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[a]) 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.[1] So far structural modeling of GPCRs has been limited to either ab initio models[2] or bovine rhodopsin (bRho)-based[3] homology models. Recently, crystal structures of a second GPCR, the β2 adrenergic receptor (ADRB2), were solved.[4–6] 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.[7] It has been stated, however, that the ground state of ADRB2 and bRho crystal structures are only suitable for discovering inverse agonists and antagonists.[8–10] 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),[5,6] 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 acceptor and donor to the ligand amine nitrogen and hydroxyl oxygen, respectively (Figure 1A).[5] The carbazole heterocycle of the ligand interacts with F6.52 via an edge-to-face π stacking, while the carbazole nitrogen atom donates an H-bond to the hydroxyl oxygen of S5.42.[5] This binding mode is in line with earlier site-directed mutagenesis studies, supporting the involvement of D3.32,[11,12] S5.42,[13] and N7.39[14] in binding of both antagonist and agonists. Both functional[15] and biophysical[16–19] studies on ADRB2 suggest that agonist binding occurs through

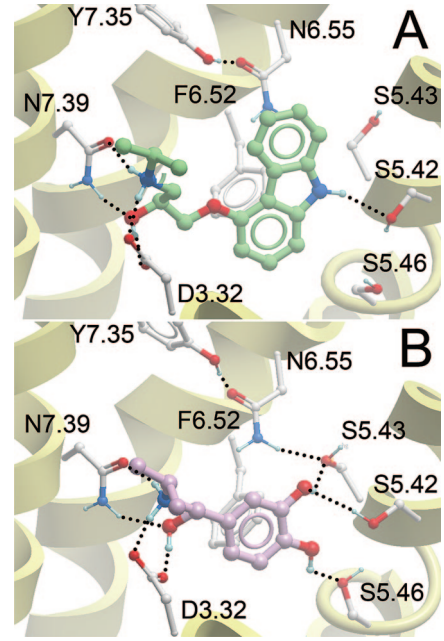

Figure 1. Binding modes of (A) *S*-carazolol (green carbon atoms; see Figure 2 for 2D representation) in the ADRB2 crystal structure and (B) *R*-isoproterenol (magenta carbon atoms; see Figure 2 for 2D representation) in the customized ADRB2 structure. The backbone of transmembrane helices 3, 5, 6, and 7 are represented by yellow ribbons (the top of TM3 is not shown for clarity). Important binding residues are depicted as ball-and-sticks with gray carbon atoms. Oxygen, nitrogen, and hydrogen atoms are colored red, blue, and cyan, respectively. H-bonds described in the text are depicted by black dots.

kinetically distinct steps involving several conformational intermediates.[20] The agonist binding model hypothesis follows at least three steps.[20,21] In the first step, the protonated amine of the agonist forms an ionic/H-bond interaction with D3.32[11] (and N7.39[14]) while the catechol ring stacks with F6.52.[22] A fast conformational change[17] 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)[13,23] in a second step. This triggers a slower[17] conformational change involving a rotamer "toggle switch"[24] and H-bond formation between the β-alcohol of the agonist and N6.55.[17,25]

* To whom correspondence should be addressed. Phone: +33-(0)390-244235. Fax: +33-(0)390-244310. E-mail: didier.rognan@pharma.u-strasbg.fr.

[a] Abbreviations: ADRB2, β2 adrenergic receptor; bRho, bovine rhodopsin; CCR5, C−C chemokine receptor type 5; CNR2, cannabinoid 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).

10.1021/jm800710x CCC: $40.75
Published on Web 08/05/2008

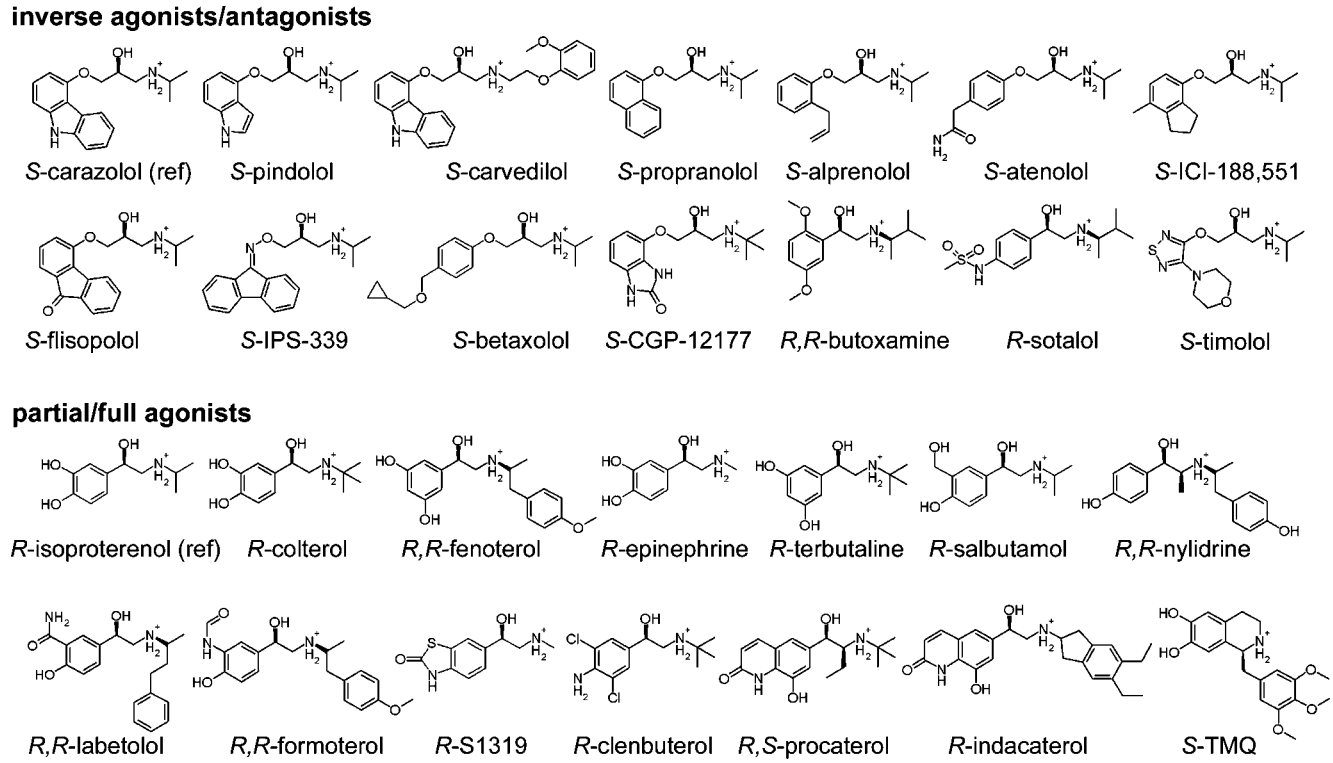

Figure 2. Structures of inverse agonists/antagonists and partial/full agonists of ADRB2. References (ref) *S*-carazolol and *R*-isoproterenol for the two different ligands classes are indicated. The other 13 inverse agonists/antagonists and 13 partial/full agonists are used as test set.

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[26,27] of inverse agonists, neutral antagonists, partial agonists, and full agonists (Figure 2) in the ADRB2 crystal structure[5] 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).[20] 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,[28] 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[13,23] 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[25,29] (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[30] 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 switch[24] or rigid-body shifts of seesaw motions of transmembrane segments[31] probably involves conformational changes like bending[24,32] and/or rigid body movements[31,33] of TM helices which are likely to be ligand-dependent and therefore difficult to predict. Comparison of experimental structures of inactive[3] and (semi)activated[34,35] states of bovine rhodopsin, however, suggest that even these structural changes are smaller than previously proposed on the basis of indirect biochemical methods.[36]

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)[37] and served as templates for three-dimensional ligand-based ROCS queries,[38] while their two-dimensional circular fingerprints were used for ECFP-4 similarity searches[39] 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-

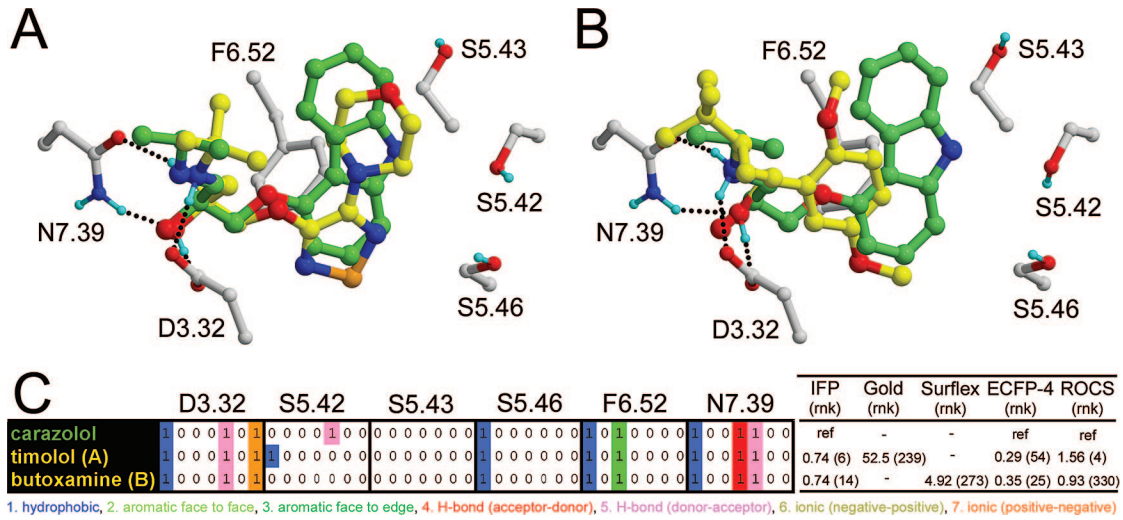

Figure 3. Docking poses selected by IFP of (A) *S*-timolol (yellow carbon atoms, docked with Gold) and (B) *R,R*-butoxamine (yellow carbon atoms, docked with Surflex) in the ADRB2 crystal structure compared to the reference binding mode of *S*-carazolol (green carbon atoms). H-Bonds between the docking 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.

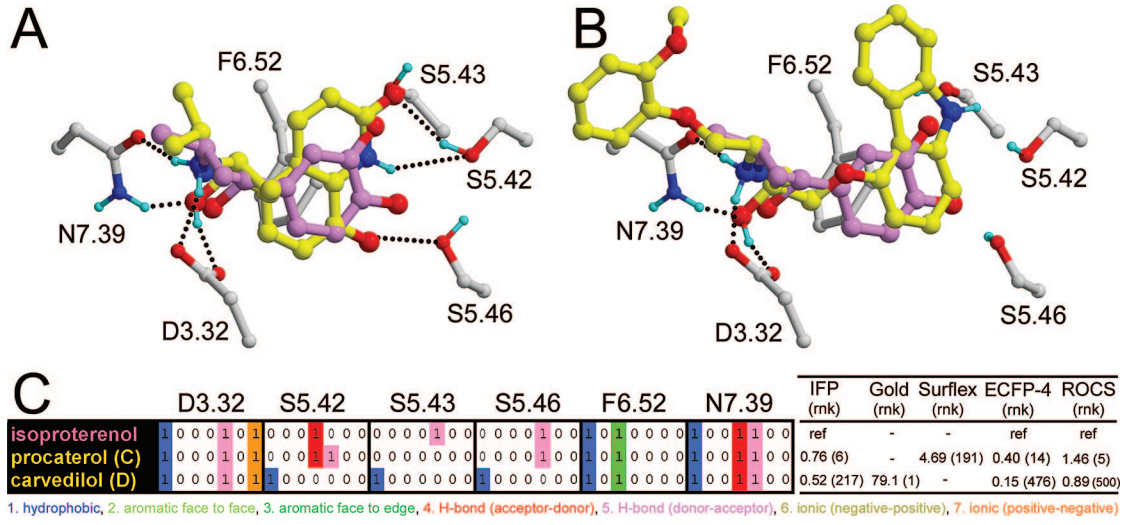

Figure 4. Docking poses of (A) *R,S*-procaterol (yellow carbon atoms, docked with Surflex) and (B) *S*-carvedilol (yellow carbon atoms, docked with Gold) in the customized ADRB2 structure compared to the reference binding mode of *R*-isoproterenol (green carbon atoms). H-Bonds between the docking pose 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.

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 docking[26,27] into the ADRB2 X-ray structure (Figure 1A) and scoring with a protein−ligand interaction fingerprint (IFP) similarity metric[37] (Figure 3) successfully recover known ADRB2 ligands among a large chemical database of chemically similar decoys but do not discriminate full and partial agonists from inverse agonists and neutral antagonists in terms of ROC values (Table 1). The enrichments over random picking at a 2% false positive rate (Table 1, Figure 5) are at least as good and often better than those previously reported in similar structure-based virtual screening experiments against ADRB2.[8,42,43] The current study furthermore shows that the

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

	full/partial agonists[b]						inverse agonists/antagonists[b]					
	ROC		EF[c]				ROC		EF[c]			
method[a]	area[d]	95% CI[e]	0.5%	1%	2%	5%	area[d]	95% CI[e]	0.5%	1%	2%	5%
Gold−Gold (ISP)	0.750	0.722−0.776	31	15	8	3	0.491	0.459−0.522	15	8	4	2
Gold−IFP (ISP)	0.991	0.983−0.996	154	77	38	17	0.702	0.672−0.730	0	0	0	3
Surflex−Surflex (ISP)	0.902	0.882−0.920	46	23	15	11	0.940	0.923−0.954	15	15	23	11
Surflex−IFP (ISP)	0.990	0.982−0.995	169	92	46	18	0.870	0.847−0.890	0	0	4	9
ECFP-4 (ISP)	0.998	0.992−1.000	169	92	50	20	0.932	0.914−0.947	0	8	31	15
ROCS (ISP)	0.984	0.975−0.991	123	85	42	17	0.927	0.909−0.942	0	15	31	15
Gold−Gold (CAZ)	0.585	0.553−0.615	0	0	0	0	0.590	0.558−0.620	0	0	4	3
Gold−IFP (CAZ)	0.940	0.923−0.954	15	8	15	12	0.885	0.863−0.904	108	54	31	14
Surflex−Surflex (CAZ)	0.796	0.769−0.820	0	8	15	11	0.973	0.961−0.982	108	77	38	18
Surflex−IFP (CAZ)	0.855	0.832−0.876	15	8	15	8	0.995	0.989−0.998	123	77	46	20
ECFP-4 (CAZ)	0.666	0.636−0.695	0	8	3	3	0.990	0.982−0.995	138	69	31	20
ROCS (CAZ)	0.685	0.656−0.714	0	0	0	0	0.970	0.957−0.979	154	85	42	17
ideal	1.000		200	100	50	20	1.000		200	100	50	20
random	0.500		1	1	1	1	0.500		1	1	1	1

[a] Gold−Gold (ISP): Gold docking, Goldscore ranking (activated receptor model). Gold−IFP (ISP): Gold docking, IFP ranking (activated receptor model). Surflex−Surflex (ISP): Surflex docking, Surflex ranking (activated receptor model). Surflex−IFP (ISP), Surflex docking, IFP ranking (activated receptor model). ECFP-4 (ISP): ECFP-4 similarity to isoproterenol. ROCS (ISP): ROCS similarity to isoproterenol. Gold−Gold(CAZ): Gold docking, Goldscore ranking (inactivated X-ray structure). Gold−IFP (CAZ): Gold docking, IFP ranking (inactivated X-ray structure). Surflex−Surflex (CAZ): Surflex docking, Surflex ranking (inactivated X-ray structure). Surflex−IFP (CAZ), Surflex docking, IFP ranking (inactivated X-ray structure). ECFP-4 (CAZ): ECFP-4 similarity to carazolol. ROCS (CAZ): ROCS similarity to isoproterenol. [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. [e] 95% confidence interval (CI).

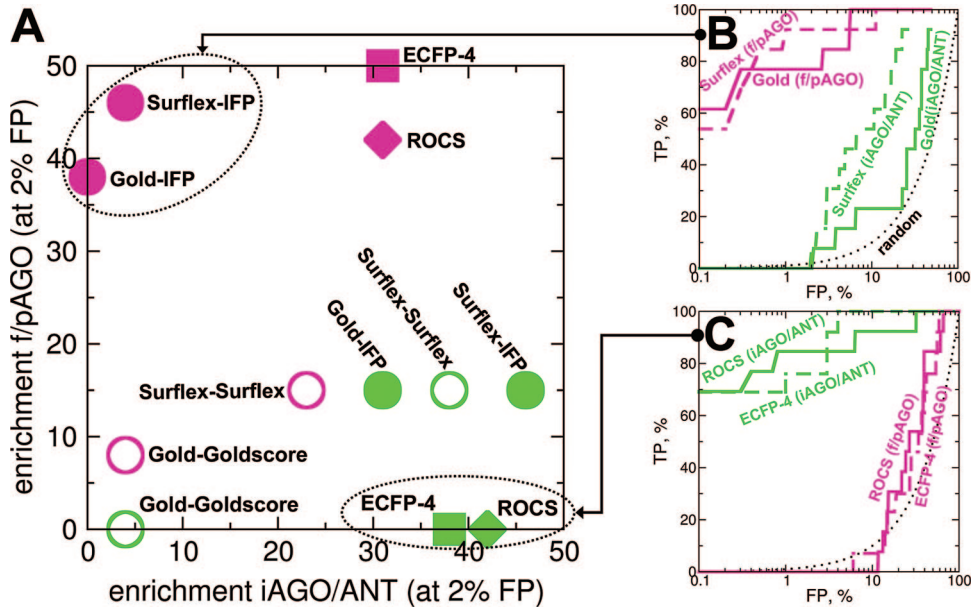

Figure 5. (A) Enrichment in full/partial agonists (f/pAGO) vs enrichment in inverse agonists/antagonists (iAGO/ANT) at a constant 2% false positive rate for different screening approaches using the inverse agonist carazolol (green) or the full agonist isoproterenol (magenta) as a reference. Enrichments 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).

inactive state of the receptor is suitable not only for retrieving inverse agonists/antagonists but also for identifying partial/full agonists with docking-based in silico screening methods (Table 1, Figure 5). This is in line with recent prospective virtual screening campaigns against "inactive" bovine rhodopsin (bRho) based receptor models of CCR5,[44] CNR2,[45] DRD2,[46] and FFAR1[47] receptors, yielding *partial or full agonists*. It has been previously argued that the inactive bRho and ADRB2 crystal structures are unsuitable templates for discovering partial or full agonists.[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.

Overcoming Scoring Deficiencies with Protein−Ligand Interaction Fingerprints (IFP). Ranking Gold poses with the Goldscore scoring function[26] yields dramatically lower ROC values and enrichments than ranking with the IFP scoring function, as was also observed in earlier studies[37,48] (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[27] 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 disfavorable 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 Suflex 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[50] 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 Suflex release,[27] 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 disfavorable 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 tetrahydroisoquinoline-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

particularly well suited in retrieving new chemotypes, provided that key molecular interactions are present.[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 (bRho)-based receptor models receptors, yielding partial or full agonists.[44–47] It has been previously argued that the inactive bRho and ADRB2 crystal structures are unsuitable templates for discovering partial or full agonists.[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.

Computational Methods

Residue Numbering and Nomenclature. The Ballesteros−Weinstein residue numbering scheme[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,[59] and assuming standard protonation states for charged residues. All nonstandard monomers, water, and the T4 lysozyme fragments[6] were removed. Atomic coordinates of polar hydrogen atoms in the protein and the ligand were then optimized by constrained docking using the "7−8 times speedup" settings of the Gold, version 3.2, program,[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 (computed 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.[26] The χ_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.[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,[13,14,23] 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,[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[39] and ROCS[38]) and receptor-based (Gold[26] and Surflex[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 SciTegic ECFP-4 circular fingerprint,[39] of any decoy to any true active is 0.56). ADRB2 ligands were manually sketched in Isis Draw.[61] Starting from the Isis Draw sketch, 2D sd structures were subsequently ionized using Filter2[62] and converted into 3D mol2 files with Corina 3.1.[63] 2D and 3D databases are available as Supporting Information.

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 2.11[27] and "3-times speedup" settings of Gold, version 3.2.[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.[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[30] 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, D3.32, V3.33, V3.36, T3.37, I3.40, 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,[37] 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 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,[64] and searched with ROCS, version 2.3.1,[38] using standard settings as well. The conformations of *S*-carazolol found in the ADRB2 X-ray structure

and *R*-isoproterenol 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).[38] Using ROCS conformations of *R*-isoproterenol and *S*-carazolol conformations (closest to the original conformations) 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[39] 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.[65] Enrichment E in true positives (TP) is reported at different false positive rates FP_x as follows:

$$E = \frac{\mathrm{TP}}{\mathrm{FP}_x}$$

Early enrichments[66] at 0.5%, 1%, 2%, and 5% FP rates were computed for each virtual screen.

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.

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 Surflex (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.

References

(1) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**, *3*, 928–944.
(2) Shacham, S.; Marantz, Y.; Bar-Haim, S.; Kalid, O.; Warshaviak, D. PREDICT modeling and in-silico screening for G-protein coupled receptors. *Proteins* **2004**, *57*, 51–86.
(3) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; et al. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **2000**, *289*, 739–745.
(4) Rasmussen, S. G.; Choi, H. J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **2007**, *450*, 383–387.
(5) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–1265.
(6) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; et al. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **2007**, *318*, 1266–1273.
(7) Topiol, S.; Sabio, M. Use of the X-ray structure of the beta2-adrenergic receptor for drug discovery. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1598–1602.
(8) 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.
(9) Archer, E.; Maigret, B.; Escrieut, C.; Pradayrol, L.; Fourmy, D. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **2003**, *24*, 36–40.
(10) Kobilka, B.; Schertler, G. F. New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol. Sci.* **2008**, *29*, 79–83.
(11) Strader, C. D.; Sigal, I. S.; Candelore, M. R.; Rands, E.; Hill, W. S. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. *J. Biol. Chem.* **1988**, *263*, 10267–10271.
(12) Strader, C. D.; Sigal, I. S.; Register, R. B.; Candelore, M. R.; Rands, E.; et al. Identification of residues required for ligand binding to the beta-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4384–4388.
(13) Liapakis, G.; Ballesteros, J. A.; Papachristou, S.; Chan, W. C.; Chen, X.; et al. The forgotten serine. A critical role for Ser-2035.42 in ligand binding to and activation of the beta 2-adrenergic receptor. *J. Biol. Chem.* **2000**, *275*, 37779–37788.
(14) Suryanarayana, S.; Kobilka, B. K. Amino acid substitutions at position 312 in the seventh hydrophobic segment of the beta 2-adrenergic receptor modify ligand-binding specificity. *Mol. Pharmacol.* **1993**, *44*, 111–114.
(15) Liapakis, G.; Chan, W. C.; Papadokostaki, M.; Javitch, J. A. Synergistic contributions of the functional groups of epinephrine to its affinity and efficacy at the beta2 adrenergic receptor. *Mol. Pharmacol.* **2004**, *65*, 1181–1190.
(16) Swaminath, G.; Deupi, X.; Lee, T. W.; Zhu, W.; Thian, F. S. Probing the beta2 adrenoceptor binding site with catechol reveals differences in binding and activation by agonists and partial agonists. *J. Biol. Chem.* **2005**, *280*, 22165–22171.
(17) Swaminath, G.; Xiang, Y.; Lee, T. W.; Steenhuis, J.; Parnot, C. Sequential binding of agonists to the beta2 adrenoceptor. Kinetic evidence for intermediate conformational states. *J. Biol. Chem.* **2004**, *279*, 686–691.
(18) Ghanouni, P.; Gryczynski, Z.; Steenhuis, J. J.; Lee, T. W.; Farrens, D. L.; et al. Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor. *J. Biol. Chem.* **2001**, *276*, 24433–24436.
(19) Yao, X.; Parnot, C.; Deupi, X.; Ratnala, V. R.; Swaminath, G.; et al. Coupling ligand structure to specific conformational switches in the beta2-adrenoceptor. *Nat. Chem. Biol.* **2006**, *2*, 417–422.
(20) Kobilka, B. K.; Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **2007**, *28*, 397–406.
(21) Pardo, L.; Deupi, X.; Govaerts, C.; Campillo, M. 3-D Structure of G Protein-Coupled Receptors. In *Ligand Design for G Protein-Coupled Receptors*; Wiley-VCH: Weinheim, Germany, 2006; pp 183−203.
(22) Strader, C. D.; Sigal, I. S.; Dixon, R. A. Structural basis of beta-adrenergic receptor function. *FASEB J.* **1989**, *3*, 1825–1832.
(23) Strader, C. D.; Candelore, M. R.; Hill, W. S.; Sigal, I. S.; Dixon, R. A. Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor. *J. Biol. Chem.* **1989**, *264*, 13572–13578.
(24) Shi, L.; Liapakis, G.; Xu, R.; Guarnieri, F.; Ballesteros, J. A. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J. Biol. Chem.* **2002**, *277*, 40989–40996.
(25) Wieland, K.; Zuurmond, H. M.; Krasel, C.; Ijzerman, A. P.; Lohse, M. J. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9276–9281.
(26) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609–623.
(27) Jain, A. N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput-Aided Mol. Des.* **2007**, *21*, 281–306.
(28) *AMBER*; University of California: San Francisco, CA 94158-2517; http://amber.scripps.edu.
(29) Kikkawa, H.; Isogaya, M.; Nagao, T.; Kurose, H. The role of the seventh transmembrane region in high affinity binding of a beta 2-selective agonist TA-2005. *Mol. Pharmacol.* **1998**, *53*, 128–134.
(30) Surgand, J. S.; Rodrigo, J.; Kellenberger, E.; Rognan, D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **2006**, *62*, 509–538.
(31) Schwartz, T. W.; Frimurer, T. M.; Holst, B.; Rosenkilde, M. M.; Elling, C. E. Molecular mechanism of 7TM receptor activation-a global toggle switch model. *Annu. Rev. Pharmacol. Toxicol.* **2006**, *46*, 481–519.
(32) Ballesteros, J. A.; Jensen, A. D.; Liapakis, G.; Rasmussen, S. G.; Shi, L.; et al. Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* **2001**, *276*, 29171–29177.
(33) Farrens, D. L.; Altenbach, C.; Yang, K.; Hubbell, W. L.; Khorana, H. G. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **1996**, *274*, 768–770.
(34) Salom, D.; Lodowski, D. T.; Stenkamp, R. E.; Le Trong, I.; Golczak, M.; et al. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16123–16128.
(35) Nakamichi, H.; Okada, T. Local peptide movement in the photoreaction intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12729–12734.

(36) Hubbell, W. L.; Altenbach, C.; Hubbell, C. M.; Khorana, H. G. Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Adv. Protein Chem.* **2003**, *63*, 243–290.
(37) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195–207.
(38) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.
(39) *Pipeline Pilot*, version 6.1; SciTegic Inc.: San Diego, CA 92123-1365; http://accelrys.com/products/scitegic.
(40) Neubig, R. R. Missing links: mechanisms of protean agonism. *Mol. Pharmacol.* **2007**, *71*, 1200–1202.
(41) Galandrin, S.; Bouvier, M. Distinct signaling profiles of beta1 and beta2 adrenergic receptor ligands toward adenylyl cyclase and mitogen-activated protein kinase reveals the pluridimensionality of efficacy. *Mol. Pharmacol.* **2006**, *70*, 1575–1584.
(42) Bhattacharya, S.; Hall, S. E.; Li, H.; Vaidehi, N. Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation. *Biophys. J.* **2008**, *94*, 2027–2042.
(43) 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.*, DOI: 10.1021/ci800064z.
(44) Kellenberger, E.; Springael, J. Y.; Parmentier, M.; Hachet-Haas, M.; Galzi, J. L.; et al. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J. Med. Chem.* **2007**, *50*, 1294–1303.
(45) Salo, O. M.; Raitio, K. H.; Savinainen, J. R.; Nevalainen, T.; Lahtela-Kakkonen, M.; et al. Virtual screening of novel CB2 ligands using a comparative model of the human cannabinoid CB2 receptor. *J. Med. Chem.* **2005**, *48*, 7166–7171.
(46) Rognan, D. Receptor-Based Rational Design: Virtual Screening. In *Ligand Design for G Protein-Coupled Receptors*; Wiley-VCH: Weinheim, Germany, 2006; pp 183−203.
(47) Tikhonova, I. G.; Sum, C. S.; Neumann, S.; Engel, S.; Raaka, B. M.; et al. Discovery of novel agonists and antagonists of the free fatty acid receptor 1 (FFAR1) using virtual screening. *J. Med. Chem.* **2008**, *51*, 625–633.
(48) de Graaf, C.; Foata, N.; Engkvist, O.; Rognan, D. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins* **2007**, *71*, 599–620.
(49) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(50) Tirado-Rives, J.; Jorgensen, W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein−ligand binding. *J. Med. Chem.* **2006**, *49*, 5880–5884.
(51) Fischer, B.; Merlitz, H.; Wenzel, W. Receptor flexibility for large-scale in silico ligand screens: chances and challenges. *Methods Mol. Biol.* **2008**, *443*, 353–364.
(52) 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.
(53) Raymond, J. W.; Gardiner, E. J.; Willett, P. RASCAL: calculation of graph similarity using a maximum common edge subgraph algorithm. *Comput. J.* **2002**, *45*, 631–644.
(54) Venhorst, J.; Nunez, S.; Terpstra, J. W.; Kruse, C. G. Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. *J. Med. Chem.* **2008**, *51*, 3222−3229.
(55) Cleves, A. E.; Jain, A. N. Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 147–159.
(56) Jain, A. N. Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 201–212.
(57) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., 3rd. Assessing scoring functions for protein−ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.
(58) Ballesteros, J.; Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing pf structure-function relations of G protein-coupled receptors. *Methods Neurosci.* **1995**, *25*, 366–428.
(59) Tripos Associates, Inc., St. Louis, MO. www.tripos.com.
(60) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(61) Elsevier MDL, San Leandro, CA 94577. www.mdl.com.
(62) *Filter*, version 2; OpenEye Scientific Software: Santa Fe, NM 87507; www.eyesopen.com.
(63) Molecular Networks GmbH, D-91052 Erlangen, Germany. www.molecular-networks.
(64) *OMEGA2*, version 2.1.1; OpenEye Scientific Software: Santa Fe, NM 87507; www.eyesopen.com.
(65) MedCalc Software, 9030 Mariakerke, Belgium. http://www.medcalc.be.
(66) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.

JM800710X