Selective Structure-Based Virtual Screening for Full and Partial Agonists of the β2 Adrenergic Receptor

Chris de Graaf and Didier Rognan*

Bioinformatics of the Drug, Institut Gilbert Laustriat, CNRS UMR 7175-LC1, Université Louis Pasteur Strasbourg, 74 Route du Rhin, 67401 Illkirch, France

Received April 22, 2008

The recently solved high-resolution X-ray structure of the β2 adrenergic receptor has been challenged for its ability to discriminate inverse agonists/antagonists from partial/full agonists. Whereas the X-ray structure of the ground state receptor was unsuitable to distinguish true ligands with different functional effects, modifying this structure to reflect early conformational events in receptor activation led to a receptor model able to selectively retrieve full and partial agonists by structure-based virtual screening. The use of a topological scoring function based on molecular interaction fingerprints was shown to be mandatory to properly rank docking poses and achieve acceptable enrichments for partial and full agonists only.

Introduction

G-protein-coupled receptors (GPCRs) constitute a superfamily of transmembrane proteins of utmost pharmaceutical importance. Knowledge of the three-dimensional structure of GPCRs can provide important insights into receptor function and receptor–ligand interactions and can be used for the discovery of new drugs. So far structural modeling of GPCRs has been limited to either ab initio models or bovine rhodopsin (bRho)-based homology models. Recently, crystal structures of a second GPCR, the β2 adrenergic receptor (ADRB2), were solved. Although the structures of ADRB2 and bRho are highly homologous, the differences might be large enough to influence the outcome of structure-based virtual screening. It has been stated, however, that the ground state of ADRB2 and bRho crystal structures are only suitable for discovering inverse agonists and antagonists. The aim of the current study is to challenge the ADRB2 structure for its ability to selectively retrieve partial/full agonists using structure-based virtual screening methods.

In the high-resolution ADRB2 crystal structure (PDB entry 2rh1), D3.32 and N7.39 form a complementary H-bond network with the ethanolamine group of the inverse agonist S-carazolol; the carboxylate oxygens of D3.32 act as H-bond acceptors for the protonated amine nitrogen and hydroxyl group of the ligand, while the amide group of N7.39 acts as H-bond donor to the ligand amine nitrogen and hydroxyl oxygen, respectively (Figure 1A). The carbazole heterecycle of the ligand interacts with F6.52 via an edge-to-face π stacking, while the carbazol0e nitrogen atom donates an H-bond to the hydroxyl oxygen of S5.42. This binding mode is in line with earlier site-directed mutagenesis studies, supporting the involvement of D3.32, S5.42, and N7.39 in binding of both antagonist and agonists. Both functional and biophysical studies on ADRB2 suggest that agonist binding occurs through kinetically distinct steps involving several conformational intermediates. The agonist binding model hypothesis follows at least three steps. In the first step, the protonated amine of the agonist forms an ionic/H-bond interaction with D3.32 (and N7.39) while the catechol ring stacks with F6.52. A fast conformational change in the receptor facilitates the formation of H-bonds of the catechol hydroxyl groups of the agonist with serines in TM5 (S5.42, S5.43, and S5.46) in a second step. This triggers a slower conformational change involving a rotamer “toggle switch” and H-bond formation between the β-alcohol of the agonist and N6.55.

* To whom correspondence should be addressed. Phone: +33-(0)390-244325. Fax: +33-(0)390-244310. E-mail: didier.rognan@pharma.u-strasbg.fr.

Abbreviations: ADRB2, β2 adrenergic receptor; bRho, bovine rhodopsin; CCR5, C–C chemokine receptor type 5; CNR2, cannabimind receptor 2; DRD2, dopamine D2 receptor; EF, enrichment factor; FFAR1, free fatty acid receptor 1; GPCR, G-protein-coupled receptor; IFP, interaction fingerprint; ROC, receiver-operator characteristics; TM, transmembrane (helix).
In the current study, we have customized the ADRB2 crystal structure to reflect the second activation step and challenged this structure for its ability to discriminate partial/full agonists from inverse agonists/antagonists.

Results and Discussion

Customization of ADRB2 Crystal Structure To Accommodate Full and Partial Agonists. We have performed automated docking simulations of inverse agonists, neutral antagonists, partial agonists, and full agonists (Figure 2) in the ADRB2 crystal structure and observed that poses of full agonists like isoproterenol satisfy the requirements of the first step in agonist binding (H-bond with D3.32 and aromatic face-to-edge stacking with F6.52). In this binding orientation the catechol hydroxyl groups of R-isoproterenol are too distant from the hydroxyl oxygens of S5.42, S5.43, and S5.46 to form H-bonds. We have subsequently customized the ADRB2 crystal structure to reflect the “second” agonist binding step. Changing the rotameric states of S5.43 (from g- to g+) and S5.46 (from g- to g+) brings the hydroxyl oxygens of these residues into the binding cavity, almost within H-bond distance from the catechol m-hydroxyl (4.2 Å) and p-hydroxyl (3.5 Å) oxygens, respectively. This receptor–ligand complex was refined by two energy minimization runs, first with experimentally guided H-bond constraints, followed by a second run without constraints (see Computational Methods). During the refinement, the ligand moves slightly toward TM5 (shift in center of mass of ~1 Å) while maintaining the H-bond network with D3.32 and N7.39. This finally results in an ADRB2–isoproterenol complex satisfying experimental data in which S5.42 donates an H-bond to the m-hydroxyl, S5.43 accepts an H-bond from the m-hydroxyl, and S5.46 accepts an H-bond from the p-hydroxyl of the catechol ring. This H-bond network is further stabilized by N6.55 and Y7.35, two residues reported to be involved in agonist binding as well (Figure 1B). The conformational changes in the receptor compared to the initial crystal structure are small: 0.3 and 0.5 Å rmsd considering binding cavity backbone and side chain atoms, respectively.

The customized ADRB2 structure proposed in this study probably represents an early intermediate state in agonist binding, which requires minor conformational changes from the ground state: rotation of two serine residues (“expelled” from the binding pocket by the hydrophobic moieties of the inverse agonist carazolol) back into the binding cavity (Figure 1). In this intermediate receptor state, the ethanolamine “head” of the agonist binds to D3.32 and N7.39, the aromatic moiety stacks with F6.52, and the polar groups on the aromatic ring bind to serine residues in TM5 (S5.42, S5.43, and S5.46), while N6.55 stabilizes the H-bond network between S5.43 and the ligand. Rearrangement of the receptor–agonist binding mode via either a rotamer toggle switch24 or rigid-body shifts of seesaw motions of transmembrane segments probably involves conformational changes like bending24,32 and/or rigid-body movements of TM helices which are likely to be ligand-dependent and therefore difficult to predict. Comparison of experimental structures of inactive and (semi)activated states of bovine rhodopsin, however, suggest that even these structural changes are smaller than previously proposed on the basis of indirect biochemical methods.

Docking ADRB2 Ligands To Inactive and Early Activated Receptor Structures. The binding poses of S-carazolol (in the ADRB2 X-ray structure, Figure 1A) and R-isoproterenol (in the customized receptor structure, Figure 1B) were used to generate reference interaction fingerprints (IFPs, see Figures 3 and 4) and served as templates for three-dimensional ligand-based ROCS queries, while their two-dimensional circular fingerprints were used for ECFP-4 similarity searches in retrospective virtual screening experiments. For evaluating the performance of ligand-based and receptor-based virtual screens, a database was prepared consisting of 13 known ADRB2 antago-
nists/inverse agonists, 13 known partial/full agonists (Figure 2), and 980 chemically similar compounds randomly selected from our in-house collection of commercially available compounds. It should be mentioned that structurally different agonists are believed to disrupt distinct combinations of ground-state stabilizing intramolecular interactions.20,40 For example, the inverse agonist propranolol inhibits signaling through G proteins and adenyl cyclase but acts as a partial agonist of arrestin-mediated activation of the ERK signaling pathway.41 For the current virtual screening evaluation, we have only considered ligands which inhibit G-protein/adenyl cyclase activation as ADRB2 partial/full agonists. In order to avoid biasing virtual screening results, caution was given to select 980 druglike decoys covering similar property ranges as the true actives but structurally different from any known active.

Gold and Surflex docking26,27 into the ADRB2 crystal structure compared to the reference binding mode of R-isoproterenol (green carbon atoms). H-Bonds between the docking pose and the receptor are depicted by black dots. Note that the coordinates of (serine) hydroxyl hydrogens are extracted from the timolol Gold docking solution (A). Rendering and color coding are the same as in Figure 1. The IFP bit strings of timolol (A) and butoxamine (B) are compared to the reference IFP of carazolol in part C. For reasons of clarity, the bit strings of only six residues (out of 33) are shown as an example. For each pose and corresponding IFP bit string, scores and ranks (between brackets) are indicated for Gold/Surflex docking ranked by IFP, Goldscore, and Surflex score. ECFP-4 and ROCS scores and ranks are given for the corresponding ligand using the same reference.

Figure 3. Docking poses selected by IFP of (A) S-timolol (yellow carbon atoms, docked with Gold) and (B) R,R-butoxamine (yellow carbon atoms, docked with Surflex) in the ADRB2 crystal structure compared to the reference binding mode of S-carazolol (green carbon atoms). H-Bonds between the docking pose and the receptor are depicted by black dots. Note that the coordinates of (serine) hydroxyl hydrogens are extracted from the timolol Gold docking solution (A). Rendering and color coding are the same as in Figure 1. The IFP bit strings of timolol (A) and butoxamine (B) are compared to the reference IFP of carazolol in part C. For reasons of clarity, the bit strings of only six residues (out of 33) are shown as an example. For each pose and corresponding IFP bit string, scores and ranks (between brackets) are indicated for Gold/Surflex docking ranked by IFP, Goldscore, and Surflex score. ECFP-4 and ROCS scores and ranks are given for the corresponding ligand using the same reference.

Figure 4. Docking poses of (A) R,S-procaterol (yellow carbon atoms, docked with Surflex) and (B) S-carvedilol (yellow carbon atoms, docked with Gold) in the customized ADRB2 structure compared to the reference binding mode of R-isoproterenol (green carbon atoms). H-Bonds between the docking pose and the receptor are depicted by black dots. Note that the coordinates of (serine) hydroxyl hydrogens are extracted from the carvedilol Gold docking solution (B). Rendering and color coding are the same as in Figure 1. The IFP bit strings of procaterol (A) and carvedilol (B) are compared to the reference IFP of isoproterenol in part C. For reasons of clarity, the bit strings of only six residues (out of 33) are shown as an example. For each pose and corresponding IFP bit string, scores and ranks (between brackets) are indicated for Gold/Surflex docking ranked by IFP, Goldscore, and Surflex score. ECFP-4 and ROCS scores and ranks are given for the corresponding ligand using the same reference.
under the ROC curve.

Table 1. Virtual Screening (VS) Accuracies of Different Methods

<table>
<thead>
<tr>
<th>methoda</th>
<th>ROC</th>
<th>EFc</th>
<th></th>
<th></th>
<th></th>
<th>inverse agonists/antagonistsb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>area</td>
<td>95% CI</td>
<td>0.5%</td>
<td>1%</td>
<td>2%</td>
<td>5%</td>
</tr>
<tr>
<td>Gold—Gold (ISP)</td>
<td>0.750</td>
<td>0.722–0.776</td>
<td>31</td>
<td>15</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Gold—IFP (ISP)</td>
<td>0.991</td>
<td>0.983–0.996</td>
<td>154</td>
<td>77</td>
<td>38</td>
<td>17</td>
</tr>
<tr>
<td>Surflex—Surflex (ISP)</td>
<td>0.902</td>
<td>0.882–0.920</td>
<td>46</td>
<td>23</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Surflex—IFP (ISP)</td>
<td>0.990</td>
<td>0.982–0.995</td>
<td>169</td>
<td>92</td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td>ECFP-4 (ISP)</td>
<td>0.998</td>
<td>0.992–1.000</td>
<td>169</td>
<td>92</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>ROCS (ISP)</td>
<td>0.984</td>
<td>0.975–0.991</td>
<td>123</td>
<td>85</td>
<td>42</td>
<td>17</td>
</tr>
<tr>
<td>Gold—Gold (CAZ)</td>
<td>0.585</td>
<td>0.553–0.615</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gold—IFP (CAZ)</td>
<td>0.940</td>
<td>0.923–0.954</td>
<td>15</td>
<td>8</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Surflex—Surflex (CAZ)</td>
<td>0.796</td>
<td>0.769–0.820</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Surflex—IFP (CAZ)</td>
<td>0.855</td>
<td>0.832–0.876</td>
<td>15</td>
<td>8</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>ECFP-4 (CAZ)</td>
<td>0.666</td>
<td>0.636–0.695</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>ROCS (CAZ)</td>
<td>0.685</td>
<td>0.656–0.714</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ideal</td>
<td>1.000</td>
<td>0.980–1.020</td>
<td>200</td>
<td>100</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>random</td>
<td>0.500</td>
<td>0.470–0.530</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>


b Individual ligands and their rankings are presented in Supporting Information Table 1. c Enrichment factor (EF): the ratio of true positive rates to false positive rates at increasing false positive rates (0.5%, 1%, 2%, and 5%). d Area under the ROC curve. * 95% confidence interval (CI).

Figure 5. (A) Enrichment in full/partial agonists (f/pAGO) vs enrichment in inverse agonists/antagonists (iAGO/ANT) at a constant 2% false positive rate for different screening approaches using the inverse agonist carazolol (green) or the full agonist isoproterenol (magenta) as a reference. Enrichments in (A) are based on individual ROC curves of f/pAGO (magenta) and iAGO/ANT (green) such as the ones presented in parts B (automated docking and IFP scoring in the agonist customized ADRB2 structure) and C (ROCS and ECFP-4 searches using carazolol as a reference). Overcoming Scoring Deficiencies with Protein—Ligand Interaction Fingerprints (IFP). Ranking Gold poses with the Goldscore scoring function yields dramatically lower ROC values and enrichments than ranking with the IFP scoring function, as was also observed in earlier studies (Table 1, Figure 5). The difference in virtual screening performance between the Surflex scoring function and IFP is smaller but still significant, and although Surflex ranks more partial/full agonists at the top of the hit list using the customized ADRB2 structure than when using the original ADRB2 crystal structure, this scoring function also cannot be used to selectively retrieve partial and full agonists (Table 1, Figure 5). Figure 3 illustrates how IFP manages to retrieve inverse agonists and antagonists.
in cases where native scoring functions fail to rank these ligands among top scorers. The docking poses of the inverse agonist S-timolol (Figure 3A, docked with Gold) and the antagonist R,R-butoxamine (Figure 3B, docked with Surflex) selected by IFP describe the same interactions with the receptor binding cavity as S-carazolol does in the ADRB2 crystal structure (Figure 1A and Figure 3A,B), including H-bond (donor—acceptor) and ionic (positive—negative) interactions with D3.32, aromatic face to edge stacking with F6.52, and H-bond interactions (acceptor—donor and donor—acceptor) with N7.39 (Figure 3C). Interestingly, the same docking poses were already selected by Gold (timolol) and Surflex (butoxamine) native scoring functions (Figure 3C). In fact, as for many other ligands, the top-ranked pose according to the Gold/Surflex score has also (one of) the highest IFP similarity value(s). This means that for the ADRB2 ligands included in this study, the scoring functions perform well in terms of binding mode prediction. The Gold (S-timolol) and Surflex (R,R-butoxamine) scores of the ligand poses are, however, relatively poor, resulting in low rankings in the hit lists (Figure 3C). The bad scores of these poses stem primarily from steric rather than polar terms of the Gold and Surflex scoring functions.27,49 Many ligands with high IFP similarity values (>0.65) but low Goldscore ranks (>150) when docked to the ADRB2 crystal structure were assigned unfavorable internal ligand energies in combination with favorable external H-bond energies (i.e., inverse agonists/antagonists timolol, ICI-188,551, CGP-12177, and butoxamine and partial/full agonists procaterol, terbutaline, and colterol; see Supporting Information Table 1 for individual rankings of ADRB2 ligands). This observation was striking for Gold poses but less pronounced for Surflex poses. Only few ligands with high IFP values but low Surflex scores when docked to the ADRB2 crystal structure were assigned a large negative crash score in combination with a large positive polar score (i.e., the antagonist butoxamine and the full agonists procaterol). Interestingly, the IFP scoring function can easily overcome scoring problems associated with ligand conformational energies or small changes in the protein conformation (ligand-induced fit)51 even without fine-tuning the settings of these docking programs. Postprocessing poses by local/full energy refinement of the corresponding complex, as now recommended in the latest Surflex release,52 will partly solve the problem but not entirely unless ensemble docking is performed on a set of previously generated protein conformers.

**Comparison of Ligand-Based and Receptor-Based Screening Methods.** Faster 2D and 3D ligand-based screening approaches are shown here, in agreement with previous reports,38,52 to be very competitive with docking in screening for GPCR ligands. As to be expected, using the inverse agonist S-carazolol as a reference enables the selective retrieval of inverse agonists and antagonists with high ROC values and enrichments over random picking (Table 1, Figure 5).

The previous examples of S-timolol and R,R-butoxamine (Figure 3) show, as to be expected, that dissimilarities in the chemical structure and/or shape from the carazolol reference can result in relatively low ECFP-4 or ROCs rankings. Timolol and carazolol share the same 1-(isopropylamino)-3-aryloxypropan-2-ol scaffold, but marked differences in the aryl moiety (Figure 2) are disadvantageous to a good 2D similarity rank, while both 3D-based methods (ROCS and IFP scoring) still rank timolol among the top scorers (Figure 3C). Alternatively, when only disconnected fragments (maximum common edge subgraph MCE)53 is common to two ligands (e.g., butoxamine and carazolol share an ethanolamine moiety and a phenyl ring but with different linking moieties; Figure 2), ROCS ranking is logically affected while ECFP-4 and IFP scores are not. Hence, the latter methods still rank butoxamine among the top scorers (Figure 3C). The current results, in agreement with a recent comparative evaluation of different virtual screening methods,54 confirm that IFP scoring performs relatively well in various scaffold hopping scenarios.

**Selectively Retrieving Partial and Full Agonists by Docking to a Customized ADRB2 Structure and IFP Scoring.** Docking into the “agonist-customized” structure (Figure 1B) and scoring by IFP appears to be the only approach that selectively gives high ROC values and enrichments for partial/full agonists only (Table 1, Figure 5). Figure 4 illustrates how IFP manages to distinguish partial/full agonists from inverse agonists/antagonists. R,S-Procaterol (Figure 4A, docked with Surflex) and IFP reference R-isoproterenol (Figure 1B and Figure 4A,B) share most of their interactions with the customized ADRB2 structure, thus yielding a high IFP score despite a weak Surflex score and rank (Figure 4C). This scenario occurs for many other partial/full agonists, ranging from closely related catecholamines to quinolinones (like procaterol) or the tetrhydroisoquinoline-diol TMQ (Figure 2), (see Table 1, Figure 5, and Supporting Information Table 1 for individual rankings of ADRB2 ligands). The antagonist S-carvedilol (Figure 4B, docked with Gold), on the other hand, has a high Goldscore because it makes a high number of contacts with the receptor binding pocket (high external VdW and H-bond contributions). Like the R-isoproterenol reference, carvedilol interacts with the three important serine residues on TM5 (S5.42, S5.43, and S5.46) but through a set of different interactions (hydrophobic contacts instead of H-bonds, Figure 4C). While not affecting the Goldscore (carvedilol is ranked first), changing the type of molecular interactions to the receptor yields a relatively low IFP score and ranking (Figure 4C), as was also determined for almost all other inverse agonists/antagonists docked in the customized ADRB2 structure (Table 1). Ligand-centric methods did not exhibit any significant selectivity for partial/full agonists when using the full agonist R-isoproterenol as a reference, ROC values being equally excellent for inverse agonists/antagonists and partial/full agonists, although higher early enrichments for partial/full agonists are observed at the very low false positive rates of 0.5% and 1% (Table 1, Figure 5, and Supporting Information Table 1).

It is clear that a rigorous comparison of ligand-based and structure-based methods in such a context is extremely difficult, since both methods use different kind of inputs and are biased by different parameters.55,56 Notably, 2D similarity search methods that are strongly influenced by “traditional” chemotypes arising from the historical medicinal chemistry of the corresponding ligands. This effect is particularly true for biogenic amine receptor ligands that still constitute the majority of entries in target-annotated ligand libraries. The present data demonstrate that the usefulness of docking-based methods when compared to ligand-based methods has largely been underestimated in previous reports focusing on GPCR ligands,38,52 mostly because of the poor performance of energy-based scoring functions.57 Although ligand-based methods appear competitive for selecting β2 receptor ligands (with the above-mentioned bias), they are less suited for predicting their functional effects. IFP ranking just looks at conserved interactions with respect to one or multiple references of known binding modes and is thus
Weinstein residue numbering scheme\textsuperscript{58} was used throughout this manuscript for GPCR TM helices.

were then optimized by constrained docking using the "coordinates of polar hydrogen atoms in the protein and the ligand receptors, yielding partial or full agonists.\textsuperscript{44–47} It has been with recent prospective virtual screening campaigns against "agonists/antagonists but also for identifying partial/full agonists.


were rotated to optimize H-bonding were saved and stored in the final set of ligand-free protein coordinates.

computing all atom positions) of the docked ligand from the original ground-state receptor models.

The conformer database was generated using ROCS Search.

The Ballesteros–Weinstein residue numbering scheme\textsuperscript{58} was used throughout this manuscript for GPCR TM helices.

The binding site was defined in Surflex (protomol file) and Gold docking. Fifteen poses were generated for each ligand.

The conformer database was generated using ROCS Search.

The conformer database was generated using ROCS Search.

Automated Docking-Based Virtual Screening. The ligand database was automatically docked into the original ADRB2 crystal structure and the customized ADRB2 structure using standard parameters of Surflex, version 2.1\textsuperscript{17} and "3-times speedup" settings of Gold, version 3.2.\textsuperscript{26} Fifteen poses were generated for each ligand. The binding site was defined in Surflex (protomol file) and Gold (gold.conf file) by providing a list of cavity residues listed in the next paragraph. All docking input files are available as Supporting Information.

Interaction Fingerprint Scoring. S-Carazolol (in the original ADRB2 X-ray structure) and R-isoproterenol (in the customized receptor structure) were used to generate reference interaction fingerprints (IFPs) as previously described.\textsuperscript{37} Seven different interaction types (negatively charged, positively charged, H-bond acceptor, H-bond donor, aromatic face-to-edge, aromatic-face-to-face, and hydrophobic interactions) were used to define the IFP. The cavity used for the IFP analysis consisted of the 30 residues earlier proposed to define a consensus TM binding pocket\textsuperscript{39} plus three additional residues at positions 3.37, 5.47, and 7.40: M1.35, M1.39, I1.42, I1.46, V2.57, V2.58, G2.61, I2.65, W3.28, T3.29, T3.32, V3.33, V3.36, T3.37, T3.40, T4.56, P4.60, Y5.38, A5.39, S5.42, S5.43, S5.46, F5.47, F6.44, W6.48, F6.51, F6.52, N6.55, Y7.35, N7.39, W7.40, Y7.43, N7.45.

Note that for each Gold docking pose, a unique subset of protein coordinates with rotated hydroxyl hydrogen atoms were used to define the IFP. Standard IFP scoring parameters,\textsuperscript{32} and a Tanimoto coefficient (Tc-IFP) measuring IFP similarity with the reference molecule pose (carazolol in the ADRB2 crystal structure (Figure 1A and Figure 3) or isoproterenol in the customized ADRB2 structure (Figure 1B and Figure 4)), were used to filter and rank the docking poses of 13 known ADRB2 antagonists/inverse agonist, 13 known ADRB2 partial/full agonists, and 980 chemically similar decoys (only poses forming an H-bond and ionic interaction with D3.32 and V3.36 are considered). The reference IFP bit strings are available as Supporting Information.

ROCS Search. The conformer database was generated using standard settings OMEGA, version 2.2.1.\textsuperscript{64} and searched with ROCS, version 2.3.1.\textsuperscript{38} using standard settings as well. The conformations of S-carazolol found in the ADRB2 X-ray structure particularly well suited in retrieving new chemotypes, provided that key molecular interactions are present.\textsuperscript{54}

Conclusions

The current study shows that the inactive state of the β2 adrenergic receptor is suitable not only for retrieving inverse agonists/antagonists but also for identifying partial/full agonists with docking-based in silico screening methods. This is in line with recent prospective virtual screening campaigns against "inactive" bovine rhodopsin (bRh)-based receptor models receptors, yielding partial or full agonists.\textsuperscript{44–47} It has been previously argued that the inactive bRh and ADRB2 crystal structures are unsuitable templates for discovering partial or full agonists.\textsuperscript{8–10} The above-reported virtual screening successes as well as the current study suggest the opposite: agonists may bind to an "early activated" state model resembling the inactive state of the receptor. Moreover, this study shows that specific experimentally driven adaptations of the "rigid" inverse agonist bound state can be sufficient to obtain a partial/full agonist-selective receptor structure. The success of this approach heavily depends on preexisting knowledge, notably in the selection of clear experimental anchors that can be used to guide the protein modeling and postprocessing docking poses. The use of a topological scoring function (IFP) to rank docking poses was notably necessary to selectively retrieve partial/full agonists by docking. It, however, opens novel perspectives in structure-based virtual screening for GPCR partial/full agonists from customized ground-state receptor models.

Residue Numbering and Nomenclature. The Ballesteros–Weinstein residue numbering scheme\textsuperscript{58} was used throughout this manuscript for GPCR TM helices.

Protein Coordinates Setup. Hydrogen atoms were added to the ADRB2-carazolol X-ray structure (PDB entry 2rh1) using the Biopolymer module in Sybyl, version 7.3,\textsuperscript{59} and assuming standard protonation states for charged residues. All nonstandard monomers, water, and the T4 lysozyme fragments\textsuperscript{6} were removed. Atomic coordinates of polar hydrogen atoms in the protein and the ligand were then optimized by constrained docking using the " times speedup" settings of the Gold, version 3.2, program,\textsuperscript{26} using a strong similarity constraint (shape overlap to the input mol2 ligand coordinates, constraint weight of 100) to lock the heavy atom positions of the ligand. The root-mean-square deviations (compared on all atom positions) of the docked ligand from the original carazolol structure was 0.4 Å. Polar hydrogens of the protein that were rotated to optimize H-bonding were saved and stored in the final set of ligand-free protein coordinates.

R-Isoproterenol was docked into this structure using "3-times speedup" settings of Gold, version 3.2.\textsuperscript{26} The χ\textsubscript{1} torsion angles of S5.43 (from g+ to g-) and S5.46 (from g- to g+), as well as the hydroxyl rotor of S5.42 (toward the m-hydroxyl oxygen of R-isoproterenol) were manually rotated so that these serine residues faced the catechol group of the top-ranked R-isoproterenol docking pose. This R-isoproterenol—ADRB2 complex was further minimized with AMBER 8 using the AMBER03 force field.\textsuperscript{60} The minimizations were performed by 1000 steps of steepest descent followed by conjugate gradient until the rms gradient of the potential energy was lower than 0.05 kcal/mol. Å. A twin cutoff (12.0, 15.0 Å) was used to calculate nonbonded electrostatic interactions and the nonbonded pair-list was updated every 25 steps. Distance restraints (upper-bound distance), supported by site-directed mutagenesis studies,\textsuperscript{13,14,15} were defined between (1) the m-hydroxyl oxygen atom of R-isoproterenol and the hydroxyl hydrogen of S5.42 (2.25 Å upper-bound distance); (2) the m-hydroxyl hydrogen atom of R-isoproterenol and the hydroxyl oxygen atom of S5.43 (2.25 Å upper-bound distance); (3) the p-hydroxyl hydrogen atom of R-isoproterenol and the hydroxyl oxygen atom of S5.46 (2.25 Å upper-bound distance); (4) the amine nitrogen atom of R-isoproterenol and the oxygen atom of the N7.39 side chain (3.25 Å upper-bound distance). This minimized complex was refined by a second AMBER energy minimization without distance restraints. Isoproterenol force-field parameters were derived using the Antechamber program,\textsuperscript{60} and partial charges for the substrates were derived using the AM1-BCC procedure in Antechamber. Both protein input coordinates (X-ray structure, customized structure) are available as Supporting Information.

Ligand Database Preparation. For the evaluation of the virtual screening performance of ligand-based (ECFP-4\textsuperscript{39} and ROCS\textsuperscript{58}) and receptor-based (Gold\textsuperscript{26} and Surflex\textsuperscript{27} automated docking) methods, a single database was prepared consisting of 13 known ADRB2 antagonists/inverse agonist, 13 known ADRB2 partial/full agonists (Figure 2), and 980 chemically similar compounds randomly selected from our in-house collection of commercially available compounds. The ADRB2 ligands were manually selected for their specificity and high affinity and chosen to span the broadest chemical diversity.

In order to avoid biasing virtual screening results, caution was given to select 980 druglike decoys covering similar property ranges (molecular weight, number of rotatable bonds, number of rings, hydrogen bond donor/acceptor counts, at least one positively charged atom) as the true actives but structurally different from any known active (the highest similarity, expressed by the Tanimoto coefficient on SciTecig ECFP-4 circular fingerprint,\textsuperscript{59} of any decoy to any true active is 0.56). ADRB2 ligands were manually sketched in Isis Draw.\textsuperscript{61} Starting from the Isis Draw sketch, 2D sd structures were subsequently ionized using Filter2\textsuperscript{62} and converted into 3D mol2 files with Corina 3.1.\textsuperscript{63} 2D and 3D databases are available as Supporting Information.
and R-isoprotanol in the customized ADRB2 structure were used as query molecules for independent ROCs runs. Hits were ranked by decreasing Combo score (combination of shape Tanimoto and the normalized color score in this optimized overlay). Using ROCs conformatons of R-isoprotanol and S-carazolol conformations (closest to the original conformation) or Corina conformers as a reference did not significantly affect the ROCs results.

**ECFP-4 2D Similarity Search.** Two-dimensional similarity searches were carried out using ECFP-4 (extended connectivity fingerprints) descriptors available in Pipeline Pilot and compared using the Tanimoto coefficient.

**Virtual Screening Analysis.** Virtual screening accuracies were first determined in terms of area under the curve of receiver-operator characteristic (ROC) plots, and its 95% confidence interval was computed with the MedCalc software. **Enrichment** characteristic (ROC) plots, and its 95% confidence interval was computed with the Tanimoto coefficient.

**Supporting Information Available:** Structures of ligand test set (2D sd and 3D mol2 coordinates) and receptor mol2 coordinates (ADRB2 crystal structure and customized structure), Gold (gold.conf and list with cavity residues) and Suflex (protomol) docking input files, reference IFP bit-strings and cavity coordinates used for IFP calculations, a table with rankings of ADRB2 ligands. This material is available free of charge via the Internet at http://pubs.acs.org.

**Acknowledgment.** The authors thank AstraZeneca (Mölndal, Sweden) for financial support. The Centre Informatique National de l’Enseignement Supérieur (CINES, Montpellier, France), the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS, Orsay, France), and the Calculation Centre of the IN2P3/CNRS (Villeurbanne, France) are acknowledged for allocation of computing time.

**References**


**AMBER; University of California: San Francisco, CA 94158-2517; http://amber.scripps.edu.


Structure-Based Virtual Screening for ADRB2 Agonists


Filter, version 2; OpenEye Scientific Software; Santa Fe, NM 87507; www.eyesopen.com.


OMEGA2, version 2.1.1; OpenEye Scientific Software; Santa Fe, NM 87507; www.eyesopen.com.


JM800710X