Published in final edited form as:

J Comput Chem. 2010 March ; 31(4): 707–720. doi:10.1002/jcc.21346.

Ligand and Structure-based Models for the Prediction of Ligand-Receptor Affinities and Virtual Screenings: Development and Application to the β_2 -Adrenergic Receptor

Santiago Vilar, Joel Karpiak, and Stefano Costanzi*

Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, DHHS, Bethesda, MD 20892, USA.

Abstract

In this study, we evaluated the applicability of ligand-based and structure-based models to quantitative affinity predictions and virtual screenings for ligands of the β_2 -adrenergic receptor, a G protein-coupled receptor (GPCR). We also devised and evaluated a number of consensus models obtained through partial least square regressions, in order to combine the strengths of the individual components. In all cases, the bioactive conformation of each ligand was derived from molecular docking at the crystal structure of the receptor. We identified the most effective models applicable to the different scenarios, in the presence or in the absence of a training set. For ranking the affinity of closely related analogs when a training set is available, a ligand-based consensus model (LI-CM) seems to be the best choice, while the structure-based MM-GBSA score seems the best alternative in the absence of a training set. For virtual screening purposes, the structure-based MM-GBSA score was found to be the method of choice. Consensus models consistently had performances superior or close to those of the best individual components, and were endowed with a significantly increased robustness. Given multiple models with no a priori knowledge of their predictive capabilities, constructing a consensus model ensures results very close to those that the best model alone would have yielded.

Keywords

docking; QSAR; CoMFA; CoMSIA; statistical analysis; G protein-coupled receptors; β_2 -adrenergic receptor; Virtual screening

Introduction

G protein-coupled receptors (GPCRs) are the largest superfamily of human integral membrane proteins and, being the most common target of pharmacological intervention, they represent a very attractive and lucrative field for pharmaceutical research. Until recently, only one GPCR, bovine rhodopsin, had been crystallized, while homology modeling was widely applied to the study of other receptors.¹ The later disclosure of the crystal structure of the β_2 -adrenergic receptor (β_2 -AR) by Kobilka and coworkers not only

*Correspondence should be addressed to S.C. (stefanoc@mail.nih.gov).

Supplementary Material

A table listing the studied ligands and their binding affinities (Table S1); A table listing the statistical parameters of all models in each of the five training sets (Table S2); A table listing the statistical parameters and the predictions of all models in each of the five test sets (Table S3); A table illustrating the classification capabilities of all models in the LDA (Table S4); A table listing the percentage of true positives retrieved by all models (Table S5). This material is available on the Internet, and it is available free of charge at <http://www.interscience.wiley.com>

confirmed the existence of the conserved seven transmembrane (TM) α -helical core, but also provided the first complex of a GPCR with a non-covalently bound ligand, the partial inverse agonist carazolol.^{2, 3} The availability of a second template also allowed for the first time a direct validation of the applicability of homology modeling and ligand docking to the study of GPCR-ligand complexes.⁴ More recently, the crystal structures of the β_1 -AR with cyanopindolol, the β_2 -AR with timolol, and the adenosine A_{2A} receptor with ZM241385 have also been solved.^{5–7} Of note, all the crystallized ligands are either inverse agonists or antagonists.

In order to study the complexes of these receptors with other ligands, *in silico* docking techniques can be used.^{8, 9} The compounds' affinity for the receptor can then be estimated through scoring functions, which rank the output structures in terms of various structural and energetic factors.¹⁰ Molecular docking, alone or in combination with pharmacophore searches, can also be conveniently used to identify novel ligands through virtual screening.^{11–15} While a ligand's conformation in the binding site might be faithfully reproduced, scoring functions are notorious for not providing an accurate correlation between the score and the actual affinity.^{16, 17} Alternatively, ligand-based predictive models can be built through QSAR (quantitative structure-activity relationships).^{18–22} These, however, are vastly dependent on the compounds included in their training set, and as a result, they are of limited applicability.^{17–24} Recently, we have successfully combined receptor-based modeling and docking based-QSAR into a consensus model to predict the potency of P2Y₁ antagonists.¹⁷ Specifically, a consensus model named LIgand and Structure-based Consensus Model (LIST-CM) was generated by Partial Least Square (PLS) regression based on the individual predictions of several individual models.

Continuing with this line of work, in this study we constructed a number of ligand and structure-based models meant to correlate the chemical structures and affinities of a set of 67 ligands of the β_2 -AR, taking advantage of the recently disclosed crystal structure that provides a more accurate framework than the homology model of the P2Y₁ receptor that we used previously. We constructed individual and consensus models, meant to benefit from the strengths of the individual components (Figure 1). In all cases, the bioactive conformation of the ligands was derived from molecular docking. We divided this set of ligands into 5 training sets with 5 respective test sets, and subjected each to thorough statistical analyses, comparing the experimental and theoretical affinities obtained with each methodology. The models are also examined using randomization tests and linear discriminant analysis (LDA). Furthermore, the models are tested in a simulated virtual screening and their ability to prioritize active compounds is examined by plotting the results in the form of receiver operative characteristics (ROC) curves.

Through this rigorous analysis, we identified the most effective individual models applicable to the different scenarios, in the presence or in the absence of a training set. We also verified that consensus models offer a correlation close to or better than that of the most accurate individual component, with a consequent higher performance stability.

Methodology

Database

A database of 94 agonists and antagonists was constructed from the literature, and binding affinities from competition experiments in the human β_2 -adrenergic receptor were recorded for these compounds. In a few cases, data from a non-human receptor was accepted, under the notion that the amino acid sequence inside the binding cavity was conserved across species. If the pK_i or pK_D values differed between papers, an average value was used. Most of the compounds contain a chiral hydroxyl group corresponding to the β -hydroxyl in

epinephrine. If the stereochemistry of the experimentally tested compound was not specified in the literature, the stereochemistry geometrically corresponding to R-epinephrine was considered in this study. As confirmed by mutagenesis studies and the experimental crystal structure, the protonated amine and the hydroxyl group of carazolol interact with Asp113 in the third transmembrane domain (TM3) of the β_2 -AR and with Asn312 in TM7, respectively. We assumed that the protonated amine and the hydroxyl group of the aliphatic chain of the different ligands would have to dock in a similar way. Thus, if the compounds displayed markedly different behavior in their final docked poses, they were eliminated as inconsistent. As a result, 27 compounds were eliminated and a database of 67 compounds was used for the generation of the different models (see Table S1 of the Supplementary Material).

Molecular Docking

All docking calculations were carried out with the Maestro 8.0 package.²⁵ Compounds in the database were automatically docked into the crystallized β_2 -AR binding site by means of Glide SP and XP 4.5.²⁶ The ligands were prepared for docking with the LigPrep tool, and the crystal structure of the receptor was prepared with the Protein Preparation Wizard workflow of Maestro 8.0.²⁵ This added hydrogens, which were subsequently minimized, and also optimized the protonation state of His residues and the orientation of hydroxyl groups, Asn residues, and Gln residues. In the receptor grid generation, no scaling factors were applied to the van der Waals radii of the receptor's atoms. The docked ligand was confined to a box centered on Val114, with a size capable of accommodating ligands with a length ≤ 15 Å, and with a box of 10 Å for the ligand diameter midpoint. The scaling factor for the van der Waals radii of the docked ligand was set to 0.80, with a partial charge cutoff of 0.15e. A post-docking minimization was performed on the output complexes, reducing the initial 10 poses per ligand to 5. The selection of the final, top-ranked pose was carried out by taking into account both energetic and structural parameters. The receptor-ligand complexes obtained from Glide were also subjected to a subsequent analysis with the Liaison 4.5²⁷ engine implemented in the Schrödinger package. In particular, we calculated the LiaScore – which is a scoring function similar to that implemented in Glide but with a flexible representation of the receptor – and a score based on the linear interaction energy method in a surface generalizes Born continuum model for solvation (LIE-SGB). The latter, combines through linear regression van der Waals, electrostatic, and cavity energy terms (see also Costanzi *et al.*).¹⁷ Although the LIE method has been originally developed to be used within a molecular dynamics framework,^{28–29} here we used as sampling method energy minimizations performed with the Truncated Newton algorithm and the OPLS_2005 force field, thus significantly reducing the required computational time. Moreover, the same receptor-ligand complexes were also subjected to MM-GBSA (molecular mechanics with generalized born surface area) calculations as implemented in Prime 2.0.³⁰ Also in this case, although the classical MM-GBSA is based on molecular dynamics,^{31–32} the version implemented in the Schrödinger package utilizes energy minimizations. The Prime calculations were carried out keeping all residues frozen, and then with a flexible region of 4, 8 and 12 Å. The linear regression models were built with the help of the STATISTICA 8.0 package,³³ using various energetic variables extracted from the docking calculations as independent variables and the experimental pK_i of the compounds as the dependent variable.

CoMFA and CoMSIA

CoMFA and CoMSIA were carried out with SYBYL 8.0.³⁴ The structural alignment of the ligands obtained in Glide 4.5²⁶, as described above, was used as input for construction of the CoMFA/CoMSIA models. Gasteiger-Hückel charges were used for all ligands. The CoMFA and CoMSIA region was identified by a grid extending 4 Å over the molecules in the three spatial directions, with a spacing of 1 Å, and a c.3 probe atom having a charge of + 1.0. For

CoMFA, two fields were considered (steric and electrostatic), and for CoMSIA, five fields were considered (steric, electrostatic, H-bond donor, H-bond acceptor and hydrophobic). The CoMFA fields were generated with the standard Tripos parameters, with a distance dependent dielectric and a cutoff of 30.0 kcal/mol. The type of transition in the cutoff region was set to “smooth”. The PLS analyses were carried out in SYBYL 8.034 with 4 to 6 components.

Molecular Descriptors Models

Molecular charges were calculated in MOE 2007.09, according to the “partial equalization of orbital electronegativity (PEOE)” method of Gasteiger; 2D and 3D descriptors were also calculated with MOE 2007.09,35 on the basis of the docked conformation obtained with Glide26 as described above. 2D-QSAR and 3D-QSAR linear regression models were built with the STATISTICA 8.0 package,33 using different combinations of descriptors as independent variables and the experimental pK_i of the compounds as the dependent variable. The forward-stepwise method was used for the selection of the descriptors to be included in the models.

Consensus Models

Various Partial Least Square Consensus Models were generated through PLS regressions in MOE,35 utilizing the calculated pK_i resulting from the individual methodologies as independent variables and the experimental pK_i of the compounds as dependent variable. All PLS regressions were based on one component.

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) was carried out using the STATISTICA 8.0 package33 on each of the main models previously described. This technique is intended to find a classifier function capable of separating the compounds into two or more classes – for example actives and non actives. Here, a pK_i cut-off of 7 was selected and a value of 1 and -1 was assigned to the dependent variable to label the compounds as active and non-active respectively.

Receiver Operating Characteristic (ROC) curves

ROC curves were constructed for each model, plotting the fraction of true positives (sensitivity) against fraction of false positives (1-specificity).

Results and Discussion

The pK_i values of 67 β_2 -AR ligands, including agonists and blockers, were collected from the literature and used for generating the models (see Table S1 of the Supplementary Material). By taking into account a variety of diverse structures, we aimed at generating a general model, rather than one confined to the treatment of a specific class of molecules. We randomly divided this database into 5 separate training and test sets, consisting of 53 and 14 compounds, respectively.

The first step of the construction of all models consisted of docking all of the compounds at the crystal structure of the β_2 -AR in complex with carazolol (pdb id: 2RH1). The obtained docking complexes and bioactive conformations of the ligands were then used as the basis for the construction of all the individual ligand-based and structure-based models, as illustrated in Figure 1. A brief explanation of the individual models with their respective components is given in Table 1.

The docking was performed by means of the Schrödinger program Glide,²⁶ which treats the ligands with a fully flexible all-atom representation, and the receptor with a precalculated rigid grid representation. As evidence of the accuracy of the docking results, Figure 2 illustrates that, for the two ligands for which crystal structures are available, the top scoring docking pose corresponds closely to the experimental pose. In particular, we obtained root mean square deviations (RMSDs) of 0.3 Å and 1.0 Å for carazolol and timolol, respectively. We performed the docking with Glide both in standard precision (SP) and extra precision (XP) mode, and finally decided to use the SP results due to the lower required computational time and the essentially similar accuracies – RMSDs of 0.4 Å and 1.0 Å from the crystal structures were obtained for carazolol and timolol with Glide XP.

Figure 3 illustrates the docking poses obtained for high affinity agonists and blockers. For all compounds, the protonated amine and the hydroxyl group interact with Asp113 in the third transmembrane domain (TM3) and with Asn312 in TM7, respectively. These interactions are widely supported by mutagenesis and crystallographic data,³⁶ and have recently been well described by Audet and Bouvier.³⁷

The aromatic moiety of both agonists and blockers stretches toward TM5 in our models. However, while most of the blockers establish direct interactions with Ser203 in TM5, the shorter nature of their aliphatic chain prevents agonists from interacting with this Ser, nor with Ser204 and Ser 207, also located in TM5. Thus, our model does not capture one of the hallmarks of agonist binding to the β -ARs best established in the literature.³⁸ Consistently with our results, Kobilka and coworkers noted that the interaction of β -AR agonists with the three critical TM5 Ser residues cannot occur in the ground state of the receptor and requires changes in the binding site.³ Molecular dynamics experiments recently published by Huber and coworkers also suggested that binding of agonists would cause significant conformational changes to the receptor.³⁹ More specifically, a modeling study published by Goddard and coworkers suggested that a counterclockwise rotation of TM5 accompanying the binding of β_2 -AR agonists would allow the aforementioned interactions between the catechol hydroxyl groups of the ligands and the three TM5 Ser residues, and may constitute a key element of the activation of the receptor.⁴⁰ A similar counterclockwise rotation has also been seen in a computational analysis recently published by Vaidehi and coworkers.⁴¹ Supporting this view, in a recent study based on molecular docking – this time with a flexible representation of to the receptor – and subsequent molecular dynamics, we too found that agonists but not blockers trigger the rotation of TM5 (Karpiak et al., manuscript submitted for publication).

The fact that for the docking experiments at the heart of this study we used a rigid representation of the receptor is the reason why here we do not capture the interaction of the agonist compounds with the three TM5 Ser residues. Instead, what we believe we detect is the initial stage of agonist binding, prior to the induction of conformational changes to the receptor. This initial binding mode appears to be very similar to the binding of blockers. As we will demonstrate (*vide infra*) this is sufficient to grant to agonists quantitative predictions as accurate as those obtained for blockers.

Prediction of affinities of the compounds through ligand and structure-based models

All of the below described individual and consensus models were generated using each of the 5 training sets. The square of the correlation coefficient (r^2) of each model in each training and test set are compared in Figure 4. For all models, the average values of r^2 , q^2 (cross-validated with the leave-one-out method), and root-mean-square error (RMSE) of prediction over the five training and test sets are provided in Table 2. Complete values for each individual training and test set are provided in Table S2 and Table S3 of the Supplementary Material.

Structure based model 1: Glide

Upon the completion of the docking, the Glide SP score, Glide XP score, and their components were given as output, along with the docking poses, from Glide.²⁶ Here, besides the standard Glide SP and XP scores, we derived through linear regression^{42, 43} a model trained with the affinities of the training set compounds, dubbed as “Glide Trained,” based on three components, which performed markedly better than the former two.

Structure-based model 2: Liaison

After docking, we also performed a calculation granting flexibility to the receptor by means of Liaison²⁷ and obtained three more regression models: one based on the Schrodinger's LiaScore, which is a version of the Glide score adapted to an all-atom representation of the protein; one based on the Linear Interaction Energy method for the calculation of the free energy of binding implemented with an implicit model for the solvation (LIE-SGB), which combined *van der Waals*, electrostatic, and cavity energy terms through a linear regression; and a final one, dubbed as “Liaison Trained”, which combined three of the LiaScore components through a linear regression. The LIE-SGB and Liaison Trained models were both trained with the affinities of the training set compounds, while the LiaScore is untrained.

Structure-based model 3: MM-GBSA

Lastly, using Schrödinger's Prime engine,³⁰ we scored the compounds on the basis of molecular mechanics with an implicit treatment of the solvation (MM-GBSA), first with all receptor residues being held frozen and subsequently allowing flexibility to spherical regions comprised within 4, 8 and 12 Å from the ligand. Here we present only the results obtained holding the receptor frozen, since these yielded consistently better correlation with the experimental affinities than the others. The MM-GBSA is an untrained model. The energy noise introduced while relaxing the structure is probably the reason behind the drop in accuracy registered when allowing flexibility to the receptor. This phenomenon has been illustrated by Cavasotto and coworkers while performing molecular docking experiments using an ensemble of receptor conformations generated through a normal-mode analysis. As the authors have shown, the inclusion of nonessential low-energy modes results in added noise to the calculation of the system's energy.^{44, 45}

Structure-based consensus model (ST-CM)

Overall, the structure-based models that performed the best were Glide Trained, Liaison Trained, LIE-SGB, and MM-GBSA. Utilizing the pK_i values predicted by these four models as input data, we generated a STructure Consensus Model (ST-CM) by means of a PLS regression.^{46, 47} The results obtained through ST-CM were consistently superior to those obtained with the individual models, demonstrating an average r^2 of about 0.60 (Figure 4 and Table 2), which outperformed the other models in each training set. With the five test sets, the individual models that performed the best were Glide Trained and MM-GBSA. The ST-CM model consistently showed the tendency to approximate the best individual model, yielding an average r^2 very close to that of the best component (MM-GBSA) and with a lower standard deviation (Table 2). This trend exemplifies the increased stability of the predictions of consensus models over those of their components (Figure 4).

Ligand-based model 1, 2, and 3: CoMFA/CoMSIA Models

CoMFA and CoMSIA are 3D-QSAR techniques used to generate predictive models on the basis of three-dimensional (3D) molecular interaction fields.^{48, 49} These techniques do not take into account the interactions between the ligands and the receptors, being based exclusively on the properties of the ligands. However, in our case, the docking of the

compounds through Glide26 was an important component for the generation of the CoMFA and CoMSIA models also, since it provided the bioactive conformations of the molecules and their structural alignment (Figure 1). The CoMFA and CoMSIA models performed very well with all the training sets, displaying, in general, the highest r^2 values in all of the training sets (Figure 4 and Table 2). Using the pK_i values predicted with CoMFA and CoMSIA, we generated a consensus model dubbed SYBYL-CM through a PLS regression. Also in this case, the r^2 of the new SYBYL-CM tended to capture the predictive ability of the best individual models in the training sets (Figure 4 and Table 2). With the test sets, the resultant stability of the two individual models was somewhat lower compared to the training sets. However, also in this case SYBYL-CM performed close to, or better than, the best individual model.

Ligand-based model 3,4, and 5: Molecular Descriptors-based Models

Using the QuaSAR-Descriptor feature of MOE,³⁵ we calculated 184 2D descriptors and 67 3D descriptors. While 2D descriptors are not dependent on the conformation of the molecules, 3D descriptors are dependent on the conformation and, in some cases, the structural alignment of the molecules, which in our case was derived through docking at the receptor (Figure 1).

Using multiple linear regression (MLR), we identified 6 2D descriptors and 6 3D descriptors that correlate well with the experimental affinity of the compounds. The r^2 values were consistently better using the 3D descriptors (Figure 4 and Table 2). Also in this case, we generated a PLS consensus model, dubbed MolDes-CM, using as input the pK_i values predicted by the individual models. MolDes-CM outperformed the two individual models in all training and most test sets, showing a significantly increased prediction stability.

Ligand-based Consensus Model (LI-CM)

With a further PLS regression, we then combined the predictions of SYBYL-CM and MolDes-CM into a Ligand-based Consensus Model (LI-CM). Once again, both in the training and the test sets, the consensus model tended to mimic or outperform the best individual model.

Ligand and Structure-based Consensus Model (LIST-CM)

After the generation of ST-CM and LI-CM, we performed a final PLS regression on the predictions of the training sets of these models and generated a consensus model dubbed LIST-CM. Once again, the predictivity of LIST-CM in the training and test sets was close to, or better than, that of the best individual model, which was consistently LI-CM (Figure 4 and Table 2). In Figure 5, the LIST-CM predictions are plotted versus the experimental affinities for the first training and test set. In the figure, the data points referring to agonists are represented with crosses, while those referring to blockers are represented with filled squares. The affinities of agonists and blockers are predicted with the same level of accuracy, as evidenced by the distance of the data points from the ideal prediction, which is represented by the diagonal of the plot.

As discussed, our study is based on the structure of the β_2 -AR in complex with carazolol, an inverse agonist. Not only the structure-based methods, but also the ligand-based methods lay their foundations on the initial docking operated treating the protein as a rigid body. In this light, a comparable accuracy in the prediction of the affinity of agonists and blockers may result somewhat surprising. What our model captures is most likely related to the first stage of the agonist binding, before the subsequent induced fit conformational changes of the receptor that lead to signaling. This seems to be sufficient, especially in the presence of a training set, to ensure for the predictions related to agonists a level of accuracy comparable

to those related to blockers. As we will demonstrate, although training is necessary for quantitative predictions, untrained structure-based models are capable of robustly separating binders from non-binders, also in this case without distinguishing agonists and blockers.

To verify that the correlation had not been randomly achieved, we scrambled the experimental affinities of the compounds generating 50 random LIST-CM models. The results of this randomization test demonstrate that LIST-CM – and thus each individual component – is not the result of a chance correlation, since the r^2 and the q^2 of the genuine model are high and significantly distant from those obtained for the random models (Figure 6).

Summarizing, in all the cases, our consensus models tended to imitate, or outperform, the best individual models. Consensus models are also endowed with a significant stability in their predictive capacity, as outlined by the low standard deviations of the r^2 in the training and test sets (Table 2). The ligand-based models always yielded a better classification than the structure-based models. As a result, LI-CM seems to be the method of choice to predict the affinity of closely related analogs, when a training set is available. In the absence of a training set, and thus faced with the impossibility of using ligand-based models or trained structure-based models, the MM-GBSA model appears to be the best option.

Linear Discriminant Analysis (LDA)

Besides assessing the ability of the models to quantitatively predict affinities, here we also explored their capability to simply discriminate between active and non-active compounds. For this purpose, we set a cutoff at a pK_i of 7 and studied the correct classification yielded by each model through LDA.^{50, 51} The predicted pK_i s from the different models were taken as the independent variables, and a value of 1 or -1 was assigned to the dependent variable to label the compounds as active or non-active.

The results of the classification produced by the different models are shown in Table 3 and are in good agreement with their respective correlation coefficients (Figure 7). Complete values for each individual training and test set are provided in Table S4 of the Supplementary Material. Also with the LDA, the various consensus models tended to equal or perform better than their components. Notably, the structure-based models, and the Glide Trained in particular, yielded an efficient classification of active and inactive compounds, despite the low r^2 . In other words, the structure-based models, although not very good at yielding precise numeric values for the affinities, proved quite capable of discriminating actives and non-actives. As a result, while LI-CM proved to be substantially superior to ST-CM at quantitatively predicting affinities, the two consensus models performed almost equivalently with the test sets in the LDA.

Simulated virtual screening experiments

In order to assess the possibility of using the above described models for virtual screening purposes, we evaluated their performances with a database of compounds containing, besides the β_2 -AR ligands, also a number of inactive molecules. In particular, to the 67 β_2 -AR ligands we added 52 dopamine receptor ligands, which are structurally related to the β_2 -AR ligands, and 225 molecules randomly retrieved from ChemBank (<http://ligand.info/>), obtaining a database of 344 compounds. Using the models obtained with the 5 previously described training sets composed of 53 β_2 -AR ligands, we estimated the affinities of the remaining 291 molecules. Moreover, we also utilized structure-based techniques that do not require training to rank the whole database of 344 compounds.

To analyze the results of our virtual screening, we used a pK_i cut-off of 7 to consequently divide the database into 25 active and 319 inactive compounds. The average percentage of

true positives within the top 25 ranks yielded by the various models is provided in Table 4. Additional information is provided in Table S5 of the Supplementary Material. We also plotted the data in the form of ROC curves,^{52–54} which are particularly suited to illustrating the tendency of a binary model to either correctly classify the compounds or predict false positives (Figure 8, panels **a**, **b**, **c**, and **d**).

Ideally, in a perfect model the top 25 ranks should be occupied by the 25 active molecules, and not by false positives. In this light, the CoMSIA model performed the best, yielding 75 % true positives in the top 25 positions, whereas the models based on molecular descriptors performed only marginally better than a random selection (represented by the dotted diagonal in the plots of Figure 8). The structure-based models, including those that do not require training, also performed well. In particular, the MM-GBSA scoring yielded an average of 48 % true positives in the top 25 positions. Given that the MM-GBSA scoring does not need training, this model seems to be particularly suited for virtual screenings, when no prior active compounds are known. This reinforces the idea that scoring functions based on docking are more suitable as a discriminating tool than as binding energy predictors.

Selecting a pK_i cut-off of 7 means considering only the compounds with the highest affinity. Thus, to probe the ability of the models to discriminate low affinity binders from non-binders, for the best trained ligand-based model (CoMSIA) and the best untrained structure-based model (MM-GBSA) we also plotted ROC curves with a pK_i cut-off of 5 (Figure 8, panels **e** and **f**). It appears evident that, when reducing the pK_i cutoff from 7 to 5, the ability of the CoMSIA model (Figure 8, panels **c** and **e**) to prioritize true positives substantially diminishes, as indicated also by the reduction of the area under the curve. On the other hand, the performance of the untrained MM-GBSA model remains unaffected (Figure 8, panels **d** and **f**). This is another indication of the inherent generality of untrained structure-based models, which, while not capable of sharply predicting the affinity of the compounds, are capable of effectively discriminating binders from non-binders, and are thus applicable to virtual screening experiments. Conversely, trained models tend to perform very well with compounds structurally similar to those included in the training set, yielding quantitative affinity predictions, but they tend to perform poorly in the analysis of diverse compounds.

Of note, all of the methodologies that we applied, including the untrained structure-based methodologies, are capable of recognizing as binders both agonists and blockers (defined here as the molecules with intrinsic activity lower than 15 %), as evident in Figure 8 (panels **c**, **d**, **e**, and **f**). The highest ranks are generally given to blockers (red data points) and not to agonists (green data points), reflecting the fact that in our dataset, agonists were generally less potent than blockers (see Table S1 of Supplementary Material). The ability of prioritizing both agonists and blockers over non-binders that we found in our models is in agreement with what was found by de Graaf and Rognan, whose model based on the unmodified crystal structure of the β_2 -AR in complex with an inverse agonist proved capable of prioritizing unselectively agonists and blockers over decoy compounds.¹⁴ This suggests the suitability of molecular docking of agonists at GPCRs crystallized in complex with blockers. The obtained complexes are likely to reflect the early stages of agonist binding, rather than the fully activated receptor-agonist complex.

In contrast, crystal structures or homology models optimized in complex with either agonists seem to acquire selectivity toward this class of compounds, as exemplified by the selective prioritization of agonists by a model of the β_2 -AR modified by de Graaf and Rognan to reflect some of the changes expected to occur with agonist binding.¹⁴ This finding is in agreement with a previous virtual screening in which we noted that a model of the free fatty acid receptor 1 (FFA1) optimized in complex with agonists yielded primarily agonists.¹²

The same seems to be true for models optimized in complex with antagonists prior to the virtual screening, as suggested by the selective retrieval of novel antagonists of the melanin-concentrating hormone receptor 1 by Cavasotto and Coworker.¹¹

Conclusions

In this study, we have explored the applicability of various structure-based and ligand-based models to the prediction of the affinity of closely related analogs, and to the retrieval of active analogs from a database where the majority of the compounds are inactive. We have also proposed effective consensus models that combine the prediction obtained with selected individual models.

The ligand-based QSAR techniques, and in particular the molecular descriptor-based methodologies, were highly effective in ranking the affinities of compounds closely related to those included in the training set (Figure 4 and Table 2). Thus, a consensus of predictions obtained with ligand-based techniques (LI-CM) seems the most appropriate way of ranking closely related analogs, when a training set with known affinities is available. It is important to remember that also our ligand-based models are based on molecular docking. It is through molecular docking, in fact, that we obtained the conformation and the alignment of the ligands, which are fundamental parameters in the construction of QSAR models. In this light, the success of these models argues in favor of the reliability of the obtained ligand binding modes. Furthermore, our data confirm the value of performing molecular docking followed by QSAR analyses.¹⁰ In particular, although applicable only in the availability of prior structure-activity relationships data (*i.e.* training sets), this combination ensures the highest possible accuracy in the prediction of the affinity of novel compounds. In the absence of a training set, the untrained structure-based MM-GBSA score seems the best alternative. Although less accurate with quantitative predictions, structure-based methodologies resulted as effective as the ligand-based methodologies in the discrimination of active and non-active compounds.

Molecular descriptor-based QSAR techniques appeared very dependent on the training set and thus were not applicable for virtual screening purposes (Figure 8). On the contrary, the CoMSIA based model performed very well in our simulated virtual screenings, especially when hunting only for the most active compounds (Figure 8, panel **c**). Untrained structure-based methodologies also performed very well in our simulated virtual screening. In particular, the MM-GBSA score was the most effective method for the identification of not only strong binders (Figure 8, panel **d**), but also of weak binders (Figure 8, panel **f**), appearing to be the method of choice for virtual screening campaigns. Notably, the MM-GBSA score does not need a training set and is therefore applicable also in the absence of previously known binders.

The inability of scoring functions for protein-ligand interactions to accurately rank compounds on the basis of their affinity is a well known fact.¹⁶⁻⁵⁵ Part of this inaccuracy is probably due to the sensitivity of the scoring functions, which can substantially amplify small docking inaccuracies. The rigidity of the protein can also affect negatively the accuracy of the results, and the incorporation of receptor flexibility could yield significantly more accurate predictions.¹¹ However, since we aimed at identifying techniques applicable to virtual screenings, we could not consider here fully flexible docking procedures – *i.e.* procedures based on a flexible representation of the receptor – due their prohibitive computational time requirements that render them inapplicable to the analysis of large databases of compounds.

Our study also revealed that consensus models in general have a performance superior to, or close to that of, the best individual component, and they are endowed with a significantly increased robustness, as indicated by the low standard deviation of the r^2 (Figure 4).

Notably, our models functioned equally well with agonists and blockers, with respect to both quantitative affinity predictions and simulated virtual screening.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

β₂-AR	β ₂ -Adrenergic Receptor
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
GPCRs	G Protein-Coupled Receptors
LDA	Linear Discriminant Analysis
LIE	Linear Interaction Energy
LI-CM	Ligand-based Consensus Model
LIST-CM	Ligand and STructure-based Consensus Model
MM-GBSA	Molecular Mechanics Generalized Born Solvent Accessibility
PLS	Partial Least Square
ROC	Receiver Operating Characteristic
SP	Standard-Precision
ST-CM	STructure-based Consensus Model
TM	Transmembrane
XP	Extra-Precision

Acknowledgments

This research was supported by Intramural Research Program of the NIH, NIDDK.

Reference List

1. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Science. 2000; 289:739. [PubMed: 10926528]
2. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. Science. 2007; 318:1258. [PubMed: 17962520]
3. Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK. Science. 2007; 318:1266. [PubMed: 17962519]
4. Costanzi S. J. Med. Chem. 2008; 51:2907. [PubMed: 18442228]
5. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX. Nature. 2008; 454:486. [PubMed: 18594507]
6. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EYT, Velasquez J, Kuhn P, Stevens RC. Structure. 2008; 16:897. [PubMed: 18547522]

7. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC. *Science*. 2008; 322:1211. [PubMed: 18832607]
8. Li YY, Zhu FQ, Vaidehi N, Goddard WA, Sheinerman F, Reiling S, Morize I, Mu L, Harris K, Ardati A, Laoui A. *J. Am. Chem. Soc.* 2007; 129:10720. [PubMed: 17691773]
9. Moore S, Jaeschke H, Kleinau G, Neumann S, Costanzi S, Jiang JK, Childress J, Raaka BM, Colson A, Paschke R, Krause G, Thomas CJ, Gershengorn MC. *J. Med. Chem.* 2006; 49:3888. [PubMed: 16789744]
10. Costanzi S, Tikhonova IG, Harden TK, Jacobson KA. *J. Comput. Aided Mol. Des.* 2008
11. Cavasotto CN, Orry AJW, Murgolo NJ, Czarniecki MF, Kocsi SA, Hawes BE, O'Neill KA, Hine H, Burton MS, Voigt JH, Abagyan RA, Bayne ML, Monsma FJ. *J. Med. Chem.* 2008; 51:581. [PubMed: 18198821]
12. Tikhonova IG, Sum CS, Neumann S, Engel S, Raaka BM, Costanzi S, Gershengorn MC. *J. Med. Chem.* 2008; 51:625. [PubMed: 18193825]
13. Engel S, Skoumbourdis AP, Childress J, Neumann S, Deschamps JR, Thomas CJ, Colson AO, Costanzi S, Gershengorn MC. *J. Am. Chem. Soc.* 2008; 130:5115. [PubMed: 18357984]
14. De Graaf C, Rognan D. *J. Med. Chem.* 2008; 51:4978. [PubMed: 18680279]
15. Cavasotto CN, Orry A. *J. W. Curr. Top. Med. Chem.* 2007; 7:1006.
16. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. *J. Med. Chem.* 2006; 49:5912. [PubMed: 17004707]
17. Costanzi S, Tikhonova IG, Ohno M, Roh EJ, Joshi BV, Colson AO, Houston D, Maddileti S, Harden TK, Jacobson KA. *J. Med. Chem.* 2007; 50:3229. [PubMed: 17564423]
18. Moro S, Bacilieri M, Cacciari B, Spalluto G. *J. Med. Chem.* 2005; 48:5698. [PubMed: 16134938]
19. Moro S, Braiuca P, Deflorian F, Ferrari C, Pastorin G, Cacciari B, Baraldi PG, Varani K, Borea PA, Spalluto G. *J. Med. Chem.* 2005; 48:152. [PubMed: 15634009]
20. Moro S, Bacilieri M, Cacciari B, Bolcato C, Cusan C, Pastorin G, Klotz KN, Spalluto G. *Bioorg. Med. Chem.* 2006; 14:4923. [PubMed: 16564691]
21. Santana L, Uriarte E, Gonzalez-Diaz H, Zagotto G, Soto-Otero R, Mendez-Alvarez E. *J. Med. Chem.* 2006; 49:1149. [PubMed: 16451079]
22. Vilar S, Estrada E, Uriarte E, Santana L, Gutierrez Y. *J. Chem. Inf. Model.* 2005; 45:502. [PubMed: 15807516]
23. Cao HY, Cao R, Zhang HB, Zheng XF, Gao DB. *Curr. Med. Chem.* 2008; 15:1462. [PubMed: 18537623]
24. Gedeck P, Lewis RA. *Curr. Opin. Drug Disc.* 2008; 11:569.
25. Maestro, version 8.0. New York, NY: Schrödinger, LLC; 2007.
26. Glide, version 4.5. New York, NY: Schrödinger, LLC; 2007.
27. Liaison, version 4.5. New York, NY: Schrödinger, LLC; 2005.
28. Aqvist J, Medina C, Samuelsson JE. *Protein Eng.* 1994; 7:385. [PubMed: 8177887]
29. JonesHertzog DK, Jorgensen WL. *J. Med. Chem.* 1997; 40:1539. [PubMed: 9154975]
30. Prime, version 2.0. New York, NY: Schrödinger, LLC; 2008.
31. Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. *Acc. Chem. Res.* 2000; 33:889. [PubMed: 11123888]
32. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. *J. Am. Chem. Soc.* 1998; 120:9401.
33. STATISTICA (data analysis software system). version 8.0. Tulsa, OK, USA: StatSoft, Inc.; 2008.
34. SYBYL. version 8.0. Tripos, Inc.; 2007. www.tripos.com
35. MOE. version 2007.09. Chemical Computing Group, Inc.; 2007. www.chemcomp.com
36. Strader CD, Fong TM, Tota MR, Underwood D, Dixon RAF. *Annu. Rev. Biochem.* 1994; 63:101. [PubMed: 7979235]
37. Audet M, Bouvier M. *Nat. Chem. Biol.* 2008; 4:397. [PubMed: 18560432]
38. Sato T, Kobayashi H, Nagao T, Kurose H. *Brit. J. Pharmacol.* 1999; 128:272. [PubMed: 10510435]
39. Huber T, Menon S, Sakmar TP. *Biochemistry.* 2008; 47:11013. [PubMed: 18821775]

40. Freddolino PL, Kalani MYS, Vaidehi N, Floriano WB, Hall SE, Trabanino RJ, Kam VWT, Goddard WA. *Proc. Natl. Acad. Sci. USA.* 2004; 101:2736. [PubMed: 14981238]
41. Bhattacharya S, Hall SE, Li H, Vaidehi N. *Biophys. J.* 2008; 94:2027. [PubMed: 18065472]
42. Breiman L, Friedman JHJR. *Statist. Soc. B.* 1997; 59:3.
43. Slinker BK, Glantz SA. *Am. J. Physiol.* 1988; 255:R353. [PubMed: 3046375]
44. Cavasotto CN, Kovacs JA, Abagyan RA. *J. Am. Chem. Soc.* 2005; 127:9632. [PubMed: 15984891]
45. Kovacs JA, Cavasotto CN, Abagyan R. *J. Comput. Theor. Nanos.* 2005; 2:354.
46. Ng C, Xiao YD, Putnam W, Lum B, Tropsha A. *J. Pharm. Sci.* 2004; 93:2535. [PubMed: 15349962]
47. Wang T, Ho G, Ye K, Strickler H, Elston RC. *Genet. Epidemiol.* 2008
48. Cramer RD, Patterson DE, Bunce JD. *J. Am. Chem. Soc.* 1988; 110:5959.
49. Klebe G, Abraham U, Mietzner T. *J. Med. Chem.* 1994; 37:4130. [PubMed: 7990113]
50. Glover F. *Decis. Sci.* 1990; 21:771.
51. Tom CH, Miller LD. *Photogramm. Eng. Rem. S.* 1984; 50:193.
52. Bradley AP. *Pattern Recogn.* 1997; 30:1145.
53. Hanley JA, Mcneil B. *J. Radiology.* 1982; 143:29.
54. Mcneil BJ, Hanley JA. *Med. Decis. Making.* 1984; 4:137. [PubMed: 6472062]
55. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. *J. Med. Chem.* 2004; 47:3032. [PubMed: 15163185]

Figure 1.

A flowchart representation of the construction of the ligand and structure-based consensus models is presented. Through molecular docking, descriptors that directly quantify the receptor-ligand interactions (right side of the scheme) or that purely describe the ligand structures (left side of the scheme) are calculated. From these, a number of individual and consensus models are built. Trained models, which require a previous knowledge of the affinity of a training set of ligands, are represented as gray boxes. Models that do not require training are represented with a black box.

Figure 2. The partial inverse agonists carazolol (**2a**) and timolol (**2b**) are docked at the β_2 -receptor. The final docked poses (in red) are superimposed on top of the ligand conformations (in green) observed in the crystal structures, illustrating the high degree of accuracy of the applied docking procedure in replicating bound conformations.

Figure 3.

Docking poses obtained for isoproterenol (a high affinity synthetic agonist) and carazolol (a very high affinity synthetic inverse agonist). In both ligands, the protonated amine and the hydroxyl group interact with Asp113 in TM3 and Asn312 in TM7. We speculate that these interactions, which characterized binding of both agonists and blockers, are the first to be established in the sequential process of ligand binding. Furthermore, while carazolol establishes a hydrogen bond with Ser203 in TM5, the shorter nature of isoproterenol does not allow it to interact with any of the TM5 Ser residues. For this interaction to occur, a substantial conformational change of the receptor is necessary.

Figure 4.

Plots of the predictive results of the various individual and consensus models for each training and test set are shown. In each plot, the r^2 is represented on the y-axis, with the training or test set number labeled on the x-axis. In each case, the consensus model captures the best behavior of the individual constituent components.

Figure 5.

Representative plots of the experimental affinity versus the LIST-CM predicted affinity (pK_i) for the first training (**5a**) and test set (**5b**). The r^2 values are 0.885 for the training and 0.844 for the test. Crosses represent data points referring to agonists, while filled squares refer to data points representing blockers. The diagonal represent an ideal prediction. As evident from the distances of the various data points from the diagonal, LIST-CM predicts with similar accuracy the affinity of both agonists and blockers.

Figure 6.

A randomization test has been performed on the LIST-CM by scrambling the activities of the ligands for the generation of 50 chance models. Here, the r^2 and cross-validated q^2 are plotted for the chance models as well as the genuine one, with each circle representing an individual model. As expected, LIST-CM (represented by the closed black circle) is very far away from the random models obtained using random dependent variables.

Figure 7.

This graph provides a comparison of the performances of the various models in the quantitative prediction of ligand affinity and in the classification of binders and non-binders. In particular, the average r^2 values of the quantitative predictions are compared with the average percentage of correct classification in the LDA for the five test sets.

Figure 8.

Analysis of the simulated virtual screening results through ROC curves. **a)** Selected trained models, trained with the first training set. **b)** All untrained structure-based models (^aXP scoring of poses generated with Glide SP. ^bXP scoring of poses generated with Glide XP). In panels **a**, **b**, **c**, and **d**, the affinity cut-off is set to a pKi of 7, while, in panels **e** and **f**, it is set to a pKi of 5.

Table 1

Explanation of the descriptors used in the models described in this paper.

Importance ^a	Descriptors	Brief description
CoMFA (PLS)^b		
1.000	Steric	Steric CoMFA field
0.558	Electrostatic	Electrostatic CoMFA field
CoMSIA (PLS)		
0.579	Steric	Steric CoMSIA field
0.384	Electrostatic	Electrostatic CoMSIA field
1.000	Hydrophobic	Hydrophobic CoMSIA field
0.726	H-bond donor	H-bond donor CoMSIA field
0.736	H-bond acceptor	H-bond acceptor CoMSIA field
SYBYL-CM (PLS 1 component)		
1.000	PRED-CoMFA	pK _i predicted for the CoMFA model
0.956	PRED-CoMSIA	pK _i predicted for the CoMSIA model
MolDes 2D (MLR)^c		
0.437	GCUT_SLOGP_1	LogP GCUT (1/3)
0.355	BCUT_SLOGP_3	LogP BCUT (3/3)
0.340	radius	Smallest vertex eccentricity in graph
1.000	logS	Log solubility in water
0.531	SlogP_VSA8	Bin 8 SlogP (0.30, 0.40)
0.358	SMR_VSA5	Bin 5 SMR (0.440, 0.485)
MolDes 3D (MLR)		
0.320	AM1_HF	Heat of formation (kcal/mol) using AM1 Hamiltonian
0.424	AM1_HOMO	Energy (eV) of HOMO using AM1 Hamiltonian
0.303	E_strain	Energy minus the energy at a near local minimum
1.000	VSA	Van der Waals surface area
0.543	dipoleY	Dipole moment (Y)
0.222	dipoleZ	Dipole moment (Z)
MolDes-CM (PLS 1 component)		
0.859	PRED-MolDes 2D	pK _i predicted for the MolDes 2D model
1.000	PRED-MolDes 3D	pK _i predicted for the MolDes 3D model
LI-CM (PLS 1 component)		
0.780	PRED-MolDes-CM	pK _i predicted for the MolDes-CM
1.000	PRED-SYBYL-CM	pK _i predicted for the SYBYL-CM
Glide Trained (MLR)		
1.000	Glide_lipo	Lipophilic contact term
0.863	Glide_esite	Polar interactions in the active site
0.651	Glide_internal	Internal energy
Liaison Trained (MLR)		

Importance ^a	Descriptors	Brief description
1.000	Lia_HBond	Hydrogen-bonding term
0.890	Lia_Lipo	Lipophilic contact term
0.697	Lia_RotB	Penalty for freezing rotatable bonds
LIE-SGB (MLR)		
0.231	Liaison_<Uvdw	<i>Van der Waals</i> term
1.000	Liaison_<Uele	Electrostatic term
0.097	Liaison_<Ucav	Cavity energy term
MM-GBSA-(0) (Linear Regression)		
1.000	MM-GBSA DG (0)	MM-GBSA free energy of binding with freeze protein
ST-CM (PLS 1 component)		
0.567	PRED-LIE-SGB	pK _i predicted for the LIE-SGB model
1.000	PRED- Glide Trained	pK _i predicted for the Glide Trained
0.648	PRED- Liaison Trained	pK _i predicted for the Liaison Trained
0.745	PRED-MMGBSA (0)	pK _i predicted for the MM-GBSA (0) model
LIST-CM (PLS 1 component)		
1.000	PRED-LI-CM	pK _i predicted for the LI-CM
0.483	PRED-ST-CM	pK _i predicted for the ST-CM

^aRelative weighting of descriptors in the first training set.

^bPLS = Partial Least Square.

^cMLR = Multiple Linear Regression.

Table 2

Average Statistical Parameters of the Various LR and PLS Models over the Five Training and Test Sets (Average $\pm$ Standard Deviation). Complete Data for all Training and Test Sets are given in Table S2 and Table S3 of the Supplementary Material.

METHOD	r^2 training	q^2 training	RMSE ^c training	r^2 test	RMSE test
CoMFA	0.961 $\pm$ 0.025	0.669 $\pm$ 0.064	0.291 $\pm$ 0.069	0.555 $\pm$ 0.278	0.808 $\pm$ 0.256
CoMSIA	0.962 $\pm$ 0.016	0.750 $\pm$ 0.026	0.293 $\pm$ 0.057	0.642 $\pm$ 0.174	0.747 $\pm$ 0.219
SYBYL-CM	0.970 $\pm$ 0.010	0.968 $\pm$ 0.010	0.261 $\pm$ 0.033	0.624 $\pm$ 0.223	0.749 $\pm$ 0.239
MolDes 2D	0.696 $\pm$ 0.027	0.611 $\pm$ 0.037	0.844 $\pm$ 0.042	0.589 $\pm$ 0.159	0.806 $\pm$ 0.151
MolDes 3D	0.762 $\pm$ 0.017	0.688 $\pm$ 0.023	0.747 $\pm$ 0.040	0.663 $\pm$ 0.183	0.723 $\pm$ 0.207
MolDes-CM	0.819 $\pm$ 0.019	0.803 $\pm$ 0.021	0.651 $\pm$ 0.039	0.761 $\pm$ 0.090	0.628 $\pm$ 0.166
LI-CM	0.942 $\pm$ 0.008	0.937 $\pm$ 0.009	0.368 $\pm$ 0.004	0.757 $\pm$ 0.132	0.607 $\pm$ 0.153
Glide Trained	0.531 $\pm$ 0.018	0.456 $\pm$ 0.017	1.050 $\pm$ 0.076	0.455 $\pm$ 0.137	0.977 $\pm$ 0.352
Liaison Trained	0.482 $\pm$ 0.049	0.395 $\pm$ 0.062	1.104 $\pm$ 0.090	0.280 $\pm$ 0.147	1.103 $\pm$ 0.299
LIE-SGB	0.412 $\pm$ 0.034	0.325 $\pm$ 0.036	1.177 $\pm$ 0.092	0.311 $\pm$ 0.080	1.102 $\pm$ 0.358
MM-GBSA (0)	0.499 $\pm$ 0.047	0.456 $\pm$ 0.050	1.083 $\pm$ 0.065	0.495 $\pm$ 0.216	0.899 $\pm$ 0.217
ST-CM	0.593 $\pm$ 0.014	0.555 $\pm$ 0.015	0.978 $\pm$ 0.067	0.476 $\pm$ 0.108	0.941 $\pm$ 0.263
LIST-CM	0.889 $\pm$ 0.006	0.878 $\pm$ 0.008	0.511 $\pm$ 0.018	0.719 $\pm$ 0.125	0.663 $\pm$ 0.147

^aSquare of the correlation coefficient.

^bCross-validated square of the correlation coefficient (leave-one-out method).

$$^c \text{Root-mean-square error, } \text{RMSE} = \sqrt{\frac{\sum (Y - Y')^2}{N}}$$

Table 3

Average Percentage of Correctly Classified Compounds yielded by each Model with the LDA (Average $\pm$ Standard Deviation), Calculated over the Five Training and Test Sets. The Significance Level (p) was < 0.05 for the All the Models. Complete Data for all Training and Test Sets are given in Table S4 of the Supplementary Material.

METHOD	% correctly classified: Training Sets	% correctly classified: Test Sets
CoMFA	95.47 $\pm$ 3.15	77.14 $\pm$ 5.97
CoMSIA	96.60 $\pm$ 1.58	72.86 $\pm$ 5.97
SYBYL-CM	97.73 $\pm$ 0.84	74.29 $\pm$ 8.14
MolDes 2D	80.76 $\pm$ 4.51	65.72 $\pm$ 9.31
MolDes 3D	79.62 $\pm$ 2.46	80.00 $\pm$ 11.73
MolDes-CM	83.77 $\pm$ 3.91	77.14 $\pm$ 12.78
LI-CM	93.59 $\pm$ 2.86	74.29 $\pm$ 6.39
Glide Trained	77.74 $\pm$ 3.63	77.14 $\pm$ 15.48
Liaison Trained	67.92 $\pm$ 2.98	62.86 $\pm$ 5.98
LIE-SGB	61.13 $\pm$ 3.43	50.00 $\pm$ 8.75
MM-GBSA (0)	72.07 $\pm$ 3.10	62.86 $\pm$ 12.78
ST-CM	72.45 $\pm$ 3.91	68.57 $\pm$ 6.39
LIST-CM	87.93 $\pm$ 3.43	70.00 $\pm$ 9.31

Table 4

Percentage of True Positives (TP %) and Enrichment Factor (EF) found within the Top 25 Scored Compounds by the Different Models. The Simulated Virtual Screening was performed on a Database containing the 67 β_2 -AR Ligands listed in the Table S1 of the Supplementary Material, 225 Molecules Randomly extracted from the ChemBank Database, and 52 Dopamine Receptor Ligands. With a pK_i Cut-off of 7, the Whole Database contained a Total of 25 Active Molecules. For the Trained Models, the Values are given as an Average $\pm$ Standard Deviation calculated over the Five Training Sets. Additional Data is given in Table S5 of the Supplementary Material.

	TP % in top 25	EF in top 25
Models with training -5 Sets Average-		
CoMFA	38 $\pm$ 2.2	5.3
CoMSIA	75 $\pm$ 9.1	10.3
SYBYL-CM	58 $\pm$ 6.7	8.0
MolDes 2D	0 $\pm$ 0.0	-
MolDes 3D	0 $\pm$ 0.0	-
MolDes-CM	0 $\pm$ 0.0	-
LI-CM	27 $\pm$ 3.3	3.7
Glide Trained	5 $\pm$ 1.8	0.7
Liaison Trained	46 $\pm$ 6.1	6.3
LIE-SGB	42 $\pm$ 7.3	5.7
ST-CM	50 $\pm$ 8.8	6.9
LIST-CM	46 $\pm$ 3.6	6.3
Models without training		
SP-Score	24	3.3
XP-Score ^a	28	3.9
XP-Score ^b	20	2.8
LiaScore	16	2.2
MMGBSA (0)	48	6.6

^a (XP score in place with SP pose.

^b Docking with XP score).