 1GZM) and the subsequent positioning of R-R-formoterol in the pharmacophore. Starting with the crystal structure of bovine rhodopsin, (1) the structurally conserved alpha-helices (cyan) and five residues at each side of each helix (orange) are extracted. A homology model (2) is constructed based on the alignment of the model receptor sequence with this template. An  $\alpha$  helix specific rotamer library is used to add a rotamer ensemble (3). The pocket is detected (4) by a Delaunay tessellation of the  $C\alpha$ -atom and average side chain atom positions. Residues are scored upon ligand binding probability (5) by multiple sequence alignment analysis, and ‘interaction’ points (6) are placed inside the pocket volume using literature described interaction geometries with densities corresponding to the residue score and rotamer probability. Next, pharmacophore features are generated (7) with a fuzzy pharmacophore algorithm applied on the interaction points. Finally, (8) ligands fulfilling all pharmacophore constraints are aligned to the pharmacophore. In this example, R-R-formoterol matches a pharmacophore comprised of 6 features.

from an in-house gap-free multiple sequence alignment of those 7TM domains<sup>38</sup> and is used together with the template sequence and structure to construct the homology model. The backbone and conserved residues are kept rigid during this stage. Placement of nonconserved residues is based upon a position-specific rotamer library.<sup>41</sup> Finally the hydrogen bond network is optimized and bumps are removed.

**Rotamer Sampling.** For all residues we include the most likely rotamers from an  $\alpha$  helix specific rotamer library<sup>42</sup> to account for possible model inaccuracy of the initial homology model. During the generation of these new rotameric states, clashes are allowed, with the intent to maximize sampling. The procedure stores the probability of each rotamer in the ensemble for later use in the protein- and pocket volume definition as well as in the placement of the interaction features. The use of an

ensemble of rotamers avoids the computational magnitude that would result from considering all possible models which can be obtained by combining all possible rotamer states for each single residue. The rotamer ensemble is subsequently used to generate both the protein- and pocket volume definition.

**Protein- and Pocket Volume Definition.** To restrict the placement of pharmacophoric features to the ligand binding pocket, both pocket and protein volumes are constructed. Traditional pocket detection techniques like grid-based sampling methods require high resolution structures, cannot deal with ensemble structures and are extremely sensitive to small deviations, and are therefore less suitable for our approach. Thus, we have chosen to describe each rotamer ensemble with the position of the  $C\alpha$ -atom and an estimated point representing the side chain. This estimated point is effectively an averaging over all

rotamers which will describe the side chain almost exactly if just one rotamer exists and will approximate the C $\beta$ -atom for a diverse rotamer ensemble.

**Residue Scoring.** It has been suggested that residues which are conserved within a subfamily yet not across the complete GPCR family are important for ligand binding.<sup>43–45</sup> To identify ligand binding residues we calculated the sequence conservation (expressed as entropy values) inside- and outside each subfamily for each residue position.<sup>38</sup> A residue score is defined by a combination of both entropy values and reflects the importance of the residue position for ligand binding.

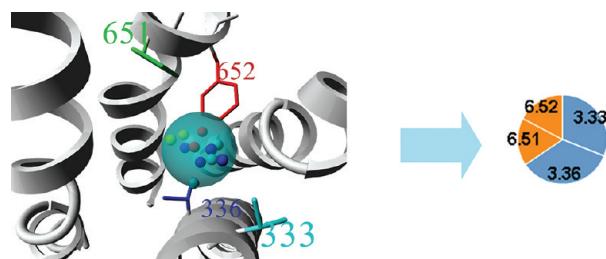
**Placement of Interaction Features.** Interaction feature points complementary to the residue properties (acceptor, donor, positive ionizable, negative ionizable, and hydrophobic) are positioned in the pocket volume for each rotamer with densities related to the residue score and rotamer likelihood. The density of interaction points is for each rotamer evenly spread over the surface as described by the corresponding interaction geometry.<sup>46–50</sup> All interaction points inside the protein volume or outside the pocket volume are removed resulting in features which are accessible and which can be used to mimic compounds which bind within the 7 TM domain.

**Pharmacophore Hypothesis Generation.** Pharmacophore features are generated at regions of high interaction feature density for a given property. It is suggested that residues which are close to each other in space can enhance the importance of those residues in ligand binding.<sup>51–54</sup> By using the interaction feature point density instead of the residue numbers, residues close in space but not necessarily in sequence can enhance each other. A measure for the importance of a feature is subsequently introduced by scoring the ratio between the number of points and the volume of a feature. These are indirectly influenced by rotamer conservation as well as residue score. A consensus hypothesis is introduced by identification of regions of high interaction feature point densities in the overlay of the interaction feature point densities of individual models. A subset of features from each pharmacophore hypothesis is selected to reduce the total number of features and improve virtual screening performance. This subset comprises the top 2 scoring features of each interaction type (acceptor, donor, positive ionizable, negative ionizable, and hydrophobic).

**Pharmacophore Comparison.** Pharmacophore hypotheses differ due to the use of different templates on which the models are based and from which the subsequent hypotheses are derived. A measure of robustness is therefore not introduced by comparison of pharmacophore feature locations in different pharmacophore hypotheses but by analysis of the residues which contribute to all single pharmacophore features. The contribution of a residue to a feature is calculated and depicted in a pie-chart (Figure 2). The pie-charts of all different pharmacophore features of the different templates can be compared to quantify the robustness and template dependence of the Snooker method to generate pharmacophore hypotheses.

**Pharmacophore Screening.** 3D conformations of a compound satisfying the pharmacophore constraints are matched to the pharmacophore hypothesis with a minimal root-mean-square deviation (rmsd) according to the fitting procedure as explained in the Materials and Methods section.

**Method Validation.** Three different experiments were performed to assess the quality of the Snooker pharmacophore hypotheses. First, human beta 2 adrenergic receptor ligands are matched to the Snooker pharmacophore hypothesis and compared to literature supported binding mode hypotheses. Second,



**Figure 2.** Combination of hydrophobic ‘interaction points’ of four different residues leads to a hydrophobic pharmacophore feature. The contribution of the different residues is visualized in a pie-chart, 4 out of the 12 interaction points can be assigned to residues 3.33 (cyan) and 3.36 (blue) and 2 out of 12 to residues 6.51 (green) and 6.52 (red). Coloring of the pie-charts is according to transmembrane membrane domain from which the residues originate.

an evaluation of whether receptor pharmacophore hypotheses are able to identify the correct ligands with a higher accuracy than ‘random’ Snooker pharmacophore hypotheses was undertaken. Third, the enrichment of active ligands is tested for multiple class A GPCRs.

**Retrospective Binding Mode Prediction.** To assess whether our approach positions the appropriate features at the correct positions, we have analyzed which residues contribute to pharmacophore features and evaluated how well the pharmacophore hypotheses perform in a pose prediction experiment. For this purpose we have built pharmacophore hypotheses of the human beta 2 adrenergic receptor and performed a search on a set of known full, partial, and inverse agonists as well as antagonists. The rmsd between the poses as reported in the literature<sup>55</sup> and the poses as generated after a short energy minimization of the initial pharmacophore matching pose is calculated and reported.

**Prospective Binding Mode Prediction.** Crystal structures of the human DRD3 and human CXCR4 receptor structure cocrystallized with a small molecule have been solved recently.<sup>56,57</sup> To assess the current status of GPCR modeling, research groups were invited to submit models of those receptors with the bound ligands.<sup>58</sup> We submitted binding mode hypotheses based on Snooker pharmacophores derived from custom-made homology models and optimized with Fleksy.<sup>17</sup> These models were scored on the accuracy of the receptor structure model as well as the predicted binding mode.

**Target-Specific Ligand Identification.** To demonstrate the target specificity of each pharmacophore hypothesis, virtual cross-screens of 15 different GPCR pharmacophore hypotheses and compound sets are performed. The target receptors and families used are listed in Table 1.

The enrichment is calculated for each combination of a compound set and pharmacophore hypothesis. Target specificity can be assumed if the hypothesis and corresponding compound set have better enrichment compared to the average enrichment of all screens using this same compound set. To show that the final results are not biased by the properties of the active compounds we performed the same experiment for a set of inactive compounds with similar physicochemical properties as the actives.

**Library Enrichment.** Shape constraints are known to improve enrichment in virtual screening.<sup>24,60</sup> To evaluate the possible performance of the Snooker pharmacophore hypotheses in library enrichment we added a shape constraint to filter the

**Table 1.** Targets Used in the Virtual Cross-Screen with Their Corresponding Receptor Families According to the GPCRDB Family Classification<sup>59</sup>

target	family
SHT7R	serotonin family
AA2AR	adenosine family
ADA2B	alpha adrenoceptors family
ADRB2	beta adrenoceptors family
AGTR1	angiotensin family
CLTR1	cysteinyl leukotriene family
DRD2	dopamine family
EDNRA	endothelin family
GASR	cholecystokinin CCK family
GHSR	thyrotropin-releasing hormone and secretagogue family
HRH3	histamine family
MCHR1	melanin-concentrating hormone receptors family
NPY5R	neuropeptide Y family
OPRM	opioid family
TA2R	prostanoid family

poses as generated in the target specificity experiment and calculated the enrichment of the set of remaining compounds.

## MATERIALS AND METHODS

**Template Selection.** Three-dimensional coordinates of the template receptors (pdb ids: 1GZM,<sup>39</sup> 1L9H,<sup>61</sup> 2RH1,<sup>62</sup> 2VT4,<sup>63</sup> 3CAP,<sup>64</sup> 3D4S,<sup>65</sup> 3DQB,<sup>66</sup> 3EML<sup>67</sup>) were obtained from the PDB and structurally aligned using the Needleman and Wunsch algorithm<sup>68</sup> as embedded in YASARA<sup>69</sup> (<http://www.yasara.org>). Residues which are 5 amino acids apart from the transmembrane helices were removed as were the ligands, waters, and other heteroatoms. Ballesteros and Weinstein numbers<sup>40</sup> of transmembrane helices are as follows: TM1: 1.33–1.56; TM2: 2.40–2.65; TM3: 3.25–3.51; TM4: 4.43–4.64; TM5: 5.38–5.63; TM6: 6.37–6.59; TM7: 7.34–7.56.

**Homology Model.** Homology models were built and optimized using WHAT IF<sup>70</sup> and upon completion residues not part of a transmembrane helix were removed.

**Rotamer Sampling.** The conformational space of all residue side chains is sampled sequentially using the dihedral angles as reported by Lovell et al.<sup>42</sup> To reduce the number of rotamers we only add a rotamer if it occurs more frequently than 5% in alpha-helices. In order to make our procedure more GPCR-specific we have introduced an additional 50% probability for the rotamer that possesses an average side chain vector most similar to the average side chain vector of the available GPCR crystal structures, and we have given a weight of 25% to this particular rotamer in the initial homology model.

**Protein- and Pocket Volume Definition.** For all residues in our model the  $C\alpha$ -atom position is calculated as well as the average vector between the  $C\alpha$  and the mean coordinate of the side chain, while taking the rotamer distribution into account. This average vector was multiplied by 1.5 and added to the mean  $C\alpha$  position. These calculated coordinates for all residues and 7 dummy points representing the end of the N-terminus and the starts and ends of the extracellular loops are subsequently used in a Delaunay tessellation.<sup>71</sup> The dummy points are positioned at the end of the vectors starting at the average  $C\alpha$ -atom position of

the 4 residues of each helix closest to the extracellular environment, and directed 6 Å toward the pocket center (pocket center: mean coordinate of the  $C\alpha$ -atoms of residues 3.32, 5.32, and 7.39) and 2 Å toward the extracellular side. After removal of all tetrahedra with any edge longer than 8.5 Å, the cavities are exposed. All surfaces of which a point 5.0 Å along the normal vector is within the original tessellation are marked as potential cavity surfaces. Next, cavity surfaces, which share an edge and have an angle  $<0.5\pi$  radians between the normal vectors, are merged. Subsequently, the pocket surface is defined as the largest surface of at least 500 Å<sup>2</sup> and within 15 Å of the pocket center. Finally, all tetrahedra which have at least one pocket surface and contain no heavy atoms are removed, a procedure which is repeated until all pocket surfaces are part of tetrahedra which contain at least one heavy atom.

**Residue Scoring.** Receptor families are defined based on sequence homology of the 7 transmembrane helices and entropy scores (reflecting sequence conservation) inside and outside the family are calculated for each position in the multiple sequence alignment<sup>38,72</sup> according to

$$E_i = - \sum_{a=1}^{20} F_{ia} \ln(F_{ia}) \quad (1)$$

$$F_{ia} = \text{Number}_{ia}/m \quad (2)$$

where  $\text{Number}_{ia}$  is the number of sequences with residue type  $a$  at alignment position  $i$ , and  $m$  is the number of total sequences.  $F_{ia}$  is the frequency of residue type  $a$  at position  $i$ , and  $E_i$  is the entropy of alignment position  $i$ . Subsequently one score is calculated for each residue by combining both scores according to eq 3, with  $E^{in}$  being the entropy inside a cluster, and  $E^{out}$  being the entropy of all sequences not part of the cluster

$$S_i = \sqrt{(E_i^{in})^2 + (\ln(20) - E_i^{out})^2} \quad (3)$$

**Placing Interaction Features.** Hydrophobic and polar interaction geometries are extracted from Rarey et al.<sup>50</sup> Charged interactions are described by a cone, the narrow end of which is positioned at the charged center as described by Kumar and Nussinov.<sup>73</sup> The interaction surface is described by the part of the sphere where a vector of 3.5 Å, starting at the narrow end of the cone, is within 38° of the vector between the  $C\beta$ -atom and the corresponding charged center (see Appendix 2, Supporting Information). The final density of interaction points is obtained after sampling a grid on the interaction surfaces. For each rotamer of each residue, sampling starts with the point which is closest to all other interaction points of the same type. Additional points are added iteratively according to

$$P_{new} = \arg_{x \in P_{left}} \max(D(x, P_{sel}) - D(x, P_{left})) \quad (4)$$

with  $P_{new}$  is the updated set of interaction points,  $x$  is a potential point to be added to the selection,  $P_{sel}$  are the points already selected,  $P_{left}$  are all not yet selected points, and  $D(x, P)$  is the average distance between  $x$  and all points in  $P$ . Addition of points stops if the number of interaction points has reached

$$S = N_{basic} \times 1.0/f_i \times R_{ij} \quad (5)$$

where  $f_i$  is the entropy based residue score of residue  $i$ ,  $R_{ij}$  is the rotamer likelihood of rotamer  $j$  of residue  $i$ , and  $N_{basic}$  is the initial density for each residue, which can be found in Appendix 3,

**Supporting Information.** After the completion of the sampling phase, all points which are not in the pocket volume or in the protein volume are removed.

**Pharmacophore Hypothesis Generation.** Pharmacophore features are generated by applying a fuzzy pharmacophore algorithm to all interaction points using an  $R_c$  value of  $2.5 \text{ \AA}$ .<sup>74</sup> Features are centered at the mean coordinate of all points contributing to the feature and the radius of each feature is set to  $2.0 \text{ \AA}$ . The weighted average vector of the center of all donor and acceptor features toward the residues which contribute to the respective feature is used as a direction vector for the polar feature. The tolerance of this directionality is defined as the minimum of the standard deviation of the vector and  $1/4\pi$  radians. The consensus hypothesis is constructed from the overlay of the interaction feature point densities based on the models using the templates 1GZM, 1L9H, 2RH1, 2VT4, 3CAP, 3D4S, 3DQB, and 3EML. Pharmacophore features are ranked according to the number of interaction points per volume. The pharmacophore which is finally used in screening comprises the top 2 ranked features of each interaction type.

**Customization of Automatically Generated Pharmacophore Hypotheses Based on Mutagenesis Data.** EDNRA. Mutagenesis data of the EDNRA receptor suggest a crucial role for Gln3.32.<sup>75</sup> This residue is not correlated to a pharmacophore feature in the consensus hypothesis due to limited accessibility of this region in models based on some of the templates. However, this region is accessible in the models which have used 3EML and 3CAP as template. Gln3.32 is related to pharmacophore features for the hypotheses based on these templates. Therefore pharmacophore hypotheses based on these two templates are screened for the EDNRA receptor.

NPY5R. For the NPY5R the residues Cys3.33 and Cys4.57 were omitted from the Delaunay tessellation. This adjustment is a consequence of the hypothesis that both form a disulfide bridge due to a slight rearrangement of TM3 and TM4 caused by the helix-disturbing residues Pro3.29, Gly4.23, and Pro4.61. Position 5.46 is known to be important for agonism in the aminergic receptors<sup>76</sup> and is also extremely conserved in the NPY5R receptor cluster. Hence, the polar pharmacophore features corresponding to this residue is promoted from rank 3 to rank 2.

TA2R. Residue position 7.36 and 7.40 have a remarkably high residue score (data not shown) and are therefore likely to be involved in ligand binding for the tromboxane A2 receptor (TA2R). A modification to the alignment is made for this receptor such that residue 7.36 matches residue 7.35 of the template structure and 7.40 matches 7.39. This effectively results in a kinked helix, as also observed in TM2 of the CXCR4 receptor structure, and positions 7.36 and 7.40 in the pocket. All ligands in the training set contain a carboxyl which seems crucial for the activity. The largest distance to a polar feature is for most ligands within  $8\text{--}12 \text{ \AA}$  of this carboxyl. We choose to use the pharmacophore based on a beta 2 adrenoceptor template (pdBID: 2RH1) instead of the consensus pharmacophore because the hypothesis derived from the homology model based on the 2RH1 template has a distance of  $\sim 12 \text{ \AA}$  between the negative ionizable feature corresponding to Arg7.40 and the acceptor feature near TM5, while this distance is larger in the consensus hypothesis and in the hypotheses based on all other templates.

OPRM. Binding mode hypotheses for opioid receptors are described in literature and depict a pharmacophore with a phenolic site deep down a subpocket near TM5, a hydrophobic region near TM5 toward the extracellular loop and an anionic site

related to residue 3.32.<sup>77–79</sup> The pharmacophore hypothesis based on the template 2VT4 seems to mimic this binding mode best with a positive ionizable and polar feature related to 3.32 and two hydrophobic features near helix V. Therefore we included the pharmacophore hypothesis based on the 2VT4 template as well.

**Compound Sets. Pose Prediction.** A maximum of 100 3D conformations per molecule was generated for all compounds depicted in Figure 2 of de Graaf et al.<sup>55</sup> using Corina<sup>80</sup> and Cyndi.<sup>81</sup> Cyndi was run with the parameters of Appendix 4, Supporting Information.

**Target Specificity.** Compounds with activities better than  $50 \text{ nM}$  were retrieved from ChEMBL02 for the 15 targets and divided into training and test sets. The test set was selected using exclusion sphere clustering on BCI fingerprints and contained the 50 most chemically diverse compounds for each target. The ‘fake’ active sets of compounds with similar properties to the actives but different architectures were selected based on the same number of positive ionizable and negative ionizable features, acceptors, donors, hydrophobic atoms, aromatic atoms, heavy atoms, rotatable bonds, and number of rings. For each known active compound in the test set the most similar (and presumed inactive) compound was chosen from all ChEMBL02 compounds that have not been tested in GPCR assays. ‘Fake’ active compounds were selected with emphasis on the same ionizable features and polar groups and with MACCS fingerprint similarities greater than 0.6 to all known actives. The 10,000 assumed inactive compounds used as decoys were selected from the 10,000 most diverse compounds from all ChEMBL02 compounds not tested in GPCR assays. This selection was carried out using the diverse molecule component with FCFP\_4 fingerprint as implemented in Pipeline Pilot.<sup>82</sup> 3D conformations were generated for all compounds using the procedure described to generate conformations for the pose prediction compound set. All ChEMBL02 compound ids of the different sets are listed in Appendix 5, Supporting Information.

**Method Validation. Retrospective Binding Mode Prediction.** The performance of Snooker in reproducing protein–ligand binding hypotheses as reported in the literature<sup>55</sup> was investigated for ADRB2. Compound poses matching 5 or more pharmacophore features were minimized using side chain optimizations and energy minimizations using the Yamber3 forcefield as embedded in YASARA.<sup>69</sup> First, a procedure of optimization of the side chains of all residues with a distance less than  $10 \text{ \AA}$  from the ligand pose using SCWALL<sup>83</sup> as implemented in YASARA followed by an energy minimization with fixed carbon atoms of the helix endings is performed twice. After this all side chains with a distance less than  $8 \text{ \AA}$  from the ligand are again optimized using SCWALL. Finally, the complex is energy minimized without constraints using the Yamber3 forcefield. The resulting poses are then compared to corresponding reference poses reported by de Graaf et al.,<sup>55</sup> by calculating the rms between all atoms within  $13 \text{ \AA}$  of the  $\text{Ca}$ -atom of Ser5.46 in the reference structures and their counterparts in the pharmacophore guided poses. Although the reference structures are obtained via a docking experiment in customized receptor models, the validity of these models is supported by their high similarity to the recently elucidated structures of the homologues ADRB1 receptor with several ligands.<sup>84</sup>

**Prospective Binding Mode Prediction.** PDB entry 2RH1 was used as a template to build the homology models. The structure was cleaned, the lysozyme protein was removed, and the bound

ligands were retained. Residues in the loop between TM5 and TM6 (residues 218–317) were discarded in the modeling process for the DRD3 receptor. For CXCR4, residues in the loop between TM6 and TM7 (residues 267–276) were discarded, and the loop between TM4-TM5 was deleted (residue Y174-A198) and replaced with the loop TM4-TM5 from PDB entry 3EML (AA2AR\_HUMAN, residues N144–N175). Homology modeling was performed using the Yasara program and its built-in modeling algorithm.<sup>85</sup> Subsequently, the initial models were manually refined and subjected to a final energy minimization step. Snooker pharmacophores were generated based on the custom built homology models by the procedure described in this manuscript. Finally, compound matches in all 5 feature pharmacophores for eticlopride in the DRD3 receptor and in all 4 feature pharmacophore for the CXCR4 receptor were optimized and scored using Fleksy.<sup>17</sup>

**Target Specificity.** The performance of Snooker with regard to target specificity was tested in a cross-screening exercise. A virtual cross-screen of 15 pharmacophore hypotheses of different GPCRs and 10000 assumed inactive compounds plus 750 active compounds (50 for each GPCR) was therefore performed. After the virtual screen, compounds were first ranked according to the number of fitted features, and subsequently by their fit value. For each set of 50 actives and 10000 assumed inactives, the area under the semilogarithmic receiver operating curve (pROC AUC) was calculated.<sup>86</sup> Finally, the 15 different pharmacophore hypotheses were ranked for each compound set using the pROC AUC. Since the pROC AUC emphasizes early enrichment we have chosen this value instead of the normal ROC AUC. To show that results are not biased toward the properties of the active compounds, 15 fake sets of active compounds were created and screened together with the 10000 assumed inactives. To construct the fake sets, one compound which has not been tested in a GPCR assay and possesses similar properties to an active was selected per active molecule. The pROC AUC and ranking is subsequently calculated for each pharmacophore hypothesis and compound as described before.

**Library Enrichment.** From all training set compounds poses we removed those which have at least one atom within 2.0 Å of the backbone of the homology model. The pose which had the best volume overlap (calculated as the smallest average tanimoto shape distance) with all filtered training set poses was defined as the reference. Next, the volume overlap between the reference and all poses of the 10050 compounds (10000 compounds not tested in a GPCR assay and 50 actives for the GPCR of interest) which match the pharmacophore were calculated. A value for the volume overlap (as tanimoto shape distance) was set at the maximal enrichment of training set actives versus decoys and including at least 25% of training set actives. Subsequently, a shape cutoff was defined by adding 0.03 to this value. All poses of the test and decoy set with a Tanimoto shape distance larger than the cutoff were removed, and the remaining compounds were ranked according to the number of matching pharmacophore features, the number of actives in the training set which hit this same pharmacophore and finally the fit value. Lastly, the AUC, pROC AUC,<sup>86</sup> enrichment at 0.5%, 1%, 2%, and 5% of the compound set were calculated.<sup>87</sup>

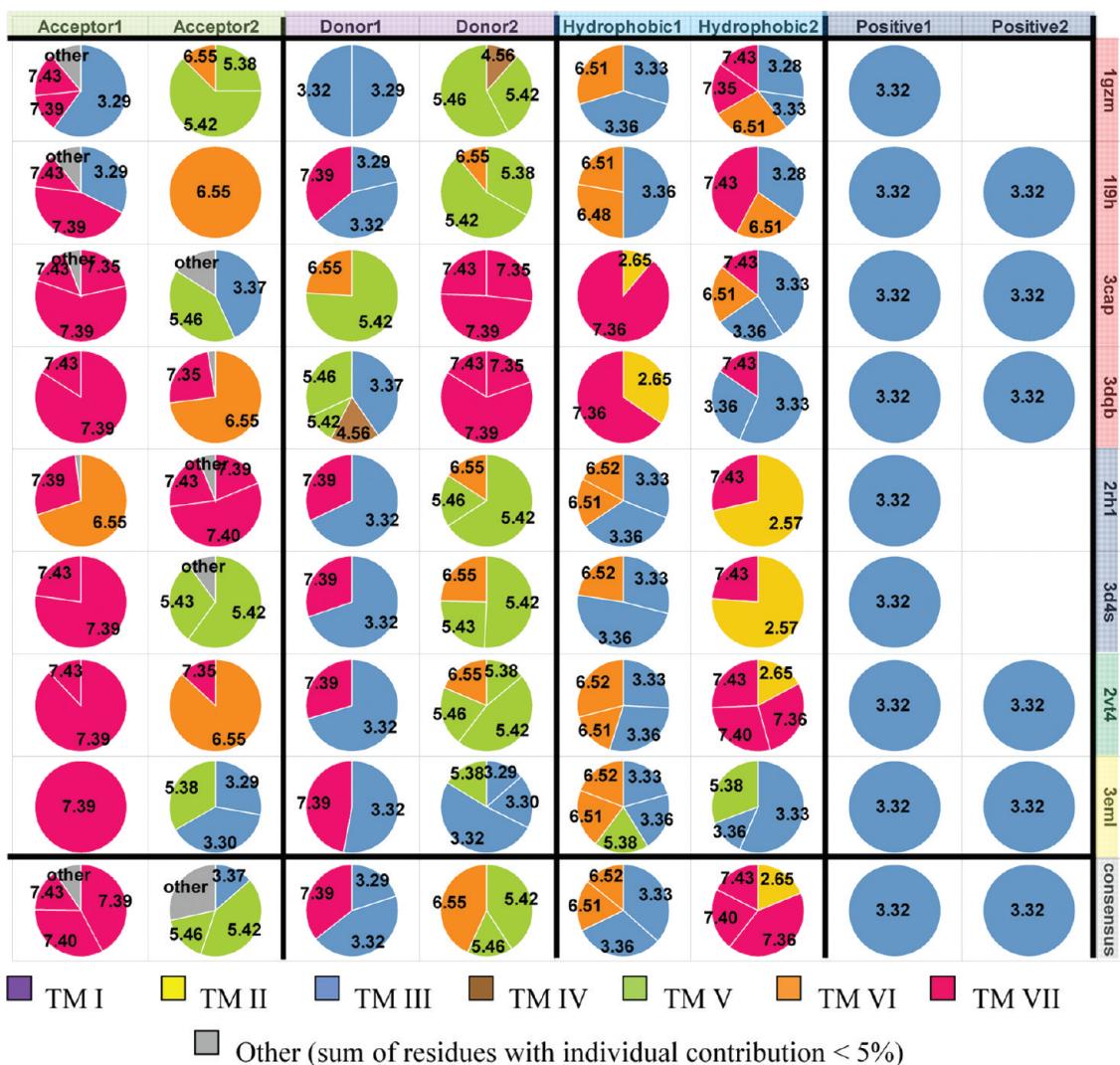
**Pharmacophore Screening.** Pharmacophore searches were performed using custom code implemented in Python and making use of RDKit.<sup>88</sup> First, ligand atoms were typed using the definitions of Appendix 6, Supporting Information. Second, all matches between pharmacophore features and ligand features

were listed after which all possible matches between a ligand and a pharmacophore were calculated. Third, all ligand atoms which match features were transformed to minimize the rms between the ligand atoms and the center of the corresponding pharmacophore feature. Fourth, the distance from each atom to the center of the feature it should match was calculated, and if this was larger than the pharmacophore radius, the weight of this pair was increased in the next transformation. This procedure is repeated three times, and a hit was defined when the average rms was smaller than the average pharmacophore radius, and when all ligand atoms were within 105% of the radius of the pharmacophore feature.

## ■ RESULTS AND DISCUSSION

**Retrospective Binding Mode Prediction.** The contributions of the different residues to the pharmacophore hypotheses deduced from the ADRB2 models derived from 8 different template structures as well as a consensus are depicted in Figure 3. The consensus pharmacophore hypothesis is derived from the overlay of the models based on the 8 different templates and contains the most robust features related to the residues identified as being important for ligand binding by multiple sequence alignment (MSA) analysis. Represented are for example known important interactions like the positive ionizable interaction of Asp3.32, the hydrophobic contact with Val3.33, and the polar interactions with Asn7.39 and Asp3.32 and with Ser5.42 and Ser5.46.<sup>76</sup> The overlay of the 8 models represents all possible interactions hypothesized by the individual models. A weighted average of all these possible interactions defines which single interactions are considered most important based on MSA analysis and robust as indicated by their presence in the majority of the models. The consensus hypothesis is newly created by revisiting the interaction feature points and recalculation of the densities, thus pharmacophore features. For example, Ser5.46 has been shown to be important for the binding of agonists,<sup>76</sup> but has only limited accessibility in the models based on the aminergic templates, probably due to the antagonist conformation in which the aminergic receptors have been crystallized. This same area of the receptor is much more accessible in the models based on the opsin and rhodopsin templates and promotes the use of models based on these templates to construct the consensus pharmacophore hypothesis. Ser5.46 is very well represented in the consensus pharmacophore hypothesis due to the inclusion of the rhodopsin templates (see Figure 3).

The pharmacophore guided poses of 16 compounds matching a consensus (sub)pharmacophore hypothesis of 5 or more features are shown in Figure 4 together with the reference as defined by de Graaf et al.<sup>55</sup> The depicted poses are the most similar poses to the reference among all possible poses which satisfy a pharmacophore of at least 5 features. Since evolutionary pressure is mainly driven by endogenous agonists<sup>89</sup> and the fact that the Snooker procedure assigns weights to residues based on calculated entropies (reflecting evolutionary pressure) from a multiple sequence alignment comprising multiple species, it is to be expected that the method is more biased toward agonism than toward antagonism. This and the fewer specific contacts for antagonists as suggested by mutagenesis experiments<sup>90</sup> might explain the relatively high number of antagonist experiments (9 out of 12 antagonists [red] compared to 1 out of 14 agonists [green]) for which no pose could be produced. Thirteen of the 16 ligands have at least one pose comparable to the reference



**Figure 3.** Contribution of the different residues to the different pharmacophore features (horizontal) for the pharmacophore hypotheses derived from the ADRB2 models based on the different templates as well as the consensus pharmacophore hypothesis (vertical). Coloring of the pie-charts is according to the transmembrane domain in which the residue is positioned. Templates corresponding to crystal structures of beta aminergic, bovine rhodopsin, and the adenosine A2A receptors are colored blue, red, and green, respectively.

(heavy atom root-mean-square deviation (rmsd)  $\leq 2.0 \text{ \AA}$ ). This indicates that Snooker pharmacophores represent the correct protein–ligand interactions for the majority of ADRB2 ligands which do fit a (sub)pharmacophore of at least 5 features.

**Prospective Binding Mode Prediction.** Binding mode hypotheses largely based on Snooker pharmacophores are generated for the human DRD3 and CXCR4 receptor and submitted for evaluation in the GPCR dock 2010 assessment.<sup>58</sup> Utilizing custom-made homology models, Snooker pharmacophores, and low resolution binding modes were generated and optimized by Fleksy. The dopamine DRD3 compound, eticlopride, fitted dominantly into one five feature pharmacophore and was consistently predicted in a binding mode similar to the pose depicted in Figure 5. Based on the five feature pharmacophore used to generate these poses, interactions with Asp3.32, His6.55, Tyr7.35, Thr7.39, Tyr7.43, Val3.33, Val5.39, and Phe6.51 can be assumed. The final poses as optimized by Fleksy indeed represent those interactions and all 5 submitted models ranked in

the top 10 in the DRD3 assessment. The best model correctly predicted 36/65 atomic contacts and 12/15 residue contacts.<sup>58</sup>

The default CXCR4 model did not show a pocket volume for the positioning of pharmacophore features due to bulky and inward directed residues in the upper part of the transmembrane domain. To generate a pocket volume the minimal edge length required for the removal of tetrahedra was reduced from 8.5 Å to 8.0 Å and 7.5 Å. This resulted in two different subpockets in the minor binding pocket between TM2, TM3, and TM7.<sup>57</sup> Pharmacophores related to both subpockets were generated and screened resulting in two distinct sets of possible poses. IT1t poses in the first subpocket 1 correlated to interactions with residues Glu7.39, Thr3.33, Phe2.57, Leu 2.61, Tyr3.32, Leu3.36, and Phe7.43 and poses in the second subpocket 2 represented interactions of the small molecule IT1t to residues Glu7.39, Tyr3.32, His3.29, Val3.28, and Leu2.61. Since the pharmacophore features from the first subpocket 1 are based on a larger number of residues originating from the transmembrane domain we optimized and submitted compound poses in this pocket for

colterol	terbutaline	isoproterenol	epinephrine
0.38	0.51	1.13	1.21
indacaterol	fenoterol	S1319	nylidrine
1.24	1.34	1.37	1.44
trimethoquinol	atenolol	sotalol	salbutamol
1.49	1.60	1.87	1.94
carvedilol	labetolol	formoterol	procaterol
2.00	2.48	2.53	3.20

**Figure 4.** Receptor structure of the reference structure (gray), reference pose (orange), most similar predicted pose (cyan), and structure-based consensus pharmacophore hypothesis of the human beta 2 adrenergic receptor. Pharmacophore features are colored cyan (hydrophobic), green (acceptor), magenta (donor), and blue (positive ionizable). Ligand names are colored red for antagonist/inverse agonist and green for agonist. The heavy atom root-mean-square deviation (rmsd) of each pose is indicated below each figure.

evaluation and ranked 14th, 17th, 28th, 29th, and 33th out of 103 predictions in the GPCR dock 2010 assessment.<sup>58</sup>

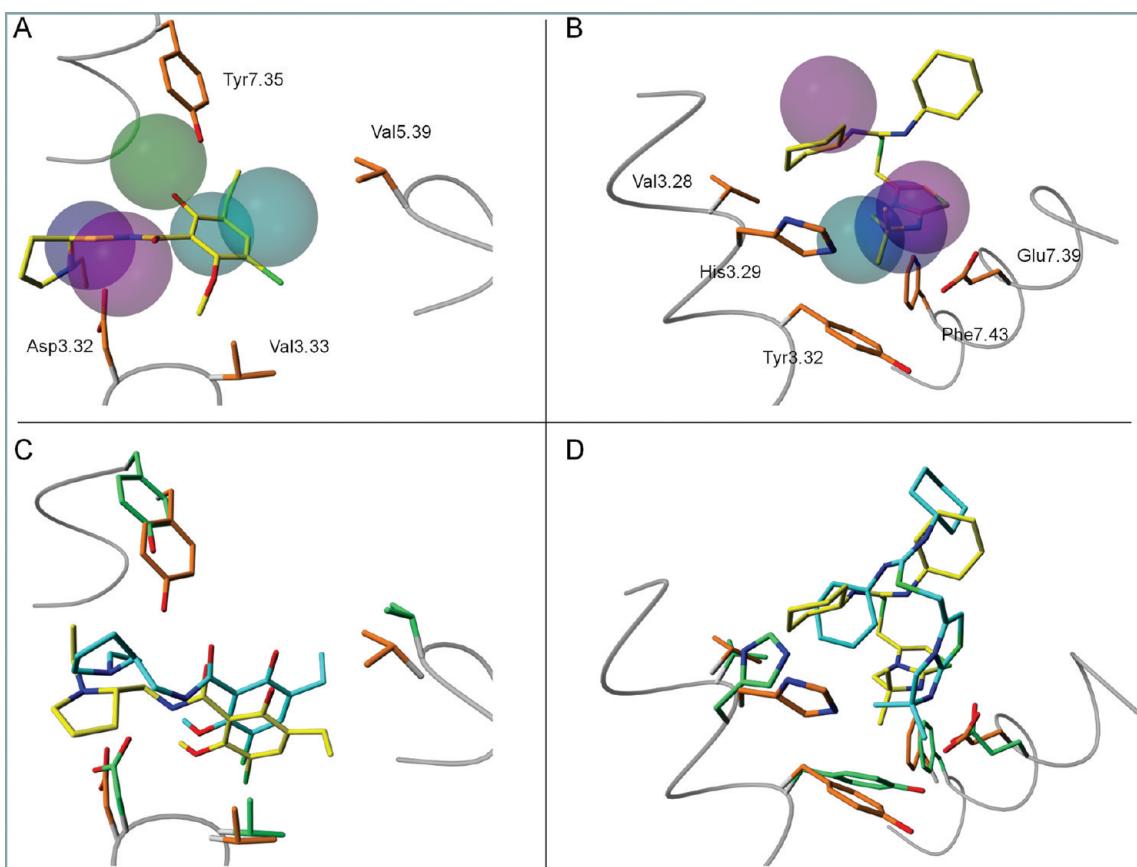
The pose prediction experiment of ADRB2 already indicated that antagonists and inverse agonists are likely to have less specific contacts as agonists. This and the observation that the cocrystallized peptide ligand in the CXCR4 crystal structure does occupy a large part of the major pocket indicates that other small molecules and agonists in particular might bind somewhere other than IT1t. Interestingly, retrospective analysis of the binding modes obtained in minor subpocket 2 (which was *not* submitted for the GPCR DOCK 2010 assessment) show structures which resemble the CXCR4-IT1t crystal structure with interactions to residues Glu7.39, Tyr3.32, His3.29, Val3.28, and Leu2.61 (see Figure 5B,D). Important interactions with W2.60 and D2.63 are however not predicted in these models, as the T2.56XP2.59 induced kink in TM2 was not correctly predicted based on the available set of GPCR crystal structure templates.<sup>58</sup>

**Identification of Receptor-Specific Ligands and Ligand Binding Residues in Cross-Screen.** Consensus pharmacophore hypotheses are generated for 15 targets using default Snooker settings, as described in the Methods section. The residue contributions to the pharmacophore hypotheses for the different models are depicted in Figure 6. The significance of each feature

for the identification of known actives for a model is indicated below the pie-chart. This significance has been calculated as the fraction of pharmacophores which contain the particular feature and match a conformation of a known active. Site-directed mutagenesis studies confirm that most of the pharmacophore features in the Snooker models relate to ligand binding residues. Most positively ionizable (and/or H-bond donor) pharmacophore features are associated with negatively charged residues which are determined to be essential for ligand binding to bioaminergic receptors ADA2B, ADRB2, DRD2, HRH3, and SHT7R (D3.32),<sup>76</sup> EDNRA (D7.35),<sup>75</sup> GHSR (E3.33),<sup>91</sup> MCHR1 (D3.32),<sup>92</sup> and OPRM (D3.32).<sup>78</sup> Likewise, negatively charged (and/or acceptor) features correspond to positively charged residues shown to be involved in ligand binding in AGTR1 (K5.42),<sup>93</sup> EDNRA (R6.55),<sup>75</sup> GHSR (R6.55),<sup>91</sup> and TA2R (R7.40).<sup>94</sup> In addition, polar residues at positions T3.36/N6.55/S7.42 in AA2AR,<sup>67,95</sup> 5.42/S5.43/S5.46 in the bioaminergic receptors,<sup>76</sup> and Q3.22 in EDNRA<sup>75</sup> are associated with donor and acceptor features and indeed play key roles in ligand binding for these receptors. The Snooker approach clearly allows an automated and fully protein-based construction of experimentally supported pharmacophore models. Interestingly, valid pharmacophore models could not only be built for receptors with existing crystal structures (ADRB2 and AA2AR) and receptors which are related to crystallized GPCRs (SHT7R, ADA2B, DRD2, HRH3) but also for receptors with low sequence similarity to GPCR crystal structure templates (AGTR1, CLTR1, EDNRA, GASR, NPYSR, GHSR, MCHR1, OPRM, TA2R).

Despite the fact that all template structures have ligands with key interactions to helix V, all features of the GHSR receptor are almost exclusively related to residues in helix III, VI, and VII. Site directed mutagenesis experiments confirm the importance of several residues (Q3.33/F6.51/R6.55/P6.58/N7.35) of helix III, VI, and VII and the relative unimportance of helix V (M5.39/V5.42/S5.43/F5.46) for ligand binding,<sup>91</sup> illustrating the ability of Snooker to presume different binding pockets by using similar template structures. The features of the CLTR1 pharmacophore hypothesis relate only to residues originating from helices III, VI, and VII. This suggests a different binding mode for ligands of this receptor compared to e.g. the amine receptors (SHT7R, ADA2B, ADRB2, DRD2, and HRH3) which do have polar interactions with helix V.<sup>76</sup> The values below the features indicate the importance of those features in retrieving active compounds, and it is remarkable to see that positive ionizable features play a crucial role in retrieving active compounds when present (as witnessed by the relatively high values). However, due to the limited number of negative ionizable residues and therefore positive ionizable features in the pocket are often two positive ionizable features related to the same residue (AA2AR, ADA2B, ADRB2, AGTR1, HRH3, OPRM). Subpharmacophores which are identical except for the positive ionizable features are therefore likely to retrieve similar compounds by assuming almost identical ligand poses. The absence of a feature in pharmacophores retrieving active compounds gives the opportunity to drop such a feature in a pharmacophore for virtual library screening and will most likely increase the hit rate.

To validate the target-specificity of the generated pharmacophore hypotheses we performed a cross-screening experiment on 15 sets of active and decoy compounds using the 15 corresponding pharmacophore hypotheses. The enrichment of all sets of active compounds is calculated for all pharmacophore hypotheses. Enrichment values of the different hypotheses are ranked



**Figure 5.** Binding mode predictions of eticlopride in the human DRD3 receptor and 1t in the CXCR4 receptor. Panel A: Human DRD3 pharmacophore and matching eticlopride pose. Panel B: Human CXCR4 pharmacophore and matching IT1t pose. Panel C: Optimized eticlopride pose and corresponding receptor structure in yellow and orange and crystal structure pose and receptor structure in green and cyan. Panel D: Optimized IT1t pose and corresponding receptor structure in yellow and orange and crystal structure pose and receptor structure in green and cyan.

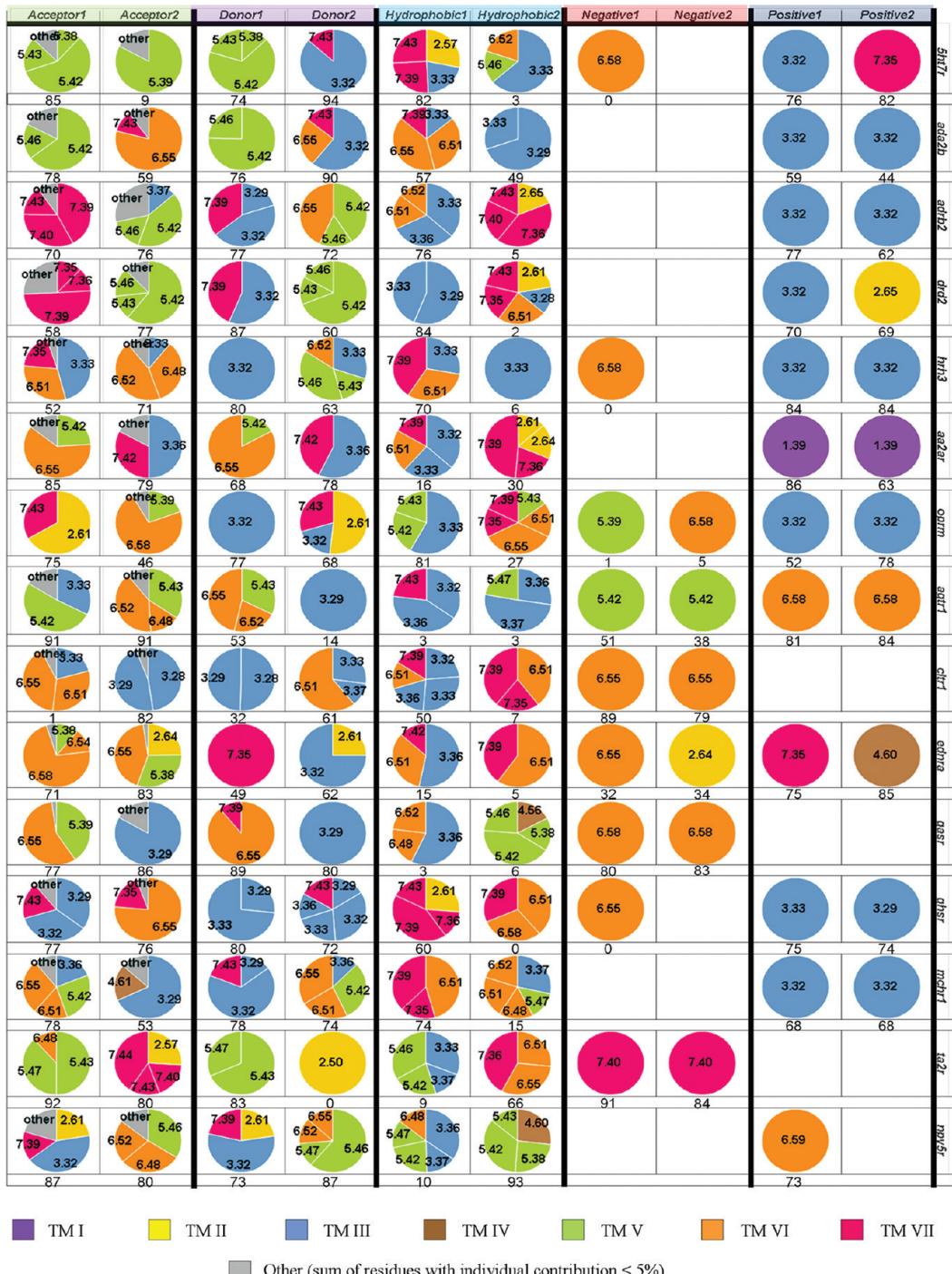
for each set of active compounds. Figure 7 shows the rank of the pharmacophore hypothesis of the target receptor which corresponds to the compound set of actives. The target consensus pharmacophore hypothesis ranks as one of the best 3 hypotheses for 10 of the 15 receptors.

**Customized Models.** A shortcoming of the automatic homology modeling procedure is the accurate prediction of helix structure, for example when kinks are present, as observed in TM2 of the recently elucidated CXCR4 receptor structure.<sup>57</sup> Performance of the consensus pharmacophore hypotheses in the cross screening experiment is poor for 4 compound sets. Mutagenesis data and multiple sequence analysis<sup>38,72</sup> indicate that minor model adjustments are required for those receptors. Gln3.32 is known to be important for ligand binding in the EDNRA receptor<sup>75</sup> and is not related to a pharmacophore feature in the original pharmacophore hypothesis. The hypotheses based on the 3CAP and 3EML template is tested for the EDNRA receptor because Gln3.32 correlates to a pharmacophore feature in these hypotheses. The modifications made to the NPY5R, TA2R, and OPRM receptor are, respectively, the adjustment of the template due to an assumed sulfide bridge and the inclusion of the third ranked polar feature instead of the second ranked polar feature, adjustment of the multiple sequence alignment and selection of a template to enable exposure of residues which are likely to be ligand binding, and selection of a template due to correspondence of the resulting pharmacophore

hypothesis with literature<sup>77–79</sup> (see Methods for details). The original consensus pharmacophore of those receptors show a reasonable amount of similarity to the modified pharmacophores and show major differences by 2–3 features (Appendix 7, Supporting Information).

Three of the manually adjusted pharmacophore hypotheses retrieve their corresponding compound set among the best 3 hypotheses in the earlier mentioned cross-screen. Thus, as in many other modeling procedures, knowledge-based input can also improve Snooker pharmacophore hypotheses,<sup>52,55</sup> but such input is not a prerequisite for this method.

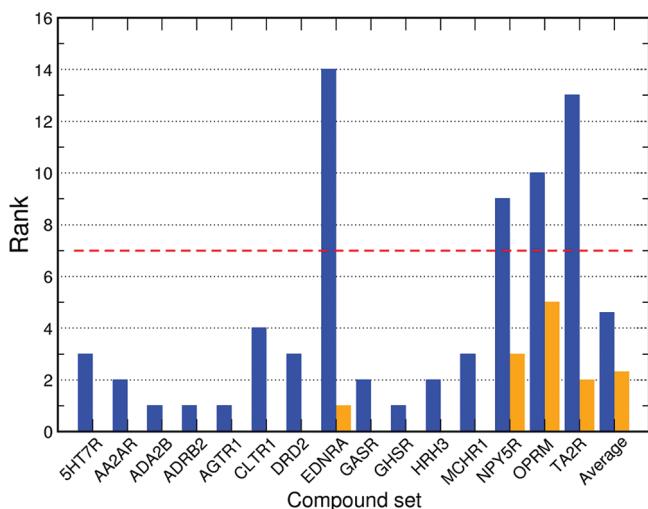
All 15 generated pharmacophore hypotheses possess at least 2 acceptors, 2 donors, and 2 hydrophobic features, and some are complemented with a maximum of 2 positive and/or negative ionizable features. The presence of ionizable features in hypotheses could potentially lead to biased enrichment — the requirement for corresponding features on ligands might function as an effective 2D filter, independent of positions in the hypotheses. To address this problem, fake active sets of compounds with chemical properties similar to the active molecules were cross-screened to determine the target specificity accomplished by the correct spatial arrangement of features. The average rank of 7.0 for the pharmacophore hypotheses corresponding to the fake active sets (compared to 2.3 for the true actives) further indicates that features have indeed the correct spatial arrangement. Ideally, compound sets should be best recognized by the pharmacophore



**Figure 6.** Pie-charts visualizing the contributions of the different residues to pharmacophore hypotheses for 15 different GPCRs. The percentage of recognized active compounds by a subpharmacophore which contain a feature is indicated below the pie-chart of that particular feature. In a scenario where 60 out of the 100 actives are retrieved, and 30 of the actives are matched in a pharmacophore where feature X is involved, the number for feature X will be 50.

hypothesis based on the homology model of the receptor at which the compounds are active. However, polypharmacology against GPCRs is widely accepted and even occurs in known antipsychotics and antidepressants. The selected compound sets include the 50 structurally most diverse compounds for each target and possibly contain a mix of partial and full agonists, antagonists, and inverse agonists. Hierarchical clustering on

compound fingerprints (Appendix 8, Supporting Information) shows the similarity of compounds selected for the different targets. Some of the compounds show activity across multiple target receptors (ChembL02 compound id. 2214 has reported activity on the SHT7R and DRD2 receptor, compound id. 1989 has reported activity on the ADA2B and DRD2 receptor and compound id. 27397 has reported activity on the ADA2B and



**Figure 7.** The rank in a list of 15 pharmacophore hypotheses at which a target pharmacophore hypothesis enriches a compound set of active molecules for that target. Blue indicates the rank of the consensus models and orange the rank of the adjusted models (see Materials and Methods). The average rank of the screen of the “fake” sets of actives is indicated with the red dashed line.

ADRB2 receptor). Although active compounds are only included in the test set of only one receptor (Appendix 5, Supporting Information), they are successfully retrieved by the pharmacophore hypotheses of the other targets on which they have reported activity (data not shown). Cross-activity inhibits the selection of sets of compounds uniquely active at only one receptor, and it is of course possible that cross activity exists, especially within receptor subfamilies, which is not yet reported in the literature. Snooker allows for the virtual screening of compounds against pharmacophores of all class A GPCRs. The outcome of such a virtual screen can be used to predict bioactivity profiles, design multitarget drugs, or to select a panel of GPCRs on which a compound should be tested to prohibit possible side effects at a later stage.

**Library Enrichment.** The ultimate goal of 3D pharmacophore hypotheses is to predict the conformation of all and only active ligands exactly as observed in the crystal structure. Shape constraints are often added to pharmacophore hypotheses to improve the enrichment of active molecules in virtual screenings of compound libraries.<sup>24,60</sup> Since only limited enrichment is observed in the cross-screening experiments we introduce a shape constraint derived from a reference pose of an active compound in the enrichment experiment to optimize enrichment factors. The reference pose is automatically extracted from a training set of active compounds and used to filter the decoys and test set on shape similarity to this reference pose (see Materials and Methods). The number of actives, decoys, and enrichment factors for the filtered and unfiltered sets are listed in Table 2.

On average the shape constraint improves enrichment values by a factor  $\sim 2.0$ . The enrichment values for those targets with the most stringent shape criteria (ADA2B, CLTR1, TA2R) are improved best. In the case of ADA2B this is due to the small size of the reference molecule as well as ADA2B active compounds. Larger compounds have by definition a higher chance to fit a pharmacophore without shape constraints and are removed by the addition of this shape constraint. As a result, the average

molecular weight of the filtered sets is reduced from 201 to 165 and 279 to 173 for the actives and decoys, respectively. For TA2R the shape constraint selects almost exclusively poses which fit the pharmacophore consisting of features 1, 2, 6, 7, and 8. This pharmacophore is rarely present in poses generated for decoy compounds. A result of the stringent cutoff is that all remaining compounds after filtering contain the same scaffold. However it should be noted that an acceptor feature and a correct orientation of the varying R-group is required to fulfill the pharmacophore criteria and shape constraint. For CLTR1 very few actives or decoys were selected. Retrospective analysis of the training set shows a high similarity among the compounds in this training set explaining the stringent cutoff value and low number of selected compounds. Both selected CLTR1 actives also show relatively high similarity to the compounds in the training set, indicating that the automatic selection of a reference and cutoff value can sometimes result in high enrichment values but with less novelty in the resulting compound sets due to a suboptimal training set. Most shape filtered sets do however show chemically diverse active compounds as desired in typical virtual screening experiments (see Appendix 9, Supporting Information). The training set of active compounds is used to derive the reference shape but also to rank the compounds which match a pharmacophore. Pharmacophores are therefore first ranked on the number of features and next on the number of training compounds which match the pharmacophore. Compounds are subsequently ordered on the pharmacophore in which they match and the fitvalue which is obtained in this pharmacophore. Using this approach we calculated 6 different performance measures for the enrichment and these are listed in Table 3. The high early enrichment values (at EF 0.5% and 1.0%) as compared to the enrichment values in Table 3 indicate that this ranking improves the result of the virtual screening.

Virtual screening methods should enrich active ligands at least 10-fold to obtain a reasonable chance of finding a true hit.<sup>96</sup> According to this criterion, Snooker shows early enrichment for 8 and 6 of the 15 receptors at 0.5% and 1% false positive rates. Interestingly, high virtual screening enrichments are obtained not only for bioaminergic receptors (ADRB2, 5HT7R, ADA2B, DRD2) but also pharmacophore models based on receptors with low sequence similarity to GPCR crystal structures (AGTR1, EDNRA, GASR, GHSR, TA2R) yield high early enrichment results. This shows that Snooker is not necessarily dependent on the availability of high resolution structural data and demonstrates the strength of the pharmacophore modeling approach. Although enrichment is mainly achieved by the additional shape restraints, this method results in a binding model of the compounds which is likely to reflect the true interaction of the compound with the receptor. Such information can be very useful in the design of experiments and compound optimization after the discovery of a new active compound. The relatively poor performance of the AA2AR pharmacophore hypothesis, partly based on structural information of the AA2AR crystal structure, can be explained by the fact that the ligand binding site is for a large part located between the extracellular loops, currently not included in Snooker pharmacophore models. It should furthermore be noted that the test sets of 50 active compounds consist of compounds with different scaffolds and possibly different binding modes, explaining the generally low global virtual screening results (AUC). Such a diverse set of active molecules is possibly more difficult to describe with a combination of a single pharmacophore hypothesis and shape restraint and results

**Table 2.** Retrospective Virtual Screening Accuracies of 15 Compound Sets Using Pharmacophores with and without Shape Constraints

receptor	actives (unfiltered)	decoys (unfiltered)	shape cutoff <sup>a</sup>	actives (filtered)	decoys (filtered)	EF (unfiltered) <sup>b</sup>	EF (filtered) <sup>c</sup>
SHT7R	16	3023	0.560	5	567	1.06	1.76
AA2AR	25	4170	0.555	9	884	1.20	2.04
ADA2B	30	5232	0.434	12	38	1.15	63.16
ADRB2	38	2253	0.576	21	452	3.37	9.29
AGTR1	22	1961	0.635	9	350	2.24	5.14
CLTR1	13	2823	0.514	2	20	0.92	20.00
DRD2	30	2918	0.560	12	845	2.06	2.84
EDNRA	44	5721	0.578	24	1917	1.54	2.50
GASR	11	1444	0.844	11	1230	1.52	1.79
GHSR	48	4439	0.588	17	508	2.16	6.69
HRH3	37	4762	0.570	19	1061	1.55	3.58
MCHR1	36	4386	0.603	13	802	1.64	3.24
NPYSR	14	1925	0.618	4	762	1.45	1.05
OPRM	24	3610	0.668	5	678	1.33	1.47
TA2R	25	675	0.538	5	1	7.41	1000.00

<sup>a</sup>Shape cutoff used to filter the test and decoy set (see Materials and Methods). <sup>b</sup>Enrichment factor (EF): ratio of “filtered actives/filtered decoys” to “all actives/all decoys”. <sup>c</sup>Enrichment factor (EF): ratio of “unfiltered actives/unfiltered decoys” to “all actives/all decoys”.

**Table 3.** Retrospective Virtual Screening Accuracies of 15 Compound Sets Using Snooker Pharmacophores and Shape Constraints<sup>d</sup>

receptor	pROC AUC <sup>a</sup>	AUC <sup>b</sup>	EF <sup>c</sup> at 0.5%	EF <sup>c</sup> at 1.0%	EF <sup>c</sup> at 2.0%	EF <sup>c</sup> at 5.0%
SHT7R	0.53	0.57	12	6	4	2
AA2AR	0.55	0.53	0	2	1	2
ADA2B	0.61	0.98	44	22	11	5
ADRB2	0.69	1.04	16	14	14	8
AGTR1	0.58	0.71	16	12	9	4
CLTR1	0.52	0.56	8	4	2	2
DRD2	0.59	0.76	16	10	8	5
EDNRA	0.66	0.90	16	8	5	5
GASR	0.56	0.62	4	10	5	3
GHSR	0.65	0.96	16	14	10	7
HRH3	0.64	0.82	8	6	4	6
MCHR1	0.60	0.72	8	4	6	4
NPYSR	0.50	0.43	0	0	1	1
OPRM	0.52	0.50	0	2	3	2
TA2R	0.54	0.71	16	8	4	2

<sup>a</sup>Area under the ROC curve. <sup>b</sup>Area under the semilogarithmic curve.

<sup>c</sup>Enrichment factor (EF): the ratio of true positive rates to false positive rates at increasing false positive rates (0.5%, 1%, 2%, and 5%). <sup>d</sup>Enrichment factors (EF) above 3 and 10 are colored orange and green, respectively.

therefore usually in less satisfying enrichments. The definition of multiple combinations of pharmacophores and shape restraints based e.g. overlays of different compound series available in the training set might improve enrichment scores but would likely limit the structural diversity of hits.

## CONCLUSION

In this paper, we present Snooker, a new structure-based approach to generate low resolution pharmacophore hypotheses

for class A GPCRs. We show that Snooker generates ADRB2 pharmacophore hypotheses which retrieve 3/12 antagonists-inverse agonists and 13/14 agonists and assumes the correct binding mode for 3 and 10 compounds, respectively. All 5 submitted eticlopride binding mode predictions in the human DRD3 receptor, which are largely based on the eticlopride matches in Snooker pharmacophore, showed to be in the top 10 of all submitted models in the GPCR dock 2010 assessment. The automated and fully protein-based construction of pharmacophore hypotheses is in line with experimental site-directed mutagenesis data on essential ligand binding residues for a diverse set of 15 class A GPCRs. For several of the more difficult targets, the default Snooker settings can be adjusted with target-based knowledge to obtain good results. Interestingly, valid pharmacophore models were built for not only 2 receptors with known crystal structures and 4 related receptors but also for 9 receptors with low sequence similarity to GPCR crystal structure templates. A virtual screening experiment using the Snooker pharmacophore hypotheses in combination with a shape constraint resulted in >10 fold enriched compound sets for 8 out of 15 targets. As such, the method is suitable to design focused compound libraries targeting a small subfamily of class A GPCRs.

## ASSOCIATED CONTENT

**S Supporting Information.** Ballesteros and Weinstein numbering of the transmembrane domains and corresponding residue numbers of the pdb-files of the OPSD, ADRB1, ADRB2, and AA2AR receptor, a depiction of the charged interaction geometry, basic interaction feature densities, Cyndi parameter settings, ChEMBL id's, pharmacophore feature definitions, default consensus pharmacophore hypotheses of the EDNRA, NPYSR, OPRM, and TA2R receptor, the clustering of all active compounds of the 15 targets used in this manuscript, and the chemical structures of the active molecules after shape filtering (Appendices 1–9). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: 0031412661383. E-mail: j.klomp@cmci.ru.nl. Corresponding author address: Department of Molecular Design and Informatics, MSD, P.O. Box 20, 5340BH Oss, The Netherlands.

## ■ ACKNOWLEDGMENT

The authors thank Peter Groenen, Ross McGuire, Scott Lusher, and Andreas Bender for critical reading of the manuscript. This work was performed within the framework of Dutch Top Institute Pharma, project "The GPCR Forum: Novel concepts and tools for established targets (project nr. D1-105)".

## ■ REFERENCES

- (1) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**, *3*, 928–944.
- (2) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727–730.
- (3) Franke, L.; Byvatov, E.; Werz, O.; Steinheilber, D.; Schneider, P.; Schneider, G. Extraction and visualization of potential pharmacophore points using support vector machines: application to ligand-based virtual screening for COX-2 inhibitors. *J. Med. Chem.* **2005**, *48*, 6997–7004.
- (4) Marrero-Ponce, Y.; Iyarreta-Veitia, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Avila, P. E.; Kirchgatter, K.; Machado, Y. Ligand-based virtual screening and in silico design of new antimalarial compounds using nonstochastic and stochastic total and atom-type quadratic maps. *J. Chem. Inf. Model.* **2005**, *45*, 1082–1100.
- (5) Schuster, D.; Maurer, E. M.; Laggner, C.; Nashev, L. G.; Wilckens, T.; Langer, T.; Odermatt, A. The discovery of new 11beta-hydroxysteroid dehydrogenase type 1 inhibitors by common feature pharmacophore modeling and virtual screening. *J. Med. Chem.* **2006**, *49*, 3454–3466.
- (6) 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*, 5448–5465.
- (7) Schneider, G.; Nettekoven, M. Ligand-based combinatorial design of selective purinergic receptor (A2A) antagonists using self-organizing maps. *J. Comb. Chem.* **2003**, *5*, 233–237.
- (8) Oloff, S.; Mailman, R. B.; Tropsha, A. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. *J. Med. Chem.* **2005**, *48*, 7322–7332.
- (9) Zhang, Q.; Muegge, I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J. Med. Chem.* **2006**, *49*, 1536–1548.
- (10) Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* **2004**, *8*, 349–358.
- (11) Congreve, M.; Murray, C. W.; Blundell, T. L. Structural biology and drug discovery. *Drug Discovery Today* **2005**, *10*, 895–907.
- (12) Jorgenson, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
- (13) Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* **2004**, *10*, 1011–1033.
- (14) de Graaf, C.; Rognan, D. Customizing G Protein-coupled receptor models for structure-based virtual screening. *Curr. Pharm. Des.* **2009**, *15*, 4026–4048.
- (15) 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*, 6843–6848.
- (16) 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*, 1799–1809.
- (17) Nabuurs, S. B.; Wagener, M.; de Vlieg, J. A flexible approach to induced fit docking. *J. Med. Chem.* **2007**, *50*, 6507–6518.
- (18) Crossley, R. The design of screening libraries targeted at G-protein coupled receptors. *Curr. Top. Med. Chem.* **2004**, *4*, 581–588.
- (19) Salo, O. M.; Raitio, K. H.; Savinainen, J. R.; Nevalainen, T.; Lahtela-Kakkonen, M.; Laitinen, J. T.; Jarvinen, T.; Poso, A. Virtual screening of novel CB2 ligands using a comparative model of the human cannabinoid CB2 receptor. *J. Med. Chem.* **2005**, *48*, 7166–7171.
- (20) Kenyon, V.; Chorny, I.; Carvajal, W. J.; Holman, T. R.; Jacobson, M. P. Novel human lipoxygenase inhibitors discovered using virtual screening with homology models. *J. Med. Chem.* **2006**, *49*, 1356–1363.
- (21) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* **2003**, *50*, 5–25.
- (22) Bissantz, C.; Logean, A.; Rognan, D. High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1162–1176.
- (23) Kratochwil, N. A.; Malherbe, P.; Lindemann, L.; Ebeling, M.; Hoener, M. C.; Muhlemann, A.; Porter, R. H.; Stahl, M.; Gerber, P. R. An automated system for the analysis of G protein-coupled receptor transmembrane binding pockets: alignment, receptor-based pharmacophores, and their application. *J. Chem. Inf. Model.* **2005**, *45*, 1324–1336.
- (24) Klabunde, T.; Giegerich, C.; Evers, A. Sequence-derived three-dimensional pharmacophore models for G-protein-coupled receptors and their application in virtual screening. *J. Med. Chem.* **2009**, *52*, 2923–2932.
- (25) Tsai, K. C.; Chen, S. Y.; Liang, P. H.; Lu, I. L.; Mahindroo, N.; Hsieh, H. P.; Chao, Y. S.; Liu, L.; Liu, D.; Lien, W.; Lin, T. H.; Wu, S. Y. Discovery of a novel family of SARS-CoV protease inhibitors by virtual screening and 3D-QSAR studies. *J. Med. Chem.* **2006**, *49*, 3485–3495.
- (26) Bologa, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.; Savchuk, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R. Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2006**, *2*, 207–212.
- (27) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug. Discovery* **2006**, *5*, 993–996.
- (28) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035–1041.
- (29) Chen, X.; Lin, Y.; Gilson, M. K. The binding database: overview and user's guide. *Biopolymers* **2001**, *61*, 127–141.
- (30) Chen, X.; Lin, Y.; Liu, M.; Gilson, M. K. The Binding Database: data management and interface design. *Bioinformatics* **2002**, *18*, 130–139.
- (31) Chen, X.; Liu, M.; Gilson, M. K. BindingDB: a web-accessible molecular recognition database. *Comb. Chem. High Throughput Screening* **2001**, *4*, 719–725.
- (32) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977–2980.
- (33) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.
- (34) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (Mother Of All Databases). *Proteins* **2005**, *60*, 333–340.
- (35) Smith, R. D.; Hu, L.; Falkner, J. A.; Benson, M. L.; Nerothin, J. P.; Carlson, H. A. Exploring protein-ligand recognition with Binding MOAD. *J. Mol. Graphics Modell.* **2006**, *24*, 414–425.
- (36) Oprea, T. I.; Allu, T. K.; Fara, D. C.; Rad, R. F.; Ostropovici, L.; Bologa, C. G. Lead-like, drug-like or "Pub-like": how different are they? *J. Comput.-Aided Mol. Des.* **2007z**, *21*, 113–119.
- (37) Okuno, Y.; Tamon, A.; Yabuuchi, H.; Niijima, S.; Minowa, Y.; Tonomura, K.; Kunimoto, R.; Feng, C. GLIDA: GPCR-ligand database for chemical genomics drug discovery--database and tools update. *Nucleic Acids Res.* **2008**, *36*, D907–912.
- (38) Sanders, M. P.; Fleuren, W. W.; Verhoeven, S.; van den Beld, S.; Alkema, W.; de Vlieg, J.; Klomp, J. P. ss-TEA: Entropy based identification

- of receptor specific ligand binding residues from a multiple sequence alignment of class A GPCRs. *BMC Bioinf.* **2011**, *12*, 332.
- (39) Li, J.; Edwards, P. C.; Burghammer, M.; Villa, C.; Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **2004**, *343*, 1409–1438.
- (40) Ballesteros, J. A. W. 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) Chinea, G.; Padron, G.; Hooft, R. W.; Sander, C.; Vriend, G. The use of position-specific rotamers in model building by homology. *Proteins* **1995**, *23*, 415–421.
- (42) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins* **2000**, *40*, 389–408.
- (43) Madabushi, S.; Gross, A. K.; Philippi, A.; Meng, E. C.; Wensel, T. G.; Lichtarge, O. Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *J. Biol. Chem.* **2004**, *279*, 8126–8132.
- (44) Oliveira, L.; Paiva, P. B.; Paiva, A. C.; Vriend, G. Sequence analysis reveals how G protein-coupled receptors transduce the signal to the G protein. *Proteins* **2003**, *52*, 553–560.
- (45) Ye, K.; Lameijer, E. W.; Beukers, M. W.; Ijzerman, A. P. A two-entropies analysis to identify functional positions in the transmembrane region of class A G protein-coupled receptors. *Proteins* **2006**, *63*, 1018–1030.
- (46) Bohm, B.; Aigner, T.; Kinne, R.; Burkhardt, H. The serine-protease inhibitor of cartilage matrix is not a chondrocytic gene product. *Eur. J. Biochem.* **1992**, *207*, 773–779.
- (47) Gunzer, F.; Bohm, H.; Russmann, H.; Bitzan, M.; Aleksic, S.; Karch, H. Molecular detection of sorbitol-fermenting Escherichia coli O157 in patients with hemolytic-uremic syndrome. *J. Clin. Microbiol.* **1992**, *30*, 1807–1810.
- (48) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243–256.
- (49) Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 212–235.
- (50) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (51) Ye, K.; Feenstra, K. A.; Heringa, J.; Ijzerman, A. P.; Marchiori, E. Multi-RELIEF: a method to recognize specificity determining residues from multiple sequence alignments using a Machine-Learning approach for feature weighting. *Bioinformatics* **2008**, *24*, 18–25.
- (52) Barillari, C.; Marcou, G.; Rognan, D. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J. Chem. Inf. Model.* **2008**, *48*, 1396–1410.
- (53) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (54) Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **1999**, *289*, 1093–1108.
- (55) de Graaf, C.; Rognan, D. Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. *J. Med. Chem.* **2008**, *51*, 4978–4985.
- (56) Chien, E. Y.; 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*, 1091–1095.
- (57) Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, 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*, 1066–1071.
- (58) Kufareva, I.; Rueda, R.; Katritch, V. GPCR Dock 2010 participants, Stevens, C. S.; Abagyan, R. Status of GPCR modeling and docking in a community wide GPCR dock 2010 assessment. *Structure* **2011**, *19*, 1108–1126.
- (59) Vroeling, B.; Sanders, M.; Baakman, C.; Borrmann, A.; Verhoeven, S.; Klomp, J.; Oliveira, L.; de Vlieg, J.; Vriend, G. GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res.* **2010**, *39*, 1–11.
- (60) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* **2010**, *53*, 539–558.
- (61) Okada, T.; Fujiyoshi, Y.; Silow, M.; Navarro, J.; Landau, E. M.; Shichida, Y. Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5982–5987.
- (62) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–1265.
- (63) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G.; Tate, C. G.; Schertler, G. F. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **2008**, *454*, 486–491.
- (64) Park, J. H.; Scheerer, P.; Hofmann, K. P.; Choe, H. W.; Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **2008**, *454*, 183–187.
- (65) Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V. P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C. A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure* **2008**, *16*, 897–905.
- (66) Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauss, N.; Choe, H. W.; Hofmann, K. P.; Ernst, O. P. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, *455*, 497–502.
- (67) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y.; 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*, 1211–1217.
- (68) Needleman, S. B.; Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453.
- (69) Krieger, E.; Koraimann, G.; Vriend, G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins* **2002**, *47*, 393–402.
- (70) Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics* **1990**, *8* (52–56), 29.
- (71) Delaunay, B. Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk* **1934**, *7*, 793–800.
- (72) Kumar, S.; Nussinov, R. Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys. J.* **2002**, *83*, 1595–1612.
- (73) Renner, S.; Schneider, G. Fuzzy pharmacophore models from molecular alignments for correlation-vector-based virtual screening. *J. Med. Chem.* **2004**, *47*, 4653–4664.
- (74) Breu, V.; Hashido, K.; Broger, C.; Miyamoto, C.; Furuchi, Y.; Hayes, A.; Kalina, B.; Loffler, B. M.; Ramuz, H.; Clozel, M. Separable binding sites for the natural agonist endothelin-1 and the non-peptide antagonist bosentan on human endothelin-A receptors. *Eur. J. Biochem.* **1995**, *231*, 266–270.
- (75) 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.
- (76) Kane, B. E.; Svensson, B.; Ferguson, D. M. Molecular recognition of opioid receptor ligands. *AAPS J.* **2006**, *8*, E126–137.
- (77) Li, J. G.; Chen, C.; Yin, J.; Rice, K.; Zhang, Y.; Matecka, D.; de Riel, J. K.; DesJarlais, R. L.; Liu-Chen, L. Y. ASP147 in the third transmembrane helix of the rat mu opioid receptor forms ion-pairing with morphine and naltrexone. *Life Sci.* **1999**, *65*, 175–185.

- (78) Surratt, C. K.; Johnson, P. S.; Moriwaki, A.; Seidleck, B. K.; Blaschak, C. J.; Wang, J. B.; Uhl, G. R. -mu opiate receptor. Charged transmembrane domain amino acids are critical for agonist recognition and intrinsic activity. *J. Biol. Chem.* **1994**, *269*, 20548–20553.
- (79) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 8.
- (80) Liu, X.; Bai, F.; Ouyang, S.; Wang, X.; Li, H.; Jiang, H. Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC Bioinf.* **2009**, *10*, 101.
- (81) Pipeline Pilot, version 8.0; Accelrys: San Diego, CA, 2010.
- (82) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **2003**, *12*, 2001–2014.
- (83) 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*, 241–244.
- (84) Krieger, E.; Joo, K.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **2009**, *77* (Suppl 9), 114–122.
- (85) Clark, R. D.; Webster-Clark, D. J. Managing bias in ROC curves. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 141–146.
- (86) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.
- (87) Landrum, G. RDKit, Q2 2010 1; Palo Alto, CA 94301, USA.
- (88) Staubert, C.; Boselt, I.; Bohnekamp, J.; Rompler, H.; Enard, W.; Schoneberg, T. Structural and functional evolution of the trace amine-associated receptors TAAR3, TAAR4 and TAAR5 in primates. *PLoS One* **5**, e11133.
- (89) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G.; 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*, 1266–1273.
- (90) Holst, B.; Frimurer, T. M.; Mokrosinski, J.; Halkjaer, T.; Cullberg, K. B.; Underwood, C. R.; Schwartz, T. W. Overlapping binding site for the endogenous agonist, small-molecule agonists, and allosteric modulators on the ghrelin receptor. *Mol. Pharmacol.* **2009**, *75*, 44–59.
- (91) Macdonald, D.; Murgolo, N.; Zhang, R.; Durkin, J. P.; Yao, X.; Strader, C. D.; Graziano, M. P. Molecular characterization of the melanin-concentrating hormone/receptor complex: identification of critical residues involved in binding and activation. *Mol. Pharmacol.* **2000**, *58*, 217–225.
- (92) Fierens, F. L.; Vanderheyden, P. M.; Gaborik, Z.; Minh, T. L.; Backer, J. P.; Hunyady, L.; Ijzerman, A.; Vauquelin, G. Lys(199) mutation of the human angiotensin type 1 receptor differentially affects the binding of surmountable and insurmountable non-peptide antagonists. *J. Renin Angiotensin Aldosterone Syst.* **2000**, *1*, 283–288.
- (93) Khasawneh, F. T.; Huang, J. S.; Turek, J. W.; Le Breton, G. C. Differential mapping of the amino acids mediating agonist and antagonist coordination with the human thromboxane A2 receptor protein. *J. Biol. Chem.* **2006**, *281*, 26951–26965.
- (94) Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Jacobson, K. A.; Gao, Z. G.; Cherezov, V.; Stevens, R. C. Structure of an agonist-bound human A2A adenosine receptor. *Science* **2011**, *332*, 322–327.
- (95) Muegge, I.; Oloff, S. Advances in virtual screening. *Drug Discovery Today: Technol.* **2006**, *3*, 405–411.
- (96) Ortore, G.; Tuccinardi, T.; Orlandini, E.; Martinelli, A. Different Binding Modes of Structurally Diverse Ligands for Human D3DAR. *J. Chem. Inf. Model.* **2010**, *50*, 2162–2175.