HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2020 July 22.

Published in final edited form as:

J Chem Inf Model. 2019 July 22; 59(7): 3177–3190. doi:10.1021/acs.jcim.9b00380.

Influence of the Structural Accuracy of Homology Models on Their Applicability to Docking-Based Virtual Screening: the β_2 Adrenergic Receptor as a Case Study

Stefano Costanzi[a,b,*], **Austin Cohen**[a], **Abigail Danfora**[a], **Marjan Dolatmoradi**[a]

[a]Department of Chemistry, American University, Washington, DC 20016, USA

[b]Center for Behavioral Neuroscience, American University, Washington, DC 20016, USA

Abstract

How accurate do structures of the β_2 adrenergic receptor (β_2AR) need to be to effectively serve as platforms for docking-based virtual screening campaigns? To answer this research question, here we targeted through controlled virtual screening experiments 23 homology models of the β_2AR endowed with different levels of structural accuracy. Subsequently, we studied the correlation between virtual screening performance and structural accuracy of the targeted models. Moreover, we studied the correlation between virtual screening performance and template/target receptor sequence identity. Our study demonstrates that docking-based virtual screening campaigns targeting homology models of the β_2AR, in the majority of the cases, yielded results that exceeded random expectations in terms of area under the receiver operating characteristic curve (ROC AUC). Moreover, with the most effective scoring method, over one third and one quarter of the models yielded results that exceeded random expectation also in terms of enrichment factors (EF1, EF5, and EF10) and BEDROC (α = 160.9), respectively. Not surprisingly, we found a detectable linear correlation between virtual screening performance and structural accuracy of the ligand-binding cavity. We also found a detectable linear correlation between virtual screening performance and structural accuracy of the second extracellular loop (EL2). Finally, our data indicate that, although there is no detectable linear correlation between virtual screening performance and template/β_2AR sequence identity, models built on the basis of templates that show high sequence identity with the β_2AR, especially within the ligand-biding cavity, performed consistently well. Conversely, models with lower sequence identity displayed performance levels

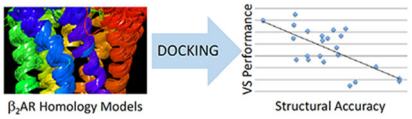
*Address correspondence to: Stefano Costanzi, Ph.D., Department of Chemistry, American University, 4400 Massachusetts Ave, NW, 20016 Washington, DC, USA, Tel: +1-202-885-1722, costanzi@american.edu.

SUPPORTING INFORMATION

Supporting information is available free of charge on the ACS Publications website and includes the following: structural accuracy of the models and template/β_2AR sequence identity (Table S1), virtual screening performance with the Glide SP method (Table S2), virtual screening performance with the Glide SP/XP-score-in-place method (Table S3), mean and median values of virtual screening performance metrics across all models (Figure S1), performance of virtual screening (Glide SP) targeting a β_2AR crystal structure using different sets of ligands and decoys (Figure S2), performance of virtual screening (Glide SP/XP-score-in-place) targeting a β_2AR crystal structure using different sets of ligands and decoys (Figure S3), plots of virtual screening performance versus structural accuracy of the models (Figure S4), correlation between EF1 and structural accuracy of the models (Figure S5), correlation between EF5 and structural accuracy of the models (Figure S6), correlation between EF10 and structural accuracy of the models (Figure S7), plots of virtual screening performance versus template/β_2AR sequence identity (Figure S8), correlation between EF1 and template/β_2AR sequence identity (Figure S9), correlation between EF5 and template/β_2AR sequence identity (Figure S10), correlation between EF10 and template/β_2AR sequence identity (Figure S11), schematic diagram of the β_2AR ligand binding cavity and corresponding residues in the dopamine D_3 receptor (Figure S12).

that ranged from very good to random, with no apparent correlation with the sequence identity itself.

Graphical Abstract



INTRODUCTION

G protein-coupled receptors (GPCRs) are highly sought-after targets for pharmaceutical intervention, due to their involvement in a wide range of physiological functions and pathological conditions. Their activity can be conveniently modulated through the administration of chemicals that, upon binding to the receptors, either stimulate their activity (agonists), lower their activity below the basal level (inverse agonists), or simply prevent their stimulation by natural ligands (neutral antagonists). Structurally, GPCRs are constituted by a single protein chain that spans the plasma membrane seven times with seven alpha helical transmembrane domains (TM1 to TM7, collectively TMs) connected by three intracellular loops (IL1 to IL3, collectively ILs) and three extracellular loops (EL1 to EL3, collectively ELs) (Figure 1). The N-terminus is located in the extracellular space, while the C-terminus is located in the cytosol.[1–4]

This work uses as a case study the β_2 adrenergic receptor (β_2AR), a prototypical GPCR for which a wealth of structural and pharmacological information is available. The β_2AR belongs to the largest family of GPCRs, also known as class A or rhodopsin family. Agonists and blockers of the receptors have a variety of pharmacological applications, including the treatment of asthma (agonists) and high blood pressure (blockers), among others. For the β_2AR, as well as for most receptors belonging to the rhodopsin family, the ligand-binding cavity (orthosteric cavity) is located at the extracellular opening of the TMs bundle. In particular, as shown in Figure 1, the orthosteric cavity of the β_2AR is lined by residues located in the upper segments of TM3, TM5, TM6, and TM7 and the second extracellular loop (EL2).[5–9]

In light of the pharmaceutical relevance of GPCRs, many efforts have been devoted and continue to be devoted to the experimental solution of their three-dimensional structures.[10–14] At the time of this writing, a total of 321 structures have been solved for 62 unique GPCRs in complex with a variety of ligands or, in rare cases, in their unliganded state (for statistics and a table listing the GPCR structure deposited in the Protein Data Bank, see the GPCRdb website at https://gpcrdb.org/structure/statistics and https://gpcrdb.org/structure/).[15]

At the same time, to enable the use of the experimentally solved structures as tools to gather insights into the structural features of the entire GPCR family, which counts over 1,000 members, efforts have been devoted to test the suitability of homology modeling for the construction of accurate models of the receptors for which experimental structures are not available. Prior to 2007, crystal structures had been published only for bovine rhodopsin. In

this context, rhodopsin-based models could be built and used for ligand-discovery purposes.[12, 16, 17] Nevertheless, there was no way to ultimately validate their accuracy in the absence of experimental structures of the modeled receptors that could serve as a reference. However, soon after the publication of the first crystal structures of the β_2AR,[7, 8] we published the first study that showed that models of this receptor endowed with a good level of accuracy could be built through rhodopsin-based homology modeling. In particular, our models showed a root mean square deviation (RMSD) of about 2 Å with respect to the backbone of the TMs of the reference β_2AR structure.[18, 19]

As GPCR structural studies expanded, a number of blind assessments have been performed in concomitance with the solution of new GPCR structures, aimed at gauging the ability of molecular modelers to forecast the structures of specific GPCRs before their solution. Overall, these assessments, together with additional studies conducted by other investigators, highlighted that high accuracy can be achieved for models based on closely related templates, while only macroscopic folding could be predicted for models based on distant templates.[20–26] Along these lines, in a recent paper, we conducted a systematic controlled homology modeling study by building models of the β_2AR based on 23 different structural templates. This study highlighted a significant correlation between template/target receptor sequence identity and accuracy of the resulting models.[27]

How accurate do β_2AR structures need to be to effectively serve as platforms for docking-based virtual campaigns? To answer this research question, here we targeted through controlled docking-based virtual screening experiments the above-mentioned 23 homology models of the β_2AR. Notably, these models are a good platform for our study, as they are endowed with different levels of structural accuracy as a consequence of the fact that they were built on the basis on different templates. The study began with the docking, at each of the 23 β_2AR models, of a molecular dataset constituted by 25 blockers and 3000 decoys, i.e. molecules endowed with high molecular similarity with respect to the ligands. Subsequently, we evaluated the virtual screening performance in each of the docking runs and studied the correlation between said virtual screening performance and the structural accuracy of the targeted models, measured relatively to the entire structure, the ligand-binding cavity, or specific domains. Finally, given that we have previously demonstrated that there is a measurable correlation between structural accuracy and template/target receptor sequence identity, we evaluated the possible presence of a correlation between virtual screening performance and the sequence identity shared by the β_2AR and the template receptors on which the models were based.

METHODOLOGY

Reference structure.

A crystal structure of the β_2AR co-crystallized in its inactive state in complex with the inverse agonist carazolol (PDB ID: 2RH1)[7, 8] was utilized as a reference structure to compare the performance of the models in docking-based virtual screening and to evaluate their structural accuracy, as detailed in the sections below. Several structures have been published for the β_2AR, in complex with a number of agonists and blockers (https://gpcrdb.org/structure/). As we previously described, we chose the 2RH1 structure as a

reference structure because it is the highest resolution (2.4 Å) structure of the β_2AR solved in the inactive state.[27]

Homology models.

The homology models targeted by the docking-based virtual screening campaigns described in this report were previously published by us.[27] The alignment of the seven TMs was performed through the structure-based alignment tool available at the GPCRdb website (http://gpcrdb.org/alignment/targetselection).[15, 28] Extracellular and intracellular loops were aligned as described in our work.[27]

Sequence identity between β_2AR and templates.

The percentages of sequence identity shared by the β_2AR and the each of the 23 homology modeling templates were calculated through Modeller,[29] by means of an "in-house" written Python script, on the basis of the sequence alignment previously published by us in the article in which we reported the construction of the models.[27] Sequence identities were calculated comparing the amino acid sequence of the 2RH1 structure with that of each template receptor, excluding from the calculation the alignment positions containing gaps. Sequence identities were calculated with respect to the whole sequence, the seven TMs, as well as individual domains, including the ligand binding cavity. For sequence identity calculations (as well as RMSD calculations), the ligand-binding cavity of the β_2AR was defined as the ensemble of residues located within 5 Å from the co-crystallized inverse agonist carazolol in the 2RH1 reference structure. Specifically, the ligand-binding cavity region comprised the following residues: Trp 109, Thr 110, Asp 113, Val 114, Val 117, Thr 118, Phe 193, Thr 195, Tyr 199, Ala 200, Ser 203, Ser 204, Ser 207, Trp 286, Phe 289, Phe 290, Asn 293, Tyr 308, and Asn 312, Tyr 316 (Figure 1). Further details regarding the definition of the other domains can be found in the article in which we reported the construction of the models.[27]

Root mean square deviation (RMSD) of the atomic coordinates between β_2AR models and crystal structure.

The structural RMSD values between the 23 β_2AR models and 2RH1 structure were previously published by us in the article in which we reported the construction of the models.[27] In that article, we reported RMSD values relative to both the α carbons and all heavy atoms (including backbone and sidechains). The latter were used in this work. Briefly, as described, the RMSD measurements were performed through Modeller,[29] by means of an "in-house" written Python script. RMSD values were calculated with respect to the whole sequence, the TMs, as well as individual domains, including the ligand binding cavity. Prior to the calculation of the RMSD values, the script performed a superposition of the α carbons of each homology model with the corresponding α carbons of the reference structure, excluding atom pairs with an RMSD of 5 Å or higher. The fitting procedure was repeated until there were no changes in the number of included atom pairs, prior to the final calculation of the RMSD values. The RMSD values relative to the ligand-binding cavity, which in the original article were obtained following the same procedure, here were recalculated after superposing exclusively the α carbons of the residues lining the binding cavity. See *Sequence identity between β_2AR and templates* for the definition of the ligand

binding cavity. Further details regarding the definition of the other domains can be found in the article in which we reported the construction of the models.[27]

Preparation of the receptor structures for docking-based virtual screening.

The reference crystal structure and the 23 homology models of the β_2AR were prepared for molecular docking with the Protein Preparation Wizard workflow, as implemented in the Schrödinger suite.[30, 31] Through this workflow, we added hydrogen atoms, calculated the protonation state of ionizable groups at pH 7, and optimized the orientation of hydroxyl groups, as well as Asn, Gln, and His residues. Finally, the structures were subjected to a restrained minimization using the OPLS3 force field and the Impact molecular mechanics engine, allowing a maximum RMSD deviation of 0.30 Å for the heavy atoms.

Ligands.

The known β_2AR ligands employed for the controlled docking-based virtual screening experiments consisted of a set of 25 known β_2AR blockers (antagonists and inverse agonists). Specifically, these were the same blockers that we used in previous controlled virtual screening studies targeting the β_2AR,[32] namely: acebutolol, AH-3474A, alprenolol, bevantolol, bupranolol, carazolol, carvedilol, CGP-12177, cicloprolol, dichloroisoproterenol, ICI-118551, ICI-89406, labetalol, metopropol (H-87), (RSS)-Nadolol, (SRRR)-nebivolol, (SRRS)-nebivolol, NIP, pindolol, pronethalol, S-propranolol, R-propranolol, sotalol, timolol, xamoterol – note: to avoid an overrepresentation of nebivolol structures, here we included only the two nebivolol isomers endowed with the highest affinity for the β_2AR, while is previous studies we docked a total of seven nebivolol isomers. Unless specified, for all blockers, the chiral carbon atom bearing the hydroxyl group was in the S chiral configuration. To treat ligands and decoys in the same way, the ligands were retrieved from the ZINC database (http://zinc.docking.org)[33, 34] in their reference protonation state at pH 7. The ligands that were not in the ZINC database were sketched from the closest ZINC entry and subsequently minimized with the MacroModel engine and the OPLS3 forcefield as implemented in the Schrödinger suite.[30]

Decoys.

The decoys employed for the controlled docking-based virtual screening experiments consisted of a set of 3000 molecules showing high similarity with the blockers described above, identified through a fingerprint-based screening conducted with the CANVAS platform of the Schrödinger suite.[30] First, MACCS fingerprints were calculated for each of the blockers as well as for about 8 million compounds from the ZINC database, in their reference protonation state at pH 7. Subsequently, through the canvasFPCombine function of CANVAS the individual fingerprints for the blockers were merged together into a single modal fingerprint.[35] The above-mentioned 8 million compounds from the ZINC database were then scanned to identify the compounds closest to the blockers, by comparing the molecular fingerprint of each ZINC compound with the modal fingerprint representing the β_2AR blockers. The 3000 ZINC compounds showing the highest Tanimoto coefficient with respect to the modal fingerprint were selected for use as decoys. The Tanimoto coefficients for the selected decoys ranged from to 0.67 to 0.59.

Identification of internal cavities for ligand docking.

Internal cavities in the receptor structures and models were identified through the SiteMap tool of the Schrödinger suite,[30, 36, 37] setting the search parameters as follows: grid spacing set at 1 Å, at least 15 points required per reported site; site maps cropped at 4.0 Å from nearest site point. Among all of the identified internal cavities, we selected for molecular docking the one comprised between the upper part of the TMs bundle and capped by the EL2 domain, which corresponds to the region to which orthosteric β_2AR ligands bind.

Molecular docking.

Receptor grids were created for the 2RH1 crystal structure and the 23 models using the Receptor Grid Generation tool in the Glide application of the Schrödinger suite.[30, 38–40] Each cubic grid was centered on the internal cavity, as identified through SiteMap, for the structure in question and was given a size that would allow the docking of a ligand as big as the entire cavity. No scaling factors were applied to the Vander der Waals (vdW) radii of the receptor atoms. Molecular docking was then performed at the 2RH1 crystal structure and each of the 23 models using the Standard Precision (SP) scoring function (Glide SP), recording the top scoring pose for each ligand. Ligands and decoys were allowed full flexibility, while the receptors were held rigid. In addition, a scaling factor of 0.8 was applied to the atoms of ligands and decoys endowed with a partial charge lower than 0.15*e*. The Glide SP poses of the docked ligands and decoys were further subjected to rescoring with the Extra Precision (XP) scoring function. The rescoring was done with both the "score in place" method, which rescored the docking complexes without allowing a geometric optimization of the ligand pose (Glide SP/XP-score-in-place), as well as the "refine" method, which rescored the docking complexes after a geometric optimization of the ligand pose (Glide SP/XP-refine).

Receiver operating characteristic (ROC) analyses.

ROC analyses were performed with the R package "enrichvs",[41, 42] treating ligands as true positives and decoys as false positives. After plotting the true positive fraction on the y-axis and the false positive fraction on the x-axis, the area under the resulting curve was calculated (ROC AUC). For an ideal virtual screening that ranks all ligands higher than decoys, the ROC analysis will yield a rectangular curve passing through the upper left corner of the plot, and the ROC AUC will have a value of 1. A virtual screening that ranks all decoys higher than ligands, will yield a rectangular curve passing through the lower right corner of the plot with an ROC AUC of 0. A virtual screening that randomly ranks ligands and decoys will yield a diagonal curve with an ROC AUC of 0.5.

Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) analyses.

BEDROC analyses were performed with the R package "enrichvs",[41, 42] treating ligands as true positives and decoys as false positives. BEDROC is a metric of virtual screening performance developed by Truchon et al., which emphasizes early recognition of ligands.[43] In BEDROC analysis, an α coefficient determines the weight put on early recognition, with higher values of α causing higher weight on early recognition. In our work, we used an α

coefficient of 160.9, which causes 80% of the BEDROC score to be determined by the top 1% of the scored dataset.[43, 44] Just as for ROC AUC, a virtual screening that ranks all ligands higher than decoys will yield a BEDROC value of 1, while a virtual screening that ranks all decoys higher than ligands will yield a BEDROC value of 0. The BEDROC value for a virtual screening that randomly ranks ligands and decoys depends on a variety of factors, including the α coefficient, the number of ligands, and the number of decoys. With our dataset of ligands and decoys and an α coefficient of 160.9, random ranking will yield a BEDROC value of 0.07.

Enrichment factors.

Enrichment factors within the top 1%, 5%, and 10% of the screened dataset (EF1, EF5, and EF10, respectively) were calculated with the R package "enrichvs".[41, 42] Essentially, EF values are obtained by taking the ratio of ligands over total number of compounds (ligands plus decoys) found within the selected percent of highest ranking compounds and dividing it by the same ratio calculated for the entire screened dataset. A virtual screening that ranks all ligands higher decoys will yield an EF1 of 100, an EF5 of 20, and an EF10 of 10. A virtual screening that ranks all decoys higher than ligands will yield EF values of 0. A virtual screening that randomly ranks ligands and decoys will yield EF values of 1.

Confidence intervals for virtual screening performance metrics.

For each virtual screening campaign 2000 iterations of bootstrap resampling of the rankings was performed through the R package "boot." Based on the bootstrapping, 95% confidence intervals for each performance metric were subsequently calculated.[41, 44, 45] Based on the calculated 95% confidence intervals, we considered a virtual screening result confidently above random expectation when the lower boundary of the 95% confidence interval exceeded the expectation for random ranking for the metric in question.

Plots, correlation analyses, calculations of mean and median values.

Linear regression analyses, calculation of the square of the correlation coefficients (R^2), as well as calculations of mean and median values were performed with R. Plotting of the data was performed with Microsoft® Excel®, except for Supporting Information Figures S4 and S8, where it was performed with R.

RESULTS

Virtual screening targets.

The controlled docking-based virtual screening campaigns object of this work targeted 23 homology models of the β_2AR previously published by us, which were built using 23 different receptor templates crystallographically solved in their inactive state.[27] For comparison, our docking-based virtual screening campaigns also targeted a reference crystal structure of the β_2AR crystallized in complex with the inverse agonist carazolol (PDB ID: 2RH1).[7, 8] Moreover, they also targeted the crystal structure of the dopamine D_3 receptor (PDB ID: 3PBL),[46] which, as illustrated below, is the template that yielded the best performing homology model.

As shown in Supporting Information Table S1, our models are endowed with different degrees of structural similarity with respect to the 2RH1 reference structure. In particular, within the ligand-binding cavity, our models displayed RMSD values with respect to the reference crystal structure ranging from 0.82 Å, for the model based on the β_1AR template, to 4.56 Å, for the model based on the $P2Y_{12}$ template – RMSD values calculated for all heavy atoms, including backbone and sidechains. Moreover, our models are based on templates endowed with different degrees of sequence identity with respect to the β_2AR. In particular, within the ligand-binding cavity, the templates on which our models are based displayed sequence identity values with respect to the β_2AR ranging from 95%, for the β_1AR template, to 5%, for the $P2Y_{12}$, CXCR4, and bovine rhodopsin templates.

Decoys.

For none of the selected decoys, a biological activity at the β_2AR has been recorded in the ChEMBL database (https://www.ebi.ac.uk/chembl/).[47] This does not exclude the possibility that some of the decoys might be found active at the β_2AR if experimentally tested. However, it is safe to assume that the majority of the decoys are *bona fide* devoid of β_2AR activity and can therefore be considered false positives for the purposes of our controlled virtual screening campaigns.

Virtual screening performance.

Following each docking run, we gauged the virtual screening performance by calculating the area under the receiver operating characteristic curve (ROC AUC), the area under the Boltzmann-enhanced discrimination of receiver operating characteristic curve (BEDROC), with an α coefficient of 160.9, and the enrichment of ligands within the top scoring 1%, 5%, and 10% of the docked molecular dataset (EF1, EF5, and EF10) – see the Methodology section for more details on the calculations and the interpretation of the results. While the ROC AUC provides a measure of the general prioritization of ligands versus decoys (true positives versus false positives) across the entire screening, BEDROC and enrichment factors provide a measure of the prioritization of ligands versus decoys within the compounds that received the highest scores – as noted in the Methodology section, an α coefficient of 160.9 causes 80% of the BEDROC value to be determined by the top 1% of the scored dataset. Hence, given that the ultimate scope of real-life virtual screening campaigns is to select a number of top scoring compounds and subject them to experimental testing with the hope of finding as many active compounds as possible, BEDROC and EFs are particularly relevant indicators of virtual screening performance.

The results of our analysis of the performance of the virtual screening campaigns are summarized in Table 1 and Supporting Information Tables S2–S3. In particular, the data provided in Table 1 are relative to the ranking yielded by Glide SP docking followed by Glide XP rescoring with a geometric optimization of the ligand pose (Glide SP/XP-refine). Conversely, the data provided in the analogous Supporting Information Tables S2–S3 are relative to the rankings yielded by Glide SP docking without further rescoring (Glide SP) and Glide SP docking followed by Glide XP rescoring without a geometric optimization of the ligand pose (Glide SP/XP-score-in-place), respectively.

The relative effectiveness of the three scoring methods used in this work can be gauged by comparing the mean and median values of the virtual performance metrics across all the β_2AR models reported in the bottom rows of Table 1 and Supporting Information Tables S2–S3 (to facilitate the comparison, a bar graph rendition of the data is provided in Supporting Information Figure S1). As the data reveals, Glide SP/XP-refine performed consistently better than Glide SP/XP-score-in-place and Glide SP without rescoring across all metrics. The only exception was noted for the ROC AUC, where Glide SP/XP-score-in-place performed slightly better than Glide SP/XP-refine.

As Table 1 illustrates, in terms of ROC AUC, the performance ranged from a maximum 0.95, for the model based on the dopamine D_3 receptor, to a minimum of 0.35, for the models based on the $P2Y_1$ receptor. The model based on the β_1AR ranked second in terms of performance, with an ROC AUC value of 0.90. The results followed a similar trend for BEDROC and enrichment factors, with the model based on the dopamine D_3 receptor and the β_1AR consistently performing better than the others, with the only exception of the model based on the CCR5 receptor, which outperformed the one based on the β_1AR in terms of EF5. For comparison purposes, Table 1 includes also the virtual screening performance registered with the 2RH1 crystal structure of the β_2AR, which, as mentioned, we used as a reference structure throughout the study. As the table illustrates, the two homology models that yielded the best performance results in virtual screening, i.e. the models based on the dopamine D_3 receptor and β_1AR, yielded results very close, and sometimes superior, to those yielded by the β_2AR crystal structure for all of the metrics.

Finally, to gauge to which extent our homology models of the β_2AR were more suited to distinguish β_2AR ligands and decoys than the templates on which the models were based, we docked our dataset of β_2AR ligands and decoys to the crystal structure of the dopamine D_3 receptor, which is the template that yielded the overall best performing β_2AR homology model. As the data shown in Table 2 indicate, the D_3-based β_2AR homology model consistently performed better than the D_3 crystal structure itself in prioritizing β_2AR ligands over decoys, for all the metrics and all the scoring methods.

Comparison with other datasets of ligands and decoys.

To gauge how the virtual screening performance registered with our dataset of ligands and decoys compared with those that would be obtained using alternative datasets of ligands and decoys, we docked at the β_2AR crystal structure (PDB ID: 2RH1) alternative sets of β_2AR ligands and decoys published by Weiss et al. (GPCR-Bench) and by Cavasotto et al. (GLL/GDD).[48, 49] The GPCR-Bench and GLL/GDD datasets are both characterized by a higher number of ligands with respect to our dataset. In particular, the GPCR-Bench and GLL/GDD datasets comprise 207 and 204 ligands, respectively, while ours comprises 25 ligands. The decoys/ligands ratios are 62 decoys per ligand for the GPCR-Bench, 40 decoys per ligand for the GLL/GDD dataset, and 121 decoys per ligand for our dataset. The methods at the basis of the selection of the decoys are also significantly different. In particular, we selected the decoys on the basis of highest topological similarity, as calculated through molecular fingerprinting. Conversely, the decoys for GPCR-Bench and GLL/GDD datasets were selected on the basis of molecular properties (molecular weight, estimated water−octanol

partition coefficient, rotatable bonds, net charge and, for GPCR-Bench only, hydrogen bond acceptors/donors) and highest topological dissimilarity (selecting only the decoys with the most dissimilar molecular fingerprints with respect to the ligands). As shown in Figure 2 for the Glide SP/XP-refine scoring method and in Supporting information Figures S2–S3 for Glide SP and Glide/XP-score-in-place, the three datasets of ligands and decoys appear similarly challenging, although there are some differences. In particular, the GLL/GDD dataset is generally the most challenging for the majority of the metrics and the majority of the scoring methods. Conversely, our dataset is generally the least challenging. This trait is particularly evident with respect to the EF5 and EF10 methods.

Correlation between virtual screening performance and structural accuracy of the models.

After gauging the performance of each docking-based virtual screening campaign, we studied the correlation between said performance and the structural accuracy of the models, assessed with respect to the entire structure, the TMs bundle, the EL2 domain, and the ligand-binding cavity – RMSD values calculated for heavy atoms, including backbone and side chains (see Methodology section for further details).

A complete set of plots of all virtual screening performance metrics versus the structural accuracy of the models is provided in Supporting Information Figure S4 for all the scoring methods. A bar graph representation of the resulting R^2 values is provided in Figure 3 for the ROC AUC and the BEDROC metrics, while analogous bar graphs for the enrichment factors (EF1, EF5, and EF10), are provided in Supporting Information Figures S5–S7.

As Figure 3 and Supporting Information Figures S5–S7 indicate, for all the metrics, the highest degree of correlation between virtual screening performance and structural accuracy of the models was found for the ligand-binding cavity and the EL2 domain. Overall, for both the structural accuracy of the ligand-binding cavity and EL2, there was a significantly stronger correlation with the ROC AUC than with the indicators of early enrichment (BEDROC and EFs). In particular, the highest R^2 values were detected for the correlation between the ROC AUC and the structural accuracy of the ligand-binding cavity – 0.55 and 0.45 for Glide SP/XP-score-in-place and Glide SP/XP-refine, respectively (Figure 4) – followed by the R^2 values for the correlation between the ROC AUC and the structural accuracy of EL2 – 0.44 and 0.41 for Glide SP/XP-score-in-place and Glide SP/XP-refine, respectively (Figure 5).

In terms of scoring methods, we found that Glide SP/XP-score-in-place and Glide SP/XP-refine performed generally better than Glide SP without rescoring. We also found that Glide SP/XP-score-in-place performed slightly better than Glide SP/XP-refine for the correlation with ROC AUC (Figure 3 panel A, and top row panels for Figures 4 and Figure 5), EF5 (Supporting Information Figure S6), and EF10 (Supporting Information Figure S7). Conversely, Glide SP/XP-refine performed slightly better than Glide SP/XP-score-in-place for the correlation with BEDROC (Figure 3 panel B, and bottom row panels for Figures 4 and Figure 5) and EF1 (Supporting Information Figure S5).

Correlation between virtual screening performance and template/β_2AR sequence identity.

Lastly, we studied the correlation between the performance of the docking-based virtual screening campaigns and the sequence identity between β_2AR and the template on which the models targeted by each campaign were based. Just as for the structural accuracy, we calculated the sequence identity values for the whole receptor, the TMs, EL2, and the ligand-binding cavity.

A complete set of plots of all virtual screening performance metrics versus the sequence identity between the β_2AR and the templates on which the models were built is provided in Supporting Information Figure S8 for all the scoring methods. A bar graph representation of the resulting R^2 values is provided in Figure 6 for the ROC AUC and the BEDROC metrics. Analogous graphs for the enrichment factors (EF1, EF5, and EF10), are provided in Supporting Information Figures S9–S11. Of note, the correlation coefficients shown in these bar graphs were obtained excluding from the calculation the β_1AR, which is indeed an outlier in terms of its very high sequence identity with respect to the β_2AR (for instance, with respect to the overall sequence, the β_1AR shares 64% of identity with the β_2AR, while the sequence identity displayed by the other templates ranges between 16% and 37%).

As Figure 6 and Supporting Information Figures S9–S11 indicate, we did not find a strong correlation between the performance of our docking-based virtual screening campaigns targeting β_2AR homology models and template/β_2AR sequence identity, with all calculated R^2 values being lower than 0.3. In fact, as evident from the plots shown in Figures 7–8 and Supporting Information Figure S8, there is not a linear correlation between the template/β_2AR sequence identity and the virtual screening performance of the resulting models. Rather, the models based on the β_1AR and the dopamine D_3 receptor – *i.e.* the two templates that share the highest sequence identity with the β_2AR overall, in the TMs, and in the ligand binding cavity – yielded excellent levels of virtual screening performance in terms of ROC AUC (Figures 7–8, panels A–B) as well as good levels of virtual screening performance in terms indicators of early enrichment, namely BEDROC (Figures 7–8, panels C–D) and EFs (Supporting Information Figure S8, panels 29–40 and panels 49–60) for both the Glide SP/XP-score-in-place and the Glide SP/XP-refine scoring methods. Conversely, for the remainder of the models, which are based on templates that share lower percentages of sequence identities with the β_2AR, no correlation between sequence identity level and virtual screening performance was detected. The separation between the two models that consistently performed well and the others is particularly evident for the ligand binding cavity, where the β_1AR and the dopamine D_3 share 95% and 55% of sequence identity with the β_2AR, respectively, while the remainder of the templates share sequence identities with the β_2AR ranging from 5% to 30% (Figure 7). The same trend is maintained for the sequence identity calculated for the TMs (Figure 8) or the overall sequence (Supporting Information Figure S8), where the β_1AR and the dopamine D_3 also share higher sequence identities with the β_2AR, although the separation between the β_1AR and D_3 receptors and the rest of the templates in terms of sequence identity with respect to the β_2AR is not as pronounced as that calculated within the ligand-binding cavity.

DISCUSSION

As evident from Table 1 and Supporting Information Tables S2–S3, a first notable result of this study is that, in our docking-based virtual screening campaigns, over half of the β_2AR homology models yielded ROC AUC values confidently above the expectation for random ranking (lower boundary of the 95% confidence interval above the expectation for random ranking) for all of the tested scoring methods. Moreover, with the best performing scoring method (Glide SP/XP-refine), over one third and one quarter of the β_2AR homology models yielded enrichment factors (EF1, EF5, and EF10) and BEDROC (α = 160.9) values confidently above the expectation for random ranking, respectively. Specifically, with the Glide SP/XP-refine scoring method, out of the 23 tested models, we registered virtual screening performance values confidently above random ranking expectations for 14 models in terms of ROC AUC, 6 in terms of BEDROC (α = 160.9), 8 in terms of EF1, 10 in terms of EF5, and 11 in terms of EF10 (Table 1).

These findings are in line with de Graaf and Rognan's observation that, despite their structural inaccuracies, GPCR models are generally useful for the identification of novel ligands.[50] Of note, our control experiment in which we docked our dataset of β_2AR ligands and decoys to the crystal structure of the dopamine D_3 receptor yielded virtual screening performance levels inferior to those obtained with the D_3–based β_2AR homology model, thus suggesting that, in the absence of the structure of a receptor, GPCR homology modeling indeed brings added value to virtual screening (Table 2). This is reasonable, in consideration of the fact that 9 out of the 20 residues that line the β_2AR binding cavity are not conserved in the dopamine D_3 receptor – for a schematic diagram of the β_2AR ligand binding cavity and the corresponding residues in the dopamine D_3 receptor, see Supporting Information Figure S12.

Not surprisingly, the data shown in Figures 3–5 indicate that there is a detectable linear correlation between the performance of the models, especially in terms of ROC AUC, and the structural accuracy of the portions of the receptor directly involved in ligand binding. In particular, for all the metrics, the highest R^2 values were detected for the ligand-binding cavity as well as the EL2 domain, which, as known, constitutes the upper lining of the β_2AR ligand-binding cavity.

The model based on the dopamine D_3 receptor, which, as the β_2AR, belongs to the family of aminergic receptors, showed the best virtual screening performance in all the metrics for both Glide SP/XP-refine and Glide SP/XP-score-in-place (Table 1 and Supporting Information Table S3). This is in line with the fact that this template yielded the most accurate model with respect to both the ligand-binding cavity and the EL2 domain, even higher than the structural accuracy of the model based on the β_1AR, which is closer to the β_2AR in terms of sequence identity. Just as this study indicates that the D_3 receptor is a good template for the construction of β_2AR, previously published studies highlighted that, in turn, the β_2AR is a good template to build models of the D_3 receptor. In particular, our finding is consistent with work published by Shoichet and coworkers that highlighted the good performance of prospective docking-based virtual screening campaigns targeting homology models of the dopamine D_3 receptor based on the β_2AR.[51] Our finding is also consistent

with the results of the GPCR Dock 2010 assessment, which highlighted that accurate models of the dopamine D_3 receptor in complex with the antagonist eticlopride could be built on the basis of adrenergic receptor templates.[22]

The fact that we found no linear correlation between template/β_2AR sequence identity and performance of the β_2AR models in virtual screening (Figure 6) is not entirely surprising. As this study demonstrates, the region whose structural accuracy affects the most virtual screening performance is the ligand-binding cavity. However, as we have previously reported, the correlation between the structural accuracy of the model of the ligand-binding cavity and template/β_2AR sequence identity calculated for the same cavity, although detectable, is not strong ($R^2 = 0.41$). Consequently, although we found a measurable correlation between structural accuracy and virtual screening performance for the ligand-binding cavity, it is reasonable that this observation did not translate into a linear correlation between the template/β_2AR sequence identity relative to the cavity and the virtual screening performance of the models.

Despite this lack of linear correlation, our work shows that, indeed, a high sequence identity relative to the ligand-binding cavity between template and modeled receptor translates into a good virtual screening performance of the resulting β_2AR homology model (Figure 7). In particular, the two models based on the two templates with particularly high sequence identity with respect to the β_2AR within the ligand-binding cavity, namely the β_1AR and the dopamine D_3 receptor, performed very well in virtual screening, especially in terms of ROC AUC. Conversely, models based on templates with lower sequence identity showed a wide spectrum of virtual screening performances, with no apparent correlation with the template/β_2AR sequence identity. As it is evident from Figure 7, there is a wide separation, in terms of percentage of sequence identity relative to the ligand-binding cavity, between the two models that performed well in virtual screening and those with unpredictable performance. In particular, the β_1AR and D_3 receptors show a sequence identity with the β_2AR of 95% and 55%, respectively, while all the others show sequence identities ranged from 5% to 30%.

Similarly, our work shows that a high sequence identity relative to the overall structure of the receptor or the TM bundle between template and modeled receptor translates into a good virtual screening performance of the resulting β_2AR homology models (Figure 8 and Supporting Information Figure S8). Again, the two models based on the β_1AR and the dopamine D_3 receptor, which show the highest sequence identity with respect to the β_2AR relative to the overall structure and the TMs, performed very well in virtual screening, especially in terms of ROC AUC. Conversely, just as noted for the ligand-binding cavity, models based on templates with lower sequence identity showed a wide spectrum of virtual screening performance, with no apparent correlation with the template/β_2AR sequence identity. Unlike for the ligand-binding cavity, there appears to be a rather small separation in terms of the percentage of sequence identity between the templates that yielded the two models that performed well in virtual screening and those that yielded models with unpredictable performance. With respect to the TM bundle, the β_1AR and D_3 receptors show a sequence identity with the β_2AR of 67% and 41%, respectively, while all the others show sequence identities comprised between 38% and 17% (Figure 8). With respect to the overall

structure of the receptor, the β_1AR and D_3 receptors show a sequence identity with the β_2AR of 64% and 37%, respectively, while all the others show sequence identities comprised between 34% and 16%. In light of these observations, it can be concluded that high template/β_2AR sequence identity relative to the overall structure of the receptor, the TM bundle, and, especially, the ligand-binding cavity can be taken as an indication of likely good virtual screening performance, in particular in terms of ROC AUC, of the resulting β_2AR homology models.

An important caveat is that the present work is a case study relative to the β_2AR. In light of the generality of the domains on which we focused our analysis, namely the TMs bundle, the ligand-binding cavity, and EL2, our study might be relevant to other GPCR systems as well. In particular, we expect it to be relevant to those GPCRs that bind small-molecule ligands in a cavity enclosed within the TMs bundle and capped by EL2, as it is the case for several class A receptors.[52] For instance, our conclusion that the overall sequence identity between template and target receptor is not a good predictor of virtual screening performance is in line with what was reported by Rataj and coworkers in a study focused on homology models of four serotonin receptors. In light of their findings, the authors hypothesized that “the three-dimensional structure of the binding pocket should bear more significance [for virtual screening performance] than overall sequence similarity.”[53] Our results entirely support and corroborate this hypothesis.

This work was not intended to probe the effectiveness of optimized β_2AR homology models or to study how to maximize virtual screening performance through the incorporation of external information. Rather, it was meant to provide an assessment of the performance of crude β_2AR homology models and determine how this is affected by the structural accuracy of the models and the sequence identity between template and modeled receptors. However, it is worth noting that it is generally recognized that the effectiveness of GPCR models as platforms for virtual screening significantly improves when the models are optimized in complex with a prototypical ligand, for instance through molecular dynamics or conformational searches, or when experimental information is incorporated in the screening, for instance through the application of filters or fingerprints meant to capture plausible receptor-ligand interactions.[25, 50, 54–56] Indeed, a recent article from Loo and coworkers, in which homology models of eight different GPCRs were evaluated for their virtual screening performance, pointed out that geometrically optimizing the models around the structure of a docked ligand significantly improved their performance in docking-based virtual screening. The authors proved this by docking into their homology models the ligand co-crystallized with a reference structure of the modeled receptor through a procedure that accounts for the flexibility of both ligands and receptors (induced fit docking). It is reasonable to expect that similar results would be obtained repeating the same procedure with different ligands of the studied receptors.[57]

As a final point, it should be noted that, as for all controlled virtual screening campaigns, the performance levels that we registered in this study are highly dependent on the ligands and decoys that we employed. This is well illustrated by the comparison of the performance of docking-based virtual screening campaigns targeting the β_2AR conducted with our dataset of ligands and decoys as well two alternative datasets of ligands and decoys previously

published by other authors (GPCR-Bench and GLL/GDD).[48, 49] In particular, the comparison shows that our dataset is somewhat less challenging than the other two. In light of these consideration, it should not be discounted that the performance levels that would be registered in prospective docking-based virtual screening campaigns targeting our β_2AR models might be lower than those reported in Table 1 and Supporting Information Tables S2–S3.

SUMMARY AND CONCLUSIONS

Our study demonstrates that controlled docking-based virtual screening campaigns targeting homology models of the β_2AR, in the majority of the cases, yielded results that confidently exceeded random expectations in terms ROC AUC (Table 1 and Supporting Information Tables S2–S3). Moreover, with the best performing scoring method, over one third and one quarter of the models yielded results that confidently exceeded random expectation also in terms of enrichment factors (EF1, EF5, and EF10) and BEDROC (α = 160.9) (Table 1). Not surprisingly, we found a detectable correlation between virtual screening performance and the structural accuracy of the ligand-binding cavity (Figure 4). Moreover, we also found a detectable, although lower, linear correlation between virtual screening performance and structural accuracy of the EL2 domain (Figure 5). As our data indicate, in both cases, the linear correlation was stronger when virtual screening performance was assessed in terms of ROC AUC values, i.e. in terms of prioritization of ligands over decoys over the entire screened dataset, than when it was assessed in terms of early enrichment metrics. Finally, our data indicate that, although there is no detectable linear correlation between virtual screening performance and template/β_2AR sequence identity, models built on the basis of templates that show high sequence identity with the β_2AR, especially within the ligand-biding cavity, performed consistently well. Conversely, models based on templates with lower sequence identity displayed performance levels that ranged from very good to random, with no apparent correlation with the sequence identity itself (Figures 7–8).

In conclusion, this study identifies the characteristics of the β_2AR homology models that determine their virtual screening performance. The lack of linear correlation between template/β_2AR sequence identity implies that, in most cases, it will not be possible to predict on the basis of sequence alignment considerations whether or not a given model will perform well in docking-based virtual screening. However, our data on the performance of the studied β_2AR homology models suggests that when the sequence identity between template and target receptor is high, especially within the ligand-binding cavity, the resulting models will perform well as docking-based virtual screening platforms. Conversely, when the sequence identity is lower, predicting the performance of the campaigns becomes unfeasible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This research was supported by: American University and the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award R15GM119084; The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Computing resources were provided by the American University High Performance Computing System, which is funded in part by a grant from the National Science Foundation (BCS-1039497).

REFERENCES

1. Pierce K; Premont R; Lefkowitz R Seven-Transmembrane Receptors. Nat Rev Mol Cell Biol 2002, 3, 639–650. [PubMed: 12209124]
2. Overington JP; Al-Lazikani B; Hopkins AL How Many Drug Targets Are There? Nat Rev Drug Discov 2006, 5, 993–996. [PubMed: 17139284]
3. Lefkowitz RJ A Brief History of G-Protein Coupled Receptors (Nobel Lecture). Angew Chem Int Ed Engl 2013, 52, 6366–6378. [PubMed: 23650015]
4. Kobilka B The Structural Basis of G-Protein-Coupled Receptor Signaling (Nobel Lecture). Angew Chem Int Ed Engl 2013, 52, 6380–6388. [PubMed: 23650120]
5. Ahn S; Kahsai AW; Pani B; Wang QT; Zhao S; Wall AL; Strachan RT; Staus DP; Wingler LM; Sun LD; Sinnaeve J; Choi M; Cho T; Xu TT; Hansen GM; Burnett MB; Lamerdin JE; Bassoni DL; Gavino BJ; Husemoen G; Olsen EK; Franch T; Costanzi S; Chen X; Lefkowitz RJ Allosteric "Beta-Blocker" Isolated from a DNA-Encoded Small Molecule Library. Proc Natl Acad Sci U S A 2017, 114, 1708–1713. [PubMed: 28130548]
6. Rasmussen S; Choi H; Rosenbaum D; Kobilka T; Thian F; Edwards P; Burghammer M; Ratnala V; Sanishvili R; Fischetti R; Schertler G; Weis W; Kobilka B Crystal Structure of the Human Beta2 Adrenergic G-Protein-Coupled Receptor. Nature 2007, 450, 383–387. [PubMed: 17952055]
7. Cherezov V; Rosenbaum D; Hanson M; Rasmussen S; Thian F; Kobilka T; Choi H; Kuhn P; Weis W; Kobilka B; Stevens R High-Resolution Crystal Structure of an Engineered Human Beta2-Adrenergic G Protein-Coupled Receptor. Science 2007, 318, 1258–1265. [PubMed: 17962520]
8. Rosenbaum D; Cherezov V; Hanson M; Rasmussen S; Thian F; Kobilka T; Choi H; Yao X; Weis W; Stevens R; Kobilka B GPCR Engineering Yields High-Resolution Structural Insights into Beta2-Adrenergic Receptor Function. Science 2007, 318, 1266–1273. [PubMed: 17962519]
9. Ring AM; Manglik A; Kruse AC; Enos MD; Weis WI; Garcia KC; Kobilka BK Adrenaline-Activated Structure of Beta2-Adrenoceptor Stabilized by an Engineered Nanobody. Nature 2013, 502, 575–579. [PubMed: 24056936]
10. Jacobson KA; Costanzi S New Insights For Drug Design From The X-Ray Crystallographic Structures of G-Protein-Coupled Receptors. Mol Pharmacol 2012, 82, 361–371. [PubMed: 22695719]
11. Isberg V; De Graaf C; Bortolato A; Cherezov V; Katritch V; Marshall FH; Mordalski S; Pin JP; Stevens RC; Vriend G; Gloriam DE Generic Gpcr Residue Numbers - Aligning Topology Maps while Minding the Gaps. Trends Pharmacol Sci 2015, 36, 22–31. [PubMed: 25541108]
12. Costanzi S; Siegel J; Tikhonova I; Jacobson K Rhodopsin And The Others: A Historical Perspective on Structural Studies of G Protein-Coupled Receptors. Curr. Pharm. Des 2009, 15, 3994–4002. [PubMed: 20028316]
13. Thal DM; Vuckovic Z; Draper-Joyce CJ; Liang YL; Glukhova A; Christopoulos A; Sexton PM Recent Advances in the Determination of G Protein-Coupled Receptor Structures. Curr Opin Struct Biol 2018, 51, 28–34. [PubMed: 29547818]
14. Tautermann CS Gpcr Structures In Drug Design, Emerging Opportunities with New Structures. Bioorg Med Chem Lett 2014, 24, 4073–4079. [PubMed: 25086683]
15. Munk C; Isberg V; Mordalski S; Harpsoe K; Rataj K; Hauser AS; Kolb P; Bojarski AJ; Vriend G; Gloriam DE Gpcrdb: The G Protein-Coupled Receptor Database - An Introduction. Br J Pharmacol 2016, 173, 2195–2207. [PubMed: 27155948]

16. Moro S; Deflorian F; Bacilieri M; Spalluto G Ligand-Based Homology Modeling as Attractive Tool to Inspect GPCR Structural Plasticity. Curr. Pharm. Des 2006, 12, 2175–2185. [PubMed: 16796562]

17. Cavasotto CN; Orry AJ; Abagyan RA Structure-Based Identification of Binding Sites, Native Ligands and Potential Inhibitors for G-Protein Coupled Receptors. Proteins 2003, 51, 423–433. [PubMed: 12696053]

18. Costanzi S On The Applicability of GPCR Homology Models to Computer-Aided Drug Discovery: a Comparison between in Silico and Crystal Structures of the Beta2-Adrenergic Receptor. J Med Chem 2008, 51, 2907–2914. [PubMed: 18442228]

19. Costanzi S Modelling G Protein-Coupled Receptors: a Concrete Possibility. Chim Oggi 2010, 28, 26–30. [PubMed: 21253444]

20. Michino M; Abola E; Participants of GPCR Dock; Brooks CR; Dixon J; Moult J; Stevens R Community-Wide Assessment of GPCR Structure Modelling and Ligand Docking: GPCR Dock 2008. Nat Rev Drug Discov 2009, 8, 455–463. [PubMed: 19461661]

21. Kufareva I; Katritch V; Participants of GPCR Dock; Stevens RC; Abagyan R Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. Structure 2014, 22, 1120–1139. [PubMed: 25066135]

22. Kufareva I; Rueda M; Katritch V; Stevens RC; Abagyan R Status of GPCR Modeling and Docking as Reflected by Community-Wide GPCR Dock 2010 Assessment. Structure 2011, 19, 1108–1126. [PubMed: 21827947]

23. Phatak SS; Gatica EA; Cavasotto CN Ligand-Steered Modeling and Docking: a Benchmarking Study in Class A G-Protein-Coupled Receptors. J Chem Inf Model 2010, 50, 2119–2128. [PubMed: 21080692]

24. Bhattacharya S; Lam AR; Li H; Balaraman G; Niesen MJ; Vaidehi N Critical Analysis of the Successes and Failures of Homology Models of G Protein-Coupled Receptors. Proteins 2013, 81, 729–739. [PubMed: 23042299]

25. Cavasotto CN; Palomba D Expanding the Horizons of G Protein-Coupled Receptor Structure-Based Ligand Discovery and Optimization Using Homology Models. Chem Commun (Camb) 2015, 51, 13576–13594. [PubMed: 26256645]

26. Obiol-Pardo C; Lopez L; Pastor M; Selent J Progress In The Structural Prediction of G Protein-Coupled Receptors: D3 Receptor in Complex with Eticlopride. Proteins 2011, 79, 1695–1703. [PubMed: 21491496]

27. Costanzi S; Skorski M; Deplano A; Habermehl B; Mendoza M; Wang K; Biederman M; Dawson J; Gao J Homology Modeling Of A Class A Gpcr In The Inactive Conformation: A Quantitative Analysis of the Correlation Between Model/Template Sequence Identity and Model Accuracy. J Mol Graph Model 2016, 70, 140–152. [PubMed: 27723562]

28. Isberg V; Mordalski S; Munk C; Rataj K; Harpsoe K; Hauser AS; Vroling B; Bojarski AJ; Vriend G; Gloriam DE GPCRdb: an Information System for G Protein-Coupled Receptors. Nucleic Acids Res 2017, 45, D356–D364.

29. Sali A; Overington J Derivation of Rules for Comparative Protein Modeling From a Database of Protein Structure Alignments. Protein Sci 1994, 3, 1582–1596. [PubMed: 7833817]

30. Schrödinger Small-Molecule Drug Discovery Suite - Release 2017-4. Schrödinger,LLC, New York, NY, USA 2017 https://www.schrodinger.com (accessed June 14, 2019)

31. Sastry GM; Adzhigirey M; Day T; Annabhimoju R; Sherman W Protein And Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. J Comput Aided Mol Des 2013, 27, 221–234. [PubMed: 23579614]

32. Costanzi S; Vilar S In Silico Screening for Agonists and Blockers of the β_2 Adrenergic Receptor: Implications of Inactive and Activated State Structures. J Comput Chem 2012, 33, 561–572. [PubMed: 22170280]

33. Irwin JJ; Shoichet BK Zinc--A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Inf Model 2005, 45, 177–182. [PubMed: 15667143]

34. Irwin JJ; Sterling T; Mysinger MM; Bolstad ES; Coleman RG Zinc: a Free Tool to Discover Chemistry for Biology. J Chem Inf Model 2012, 52, 1757–1768. [PubMed: 22587354]

35. Duan J; Dixon SL; Lowrie JF; Sherman W Analysis and Comparison of 2D Fingerprints: Insights into Database Screening Performance Using Eight Fingerprint Methods. J Mol Graph Model 2010, 29, 157–170. [PubMed: 20579912]
36. Halgren T New Method for Fast and Accurate Binding-Site Identification and Analysis. Chem Biol Drug Des 2007, 69, 146–148. [PubMed: 17381729]
37. Halgren TA Identifying and Characterizing Binding Sites and Assessing Druggability. J Chem Inf Model 2009, 49, 377–389. [PubMed: 19434839]
38. Friesner RA; Banks JL; Murphy RB; Halgren TA; Klicic JJ; Mainz DT; Repasky MP; Knoll EH; Shelley M; Perry JK; Shaw DE; Francis P; Shenkin PS Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 2004, 47, 1739–1749. [PubMed: 15027865]
39. Halgren TA; Murphy RB; Friesner RA; Beard HS; Frye LL; Pollard WT; Banks JL Glide: a New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem 2004, 47, 1750–1759. [PubMed: 15027866]
40. Friesner RA; Murphy RB; Repasky MP; Frye LL; Greenwood JR; Halgren TA; Sanschagrin PC; Mainz DT Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. J Med Chem 2006, 49, 6177–6196. [PubMed: 17034125]
41. R Core Team. R: A Language and Environment for Statistical Computing - Version 3.2.3 (2015-12-10). R Foundation For Statistical Computing 2015 https://www.r-project.org/ (accessed June 14, 2019).
42. Yabuuchi H Package 'Enrichvs', Version 0.05. The Comprehensive R Archive Network 2011 https://cran.r-project.org/web/packages/enrichvs/ (accessed June 14, 2019).
43. Truchon JF; Bayly CI Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. J Chem Inf Model 2007, 47, 488–508. [PubMed: 17288412]
44. Lenselink EB; Jespers W; Van Vlijmen HW; Ap IJ; Van Westen GJ Interacting with GPCRs: Using Interaction Fingerprints for Virtual Screening. J Chem Inf Model 2016, 56, 2053–2060. [PubMed: 27626908]
45. Canty A; Ripley B Boot, Version 1.3-20. The Comprehensive R Archive Network 2017 https://cran.r-project.org/web/packages/boot/ (accessed June 14, 2019).
46. Chien EY; Liu W; Zhao Q; Katritch V; Han GW; Hanson MA; Shi L; Newman AH; Javitch JA; Cherezov V; Stevens RC Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. Science 2010, 330, 1091–1095. [PubMed: 21097933]
47. Gaulton A; Hersey A; Nowotka M; Bento AP; Chambers J; Mendez D; Mutowo P; Atkinson F; Bellis LJ; Cibrian-Uhalte E; Davies M; Dedman N; Karlsson A; Magarinos MP; Overington JP; Papadatos G; Smit I; Leach AR The ChEMBL Database in 2017. Nucleic Acids Res 2017, 45, D945–D954. [PubMed: 27899562]
48. Weiss DR; Bortolato A; Tehan B; Mason JS GPCR-Bench: a Benchmarking Set and Practitioners' Guide for G Protein-Coupled Receptor Docking. J Chem Inf Model 2016, 56, 642–651. [PubMed: 26958710]
49. Gatica EA; Cavasotto CN Ligand and Decoy Sets for Docking to G Protein-Coupled Receptors. J Chem Inf Model 2012, 52, 1–6. [PubMed: 22168315]
50. De Graaf C; Rognan D Customizing G Protein-Coupled Receptor Models for Structure-Based Virtual Screening. Curr Pharm Des 2009, 15, 4026–4048. [PubMed: 20028320]
51. Carlsson J; Coleman RG; Setola V; Irwin JJ; Fan H; Schlessinger A; Sali A; Roth BL; Shoichet BK Ligand Discovery from A Dopamine D3 Receptor Homology Model and Crystal Structure. Nat Chem Biol 2011, 7, 769–778. [PubMed: 21926995]
52. Lee Y; Basith S; Choi S Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. J Med Chem 2018, 61, 1–46. [PubMed: 28657745]
53. Rataj K; Witek J; Mordalski S; Kosciolek T; Bojarski AJ Impact of Template Choice on Homology Model Efficiency in Virtual Screening. J Chem Inf Model 2014, 54, 1661–1668. [PubMed: 24813470]

54. Kooistra AJ; Roumen L; Leurs R; De Esch IJ; De Graaf C From Heptahelical Bundle to Hits from the Haystack: Structure-Based Virtual Screening for GPCR Ligands. Methods Enzymol 2013, 522, 279–336. [PubMed: 23374191]

55. Ciancetta A; Cuzzolin A; Deganutti G; Sturlese M; Salmaso V; Cristiani A; Sabbadin D; Moro S New Trends in Inspecting GPCR-Ligand Recognition Process: the Contribution of the Molecular Modeling Section (MMS) at the University of Padova. Mol Inform 2016, 35, 440–448. [PubMed: 27546048]

56. Warszycki D; Rueda M; Mordalski S; Kristiansen K; Satala G; Rataj K; Chilmonczyk Z; Sylte I; Abagyan R; Bojarski AJ From Homology Models to a Set of Predictive Binding Pockets-A 5-HT1a Receptor Case Study. J Chem Inf Model 2017, 57, 311–321. [PubMed: 28055203]

57. Loo JSE; Emtage AL; Ng KW; Yong ASJ; Doughty SW Assessing GPCR Homology Models Constructed from Templates of Various Transmembrane Sequence Identities: Binding Mode Prediction and Docking Enrichment. J Mol Graph Model 2018, 80, 38–47. [PubMed: 29306746]

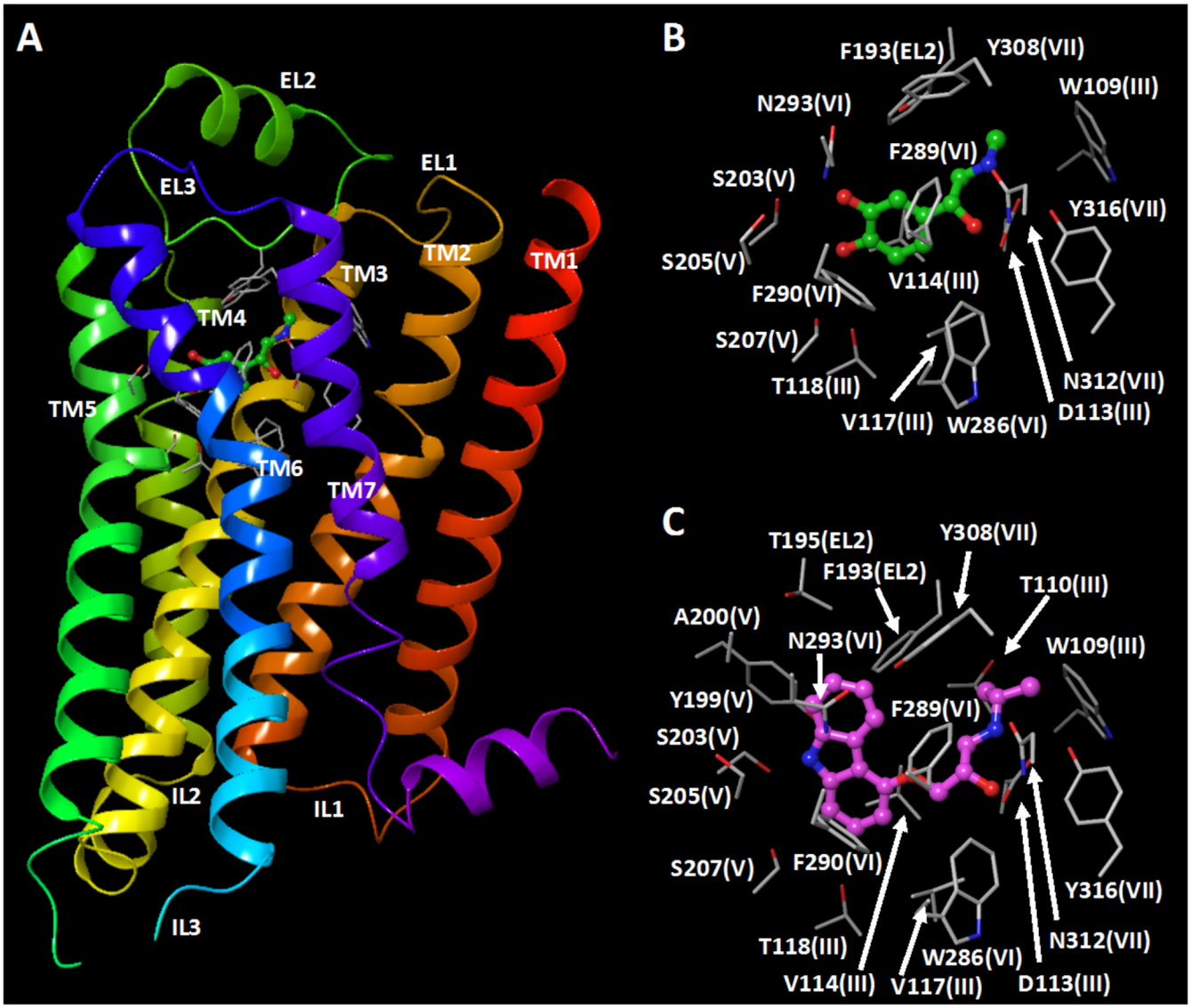


Figure 1.

Overall topology of the β_2AR bound to the agonist epinephrine (panel A, PDB ID: 4LDO), and details of the ligand-binding cavity with bound epinephrine (panel B, PDB ID: 4LDO) and the inverse agonist carazolol (panel C, PDB ID: 2RH1).[7–9] In panel A, the backbone structure of the receptor is represented as a ribbon, with a spectrum of colors ranging from red at the N-terminus to purple at the C-terminus. Labels indicate each of the seven TMs as well as the ELs and ILs. The agonist epinephrine is shown with green carbons (panels A and B). The inverse agonist carazolol is shown with magenta carbons. The residues lining the binding pocket for the bound ligands are shown with gray carbons (panels A-C).

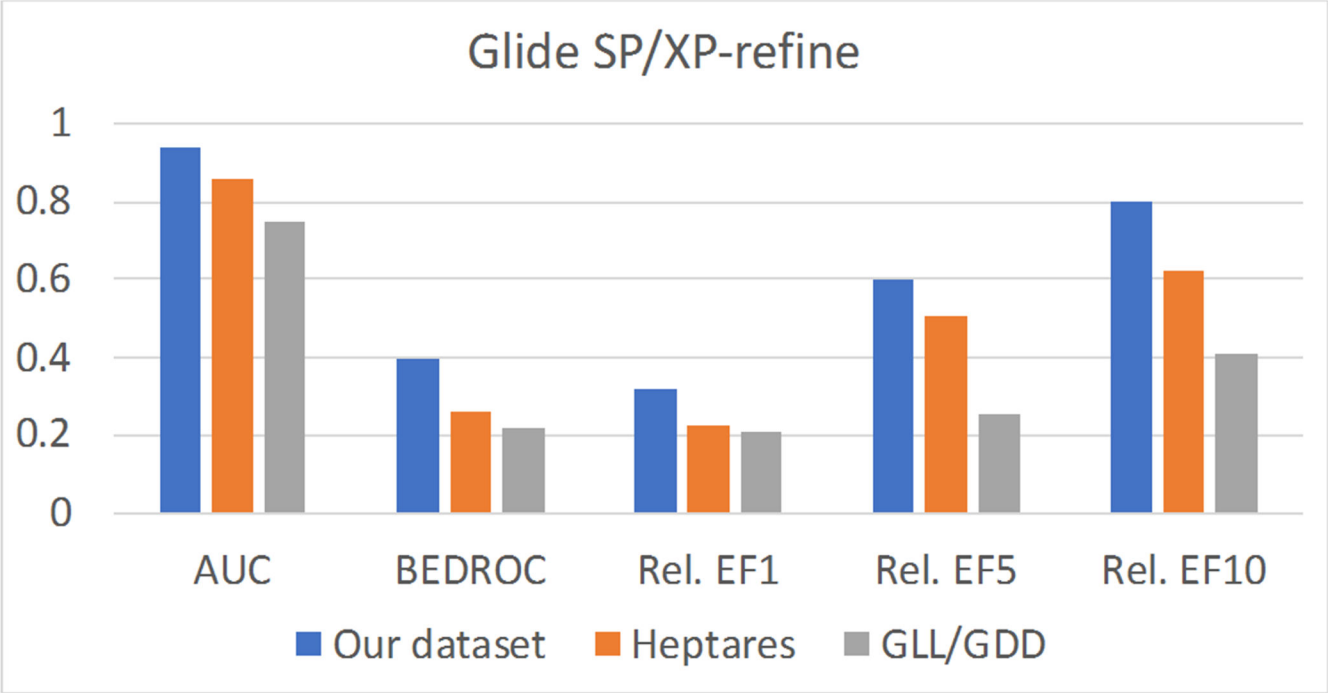


Figure 2.

Bar graph showing the performance of virtual screening experiments (Glide SP/XP-refine rankings) targeting a crystal structure of the β_2AR (PDB ID: 2RH1) using different sets of ligands and decoys. Blue bars refer to the dataset of ligands and decoys used in this work; orange bars refer to the Heptares GPCR-Bench dataset for the β_2AR; gray bars refer to the GLL/GDD dataset for the β_2AR. The presented performance metrics include ROC AUC, BEDROC (α = 160.9), and enrichment factors within the top 1% (EF1), 5% (EF5), and 10% (EF10) of the screened dataset. To facilitate the comparison, the enrichment factors are expressed as relative values with respect to the highest value for the metric achievable with each dataset (maximum EF1: 100 for our dataset, 62 for the GPCR-Bench dataset; 40 for the GLL/GDD dataset; maximum EF5: 20 for all datasets; maximum EF10: 10 for all datasets). For analogous figures relative to the data obtained with Glide SP rankings and Glide SP/XP-score-in-place rankings, see Supporting Information Figures S2–S3.

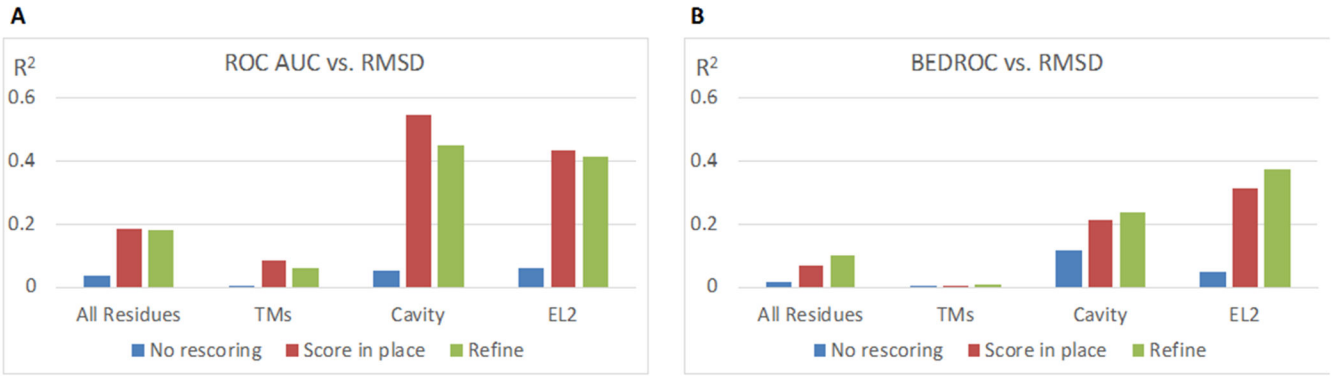


Figure 3.

Values of the square of the correlation coefficients (R^2) for the correlation between the performance of β_2AR models in the controlled virtual screening experiments and the structural accuracy of the models, expressed in terms of RMSD from a reference β_2AR crystal structure (PDB ID: 2RH1). The RMSD values were calculated relatively to the whole receptor (all residues) the TMs bundle, the ligand-binding cavity (Cavity), or the EL2 domain. The virtual screening performance was gauged in terms of ROC AUC and BEDROC (α = 160.9). Blue bars: Glide SP rankings; red bars: Glide SP/XP-score-in-place rankings; green bars: Glide SP/XP-refine rankings. For analogous figures reporting the virtual screening performance in terms of enrichment factors within the top 1%, 5% and 10% of the screened dataset (EF1, EF5, and EF10), see Supporting Information Figures S5–S7.

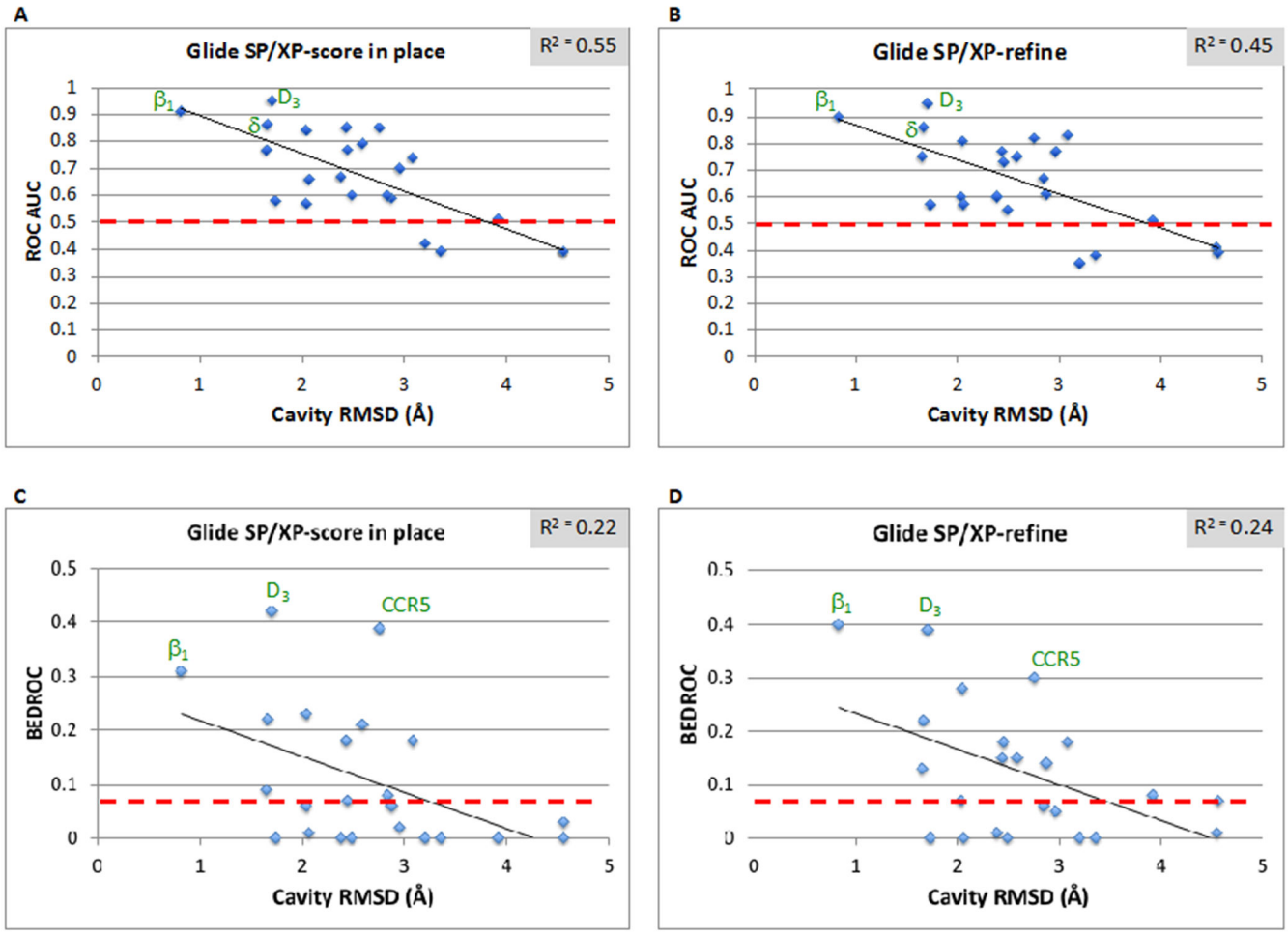


Figure 4.

Scatter plots showing the relationship between the performance of β_2AR models in the controlled virtual screening experiments and the structural accuracy of the models within the ligand-binding cavity, expressed in terms of RMSD from a reference crystal structure (PDB ID: 2RH1). The virtual screening performance was gauged in terms ROC AUC (panels A and C) and BEDROC (α = 160.9, panels C and D). Data relative to the Glide SP/XP-score-in-place and Glide SP/XP-refine rankings are shown in right and left panels, respectively. In each panel, the linear regression analysis curve and the associated R^2 values are shown. Moreover, a red dashed line marks the performance expectation for random ranking (ROC AUC = 0.50 and BEDROC = 0.07). Green labels indicate the templates on which the models that yielded the best virtual screening performance were based.

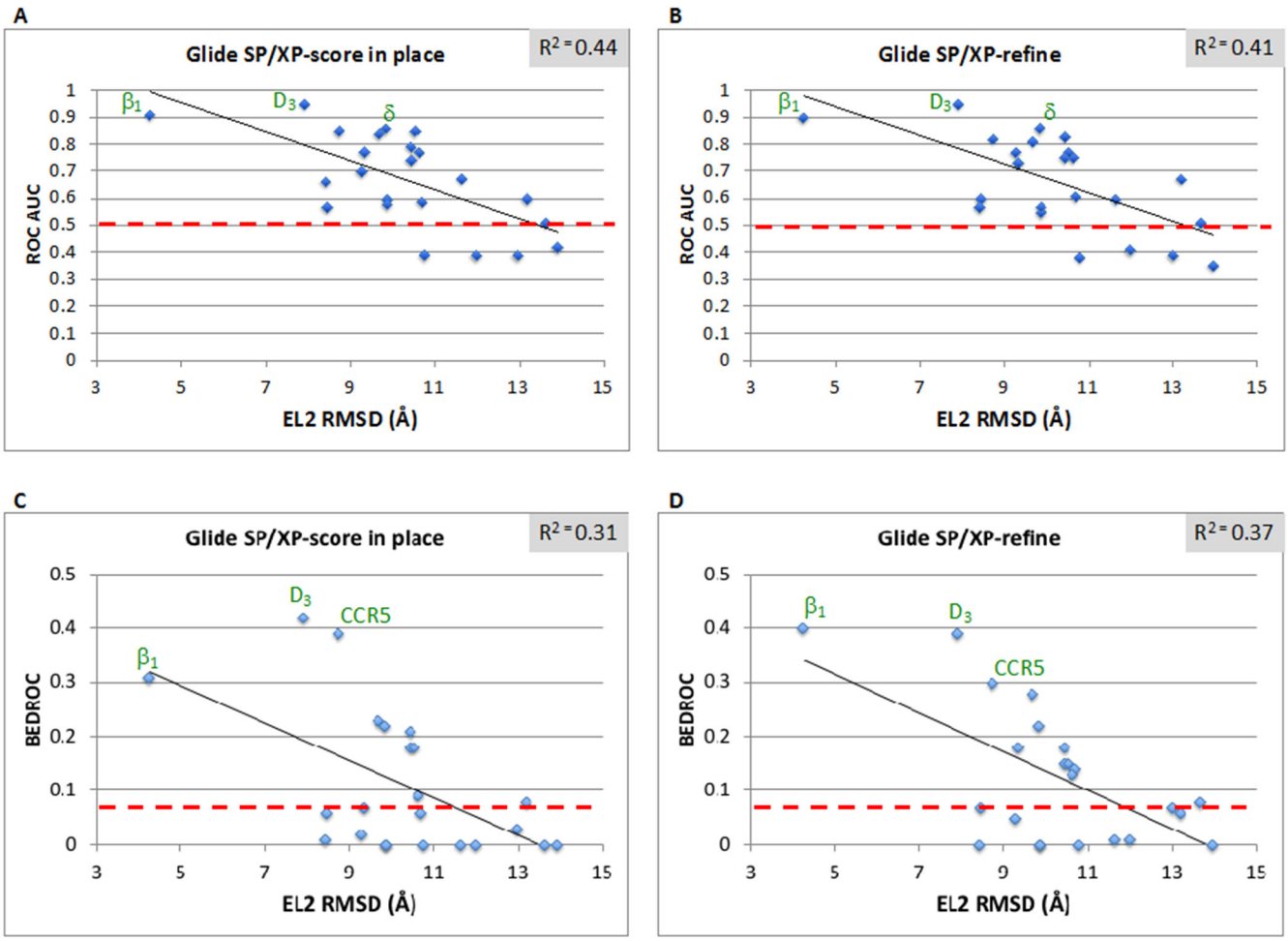


Figure 5.

Scatter plots showing the relationship between the performance of β_2AR models in the controlled virtual screening experiments and the structural accuracy of the models in the EL2 domain, expressed in terms of RMSD from a reference crystal structure (PDB ID: 2RH1). The virtual screening performance was gauged in terms of ROC AUC (panels A and C) and BEDROC (α = 160.9, panels C and D). Data relative to the Glide SP/XP-score-in-place and Glide SP/XP-refine rankings are shown in right and left panels, respectively. In each panel, the linear regression analysis curve and the associated R^2 values are shown. Moreover, a red dashed line marks the performance expectation for random ranking (ROC AUC = 0.50 and BEDROC = 0.07). Green labels indicate the templates on which the models that yielded the best virtual screening performance were based.

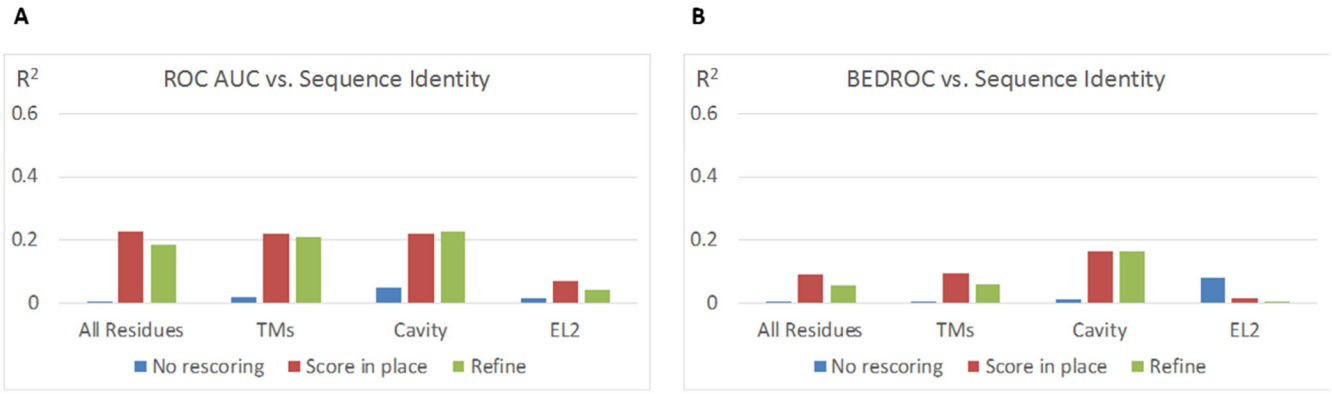


Figure 6.

Values of the square of the correlation coefficients (R^2) for the correlation between the performance of β_2AR models in the controlled virtual screening experiments and the sequence identity shared by the β_2AR and the templates on which the models were based. The sequence identities were calculated relatively to the whole receptor (all residues), the TMs bundle, the ligand-binding cavity (Cavity), or the EL2 domain. The virtual screening performance was gauged in terms of ROC AUC and BEDROC (α = 160.9). The R^2 values given in panel B were calculated excluding the β_1AR, which shares a sequence identity with the β_2AR much higher than that displayed by the other receptors. Blue bars: Glide SP rankings; red bars: Glide SP/XP-score-in-place rankings; green bars: Glide SP/XP-refine ranking. For analogous figures reporting the virtual screening performance in terms of enrichment factors within the top 1%, 5% and 10% of the screened dataset (EF1, EF5, and EF10), see Supporting Information Figures S9–S11.

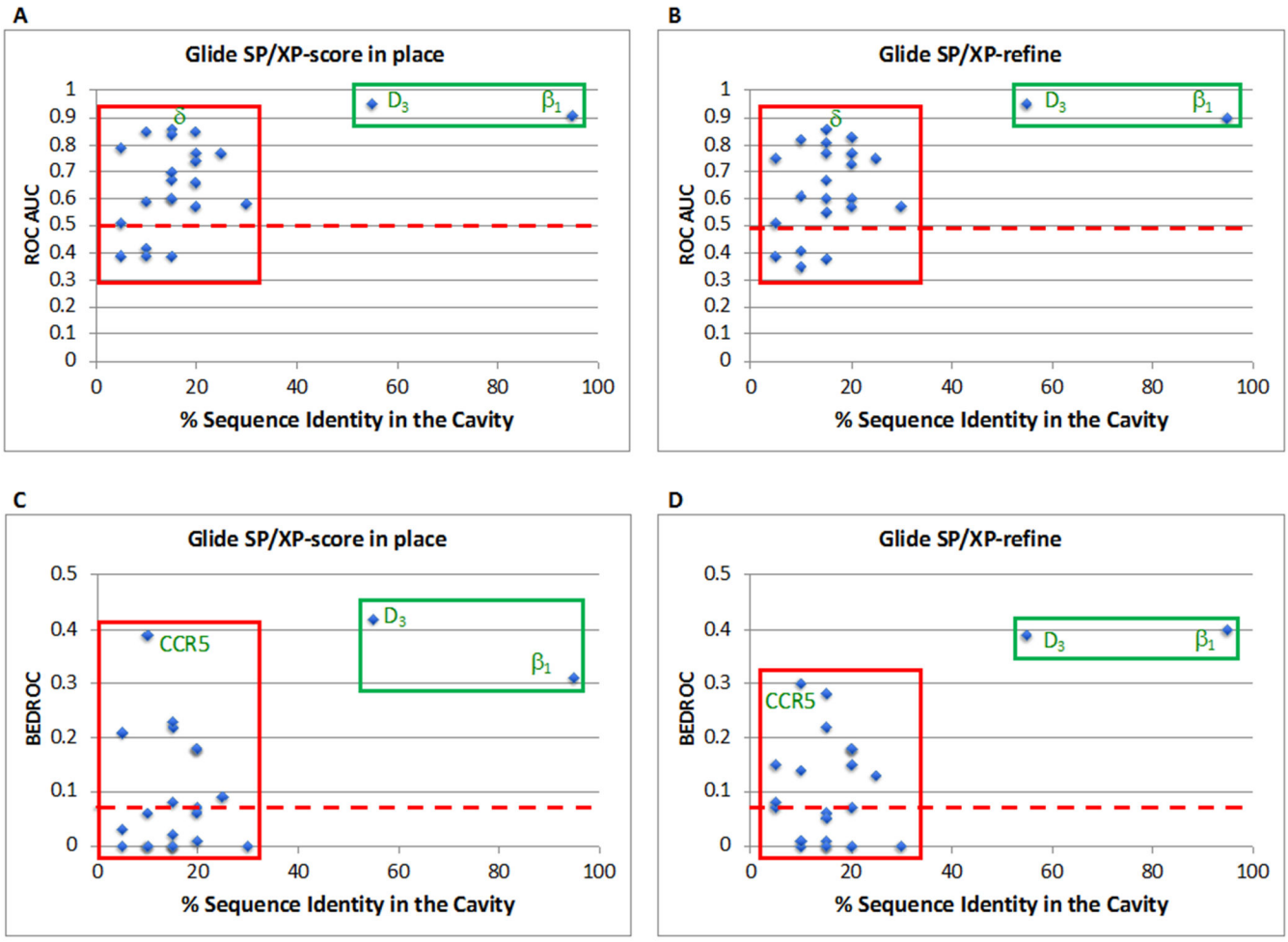


Figure 7.

Scatter plots showing the relationship between the performance of β_2AR models in the controlled virtual screening experiments and the sequence identity, relative to the residues that line the ligand-binding cavity in the β_2AR, shared by β_2AR and the templates on which the models were based. The virtual screening performance was gauged in terms of ROC AUC (panels A and C) and BEDROC (α = 160.9, panels C and D). Data relative to the Glide SP/XP-score-in-place and Glide SP/XP-refine rankings are shown in right and left panels, respectively. In each panel, the linear regression analysis curve and the associated R^2 values are shown. Moreover, in each panel, a red dashed line marks the performance expectation for random ranking (ROC AUC = 0.50 and BEDROC = 0.07). Green labels indicate the templates on which the models that yielded the best virtual screening performance were based. A green box encloses the models based on the two templates endowed with the highest sequence identity with respect to the β_2AR, while a red box encloses all the other models.

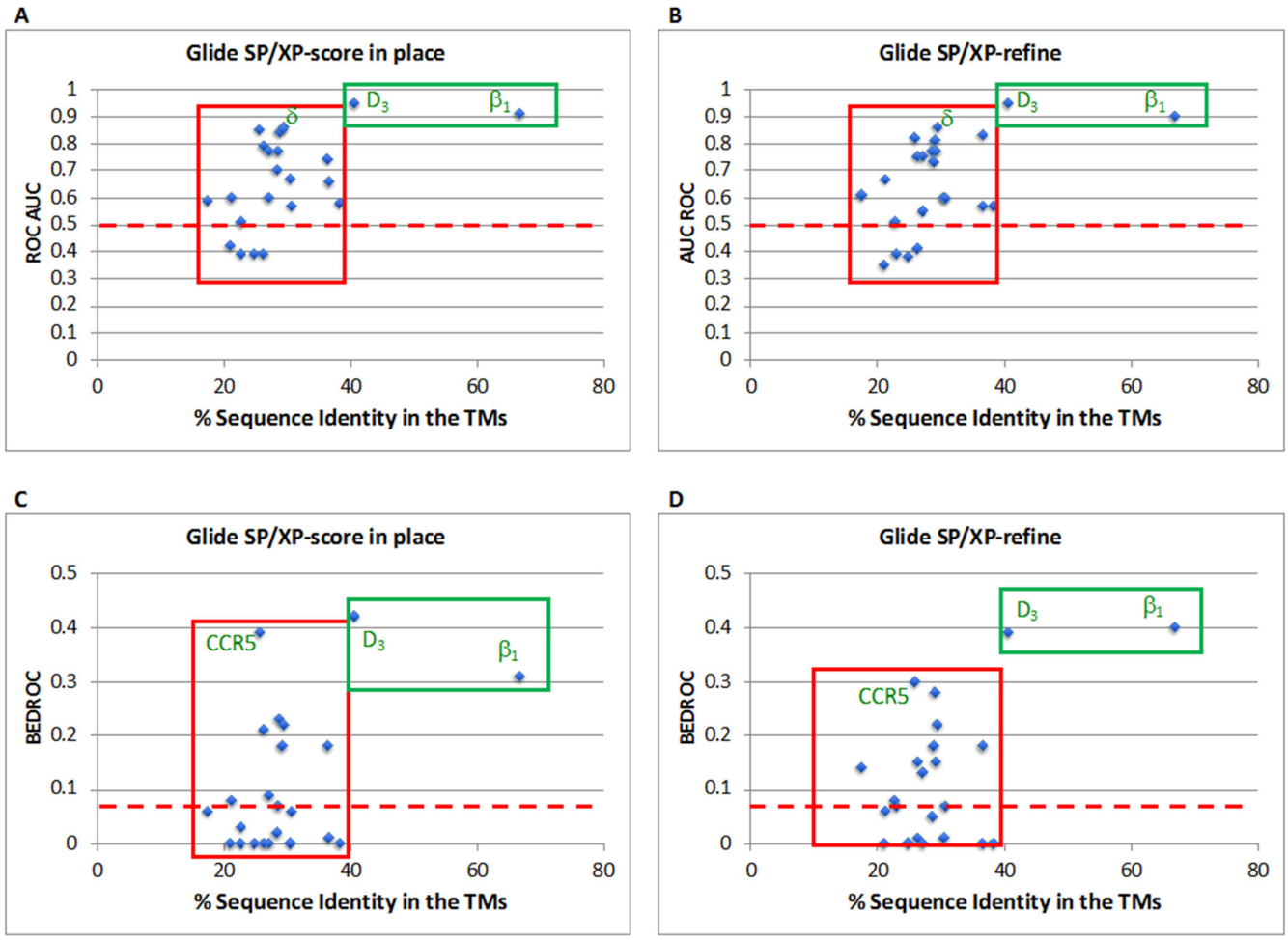


Figure 8.

Scatter plots showing the relationship between the performance of β_2AR models in the controlled virtual screening experiments and the sequence identity in the transmembrane domains shared by β_2AR and the templates on which the models were based. The virtual screening performance was gauged in terms of ROC AUC (panels A and C) and BEDROC (α = 160.9, panels C and D). Data relative to the Glide SP/XP-score-in-place and Glide SP/XP-refine rankings are shown in right and left panels, respectively. In each panel, the linear regression analysis curve and the associated R^2 values are shown. Moreover, in each panel, a red dashed line marks the performance expectation for random ranking (ROC AUC = 0.50 and BEDROC = 0.07). Green labels indicate the templates on which the models that yielded the best virtual screening performance were based. A green box encloses the models based on the two templates endowed with the highest sequence identity with respect to the β_2AR, while a red box encloses all the other models.

Table 1.

Virtual screening performance of one β_2AR crystal structure, and 23 β_2AR homology models based on the Glide SP/XP-refine rankings. The presented performance metrics include ROC AUC, BEDROC (α = 160.9), and enrichment factors within the top 1% (EF1), 5% (EF5), and 10% (EF10) of the screened dataset. For each metric, the 95% confidence interval, calculated through bootstrapping resampling of docking rankings, is shown within parentheses. Moreover, for each metric, the results for which the lower boundary of the 95% confidence interval does not exceed the expectation for random ranking (ROC AUC: 0.50; BEDROC: 0.07; EFs: 1) are underlined and marked in red. For crystal structures and homology model templates, PDB IDs are given in parentheses. Moreover, for convenience, for each model the RMSD of the ligand binding cavity with respect to the 2RH1 crystal structure is provided. For analogous tables relative to Glide SP and Glide SP/XP-score-in-place scoring methods, see Supporting Information Tables S2–S3.

β_2 Crystal Structure											
		ROC AUC		**BEDROC**		**EF1**		**EF5**		**EF10**	
β_2 (2RH1)		0.94	(0.90-0.97)	0.40	(0.21-0.60)	32.00	(13.39-50.39)	12.00	(8.23-15.67)	8.00	(6.34-9.61)
β_2 Homology Models											
Template	**RMSD Cavity (Å)**	**AUC**		**BEDROC**		**EF1**		**EF5**		**EF10**	
A_{2A} (4EIY)	3.08	0.83	(0.72-0.93)	0.18	(0.02-0.34)	12.00	(−2.66-23.75)	9.60	(5.62-13.57)	6.40	(4.39-8.12)
AT_1 (4YAY)	2.49	0.55	(0.44-0.66)	0.00	(0.00-0.00)	0.00	(0.00-0.00)	0.80	(−0.62-2.45)	1.20	(−0.20-2.48)
β_1 (4BVN)[a]	0.82	0.90	(0.83-0.97)	0.40	(0.19-0.61)	32.00	(16.02-50.49)	11.20	(7.19-15.17)	7.20	(5.51-8.92)
CCR5 (4MBS)	2.75	0.82	(0.70-0.94)	0.30	(0.11-0.49)	24.00	(7.47-40.75)	12.80	(8.95-16.63)	6.80	(4.91-8.63)
CXCR4 (3ODU)	2.59	0.75	(0.64-0.85)	0.15	(0.01-0.29)	13.00	(−1.08-25.12)	5.60	(2.01-9.24)	4.00	(1.79-5.70)
D_3 (3PBL)	1.70	0.95	(0.90-1.01)	0.39	(0.21-0.56)	36.00	(20.23-55.20)	18.40	(16.29-20.64)	9.20	(8.13-10.32)
δ (4N6H)	1.67	0.86	(0.79-0.92)	0.22	(0.08-0.37)	24.00	(7.97-40.07)	7.20	(3.22-10.83)	5.50	(3.60-7.59)
δ (4EJ4)[b]	2.38	0.60	(0.48-0.72)	0.01	(−0.01-0.03)	0.00	(−3.55-2.78)	2.40	(−0.21-5.01)	2.00	(0.54-3.71)
FFA1 (4PHU)	2.88	0.61	(0.51-0.72)	0.14	(−0.01-0.29)	12.00	(−1.00-25.10)	4.00	(0.50-7.07)	2.40	(0.68-4.12)
H_1 (3RZE)	1.74	0.57	(0.45-0.68)	0.00	(0.00-0.00)	0.00	(0.00-0.00)	0.00	(−1.19-0.88)	1.60	(0.26-3.14)
κ (4DJH)	2.45	0.73	(0.63-0.83)	0.18	(0.02-0.35)	16.00	(2.38-30.83)	5.60	(1.97-9.13)	4.00	(2.01-5.93)
LPA_1 (4Z35)	3.35	0.38	(0.29-0.47)	0.00	(0.00-0.00)	0.00	(0.00-0.00)	0.80	(−0.76-2.36)	0.80	(−0.28-1.89)
M_2 (3UON)	2.04	0.60	(0.47-0.72)	0.07	(−0.05-0.18)	4.00	(−3.65-11.70)	2.40	(−0.25-5.10)	1.60	(0.11-3.07)
M_3 (4U15)[c]	2.06	0.57	(0.48-0.66)	0.00	(0.00-0.00)	0.00	(0.00-0.00)	0.80	(−0.78-2.38)	0.40	(−0.72-1.23)
μ (4DKL)[b]	2.04	0.81	(0.71-0.92)	0.28	(0.11-0.45)	24.00	(9.71-41.20)	8.80	(4.69-12.83)	6.00	(4.10-7.91)

β_2 Crystal Structure											
		ROC AUC		**BEDROC**		**EF1**		**EF5**		**EF10**	
NOP (4EA3)	1.66	0.75	(0.68-0.82)	0.13	(−0.03-0.29)	8.00	(−2.55-18.85)	2.40	(−0.09-5.12)	1.60	(0.13-3.05)
OX_2 (4SOV)	2.44	0.77	(0.68-0.85)	0.15	(0.03-0.26)	20.00	(6.04-36.02)	5.60	(1.91-9.18)	3.60	(1.84-5.65)
$P2Y_1$ (4XNV)	3.20	0.35	(0.27-0.43)	0.00	(0.00-0.00)	0.00	(0.00-0.00)	0.00	(−0.06-0.06)	0.40	(−0.37-1.18)
$P2Y_{12}$ (4NTJ)	4.56	0.39	(0.29-0.49)	0.07	(−0.05-0.19)	4.00	(−3.69-11.83)	0.80	(−0.70-2.30)	0.40	(−0.38-1.18)
PAR1 (3VW7)	2.84	0.67	(0.55-0.79)	0.06	(−0.02-0.14)	4.00	(−4.73-11.72)	3.20	(0.04-6.04)	3.60	(1.81-5.72)
Rhod. (1U19)[d]	3.92	0.51	(0.37-0.65)	0.08	(−0.02-0.18)	8.00	(−2.57-18.90)	3.20	(0.21-6.16)	2.40	(0.52-3.99)
Rhod. (2Z73)[e]	4.55	0.41	(0.29-0.54)	0.01	(−0.01-0.02)	0.00	(−0.48-0.46)	0.80	(−0.80-2.40)	1.60	(0.17-3.04)
$S1P_1$ (3V2Y)	2.96	0.77	(0.69-0.85)	0.05	(0.00-0.10)	4.00	(−8.32-13.11)	5.60	(2.08-9.05)	4.00	(2.13-5.99)
Mean[f]	2.62	0.66		0.12		10.65		4.87		3.33	
SD[f,g]	0.91	0.18		0.12		11.17		4.70		2.49	
Median[f]	2.49	0.67		0.08		8.00		3.20		2.40	
IQR[f,h]	0.98	0.23		0.17		18.00		5.60		3.15	

[a] Source organism: turkey.

[b] Source organism: mouse.

[c] Source organism: rat.

[d] Source organism: cattle.

[e] Source organism: squid.

[f] Calculated for homology models, excluding the 2RH1 crystal structure.

[g] SD: standard deviation.

[h] IQR: interquartile range.

Table 2.

Performance of the crystal structure of the dopamine D_3 receptor (PDB ID: 3PBL) and the homology model of the β_2AR based on the dopamine D_3 receptor in controlled virtual screening experiments with the Glide SP, Glide SP/XP-score-in-place, and the Glide XP-refine scoring methods. The presented performance metrics include ROC AUC, BEDROC (α = 160.9), and enrichment factors within the top 1% (EF1), 5% (EF5), and 10% (EF10) of the screened dataset. For each metric, the 95% confidence interval, calculated through bootstrapping resampling of the docking rankings, is shown within parentheses.

	ROC AUC		BEDROC		EF1		EF5		EF10	
Glide SP										
D_3	0.62	(0.50-0.74)	0.13	(−0.03-0.30)	8.00	(−2.94-18.85)	3.20	(0.33-6.22)	2.00	(0.42-3.59)
D_3-based β_2AR model	0.80	(0.70-0.91)	0.16	(0.01-0.31)	13.00	(−0.97-25.54)	4.80	(1.29-8.26)	5.60	(3.74-7.66)
Glide SP/XP-score-in-place										
D_3	0.87	(0.81-0.92)	0.17	(0.01-0.33)	12.00	(−1.89-23.78)	8.00	(4.04-11.85)	5.20	(3.21-7.19)
D_3-based β_2AR model	0.95	(0.91-0.99)	0.42	(0.25-0.58)	44.00	(24.26-62.22)	16.00	(12.75-19.20)	8.80	(7.48-10.10)
Glide SP/XP-refine										
D_3	0.82	(0.74-0.90)	0.17	(0.00-0.35)	12.00	(−0.58-24.90)	6.40	(2.55-9.79)	5.20	(3.19-7.21)
D_3-based β_2AR model	0.95	(0.90-1.01)	0.39	(0.21-0.56)	36.00	(20.23-55.20)	18.40	(16.29-20.64)	9.20	(8.13-10.32)

J Chem Inf Model. Author manuscript; available in PMC 2020 July 22.