

## A Prospective Cross-Screening Study on G-Protein-Coupled Receptors: Lessons Learned in Virtual Compound Library Design

Marijn P. A. Sanders,<sup>†,‡,∞</sup> Luc Roumen,<sup>‡,‡</sup> Eelke van der Horst,<sup>§,‡,×</sup> J. Robert Lane,<sup>§,△</sup> Henry F. Vischer,<sup>‡</sup> Jody van Offenbeek,<sup>‡,||</sup> Henk de Vries,<sup>§</sup> Stefan Verhoeven,<sup>⊥</sup> Ken Y. Chow,<sup>‡</sup> Folkert Verkaar,<sup>‡,||</sup> Margot W. Beukers,<sup>§</sup> Ross McGuire,<sup>⊥,○</sup> Rob Leurs,<sup>‡</sup> Adriaan P. IJzerman,<sup>§</sup> Jacob de Vlieg,<sup>†,⊥,∞</sup> Iwan J. P. de Esch,<sup>‡</sup> Guido J. R. Zaman,<sup>||,●</sup> Jan P. G. Klomp,<sup>\*,⊥,◆</sup> Andreas Bender,<sup>\*,§,◇</sup> and Chris de Graaf<sup>‡,\*,‡</sup>

<sup>†</sup>Computational Drug Discovery Group, Radboud University Nijmegen Medical Centre, Geert Grooteplein, Nijmegen, The Netherlands

<sup>‡</sup>Leiden/Amsterdam Center of Drug Research (LACDR), Division of Medicinal Chemistry, Faculty of Sciences, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

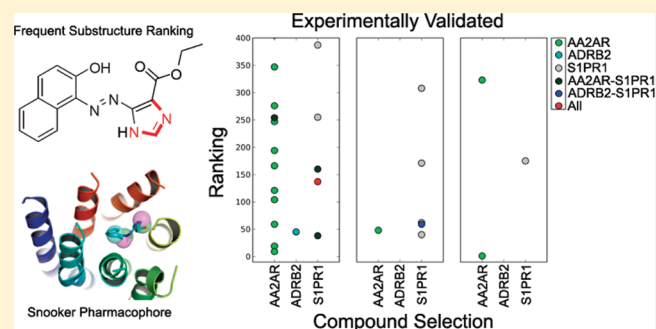
<sup>§</sup>Leiden/Amsterdam Center of Drug Research (LACDR), Division of Medicinal Chemistry, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

<sup>||</sup>Department of Molecular Pharmacology and DMPK, MSD, Molenstraat 110, 5342 CC Oss, The Netherlands

<sup>⊥</sup>Department of Molecular Design and Informatics, MSD, Molenstraat 110, 5342 CC Oss, The Netherlands

**S** Supporting Information

**ABSTRACT:** We present the systematic prospective evaluation of a protein-based and a ligand-based virtual screening platform against a set of three G-protein-coupled receptors (GPCRs): the  $\beta$ -2 adrenoreceptor (ADRB2), the adenosine A<sub>2A</sub> receptor (AA2AR), and the sphingosine 1-phosphate receptor (S1PR1). Novel bioactive compounds were identified using a consensus scoring procedure combining ligand-based (frequent substructure ranking) and structure-based (Snooker) tools, and all 900 selected compounds were screened against all three receptors. A striking number of ligands showed affinity/activity for GPCRs other than the intended target, which could be partly attributed to the fuzziness and overlap of protein-based pharmacophore models. Surprisingly, the phosphodiesterase 5 (PDE5) inhibitor sildenafil was found to possess submicromolar affinity for AA2AR. Overall, this is one of the first published prospective chemogenomics studies that demonstrate the identification of novel cross-pharmacology between unrelated protein targets. The lessons learned from this study can be used to guide future virtual ligand design efforts.

**■ INTRODUCTION**

Chemogenomics is a relatively novel research area aimed at systematically studying the biological effect of a large number of small molecules (ligands) on a large number of macromolecular targets (gene products).<sup>1–4</sup> Since experimental measurements of a large number of ligand–target interactions are time-consuming and cost intensive, they are often complemented by high-throughput in silico chemogenomics approaches. These virtual screening methods are typically divided into ligand-based and target-based approaches, and both approaches can be employed to profile either a ligand against a set of proteins or a set of ligands against one specific protein target.<sup>5,6</sup> Both methods have been successfully applied in the drug discovery context<sup>5,6</sup> with the aim to guide the development of a compound with the desired bioactivity profile.<sup>7–10</sup>

What needs to be kept in mind is that the applicability of in silico chemogenomics models depends on the quality and completeness of the training sets that are used for model construction and validation.<sup>11</sup> Bioactivity data of small molecules in particular are often incomplete, since molecules are not usually screened systematically against a large panel of protein targets but often only on a select set comprising the target of interest and a limited number of proteins over which selectivity needs to be achieved.<sup>12</sup> Furthermore, most scientific studies focus on the presentation of “active” molecules and (potentially) “inactive” molecules are often not reported, leading to a selection bias of published bioactivity data.<sup>12</sup> Even in annotated ligand databases such as ChEMBLdb,<sup>13</sup>

Received: February 28, 2012

Published: May 7, 2012

DrugBank,<sup>14</sup> BindingDB,<sup>15–17</sup> PDBeBind,<sup>18,19</sup> MOAD,<sup>20,21</sup> WOMBAT,<sup>22</sup> and GLIDA<sup>23</sup> protein–ligand interaction matrices are incomplete.<sup>12</sup>

Computational methods have, however, been successfully used to fill the gap in experimental ligand–target affinity matrices<sup>24</sup> and to identify new drug–target associations.<sup>5,25</sup> From generalization of this analysis, the high number of active molecules in target-annotated chemical databases even allows for the identification of molecular features that determine binding to specific proteins and protein classes.<sup>26</sup> The principle that similar receptors bind similar ligands<sup>27</sup> is used in ligand-based virtual screening methods that extrapolate from known active compounds<sup>28</sup> and aim to identify structurally diverse compounds having similar bioactivities<sup>29</sup> by use of techniques like substructure mining,<sup>26</sup> molecular fingerprint similarity searches,<sup>30</sup> and ligand-based pharmacophore models.<sup>31</sup> These techniques are generally faster than methods utilizing the structure of the protein target, such as molecular docking<sup>32</sup> and protein structure-based pharmacophore models,<sup>33</sup> and in most cases they can be trained on larger data sets. Structure-based methods on the other hand are more suitable for finding ligands that are structurally novel and offer insight in the atomic details of protein–ligand interactions.

The recent elucidation of GPCR structures enables *in silico* screens based on structure-based approaches for this protein family,<sup>34</sup> which is particularly relevant for the identification of new ligands and the prediction of their binding modes.<sup>35</sup> The research project described herein is hence an effort to utilize, and evaluate, the increasing amount of information on the structural side (such as X-ray structures) as well as the bioactivity database side (such as the ChEMBL database<sup>13</sup>). Accordingly, we in parallel applied two different virtual screening approaches in the current work, the first one being based on the recently elucidated receptor structures and the second one being based on ligand bioactivity information, in order to select potentially bioactive compounds.

More specifically, our work focused on the discovery of novel bioactive compounds against a panel of three pharmacologically relevant receptors, namely, the  $\beta$ -2 adrenoceptor (ADRB2), the adenosine A<sub>2A</sub> receptor (AA2AR), and the sphingosine 1-phosphate receptor (S1PR1). While ADRB2 plays an important role in cardiovascular disorders,<sup>36</sup> asthma,<sup>36</sup> and other pulmonary disease states,<sup>36</sup> AA2AR is involved in coronary disease<sup>36</sup> and Parkinson's disease.<sup>37</sup> The S1PR1 on the other hand is an important target in the treatment of autoimmune diseases<sup>36</sup> and, potentially, cancer.<sup>38</sup> ADRB2 and AA2AR were the first druggable GPCRs for which crystal structures have been obtained<sup>39,40</sup> and are representative of the selected few GPCRs for which protein crystal structures are available. It was only very recently (February 17, 2012), during the preparation of the current manuscript, that the ligand bound crystal structure of S1PR1 was released.<sup>41</sup> Interestingly, the S1PR1 homology model based pharmacophore constructed in the current study (prior to the release of this X-ray structure) contains the most essential protein–ligand interactions observed in the experimentally determined S1PR1 structure (as will be discussed).<sup>41</sup> One particular reason why a structure-based model of S1PR1 was relevant in the context of the current study is that while considerable ligand information is available for ADRB2 and AA2AR, this is not the case for S1PR1. Hence, structure-based confirmation would be of great importance for S1PR1. Because of the diversity and number of ligands available, when going from the AA2AR via the ADRB2 to the S1PR1 receptor, one would

expect the relative importance of the ligand-based model to decrease and that of the structure-based model to increase for identifying novel bioactive compounds. An additional, practical consideration for selecting these receptors was our ability to perform experimental assays for the targets mentioned to prospectively validate our models.

In a prospective manner, we selected 300 compounds per receptor target and tested *all compounds* for bioactivity on *all three targets* in a cross-screening experiment. This approach allowed us to address the primary aim of our study, namely, to validate and compare structure-based and ligand-based screening approaches against complete biological data (i.e., for all ligands against all targets). This would give insights into what degree hits are (also) active against a different receptor than the target they were selected for. The overarching goal was to determine which method is superior in virtual compound screening and also to gauge whether there might exist any hidden relationship (such as shared information, resulting in mutual enrichment of bioactive compounds) between the different models employed.

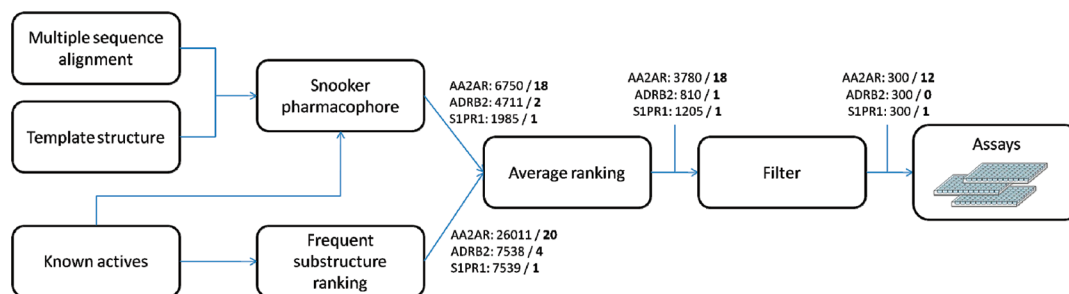
## RESULTS AND DISCUSSION

The aim of the current study was to evaluate the performance of (combined) ligand- and structure-based virtual screening approaches (Figure 1), optimized using retrospective screening simulations (Figures 2 and 3) and applied in a prospective, experimental *all-against-all* validation (Figures 4–8 and Tables 1 and 2).

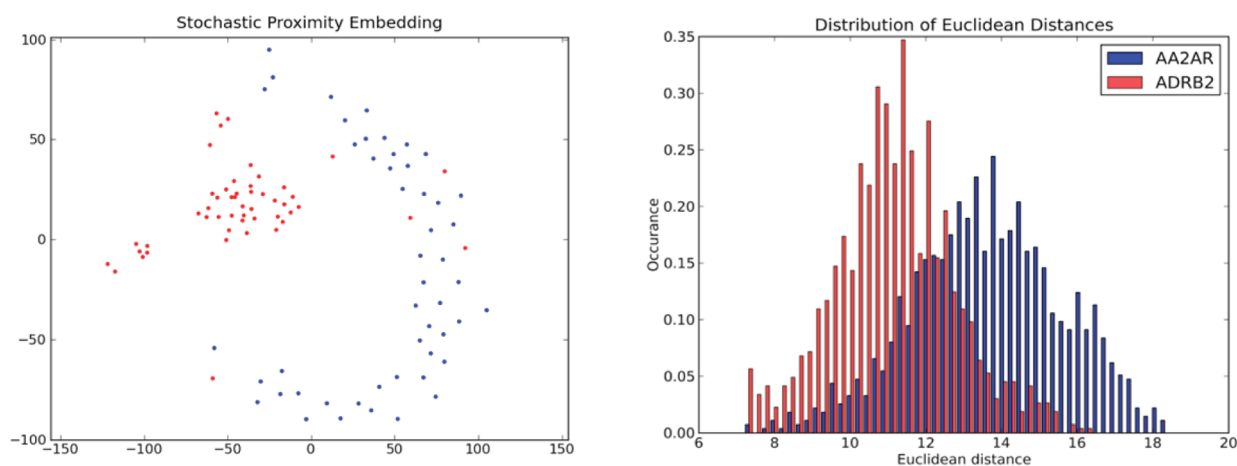
**Chemical Diversity of the Ligand Training Sets.** A large diversity space of active compounds is desired in drug design projects because it allows the selection and synthesis of druglike compounds with good solubility, ADMET properties, and good selectivity toward only the desired targets. In order to understand similarities between the targets investigated in this work in more detail, we first analyzed the diversity of the bioactive ligand sets employed for training. The diversity of the 50 most dissimilar compounds for AA2AR and ADRB2 in BCI fingerprint space is presented in Figure 2. ADRB2 ligands are more similar to each other (average Euclidean distance of  $\sim 11$ ) and form a rather tight cluster, as compared to AA2AR ligands (with an average Euclidean distance of  $\sim 14$ ). This suggests that it is more challenging to find novel ligands for the ADRB2 receptor than for the AA2AR receptor. S1PR1 compounds include primarily compounds that are chemically similar to the endogenous ligand sphingosine 1-phosphate. Diversity selection of S1PR1 compounds is therefore not very informative, and results based on S1PR1 training set compounds are expected to be biased toward these chemical series.

**Pharmacophore Models Match Experimentally Supported Ligand Binding Modes.** In order to generate insight into our structure-based pharmacophore models, we first investigated binding modes of ligands and related this back to previous crystal structures as well as mutation studies.

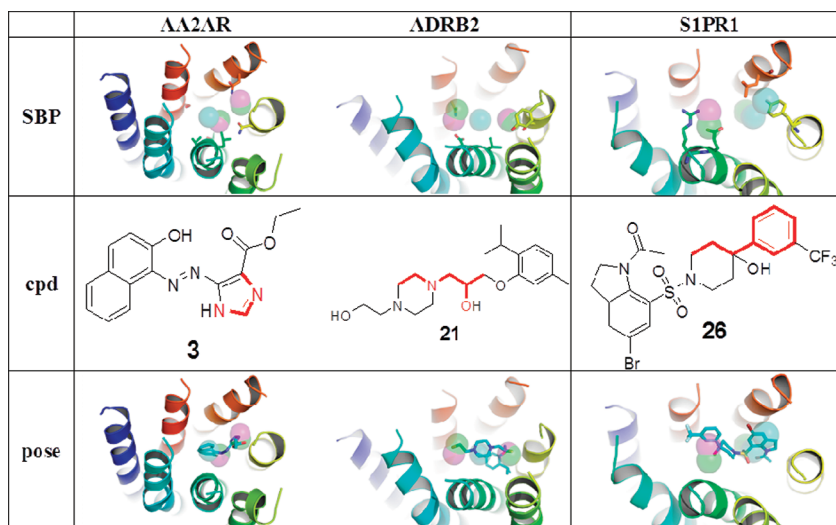
It can be seen (Figure 3) that several polar residues, namely, T88<sup>3,36</sup>, Q89<sup>3,37</sup>, N253<sup>6,55</sup>, E169<sup>45,53</sup>, and S277<sup>7,42</sup>, form H-bond interactions with cocrystallized AA2AR ligands in crystal structures,<sup>40,41</sup> as well as being important in agonist (and xanthine antagonist) binding according to site-directed mutagenesis studies.<sup>42–46</sup> The structure-based pharmacophore model of AA2AR used in our virtual screen includes interaction features derived from three of these polar residues, namely, T88<sup>3,36</sup>, N253<sup>6,55</sup>, and S277<sup>7,42</sup>.



**Figure 1.** Virtual screening flowchart. Numbers indicate the number of compounds that pass the respective structure-based and ligand-based bioactivity models. Bold numbers indicate the number of actives included in the selections. Note that out of the 300 compounds selected for each of the targets, some overlap between the hit lists had to be removed and not all compounds were available as physical samples, leading to slight modifications of the hit lists suggested by the virtual screening methods.



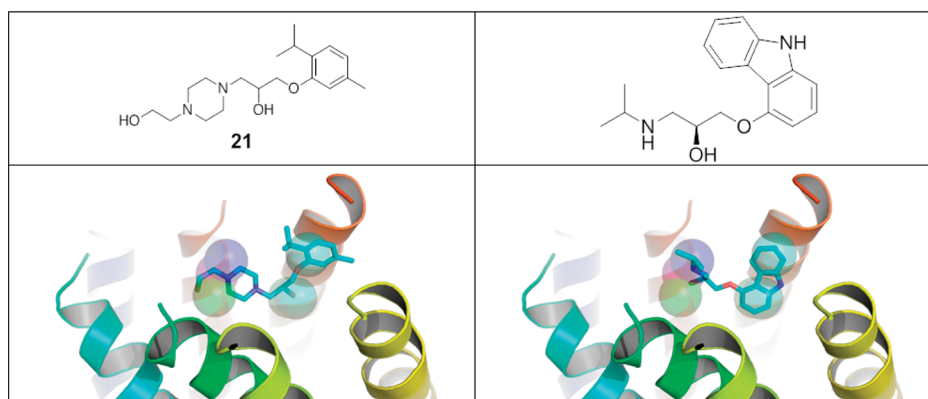
**Figure 2.** Visualization of ligand similarities for the known actives on the adenosine  $A_{2A}$  and the  $\beta$ -2 adrenergic receptor, represented by (left) stochastic proximity embedding and (right) distribution of Euclidean distances. While, on average, ADRB2 ligands (red) are more similar to each other, some atypical ligands can also be found that resemble compounds active on AA2AR (blue).



**Figure 3.** Important interacting residues for the different receptors with the pharmacophores and fitted compounds. AA2AR: related to acceptor features, N181<sup>5.42</sup> and N253<sup>6.55</sup>, T88<sup>3.36</sup> and S277<sup>7.42</sup>; related to donor features, N181<sup>5.42</sup> and N253<sup>6.55</sup>, T88<sup>3.36</sup> and S277<sup>7.42</sup>; hydrophobic features, V84<sup>3.32</sup> and L85<sup>3.33</sup>. ADRB2: acceptors, Y199<sup>5.38</sup> and S203<sup>5.42</sup> and S207<sup>5.46</sup>, N312<sup>7.39</sup>; donors, D113<sup>3.32</sup>, Y199<sup>5.38</sup>, and S203<sup>5.42</sup> and S207<sup>5.46</sup>; hydrophobic feature, V114<sup>3.33</sup>. S1PR1: acceptors, R120<sup>3.28</sup>, E121<sup>3.29</sup>; donor, R120<sup>3.28</sup>; hydrophobic features, Y202<sup>5.39</sup> and L276<sup>6.55</sup>, Y202<sup>5.39</sup> and L276<sup>6.55</sup>. For each compound, the highest scoring substructure that matched the compound is highlight in red.

ADRB2 ligands share an essential positively charged amine, as well as an aromatic ring separated by a distance of about 5 Å (partial and full agonists) to 7 Å (inverse agonists and antagonists). The ADRB2 crystal structure<sup>39</sup> shows protein–ligand

interactions to residues D113<sup>3.32</sup>, S203<sup>5.42</sup>, N312<sup>7.39</sup>, and F290<sup>6.52</sup>, which is in line with site-directed mutagenesis studies.<sup>47–50</sup> Furthermore, mutation of residues S204<sup>5.43</sup>, S207<sup>5.46</sup>, N293<sup>6.55</sup>, and Y308<sup>7.35</sup> were shown to affect partial and full agonist



**Figure 4.** Similarity between the structure based pharmacophore of S1PR1 and ADRB2 ligands, illustrated by compound 21 and carazolol (in the conformation as found in the crystal structure of ADRB2<sup>39</sup>). Both compounds match the five-feature pharmacophore of S1PR1 and also fit in the ADRB2 receptor pharmacophore (not shown).

**Table 1. Number of Molecules in Training Set and Test Set As Used for the Construction of the Substructure-Based Model for AA2AR, ADRB2, and S1PR1<sup>a</sup>**

	training set	test set <sup>b</sup>	total
AA2AR <sup>c</sup>	179	76	255
ADRB2 <sup>c</sup>	301	129	430
S1PR1 <sup>d</sup>	172	74	246

<sup>a</sup>Ligands for all three receptors were retrieved from ChEMBL.<sup>13</sup> <sup>b</sup>The test set is 30% of the total set of selected ligands. <sup>c</sup>Ligands with either  $K_i$ ,  $IC_{50}$ , or  $EC_{50}$  of 10 nM or less. <sup>d</sup>Ligands with either  $K_i$ ,  $IC_{50}$ , or  $EC_{50}$  of 10  $\mu$ M or less.

binding.<sup>51–53</sup> The ADRB2 pharmacophore model contains a hydrophobic contact with V114<sup>3,33</sup> and polar interactions with N293<sup>7,39</sup>, D113<sup>3,32</sup>, S203<sup>5,42</sup>, and S207<sup>5,46</sup> as supported by ligand cocrystallized ADRB2 crystal structures and site-directed mutagenesis studies<sup>47–53</sup> (Figures 3 and 4).

S1PR1 receptor ligands are characterized by the presence of a polar head region, which contains negatively and positively charged groups, and a long hydrophobic tail region.<sup>54</sup> On the basis of these chemical ligand properties, computational modeling and site-directed mutagenesis studies have identified important S1PR1 receptor–ligand interactions, including ionic interactions between the negatively charged phosphate oxygens of sphingosine 1-phosphate (S1P) and R120<sup>3,28</sup> and between the protonated amine of S1P and E121<sup>3,29,55,56</sup>. In addition, L276<sup>6,55</sup> has been identified as a residue that contributes to ligand selectivity for S1PR1 over S1PR3 (containing a leucine residue at this position).<sup>57</sup> The structure-based pharmacophore model for S1PR1 indeed includes features for R120<sup>3,28</sup>, E121<sup>3,29</sup>, and L276<sup>6,55</sup>, as supported by site-directed mutagenesis studies<sup>55,56</sup> (Figures 3 and 4). Interestingly these essential S1PR1–ligand interaction features are indeed observed in the recently solved S1PR1 crystal structure (released during the preparation of this manuscript),<sup>41</sup> corroborating our S1PR1 homology model<sup>58</sup> and the structure-based pharmacophore model constructed in the current study.

**Retrospective Virtual Screening Validation.** In order to assess the likelihood of prospective performance of our structure-based and ligand-based virtual screening protocols, first a retrospective validation was performed, using the 50 diverse actives for AA2AR and ADRB2 vs the 50 308 assumed inactive compounds of the compound library. Here significant early enrichment was observed, as shown in Figure 5.

Retrospective enrichment for AA2AR is poor for the structure-based method, probably (i) because of the fact that a large portion of AA2AR ligands bind in the extracellular domain that is not represented in Snooker pharmacophores and (ii) because it is likely that the A2A adenosine receptor possesses no conserved ligand binding site.<sup>40</sup> The ligand-based method on the other hand exhibits excellent performance for this receptor; however, it might be biased toward known chemistry in our retrospective virtual screening studies. A combination of both methods is hence likely to result in a reduced number of identified actives when compared to the substructure-based method alone; however, it would be assumed to show more novelty among identified active compounds (as well as potentially a different set of ligand–protein interactions), compared to the training set molecules.

In contrast to AA2AR ligands, the ADRB2 ligands bind largely within the transmembrane (TM) domain and share a common binding mode within a buried pocket. Therefore, structure-based searches perform better on ADRB2 than on AA2AR (Figure 5). They not only are able to capture the actives faster but also retrieve a higher percentage of actives. In this particular case, a combination of structure-based and ligand-based methods for this receptor outperforms both the individual ligand-based and structure-based virtual screening methods (Figure 5).

**Prospective Cross-Screening.** For each of the three bioactivity models a total of 300 compounds were selected. Duplicates were removed, and in the case of unavailability of the selected compound it was replaced with the next compound in the list. The final selection of compounds (including cross-receptor overlap) is visualized in Figure 6 (and Supporting Information Figures 5–7) for all of the models. Figure 6 shows that the ADRB2 model and the S1PR1 models select the same compounds more often than other pairs of activity models. Overlap in “hit selection space” can indeed have implications on the results of prospective all-against-all virtual screening studies, as described in more detail below.

Using our combined ligand- and structure-based virtual screening approach, we have successfully identified 18 AA2AR, 6 ADRB2, and 3 S1PR1 ligands (Figures 7 and 8). ADRB2 and AA2AR were tested for ligand displacement, detecting both agonists and antagonists that bind to the orthosteric binding site. Both protein-based pharmacophore models (derived from orthosteric ligand binding sites of ADRB2 and AA2AR; see Figure 3) and ligand-based models (based on substructures of



Table 2. AA2AR and ADRB2 Receptor Binding Affinities and S1PR1 Potencies of Validated Hits Selected by a Combined Ligand-Based (Frequent Substructure Ranking<sup>59</sup>) and Structure-Based (Snooker<sup>60</sup>) Virtual Screening Protocol (Figure 1)<sup>4</sup>

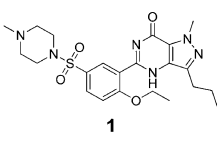
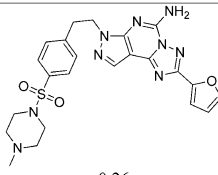
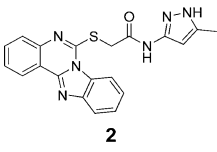
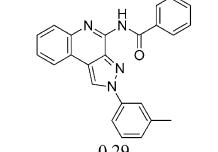
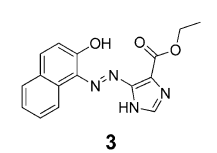
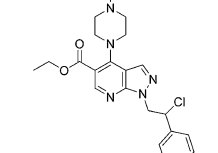
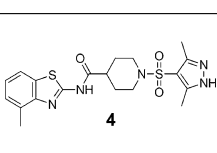
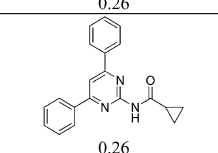
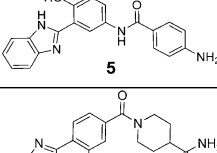
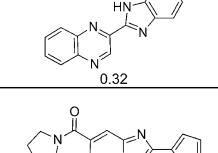
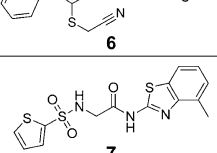
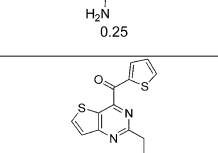
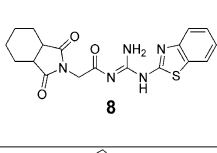
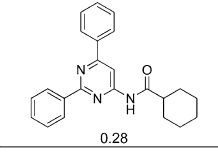
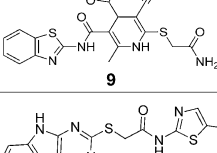
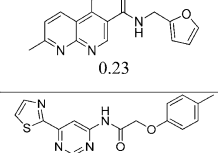
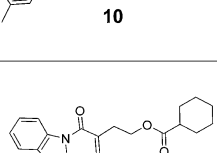
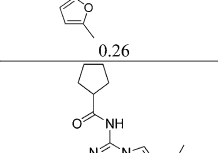
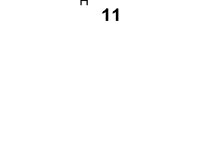
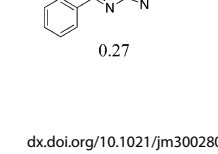


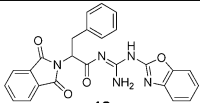
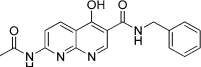
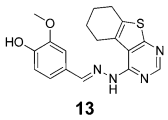
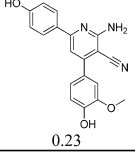
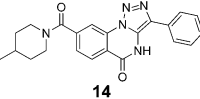
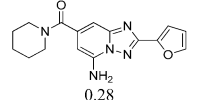
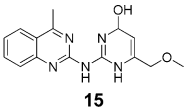
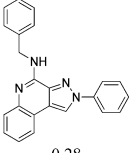
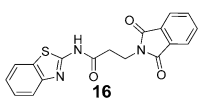
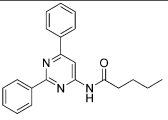
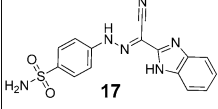
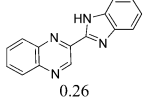
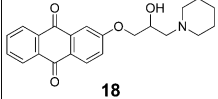
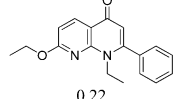
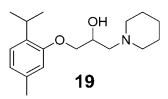
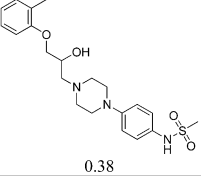
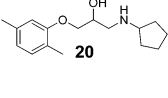
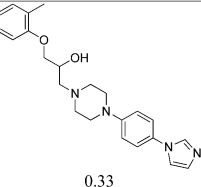
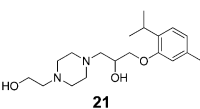
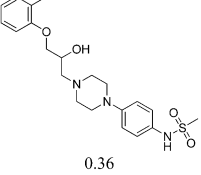
Compound	Rank AA2AR <sup>a</sup>	Rank ADRB2 <sup>b</sup>	Rank S1PR1 <sup>c</sup>	Receptor	Affinity (pKi) <sup>d,e,f</sup> / Potency (pEC <sub>50</sub> ) <sup>f</sup>	Closest Chembl <sup>g</sup>
 <b>1</b>	2285	-	160	AA2AR	6.7 <sup>a</sup>	 0.26
 <b>2</b>	347	-	-	AA2AR	6.2 <sup>a</sup>	 0.29
 <b>3</b>	247	-	-	AA2AR	5.9 <sup>a</sup>	 0.26
 <b>4</b>	254	-	1110	AA2AR	6.0 <sup>a</sup>	 0.26
 <b>5</b>	405	-	-	AA2AR	5.8 <sup>a</sup>	 0.32
 <b>6</b>	-	-	255	AA2AR	6.0 <sup>a</sup>	 0.25
 <b>7</b>	121	-	-	AA2AR	5.8 <sup>a</sup>	 0.27
 <b>8</b>	59	-	-	AA2AR	5.3 <sup>a</sup>	 0.28
 <b>9</b>	19	-	-	AA2AR	5.2 <sup>a</sup>	 0.23
 <b>10</b>	9	-	-	AA2AR	5.4 <sup>a</sup>	 0.26
 <b>11</b>	-	-	387	AA2AR	5.6 <sup>a</sup>	 0.27

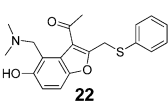
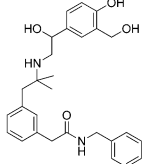
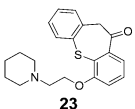
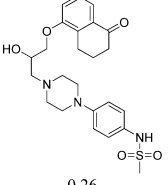
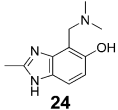
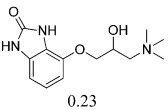
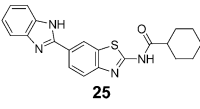
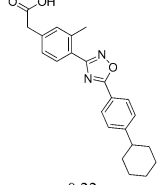
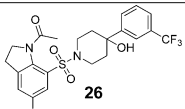
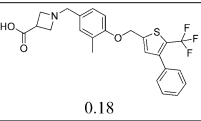
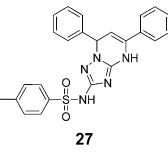
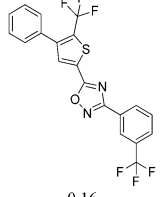
Table 2. continued

Compound	Rank AA2AR <sup>a</sup>	Rank ADRB2 <sup>b</sup>	Rank S1PR1 <sup>c</sup>	Receptor	Affinity (pKi) <sup>d,e</sup> / Potency (pEC <sub>50</sub> ) <sup>f</sup>	Closest Chembl <sup>g</sup>
 <b>12</b>	657	415	137	AA2AR	5.4 <sup>a</sup>	 0.24
 <b>13</b>	166	-	-	AA2AR	5.3 <sup>a</sup>	 0.23
 <b>14</b>	3636	-	38	AA2AR	5.6 <sup>a</sup>	 0.28
 <b>15</b>	194	-	-	AA2AR	6.2 <sup>a</sup>	 0.28
 <b>16</b>	104	-	-	AA2AR	5.3 <sup>a</sup>	 0.27
 <b>17</b>	276	-	-	AA2AR	5.3 <sup>a</sup>	 0.26
 <b>18</b>	-	45	-	AA2AR	5.3 <sup>a</sup>	 0.22
 <b>19</b>	-	-	40	ADRB2	6.1 <sup>b</sup>	 0.38
 <b>20</b>	-	-	171	ADRB2	5.7 <sup>b</sup>	 0.33
 <b>21</b>	-	624	59	ADRB2	4.4 <sup>b</sup>	 0.36

orthosteric ADRB2 and AA2AR ligands) are constructed based on information on the orthosteric ADRB2 and AA2AR binding cavities. S1PR1 was tested in a functional assay in agonistic format, and experimentally confirmed S1PR1 hits might in principle also bind to binding sites other than the cavity included in the S1PR1-based pharmacophore model (Figure 3). Many of the validated hits are chemically dissimilar to any known ADRB2,

AA2AR, or S1PR1 ligand (ECFP-4 Tanimoto similarity below 0.4 (loose cutoff)<sup>61</sup> or even below 0.26 (strict cutoff);<sup>62</sup> see Table 2). Structures of the new ligands are visualized in Table 2, and plots of compound ranks versus receptor activity are displayed in Figure 7, also distinguishing between the in silico models used to select each bioactive compound identified in this study. As can be seen in Figure 7,

Table 2. continued

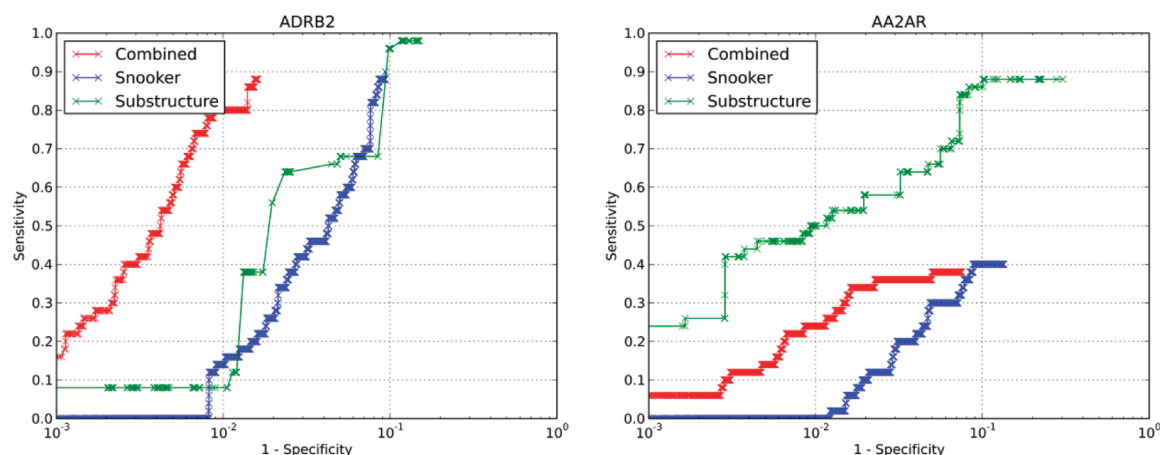
Compound	Rank AA2AR <sup>a</sup>	Rank ADRB2 <sup>b</sup>	Rank S1PR1 <sup>c</sup>	Receptor	Affinity (pK <sub>i</sub> ) <sup>d,e,f</sup> / Potency (pEC <sub>50</sub> ) <sup>f</sup>	Closest ChEMBL <sup>g</sup>
 22	-	-	62	ADRB2	4.4 <sup>b</sup>	 0.20
 23	-	-	308	ADRB2	3.1 <sup>b</sup>	 0.26
 24	48	-	-	ADRB2	4.6 <sup>b</sup>	 0.23
 25	1	-	-	S1PR1	4.7 <sup>c</sup>	 0.22
 26	-	-	175	S1PR1	4.5 <sup>c</sup>	 0.18
 27	323	-	-	S1PR1	4.3 <sup>c</sup>	 0.16

<sup>a</sup>Average rank based on AA2AR proteins-based pharmacophore and frequent AA2AR ligand substructure rankings (Figure 1). <sup>b</sup>Average rank based on ADRB2 protein-based pharmacophore and frequent ADRB2 ligand substructure rankings (Figure 1). <sup>c</sup>Average rank based on S1PR1 protein-based pharmacophore and frequent S1PR1 ligand substructure rankings (Figure 1). <sup>d</sup>pK<sub>i</sub> values are calculated by displacement of [<sup>3</sup>H]ZM241385 binding on membranes of HEK293T cells stably expressing the AA2AR (*n* = 3, each performed in duplicate). <sup>e</sup>pK<sub>i</sub> values are calculated by displacement of [<sup>3</sup>H]dihydroalprenolol binding on membranes of HEK293T cells transiently expressing the ADRB2 (*n* = 3, each performed in triplicate). <sup>f</sup>pEC<sub>50</sub> values are calculated by induction of β-arrestin2 recruitment in CHO-K1 S1PR1 β-arrestin EFC cells (*n* = 1, each performed in triplicate). <sup>g</sup>ECFP-4 circular fingerprint Tanimoto similarity to closest known AA2AR/ADRB2/S1PR1 active in ChEMBLdb. A similarity higher than 0.40 is considered significant.<sup>61</sup> <sup>h</sup>Closest chemical similarity to any known AA2AR/ADRB2/S1PR1 ligand is given for each validated hit.

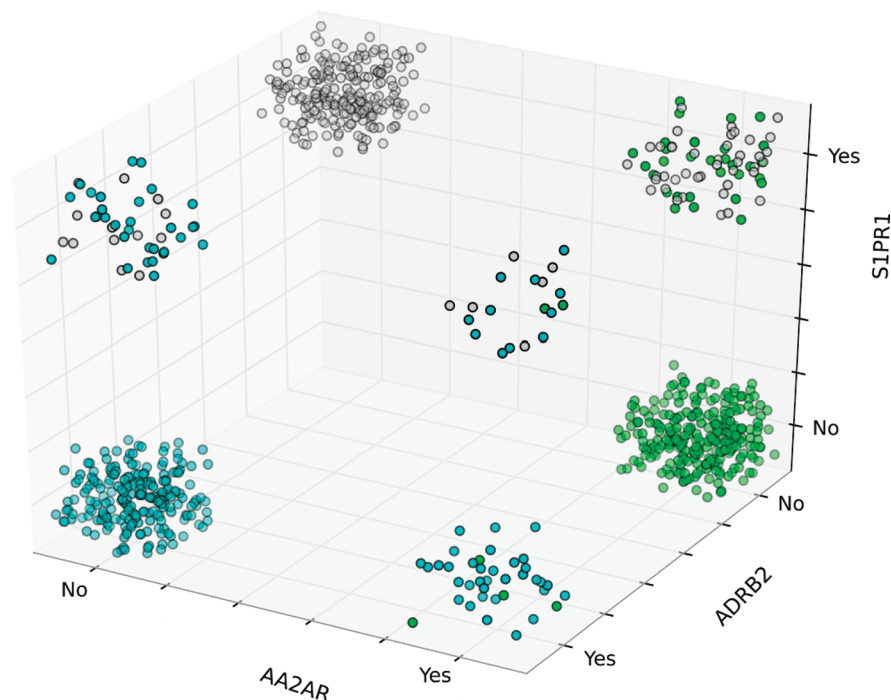
AA2AR showed highest overall hit rates, which agrees with our initial hypothesis. However, a significant number of confirmed AA2AR hits also came from the S1PR1 model, which was rather surprising to us and which is discussed in more detail below. Experimentally determined radioligand displacement (AA2AR and ADRB2) and dose–response curves (S1PR1) are shown in Figure 8.

The virtual screening procedure was very successful for AA2AR and selected 12 out of the 18 active compounds for this receptor directly (including compound 3, for which the proposed binding mode in AA2AR is depicted in Figure 3). Another three compounds were contained in the virtual AA2AR hit list; however, they were actually selected for testing because they had a better rank in the virtual screening for

another receptor (namely, S1PR1). Surprisingly, one of those three compounds is the phosphodiesterase 5 (PDE5) inhibitor sildenafil (compound 1) that has submicromolar affinity for the AA2AR receptor (Table 2, Figure 2). This is in line with previous *in vivo* studies suggesting that spinal adenosine receptors may play a role in sildenafil-induced antinociception<sup>63</sup> and cardioprotective signaling.<sup>64</sup> Comparison of the ZM241385 bound AA2AR<sup>40</sup> and sildenafil bound PDE5<sup>65</sup> cocrystal structures reveals that both pockets possess similar interaction features for ligand binding. ZM241385 makes hydrogen bonding interactions with the amide side chain of N253<sup>6,55</sup> equivalent to the interaction of sildenafil with Q817 in PDE5. Similarly, the bicyclic rings of both ZM241385 and sildenafil are locked in a hydrophobic clamp between F168<sup>45,52</sup>, I274<sup>7,39</sup>, and



**Figure 5.** Receiver operating curves (ROC) for the ligand-based and structure-based virtual screening methods employed in this work. While the substructure-based method, on average, is able to achieve higher enrichment than the employed structure-based method, the extent of this difference is very much dependent on the target receptor considered, with AA2AR showing significantly higher retrieval of active compounds than ADRB2. Consensus scoring outperforms each individual method in the case of ADRB2.

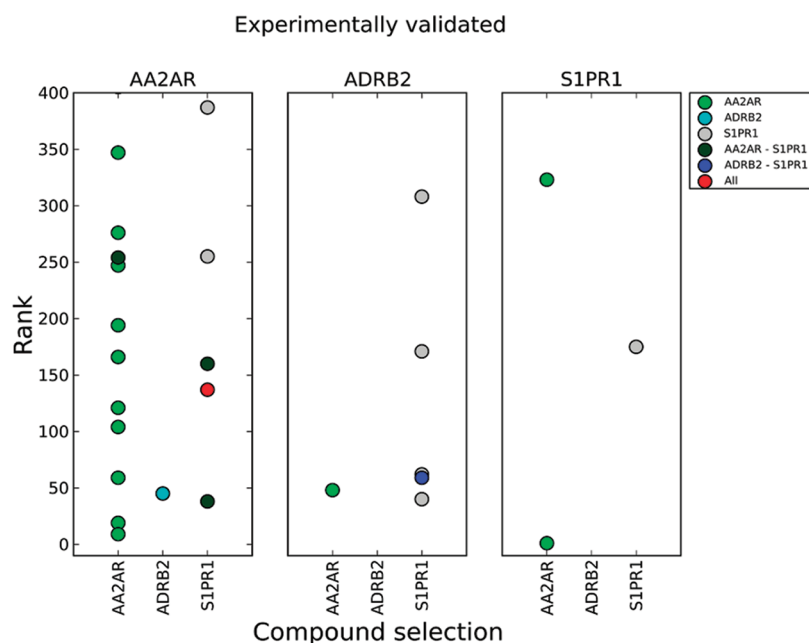


**Figure 6.** Hit selection space based on the combined protein-based pharmacophore and ligand substructure virtual screens (3780 for AA2AR (blue), 810 for ADRB2 (green), and 1205 for S1PR1 (gray)). Most compounds are selected, since they are expected to have activity on only one of the three receptors. Some compounds are, however, expected to have activity on multiple receptors (front-upper, left-upper, right-upper, and front-lower corner), and a small set (front upper corner) is even expected to have activity on all three targets. Compounds are colored according to the receptor in which they receive the best rank.

L249<sup>6,51</sup> (AA2AR) and F820, V782, and F786 (PDE5), respectively (Supporting Information Figure 6). Our unbiased computer-aided discovery of sildenafil as a true AA2AR ligand is therefore an interesting result, indicating that systematic *in silico* chemogenomics studies can be used to identify new interesting cross-pharmacology between unrelated protein targets. In this context the concepts of “GPCR-like compounds” and “privileged scaffolds”<sup>66</sup> is also likely reflected in our selection procedure, given the observation that compounds active on the AA2AR receptor were retrieved when explicitly selecting for its ligands, as well in the ADRB2 and S1PR1 compound selection rounds.

For the ADRB2 receptor 6 confirmed hits could be identified (including compound **21**, for which the proposed binding mode in ADRB2 is shown in Figure 3), which is a much smaller number of hits than in the case of AA2AR. This is in agreement with the restricted chemical diversity of ligands for this receptor, compared to those of the AA2AR receptor. The active compounds that were identified against ADRB2 interestingly resulted from the compound selection for the AA2AR and S1PR1 receptors. The fact that we were not able to identify any ADRB2 ligands from the ADRB2 virtual screening run might of course be related to the fact that all compounds similar to known ADRB2 actives were excluded from the compound set;





**Figure 7.** Novel active compounds found using the respective models for the three receptors in this study (adenosine  $A_{2A}$  receptor,  $\beta$ -2 adrenergic receptor, and sphingosine 1-phosphate receptor). The top of the plot shows the receptor for which the ligands were experimentally found to be active against, while the bottom of the plot shows the model that was used in silico to select each respective compound.

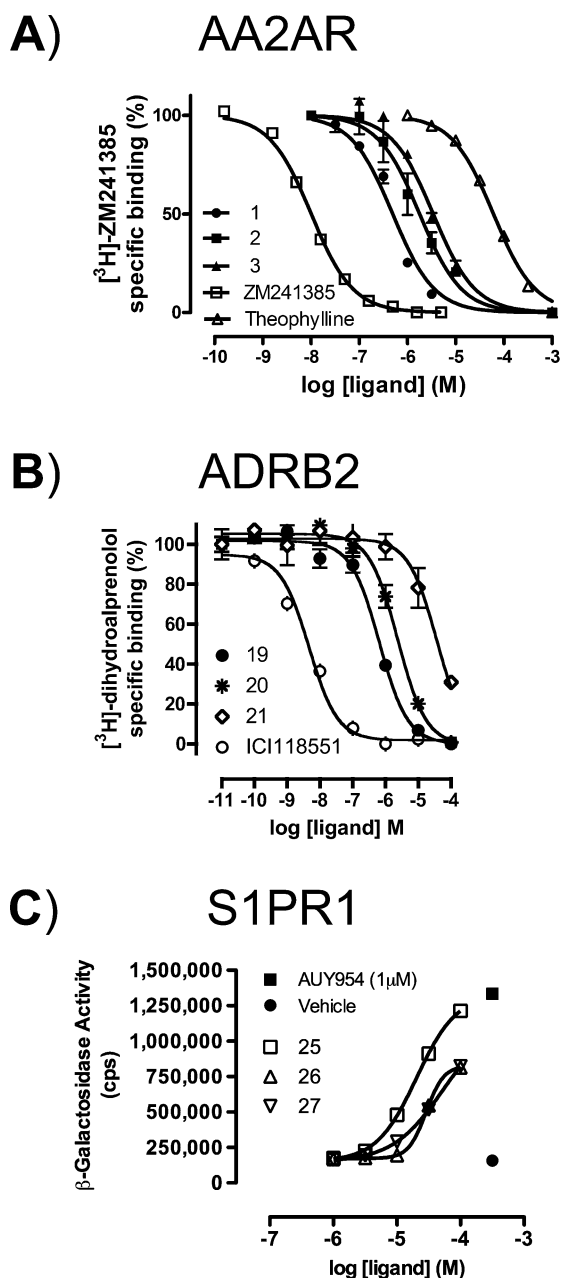
however, the ability to identify compounds active on one receptor using a model for another (albeit related) receptor remains remarkable (as well as consistent throughout the receptors employed in this study). Notably, five of the ADRB2 bioactives were retrieved with the S1PR1 receptor pharmacophore. One possible reason we identified is that important hydrogen bonding and charge interactions are present in both ligand–receptor interactions (which correspond to a similar distance between D113<sup>3,32</sup> and Y199<sup>5,38</sup> in ADRB2, compared to R120<sup>3,28</sup> and E121<sup>3,29</sup> in S1PR1; Figures 3 and 4). In addition to these polar features, the pharmacophore model consists of hydrophobic features that are located in the vicinity of L276<sup>6,55</sup>, which forms the binding site of the aliphatic tail of the ligand in the S1PR1 cocrystal structure.<sup>41</sup> The hydrophobic features are situated at the conjunction of L276<sup>6,55</sup> and Y202<sup>5,39</sup> in the S1PR1 homology model (Figures 3 and 4). However, the side chain of Y202<sup>5,39</sup> is located further from the ligand binding site (closest distance to the ligand is 6 Å) as a result of the different helical conformation of TMS (the reason being that S1PR1 lacks the conserved P<sup>5,50</sup> residue that induces a helical kink in other GPCR crystal structures<sup>41</sup>). Another aromatic residue, Y125<sup>3,33</sup>, recently shown to play a role in S1PR1 ligand binding based on site-directed mutagenesis studies<sup>58</sup> and involved in S1PR1–ligand interactions in the S1PR1 crystal structure<sup>41</sup> is also very close to the location of some of the hydrophobic pharmacophore features in our S1PR1 homology model. While hydrophobic interactions originate in ADRB2 mainly from V114<sup>3,33</sup> and in S1PR1 from Y202<sup>5,39</sup> and L276<sup>6,55</sup>, they are in both cases located at fairly similar positions, further rationalizing the relatively high degree of overlap of the in silico hit lists of ADRB2 and S1PR1. The match of compound 21 in the S1PR1 pharmacophore is displayed in Figure 4, including a fit of the crystal structure conformation of carazolol of the ADRB2 receptor<sup>39</sup> in this pharmacophore. The complementarity of the S1PR1 pharmacophore to known ADRB2 ligands, in combination with the high similarity of the S1PR1 pharmacophore to the ADRB2 pharmacophore, is the best explanation of the ability of

the S1PR1 structure-based pharmacophore to retrieve ligands bioactive against ADRB2. Although compounds 21–24 have relatively low affinity for ADRB2, these molecules offer interesting starting points for hit optimization and give insights into molecular determinants of ADRB2 binding. The low affinity of hit 21 suggests that small polar groups close to the essential basic nitrogen moiety are not favorable for ADRB2 binding. Compound 24 on the other hand has a relatively high ligand efficiency (LE = 0.42) and shares a 2-((dimethylamino)-methyl)phenol scaffold with compound 22 that can be used for future ADRB2 ligand optimization studies.

For S1PR1, three agonists were identified, of which compound 26 originated from the ligands selected explicitly for this receptor. The proposed binding mode in the S1PR1 binding mode of 26 is presented in Figure 3. Overall, two of the three compounds active against S1PR1 were actually selected by the AA2AR model, again a surprising result of our prospectively validated all-against-all virtual screening study. Although the potency of the validated S1PR1 agonists 26–28 is relatively low, so far only very few S1PR1 ligands have been reported that lack a negatively ionizable functional group. The validated hits 26–28 therefore are potential interesting new starting points for S1PR1 ligand optimization studies.

**Lessons Learned.** The following lessons can be learned from our systematic prospective evaluation of a protein-based and a ligand-based virtual screening platform against ADRB2, AA2AR, and S1PR1:

1. *Systematic validation of structure-based and ligand-based virtual screening approaches requires experimental biological activity data for all ligands against all investigated protein targets.* With the increasing amount of experimental information on protein targets and their ligands, complementary computational tools are required that are able to mine this large matrix of experimental data to predict novel protein–ligand interactions. Complete all-against-all data sets containing the biological activities of multiple molecules for a set of protein targets (like the one generated in our study) are



**Figure 8.** (A) Compounds 1–3 selected from the AA2AR screen dose-dependently displacing [<sup>3</sup>H]ZM241385 from AA2AR. (B) Compounds 19–21 selected from the ADRB2 screen dose-dependently displacing [<sup>3</sup>H]dihydroalprenolol from ADRB2. (C) Compounds 25–27 selected from the S1PR1 screen dose-dependently inducing β-arrestin2 recruitment. PathHunter CHO-K1 S1PR1 β-arrestin EFC cells genetically engineered to increase β-galactosidase activity following the recruitment of β-arrestin2 to S1PR1 receptor were stimulated with increasing concentrations of 25, 26, 27 for 2 h before measurement of β-galactosidase activity. A single concentration of AUY954 (1 μM) was taken along to measure maximal β-arrestin2 recruitment.

essential for testing in silico methods in real life screening scenarios and for evaluating their ability to predict, for example, ligand cross-pharmacology and toxicity (i.e., interactions with antitargets).

2. *Retrospective validation of virtual screening methods can be used to assess the limitations of (combining) ligand-based and protein-based discovery of chemically novel ligands.* The rather diverse AA2AR ligand training set

facilitated the construction of a robust ligand-based model that was successfully used to discover a diverse set of novel AA2AR ligands. The lower chemical diversity of the ADRB2 and S1PR1 training sets made the discovery of novel ligands for these receptor more challenging and required a combination of ligand- and protein-based virtual screening models.

3. *Overlap in protein-based pharmacophore models can yield experimentally validated hits that are not (only) active against the receptor they were selected for but (also) against a different member of the panel.* However, by use of protein-based virtual screening methods (complementary to ligand-based virtual screening), protein–ligand binding mode hypotheses can be generated to identify conserved and selective interaction features in different protein targets and to guide hit optimization studies.

## CONCLUSIONS

We performed compound selections for three receptors (AA2AR, ADRB2, and S1PR1) based on a combined structure (pharmacophore) and ligand (substructure) based approach. Retrospective analysis of both individual methods on the AA2AR and ADRB2 receptor indicated that virtual screening accuracy is to a large extent dependent on the chemical diversity among the bioactive compound class. The rather diverse AA2AR ligand training set facilitated the construction of a robust ligand-based model that turned out to be superior to the protein-based model, while a combination of a ligand- and protein-based model gave the best results for ADRB2.

Using a consensus scoring procedure combining the ligand-based and structure-based approaches, we selected 300 compounds for each of the three GPCR targets and experimentally tested all 900 in silico hits against all other receptors in the study. Novel bioactive compounds were identified using a consensus scoring procedure combining ligand-based and structure-based tools (18 actives for AA2AR, 6 actives for ADRB2, and 3 actives for S1PR1). While these variations in hit rates might not be surprising by themselves, what is certainly remarkable is the high number of experimentally validated hits that were not active against the receptor they were selected for but rather against a different member of the panel. Interestingly, one of those ligands is the phosphodiesterase 5 (PDE5) inhibitor sildenafil that was found to have submicromolar affinity for the AA2AR adenosine receptor, demonstrating that our systematic in silico chemogenomics studies can be used to identify new interesting cross-pharmacology between unrelated protein targets.

Ideally, the chemogenomics approach presented here should be extended with more compounds and a larger panel of targets and should preferably be tested using the same assay technology. Such endeavors are already frequently employed for other protein families, such as for kinases, and they are hoped to provide more complete information for both hit identification and optimization than the current state-of-the-art single-target screens are able to do.<sup>67</sup> The lessons learned from this exercise can be used to guide future virtual ligand design efforts.

## EXPERIMENTAL SECTION

The virtual screening setup employed in this work consists of (1) a structure-based protocol (based on inverse pharmacophores), (2) a ligand-based protocol (based on frequent substructure mining), and (3) a merging of the results of the two followed by a prospective validation. This setup is schematically displayed in Figure 1, and its parts will be described in detail in the below sections.

**Structure-Based Pharmacophore Screening.** Structure-based pharmacophores describing the negative image of the pocket, located between the transmembrane helices, were generated for all three GPCRs with the previously described Snooker program.<sup>60</sup> Snooker is a structure-based approach to generate pharmacophore hypotheses for compounds binding to the TM domain of GPCRs. Snooker does not require prior knowledge of ligands but selects residues based on their probability of being involved in ligand binding based on their sequence conservation (expressed as entropy values) inside and outside each GPCR subfamily.<sup>68</sup> Subsequently, protein properties of selected residues are projected to complementary ligand space, and pharmacophore features are generated at positions where protein–ligand interactions are likely.<sup>60</sup> For AA2AR and ADRB2 a consensus model based on eight different templates (PDB codes 1GZM, 1L9H, 2RH1, 2VT4, 3CAP, 3D4S, 3DQB, 3EML) was constructed,<sup>60</sup> whereas for S1PR1 a model was generated based on an experimentally validated agonist-bound homology model.<sup>58</sup> It was only very recently, during the preparation of the current manuscript, that the ligand bound crystal structure of S1PR1 was published<sup>41</sup> showing the same essential receptor–ligand interactions as predicted in our receptor structure-based pharmacophore model (constructed prior to the release of the S1PR1 crystal structure).<sup>41</sup> To gain specificity, directionality was added to the polar pharmacophore features as the average vector between the polar feature center and the C $\alpha$  atom of the residues that constitute the respective features.

**Structure-Based Pharmacophore (SBP) Ligand Training Set.** The structure-based pharmacophores for virtual screening are selected by a training procedure that utilizes ligand information. Known active compounds for the three target proteins investigated, i.e., the  $\beta$ -2 adrenoceptor (ADRB2), the adenosine A<sub>2A</sub> receptor (AA2AR), and the sphingosine 1-phosphate receptor (S1PR1), were extracted for all species from the ChEMBL database (release August 2009) independently of the functional class (agonist, partial agonist, antagonist, inverse agonist) using an activity cutoff on K<sub>i</sub>, IC<sub>50</sub>, or EC<sub>50</sub> of less than 50 nM. To reduce the bias introduced by the deposition of compound series, diverse subsets of 50 compounds were generated for the AA2AR and ADRB2 receptors by exclusion sphere clustering on Tanimoto distances between the BCI fingerprints of the compounds. For the S1PR1 receptor all 43 compounds reported in ChEMBL were used as a training set.

**Pharmacophore Model Generation and Training.** Structure-based pharmacophores for all three targets were subsequently trained with their corresponding SBP ligand training sets of 50 known actives. In this training procedure, the subsets of pharmacophore features that are able to retrieve active compounds were ranked first according to the number of features in the pharmacophore and subsequently according to the number of known actives that they were able to retrieve. To optimize screening outcome, the ranking of pharmacophores was manually adjusted based on crystal structure information and mutagenesis data (see Supporting Information Tables 1–4). Finally, shape restraints were defined by the Tanimoto distance of the compounds matching the ADRB2 and AA2AR pharmacophores to the cocrystallized ligands carazolol and ZM241385, respectively.<sup>39,40</sup> For S1PR1, shape restraints were defined as the Tanimoto distance of compounds matching the S1PR1 pharmacophore to the average pose of known actives from the S1PR1 training set in the same pharmacophore. Tanimoto distance cutoffs were set to those values at which the enrichment of known actives over the 50 308 database compounds was optimal (0.700 for AA2AR, 0.675 for ADRB2, and 0.600 for S1PR1).

**Pharmacophore Screening Settings.** Pharmacophore screening was performed via RDKit using the procedure described by Sanders et al.<sup>60</sup> Projected points were also calculated in RDKit and treated in the same manner as other feature types, with the exception that a heavy-atom–projected-point pair could not be separated and had to match a feature–projected-feature pair. The angular difference between the vectors describing the heavy-atom–projected-point pair and the feature–projected-feature pair was required to lie within 45° of each other.

**Preparation of Prospective Pharmacophore Screening Library.** The compound library used for screening (described in detail below) was prepared as follows for the structure-based virtual screening protocol. Initial three-dimensional conformations were generated with CORINA,<sup>69</sup> and multiple 3D conformations were created with a genetic algorithm, Cyndi,<sup>70</sup> employing a population size of 200 and finally outputting 100 conformations per molecule. Given that the algorithm used here is designed for screening millions of compounds of typical in-house or vendor libraries, no further force field minimization of compounds has been performed.

**Ligand (Frequent Substructure) Based Virtual Screening.** For the ligand-based screening study, the substructure-based screening method of van der Horst et al.<sup>59</sup> was employed. This method performs screening by searching library compounds for the occurrence of substructures characteristic for bioactivity against a particular receptor. These substructures are derived from existing ligands of the target investigated and are selected for their ability to distinguish ligands for this target from ligands for other, secondary proteins.

**Ligand-Based (LB) Training Set.** For the human adenosine A<sub>2A</sub> receptor (AA2AR),  $\beta$ <sub>2</sub> adrenoceptor (ADRB2), and the S1P<sub>1</sub> lysophospholipid receptor (S1PR1), ligand structures and activity data were retrieved from the ChEMBL database,<sup>13</sup> selecting compounds with activity, i.e., a K<sub>i</sub>, IC<sub>50</sub>, or EC<sub>50</sub> of 10 nM or less for the adenosine A<sub>2A</sub> receptor and 10  $\mu$ M or less for the other two receptors. For the A<sub>2A</sub> ligands, we chose a stricter selection, since this resulted in a set of high-affinity ligands that was still of comparable size to the other two ligand sets. Subsequent manual inspection was performed to ensure validity of the LB training set. All source sets were split into a training set and a test set using the “Diverse Molecules” component in PipelinePilot, version 6.1, using a 30% test set in combination with FCFP-4 fingerprints (both training and test set sizes are provided in Table 1). For analysis of the substructures, all training sets were contrasted against a background compound distribution, consisting of 10 000 randomly selected compounds from the druglike subset of the ZINC database (accessed February 12, 2010).<sup>71</sup> Chemical structures were represented as graphs indicating the type of the bonds, e.g., single, double, or aromatic.

**Generation of “Frequently Occurring Substructures”.** Frequent substructure sets were generated using the frequent graph miner Gaston.<sup>72</sup> For each substructure, the number of molecules containing that particular substructure was calculated. The difference between the relative occurrence (fraction) of a substructure in the antagonists set and the background set is the score contribution of that substructure. Substructures were ranked according to the score contribution in descending order, and the 50 highest-ranking substructures were selected for the virtual screening model.

**Substructure-Based Virtual Screening: Ranking of Compounds.** Substructure scores were employed to rank database compounds (which might contain multiple substructures per compound). For this ranking, the score for a compound was calculated as the sum of all score contributions of substructures present in each particular structure.

**Virtual Screening Library.** A diverse subset of the MSD/Organon (Oss, The Netherlands) library consisting of 50 308 compounds was used for virtual screening. In order to characterize the library, the overlap of the available library with the ZINC purchasable compound set (~23 700 000) was determined. The latter consists of compounds with a MW  $\leq$  500 Da from 26 vendors. To determine the overlap between the two libraries, structures were converted into unique hash codes without considering stereochemistry. From the 50 308 MSD compounds, 85% occurred in the 23 691 219 ZINC<sup>71</sup> compound database and 60% occurred in the 6 981 556 CoCoCo<sup>73</sup> compound database (Supporting Information Figure 1). The MSD compounds possess physicochemical properties similar to those within the ZINC and CoCoCo compound databases, placing emphasis on druglikeness<sup>74,75</sup> (Supporting Information Figure 2). Compounds that occurred in the training or test sets were removed from the screening library, as well as compounds that had already been tested against one of the targets. In addition, for the AA2A receptor, compounds with a typical adenosine receptor ligand scaffold, such as xanthines, have been



removed in order to improve the likelihood of discovering structurally novel compounds through this work.

**Final Compound Selection.** After compound selection, we removed compounds very similar to known active structures in the cases of the AA2AR and ADRB2 virtual hit list, namely, those with a Tanimoto coefficient ( $T_c$ ) of  $>0.5$  to known bioactive compounds (as annotated in ChEMBL and employing ECFP<sub>4</sub> fingerprints). This step was performed to ensure that sufficiently novel active compounds would be discovered in this work, which is particularly relevant in the case of the AA2AR receptor for which already chemically diverse sets of ligands are known (Figure 2). A selection of 300 compounds for each receptor was made after determination of the average rank of each compound that matched a pharmacophore and had a positive score in the frequent substructure procedure. Where compounds were not available for testing, the next best compound was selected. The compounds selected by virtual screening were part of a diverse subset of the in-house chemical library of MSD/Organon (Oss, The Netherlands) and originally purchased from available screening collections of eight vendors: Asinex ([www.asinex.com](http://www.asinex.com)), Bionet ([www.keyorganics.co.uk](http://www.keyorganics.co.uk)), Chembridge ([www.chembridge.com](http://www.chembridge.com)), Chemdiv ([www.chemdiv.com](http://www.chemdiv.com)), Contract Chemicals ([www.contract-chemicals.com](http://www.contract-chemicals.com)), Interbioscreen ([www.ibscreen.com](http://www.ibscreen.com)), Orion ([www.orionscientific.in](http://www.orionscientific.in)), and Specs ([www.specs.net](http://www.specs.net)). Purity of compounds was equal to or greater than 95% as verified by LC–MS experiments performed by the vendors.

**Experimental Validation. Adenosine A<sub>2A</sub> Receptor.** HEK293 cells stably expressing the human AA2AR receptor (gift from Dr. Wang, Biogen, Cambridge, MA) were used to determine the affinity of compounds in a radioligand binding assay with [<sup>3</sup>H]-ZM241385 as the radioligand. Membranes containing 40  $\mu$ g of protein were incubated in a total volume of 100  $\mu$ L of Tris-HCl (50 mM, pH 7.4) and [<sup>3</sup>H]-ZM241385 (final concentration of 1.7 nM) for 2 h at 25 °C in a shaking water bath. Nonspecific binding was determined in the presence of 100  $\mu$ M CGS21680. The incubation was terminated by filtration over prewetted Whatman GF/B filters under reduced pressure with a Brandel harvester. Filters were washed three times with ice-cold buffer and placed in scintillation vials. Emulsifier Safe (3.5 mL) was added, and after 2 h radioactivity was counted in a TriCarb 2900TR liquid scintillation counter. Compounds that inhibited binding by  $\geq 50\%$  at 10  $\mu$ M were subject to testing in concentration–response curves.

**$\beta$ -2 Adrenoreceptor.** HEK293T cells were cultured and transiently transfected with 2.5  $\mu$ g of ADRB2-pcDNA3.1+ (obtained from the Missouri S&T cDNA Resource Center) per  $10^6$  cells using 12  $\mu$ g of linear 25 kDa polyethylenimine (Polysciences, Warrington, PA, U.S.) as described previously.<sup>76</sup> Cells were harvested 48 h after transfection, and membrane fractions were prepared as described previously.<sup>76</sup> ADRB2-expressing membranes were incubated at room temperature in 96-well plates in binding buffer (50 mM HEPES, pH 7.4, 1 mM CaCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 100 mM NaCl, and 0.5% (w/v) BSA) with 1 nM [<sup>3</sup>H]-dihydroalprenolol (DHA, 104.4 Ci/mmol from PerkinElmer Life Sciences) and 10  $\mu$ M or increasing concentrations of compounds. After 1 h, incubations were terminated by rapid filtration through Unifilter GF/C plates (PerkinElmer Life Sciences) presoaked in 0.5% polyethylenimine and washed with ice-cold binding buffer supplemented with 500 mM NaCl. Radioactivity was measured using a MicroBeta Trilux (PerkinElmer Life Sciences). Compounds that inhibited binding by  $\geq 50\%$  at 10  $\mu$ M were subject to testing in concentration–response curves. Nonlinear regression analysis of data and calculation of  $K_d$  and  $K_i$  values was performed using GraphPad Prism, version 4, software.

**Sphingosine 1-Phosphate Receptor.** The S1PR1 assay was performed using the PathHunter enzyme fragment complementation  $\beta$ -arrestin recruitment technology as described previously.<sup>77</sup> CHO-K1 S1PR1  $\beta$ -arrestin EFC cells (DiscoverRx, Fremont, CA) were cultured in Dulbecco's modified Eagle's medium F-12 (Invitrogen, Carlsbad, CA), supplemented with 10% heat-inactivated fetal calf serum (Cambrex, Verviers, Belgium), 100 U/mL penicillin, 100  $\mu$ g/mL streptomycin, 300  $\mu$ g/mL hygromycin B, and 800  $\mu$ g/mL Geneticin (Invitrogen). Cells were seeded at a density of 10 000 cells per well of

a 384-well culture plate (PerkinElmer, Boston, MA) in 20  $\mu$ L of OPTI-MEM (Invitrogen). After overnight incubation at 37 °C in a humidified incubator (5% CO<sub>2</sub>, 95% humidity), 4  $\mu$ L of compound dilution was added to cells and the plate was returned to the incubator for 2 h, followed by incubation at room temperature for 1 h. Cells were lysed using 8  $\mu$ L of PathHunter detection reagent (DiscoverRx). Plates were incubated in the dark for 2 h at room temperature before measurement of  $\beta$ -galactosidase activity (chemiluminescence) on an Envision multilabel plate reader (PerkinElmer Life Sciences). Compounds that induced  $\geq 30\%$   $\beta$ -arrestin recruitment compared to the reference compound AU954 were selected and tested in dose–response curves.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Pharmacophore definitions of AA2AR, ADRB2, and S1PR1; lists of ranked pharmacophores according to their expected retrieval of actives for all three targets; overview of property distributions of the 50.308 MSD compounds and the compounds available in the ZINC and CoCoCo database; overlap analysis between the 50.308 MSD compounds and ZINC and CoCoCo based on ECFP<sub>4</sub> fingerprints; plots of compound selections and their respective ranking in the structure-based and ligand-based ranking protocols. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### ✉ Corresponding Author

\*For J.P.G.K.: phone, 0031 243505766; e-mail, [jan.klomp@leadpharma.com](mailto:jan.klomp@leadpharma.com). For A.B.: phone, 0044 1223 762983; e-mail, [Andreas.Bender@cantab.net](mailto:Andreas.Bender@cantab.net). For C.d.G.: phone, 0031 20 5987553; e-mail, [C.de.Graaf@vu.nl](mailto:C.de.Graaf@vu.nl).

### 📍 Present Addresses

<sup>∞</sup>Netherlands eScience Center, Science Park 140, 1098XG, Amsterdam, The Netherlands.

<sup>×</sup>Bioinformatics Laboratory, Clinical Epidemiology, Biostatistics and Bioinformatics (KEBB), Academic Medical Center (AMC), Meibergdreef 9, J1B-207, 1100 DE Amsterdam, The Netherlands.

<sup>△</sup>Drug Discovery Biology, Department of Pharmacy and Pharmaceutical Sciences, Monash University, Parkville, Victoria 3052, Australia.

<sup>○</sup>BioAxis Research BV, Bergse Heihoek 56, 5351 SL, Berghem, The Netherlands.

<sup>●</sup>Netherlands Translational Research Center B.V., P.O. Box 280, 5340 AG Oss, The Netherlands.

<sup>◆</sup>Lead Pharma Medicine, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands.

<sup>◇</sup>Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom.

### ✍ Author Contributions

<sup>#</sup>These authors contributed equally to this work.

### 📝 Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Dutch Top Institute Pharma (TI Pharma), Project No. D1-105, and The Netherlands Organisation for Scientific Research (NWO) through a VENI grant (No. 700.59.408 to C.d.G.).

## ■ ABBREVIATIONS USED

AA2AR, adenosine A<sub>2A</sub> receptor; ADRB2, adrenergic  $\beta$ -2 receptor; BCI, Barnard Chemical Information; ECFP, extended connectivity fingerprint; FCFP, functional class fingerprint; GPCR, G-protein-coupled receptor; LB, ligand based; PDE5, phosphodiesterase 5; S1PR1, sphingosine 1-phosphate receptor; S1P, sphingosine 1-phosphate; T<sub>C</sub>, Tanimoto coefficient; TM, transmembrane (helix); SBP, structure based pharmacophore; VS, virtual screening

## ■ REFERENCES

- (1) Kubinyi, H., Muller, G., Eds. *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective*; Wiley-VCH: Weinheim, Germany, 2004.
- (2) Bredel, M.; Jacoby, E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* **2004**, *5*, 262–275.
- (3) Harris, C. J.; Stevens, A. P. Chemogenomics: structuring the drug discovery process to gene families. *Drug Discovery Today* **2006**, *11*, 880–888.
- (4) Caron, P. R.; Mullican, M. D.; Mashal, R. D.; Wilson, K. P.; Su, M. S.; Murcko, M. A. Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 464–470.
- (5) Rognan, D. Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.* **2007**, *152*, 38–52.
- (6) Rognan, D. Structure-based approaches to target fishing and ligand profiling. *Mol. Inf.* **2010**, *29*, 176–187.
- (7) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijter, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181.
- (8) Badrinarayan, P.; Sastry, G. N. Virtual high-throughput screening in new lead identification. *Comb. Chem. High Throughput Screening* **2011**, *14*, 840–860.
- (9) Seifert, M. H.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 298–307.
- (10) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. Virtual screening in drug discovery—a computational perspective. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.
- (11) Briano, F.; Carrasco, M. C.; Oprea, T. I.; Mestres, J. Cross-pharmacology analysis of G protein-coupled receptors. *Curr. Top. Med. Chem.* **2011**, *11*, 1956–1963.
- (12) Mestres, J.; Gregori-Puigjane, E.; Valverde, S.; Sole, R. V. Data completeness—the Achilles heel of drug–target networks. *Nat. Biotechnol.* **2008**, *26*, 983–984.
- (13) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107.
- (14) 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–D1041.
- (15) Chen, X.; Lin, Y.; Gilson, M. K. The binding database: overview and user’s guide. *Biopolymers* **2001**, *61*, 127–141.
- (16) Chen, X.; Lin, Y.; Liu, M.; Gilson, M. K. The binding database: data management and interface design. *Bioinformatics* **2002**, *18*, 130–139.
- (17) Chen, X.; Liu, M.; Gilson, M. K. BindingDB: a Web-accessible molecular recognition database. *Comb. Chem. High Throughput Screening* **2001**, *4*, 719–725.
- (18) 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.
- (19) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.
- (20) 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.
- (21) 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.
- (22) 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.* **2007**, *21*, 113–119.
- (23) Okuno, Y.; Tamon, A.; Yabuuchi, H.; Nijijima, 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–D912.
- (24) Krejsa, C. M.; Horvath, D.; Rogalski, S. L.; Penzotti, J. E.; Mao, B.; Barbosa, F.; Migeon, J. C. Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 470–480.
- (25) Bender, A.; Young, D. W.; Jenkins, J. L.; Serrano, M.; Mikhailov, D.; Clemons, P. A.; Davies, J. W. Chemogenomic data analysis: prediction of small-molecule targets and the advent of biological fingerprint. *Comb. Chem. High Throughput Screening* **2007**, *10*, 719–731.
- (26) van der Horst, E.; Okuno, Y.; Bender, A.; AP, I. J. Substructure mining of GPCR ligands reveals activity-class specific functional groups in an unbiased manner. *J. Chem. Inf. Model.* **2009**, *49*, 348–360.
- (27) Klabunde, T. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.* **2007**, *152*, 5–7.
- (28) van Westen, G. J. P.; Wegner, J. K.; IJzerman, A. P.; van Vlijmen, H. W. T.; Bender, A. Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. *Med. Chem. Commun.* **2011**, *2*, 16–30.
- (29) Geppert, H.; Vogt, M.; Bajorath, J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–216.
- (30) Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C.; Glick, M.; Davies, J. W. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *J. Chem. Inf. Model.* **2009**, *49*, 108–119.
- (31) 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.
- (32) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153* (Suppl. 1), S7–S26.
- (33) Sanders, M. P. A.; McGuire, R.; Roumen, L.; De Esch, I. J.; de Vlieg, J.; Klomp, J. P. G.; de Graaf, C. From the protein’s perspective: the benefits and challenges of protein structure-based pharmacophore modeling. *MedChemComm* **2011**, *3*, 28–38.
- (34) de Graaf, C.; Rognan, D. Customizing G Protein-coupled receptor models for structure-based virtual screening. *Curr. Pharm. Des.* **2009**, *15*, 4026–4048.
- (35) Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: a Web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38*, W609–W614.
- (36) Rask-Andersen, M.; Almen, M. S.; Schioth, H. B. Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discovery* **2011**, *10*, 579–590.
- (37) Szabo, N.; Kincses, Z. T.; Vecsei, L. Novel therapy in Parkinson’s disease: adenosine A(2A) receptor antagonists. *Expert Opin. Drug Metab. Toxicol.* **2011**, *7*, 441–455.
- (38) Dorsam, R. T.; Gutkind, J. S. G-Protein-coupled receptors and cancer. *Nat. Rev. Cancer* **2007**, *7*, 79–94.



- (39) 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.
- (40) 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.
- (41) Hanson, M. A.; Roth, C. B.; Jo, E.; Griffith, M. T.; Scott, F. L.; Reinhart, G.; Desale, H.; Clemons, B.; Cahalan, S. M.; Schuerer, S. C.; Sanna, M. G.; Han, G. W.; Kuhn, P.; Rosen, H.; Stevens, R. C. Crystal structure of a lipid G protein-coupled receptor. *Science* **2012**, *335*, 851–855.
- (42) 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.
- (43) Jiang, Q.; Van Rhee, A. M.; Kim, J.; Yehle, S.; Wess, J.; Jacobson, K. A. Hydrophilic side chains in the third and seventh transmembrane helical domains of human A2A adenosine receptors are required for ligand recognition. *Mol. Pharmacol.* **1996**, *50*, 512–521.
- (44) Kim, J.; Wess, J.; van Rhee, A. M.; Schoneberg, T.; Jacobson, K. A. Site-directed mutagenesis identifies residues involved in ligand recognition in the human A2a adenosine receptor. *J. Biol. Chem.* **1995**, *270*, 13987–13997.
- (45) Kim, J.; Jiang, Q.; Glashofer, M.; Yehle, S.; Wess, J.; Jacobson, K. A. Glutamate residues in the second extracellular loop of the human A2a adenosine receptor are required for ligand recognition. *Mol. Pharmacol.* **1996**, *49*, 683–691.
- (46) Moro, S.; Gao, Z. G.; Jacobson, K. A.; Spalluto, G. Progress in the pursuit of therapeutic adenosine receptor antagonists. *Med. Res. Rev.* **2006**, *26*, 131–159.
- (47) Strader, C. D.; Sigal, I. S.; Candelore, M. R.; Rands, E.; Hill, W. S.; Dixon, R. A. 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.
- (48) Strader, C. D.; Sigal, I. S.; Register, R. B.; Candelore, M. R.; Rands, E.; Dixon, R. A. Identification of residues required for ligand binding to the beta-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4384–4388.
- (49) Liapakis, G.; Ballesteros, J. A.; Papachristou, S.; Chan, W. C.; Chen, X.; Javitch, J. A. The forgotten serine. A critical role for Ser-203S.42 in ligand binding to and activation of the beta 2-adrenergic receptor. *J. Biol. Chem.* **2000**, *275*, 37779–37788.
- (50) 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.
- (51) 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.
- (52) 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.
- (53) 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.
- (54) Pham, T. C.; Fells, J. I., Sr.; Osborne, D. A.; North, E. J.; Naor, M. M.; Parrill, A. L. Molecular recognition in the sphingosine 1-phosphate receptor family. *J. Mol. Graphics Modell.* **2008**, *26*, 1189–1201.
- (55) Parrill, A. L.; Wang, D.; Bautista, D. L.; Van Brocklyn, J. R.; Lorincz, Z.; Fischer, D. J.; Baker, D. L.; Liliom, K.; Spiegel, S.; Tigyi, G. Identification of Edg1 receptor residues that recognize sphingosine 1-phosphate. *J. Biol. Chem.* **2000**, *275*, 39379–39384.
- (56) Wang, D. A.; Lorincz, Z.; Bautista, D. L.; Liliom, K.; Tigyi, G.; Parrill, A. L. A single amino acid determines lysophospholipid specificity of the S1P1 (EDG1) and LPA1 (EDG2) phospholipid growth factor receptors. *J. Biol. Chem.* **2001**, *276*, 49213–49220.
- (57) 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.
- (58) van Loenen, P. B.; de Graaf, C.; Verzijl, D.; Leurs, R.; Rognan, D.; Peters, S. L.; Alewijnse, A. E. Agonist-dependent effects of mutations in the sphingosine-1-phosphate type 1 receptor. *Eur. J. Pharmacol.* **2011**, *667*, 105–112.
- (59) van der Horst, E.; van der Pijl, R.; Mulder-Krieger, T.; Bender, A.; Ijzerman, A. P. Substructure-based virtual screening for adenosine A2A receptor ligands. *ChemMedChem* **2011**, *6*, 2302–2311.
- (60) Sanders, M. P.; Verhoeven, S.; De Graaf, C.; Roumen, L.; Vroling, B.; Nabuurs, S. B.; de Vlieg, J.; Klomp, J. P. Snooker: a structure-based pharmacophore generation tool applied to class A GPCRs. *J. Chem. Inf. Model.* **2011**, *51*, 2277–2292.
- (61) Wawer, M.; Bajorath, J. Similarity-potency trees: a method to search for SAR information in compound data sets and derive SAR rules. *J. Chem. Inf. Model.* **2010**, *50*, 1395–409.
- (62) Steffen, A.; Kogej, T.; Tyrchan, C.; Engkvist, O. Comparison of molecular fingerprint methods on the basis of biological profile data. *J. Chem. Inf. Model.* **2009**, *49*, 338–347.
- (63) Lee, H. G.; Kim, W. M.; Choi, J. I.; Yoon, M. H. Roles of adenosine receptor subtypes on the antinociceptive effect of sildenafil in rat spinal cord. *Neurosci. Lett.* **2010**, *480*, 182–185.
- (64) Salloum, F. N.; Das, A.; Thomas, C. S.; Yin, C.; Kukreja, R. C. Adenosine A(1) receptor mediates delayed cardioprotective effect of sildenafil in mouse. *J. Mol. Cell. Cardiol.* **2007**, *43*, 545–551.
- (65) Wang, H.; Liu, Y.; Huai, Q.; Cai, J.; Zoraghi, R.; Francis, S. H.; Corbin, J. D.; Robinson, H.; Xin, Z.; Lin, G.; Ke, H. Multiple conformations of phosphodiesterase-5: implications for enzyme function and drug development. *J. Biol. Chem.* **2006**, *281*, 21469–21479.
- (66) Gozalbes, R.; Rolland, C.; Nicolai, E.; Paugam, M.; Coussy, L.; Horvath, D.; Barbosa, F.; Mao, B.; Revah, F.; Froloff, N. QSAR strategy and experimental validation for the development of a GPCR focused library. *QSAR Comb. Sci.* **2005**, *24*, 508–516.
- (67) Goldstein, D. M.; Gray, N. S.; Zarrinkar, P. P. High-throughput kinase profiling as a platform for drug discovery. *Nat. Rev. Drug Discov.* **2008**, *7*, 391–397.
- (68) 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.
- (69) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547.
- (70) 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.
- (71) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
- (72) Nijssen, S.; Kok, J. N. A Quickstart in Frequent Structure Mining Can Make a Difference. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004*; Kohavi, R., Gehrke, J., DuMouchel, W., Ghosh, J., Eds.; ACM: New York, 2004; pp 647–652.
- (73) Del Rio, A.; Barbosa, A. J.; Caporuscio, F.; Mangiatordi, G. F. CoCoCo: a free suite of multiconformational chemical databases for high-throughput virtual screening purposes. *Mol. BioSyst.* **2010**, *6*, 2122–2128.
- (74) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.
- (75) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. A comparison of physicochemical property profiles of

development and marketed oral drugs. *J. Med. Chem.* **2003**, *46*, 1250–1256.

(76) Verzijl, D.; Storelli, S.; Scholten, D. J.; Bosch, L.; Reinhart, T. A.; Streblow, D. N.; Tensen, C. P.; Fitzsimons, C. P.; Zaman, G. J.; Pease, J. E.; de Esch, I. J.; Smit, M. J.; Leurs, R. Noncompetitive antagonism and inverse agonism as mechanism of action of nonpeptidergic antagonists at primate and rodent CXCR3 chemokine receptors. *J. Pharmacol. Exp. Ther.* **2008**, *325*, 544–555.

(77) van der Lee, M. M. C.; Bras, M.; van Koppen, C. J.; Zaman, G. J. R. beta-Arrestin recruitment assay for the identification of agonists of the sphingosine 1-phosphate receptor EDG1. *J. Biomol. Screening* **2008**, *13*, 986–998.