Journal of Computer-Aided Molecular Design
https://doi.org/10.1007/s10822-019-00222-y

# Predicting binding poses and affinity ranking in D3R Grand Challenge using PL-PatchSurfer2.0

**Woong-Hee Shin**[1,2] · **Daisuke Kihara**[1,3,4,5]

Received: 12 June 2019 / Accepted: 28 August 2019


**Abstract**
Computational prediction of protein–ligand interactions is a useful approach that aids the drug discovery process. Two major tasks of computational approaches are to predict the docking pose of a compound in a known binding pocket and to rank compounds in a library according to their predicted binding affinities. There are many computational tools developed in the past decades both in academia and industry. To objectively assess the performance of existing tools, the community has held a blind assessment of computational predictions, the Drug Design Data Resource Grand Challenge. This round, Grand Challenge 4 (GC4), focused on two targets, protein beta-secretase 1 (BACE-1) and cathepsin S (CatS). We participated in GC4 in both BACE-1 and CatS challenges using our molecular surface-based virtual screening method, PL-PatchSurfer2.0. A unique feature of PL-PatchSurfer2.0 is that it uses the three-dimensional Zernike descriptor, a mathematical moment-based shape descriptor, to quantify local shape complementarity between a ligand and a receptor, which properly incorporates molecular flexibility and provides stable affinity assessment for a bound ligand–receptor complex. Since PL-PatchSurfer2.0 does not explicitly build a bound pose of a ligand, we used an external docking program, such as AutoDock Vina, to provide an ensemble of poses, which were then evaluated by PL-PatchSurfer2.0. Here, we provide an overview of our method and report the performance in GC4.

**Keywords** PL-PatchSurfer · D3R Grand Challenge · Protein–ligand interaction · Virtual screening · BACE-1 · CatS

## Introduction

Computer-aided drug discovery (CADD) has been widely used in modern drug discovery [1, 2]. CADD methods can be roughly classified into two categories based on their purposes. The first category predicts a binding pose and the binding energy of a compound within a binding pocket of a target protein. Methods in this category take three-dimensional structures of the protein and the ligand as input, generate conformations of the ligand molecules in the pocket, and predict the binding Gibbs free energy of the conformations. Most popular methods in this category use rather simplified empirical scoring functions to assess the binding energy [3–6]. A free energy calculation using molecular dynamics simulation such as the free energy perturbation (FEP) [7] provides a more rigorous option for binding energy computation [8].

The other category is virtual screening (VS). VS performs high throughput screening of compounds, where the goal is to find highly promising ligands that could be seed compounds to develop drug molecules for a target disease or protein from a large library. VS software can be further classified into structure-based methods, which use a target protein's structure information, and ligand-based methods, which use the information of known drugs. Various types of scoring functions have been developed. For structure-based VS, scores used include receptor pharmacophore search [9] and protein–ligand docking [10]. For ligand-based VS, scores such as fingerprint comparison [11],

✉ Daisuke Kihara
dkihara@purdue.edu

1 Department of Biological Science, Purdue University, West Lafayette, IN 47907, USA

2 Department of Chemistry Education, Sunchon National University, Suncheon 57922, Republic of Korea

3 Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA

4 Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN 47907, USA

5 Department of Pediatrics, University of Cincinnati, Cincinnati, OH 45229, USA

Published online: 10 September 2019

molecular shape-based comparison [12], QSAR [13], 2D molecular graph comparison [14], and ligand pharmacophore search [15].

Recently, machine learning including deep neural network has become popular in CADD. Machine learning methods can take advantage of an increasing amount of data accumulated over the years and can combine various types of data to make predictions [16]. They also consider non-additive contributions of functional groups and perform screening relatively faster than speed than conventional energy-based methods [17–20].

Since various software packages have been developed for different categories of CADD, it is difficult to objectively compare existing software and understand their strengths and weaknesses. Drug Design Data Resource (D3R), directed by Drs. Amaro and Gilson of the University of California San Diego, work toward this goal, aiming to advance current CADD methodologies by organizing blind, community-wide prediction challenges [21]. The challenge is composed of three parts: binding pose prediction, free energy prediction, and affinity ranking of ligands. In 2018, D3R organized the fourth Grand Challenge (GC4), targeting beta-secretase 1 (BACE-1) and cathepsin S (CatS). Ligands to be docked or ranked were provided by two pharmaceutical companies, Novartis and Janssen Pharmaceuticals.

Here, we discuss the performance of our group in GC4. We participated in GC4 challenges using our own structure-based virtual screening program, PL-PatchSurfer2.0 [22–24]. The unique feature of this program is that it uses molecular surface patches to quantify physicochemical complementarity between a ligand and a receptor. Physicochemical characteristics of a molecular surface patch are encoded as vectors of three-dimensional Zernike descriptors (3DZD), a rotationally invariant 3D shape descriptor, which is based on a mathematical series expansion of a 3D function [25]. 3DZD has been successfully applied for global shape comparison of biomolecules [26–28], local patch matching for protein–protein docking prediction [29, 30], ligand-based VS [31], binding ligand prediction by pocket–pocket comparison [32, 33]. In our previous work, we showed PL-PatchSurfer2.0 performed better than other receptor-based VS methods, especially when using apo structures and template-based modeled structures as receptor structures [22].

In GC4, we participated in two binding pose prediction challenges of BACE-1, as well as the affinity ranking of both BACE-1 and CatS. For affinity ranking, we placed 8th out of 54 participants and 27th out of 55 participants in terms of Kendall's $\tau$ correlation coefficients for BACE-1 and CatS, respectively. In the binding pose prediction category, our performance was ranked 61st out of 70 submissions.

## Methods

### Overview of PL-PatchSurfer2.0

PL-PatchSurfer2.0 is a virtual screening program that represents a molecular surface as a set of overlapping surface patches using 3DZD [22–24] (Fig. 1). The surface of a molecule is computed with APBS [34]. The surface is segmented to patches by cutting the whole surface using 5 Å radius spheres with 3 Å overlap. Four physicochemical features are considered to characterize a surface patch: shape, the electrostatic potential, hydrogen bond donors/acceptors, and hydrophobicity. APBS calculates the electrostatic potential on surface points by solving the Poisson–Boltzmann equation. Hydrophobicity on surface points is calculated as a field-type [35] by taking atomic hydrophobicity from XLogP3 [36]. Hydrogen bond characters on each surface point are taken from the nearest atom. The values mapped on the surface patch are then converted to 3DZD. 3DZD is a vector, which is a set of coefficients from a series expansion of a 3D function using Zernike–Canterakis basis functions [25]. A unique feature of 3DZD is its rotational invariance, thus there is no need to rotate a surface patch to compare against another patch. A Similarity of two 3DZDs can be simply computed as the Euclidean distance between them. Compatibility of a ligand and a binding pocket is quantified by the local patch similarity search using the Euclidean distance of 3DZD of four physicochemical properties and the similarity of the location of matching patches in the ligand and the pocket surface. The local patch representation with 3DZD already accommodates a certain level of molecular flexibility; however, alternative conformations of small ligand molecules that arise by rotational isomers of chemical bonds result in incompatibly large differences in surface shapes. Therefore, PL-PatchSurfer2.0 first generates multiple conformations of compounds in a library, and each of the conformations is evaluated for compatibility with a target binding pocket.

### D3R Grand Challenge 4

The fourth round of D3R, GC4, was composed of two subchallenges. Subchallenge 1 had BACE-1 as the target protein and consisted of three stages, 1a, 1b, and 2. BACE-1 is a transmembrane aspartic-acid protease, which is crucial for generating beta-amyloid, making it an interesting therapeutic target for Alzheimer's disease [37]. Stage 1a was to predict crystallographic poses of 20 ligands, to rank 154 ligands, and to calculate the absolute binding free energies of 34 ligands. The compound lists

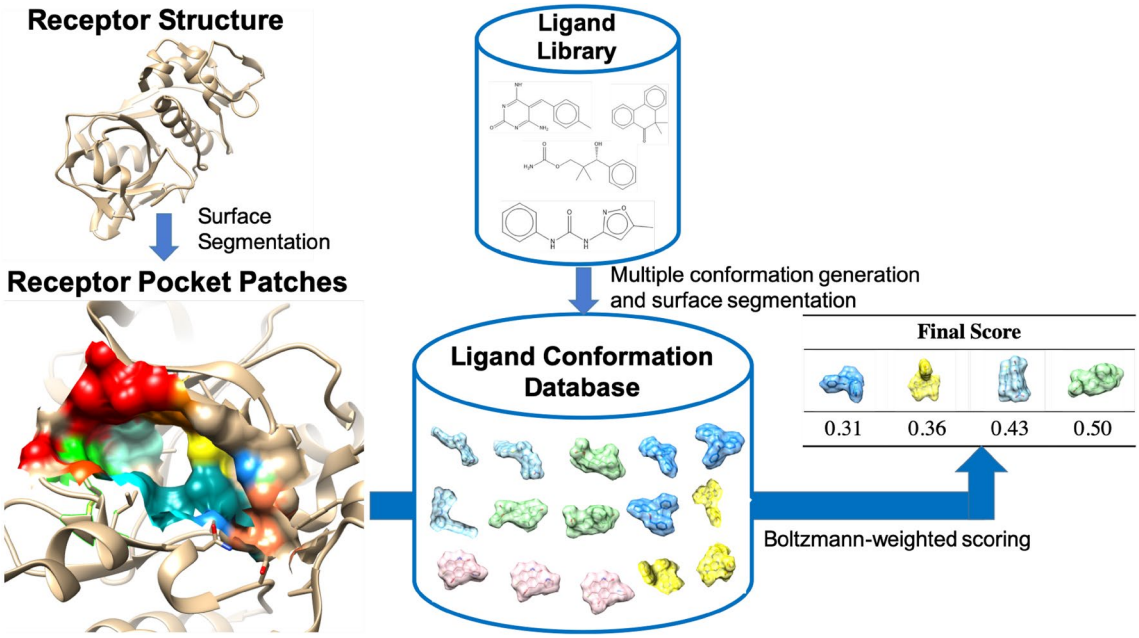


**Fig. 1** A schematic illustration of PL-PatchSurfer2.0. A surface of the target binding pocket of the receptor is segmented into patches. For a compound library, multiple conformations of compounds (up to 50) are generated. The surface of a ligand conformation is segmented into a couple of patches. Four characteristics, shape, electrostatics, hydrogen bond, and hydrophobicity of patches are computed and further converted to 3DZD. The complementarity between the pocket and the ligand conformation is computed by a weighted sum of Euclidean distances of 3DZD and difference in geodesic distance distribution of surface patches ($PS(P, C)$ in Eq. 1). The scores were further averaged using the Boltzmann-weighted scheme. The ligands in the library are ranked in ascending order of Boltzmann-weighted scores

were provided by Novartis. A FASTA format sequence of BACE-1, chain A of 5YGX, which is a PDB entry [38] to be used as a reference frame crystal structure, and the three ligand sets of SMILES strings were provided to participants. In the binding pose prediction category, the participants needed to submit atomic coordinates of complexes of the receptor (in the PDB format) and the ligand (in the MOL format). Thus, participants of this category had to model the three-dimensional structure of the protein or identify a crystal structure from the PDB database suitable for modeling, and perform ligand docking. The organizers calculated RMSD of the submitted poses after the submitted receptor structure (5YGX A chain) was superimposed to a reference structure, and the performance of each participating group was evaluated in terms of the RMSD.

For the affinity ranking category, the participants were asked to sort the ligands according to the predicted affinity and turn in the ranking of ligands with scores. The performance was evaluated by Kendall's $\tau$ correlation coefficients and Spearman's $\rho$ correlation coefficients, between the submitted compound rank and the rank of ligands sorted by $IC_{50}$. The two coefficients are measures of rank correlation, the similarities of the orderings of the data when ranked by each of the quantities, range from $-1$ (perfectly reversed ranking) to 1 (perfect ranking). The difference between $\tau$ and $\rho$ is how they are calculated. Spearman's $\rho$ correlation coefficients are calculated based on deviations. On the other hand, Kendall's $\tau$ correlation coefficients are proportional to a difference between the number of concordant pairs and discordant pairs. In statistics, two variables $(X_1, Y_1)$ and $(X_2, Y_2)$ are called concordant if both $X_1 > X_2$ and $Y_1 > Y_2$ or if both $X_1 < X_2$ and $Y_1 < Y_2$. In the case of affinity ranking, X and Y are ranks of a compound.

Stage 1b was a self-docking experiment. The aim of stage 1b was to predict docking poses of the 20 ligands, the same compounds given in stage 1a, in their cognate receptor structures. After the stage 1a deadline had passed, 20 pairs of a crystal structure of BACE-1 without bound ligand coordinates and the SMILES string of the ligand were provided. The evaluation process was the same as stage 1a, where the RMSD of the submitted ligand conformation was used as the evaluation metric.

Stage 2 was to predict the affinity ranking of the 154 ligands and the absolute binding energies of the 34 ligands, which were the same ligand sets as used in stage 1. A difference in stage 2 was that 20 additional crystal structures of BACE-1 were provided, which bind different ligands from the target ligands in this challenge. Our group participated in binding pose prediction and affinity ranking of compounds in stage 1a, stage 1b, and stage 2 of subchallenge 1, but we did not participate in the absolute binding free energy

**Table 1** Summary of D3R GC4 challenges that our group participated

| Subchallenge | Stage | Task | Used receptor PDB ID | Number of ligands | Procedure |
|---|---|---|---|---|---|
| 1: BACE-1 | 1a | BP, AR, FE[a] | BP: 5YGX, a reference structure given by the organizers<br>AR: 6EJ2, the top hit by our BLAST search | BP: 20<br>AR: 154<br>FE: 34 | BP: 20 ligand poses per ligand was predicted by AutoDock Vina and rescored by the PL-PatchSurfer2.0 score<br>AR: Multiple ligand conformations generated by Multi-Conf-Dock. A ligand was scored by the Boltzmann-weighted scoring scheme |
| | 1b | BP: predict binding poses of ligands to their cognate protein structures | 20 receptor structures given by the organizers | 20 ligands | |
| | 2 | AR, FE[a] | 20 receptor structures of stage 1b and 6EJ2 | AR: 154<br>FE: 34 | Ensemble average of ligands using the 21 receptor structures |
| 2: CatS | | AR, FE[a] | 3IEJ, the top hit by BLAST search | AR: 459<br>FE: 39 | Multiple ligand conformations generated by OMEGA. Each ligand was scored by the Boltzmann-weighted scoring scheme |

*BP* binding pose prediction, *AR* affinity ranking, *FE* absolute binding free energy prediction

[a]Categories that we did not participate in

prediction (Table 1, stages which we did not participated in were indicated with [a]).

Subchallenge 2 used a different target receptor, CatS. This protein is involved in protein degradation and highly expressed in antigen-presenting cells and can be a target for regulating immune hyper-responsiveness [39]. The goal of subchallenge 2 was to predict absolute binding energies of 39 compounds and to rank 459 compounds by their binding energies. The compounds were provided by Janssen Pharmaceuticals. As in stage 1, a FASTA format sequence of CatS and SMILES strings of the compounds were provided to participants. Our group participated in the affinity ranking part of this subchallenge 2.

## Affinity ranking of compounds using PL-PatchSurfer2.0

For the ranking of compounds for a target structure, we followed the standard protocol of PL-PatchSurfer2.0 [22, 23] (Fig. 1). A ligand conformation database for the provided compound was prepared as follows: Ligand SMILES strings were first converted to a single 3D structure by OpenBabel in the MOL2 format [40]. Hydrogens and Gasteiger atomic charges were added to the 3D structure. Then, multiple conformations of ligands were generated by Multiconf-Dock [5] for suschallenge 1 and OMEGA [41] for subchallenge 2. Multiconf-Dock samples ligand conformations by rotating rotatable torsions with pre-determined rotation angles using the DOCK scoring function, which considers the van der Waals and the Coulomb forces [5]. OMEGA is another ligand conformer generation method, which employs a torsional library taken from small-molecule crystal structures [41]. The programs generated up to 50 ligand conformations.

To screen the compounds for stage 1 of subchallenge 1 and subchallenge 2, we used PDB entries 6EJ2 and 3IEJ as the receptor structures, respectively. They were the top hits with 100% sequence identity and 100% alignment coverage to the query FASTA sequence of BACE-1 and CatS, which were identified by BLAST [42] against the PDB database [38]. 6EJ2 is a complex structure of BACE-1 with compound 28 (PDB ID: B7E). The binding pocket of 6EJ2 was defined by removing the cognate ligand, compound 28, and ray-casting from the geometrical center of the cognate ligand to the pocket wall. Then, the protonation states and atomic charges were assigned with PDB2PQR [43] to the receptor. As for ligands in the library, multiple conformations were generated using MultiConf-Dock [5]. Similarly, for stage 2, the 20 crystal structures provided in stage 1b were aligned to 6EJ2 using TM-align [44] to define their binding pockets. After aligning the structures, hydrogens and atomic charges were added by PDB2PQR, and compound 28 was used to define binding pockets for the 20 structures. For subchallenge 2, the binding pocket of 3IEJ was defined by ray-casting from its bound ligand (cathepsin inhibitor, PDB ID: 599). OMEGA [41] was employed to generate multiple conformations (up to 50) of the 459 ligands.

To assess the complementarity between a ligand and a pocket, we used the Boltzmann-weighted scheme in PL-PatchSurfer2.0 [22]. A Boltzmann-weighted score of ligand $L$ with pocket $P$ is calculated as follows:

$$\text{Boltzmann-Weighted Ligand Score (P, L)} = \frac{\sum_{C}^{N_{conf}} PS(P,C) \times \exp[-\beta \times PS(P,C)]}{\sum_{C}^{N_{conf}} \exp[-\beta \times PS(P,C)]} \quad (1)$$

where $PS(P,C)$ is a conformer score between receptor pocket $P$ and ligand conformation $C$. The conformer score is a weighted linear sum of Euclidean distance of 3DZD of four physico-chemical characteristics, as described in the *Overview of PL-PatchSurfer2.0* section and the difference in pairwise geodesic distances distribution of patches (i.e. relative position of corresponding patches) between a pocket and a ligand. $N_{conf}$ is the number of conformations of a ligand generated and $\beta$ is a parameter, which is set to 1. Thus, this score summarizes the compatibility of each conformer of the ligand.

In subchallenge 1, stage 1 and subchallenge 2, the ligands for affinity ranking (154 ligands for subchallenge 1 and 459 ligands for subchallenge 2) were ordered according to the order of Boltzmann-weighted scores and submitted as our answers. In stage 2 of subchallenge 1, each of 154 ligands had 21 scores, one score for one of 20 crystal structures plus one (computed for 6EJ2) from stage 1. Using the 21 scores, we took an ensemble docking strategy [6], where we averaged the 21 scores for each ligand. The ligands were further ranked by the scores and submitted to the D3R GC4 server.

### Applying PL-PatchSurfer2.0 for predicting docking poses

Since PL-PatchSurfer2.0 does not explicitly output a docking pose, we needed an external program to output binding conformations to participate in the binding pose prediction in subchallenge 1. For this purpose, we used AutoDock Vina [3] to generate binding conformations in a binding pocket. The receptor and ligand structures were prepared using python scripts in AutoDockTools. An initial 3D conformation of 20 ligands was generated by OpenBabel [40]. A binding box with a 25 Å length for each side was centered at the geometrical center of the cognate ligand of 5YGX, the provided reference structure for stage 1a. For stage 1b, 20 crystal structures were aligned with 5YGX using TM-align [44] and the same center coordinate was used. The compounds were treated as flexible during the simulation and AutoDock Vina generated 20 docking poses per ligand.

Evaluation of a binding pose with PL-PatchSurfer2.0 was executed as follows: (1) generating molecular surfaces of the binding pocket and the docked poses and segmenting the surfaces to the patches; (2) the closest patches from the pocket and the ligand were matched to define a set of patch pairs; (3) the patch pair set was evaluated with the scoring function of PL-PatchSurfer2.0, and (4) up to 5 poses of each ligand were submitted to D3R GC4 server, numbered from 1 (best) to 5 (worst).

All the tasks and the protocols we performed are summarized in Table 1.

## Results and discussion

### Performance on ranking compounds of BACE-1 and CatS

Performance of ligand ranking is evaluated by two correlation coefficients, Kendall's τ and Spearman's ρ, which were computed between a submitted compound rank and rank by $IC_{50}$. Figure 2 shows a bar plot of Kendall's τ correlation coefficients of participated structure-based scoring methods. Comparing the coefficients with other structure-based methods, our method ranked 3rd (0.28) out of 26 submissions (Fig. 2a) for BACE-1 stage 1, 8th (0.28) out of 48 submissions (Fig. 2b) for BACE-1 stage 2, and 16th (0.21) out of 43 submissions for CatS in subchallenge 2 (Fig. 2c). In the parentheses shown are the correlation values. The ranking of structure-based methods by Spearman's ρ correlation coefficients resulted in the same rank (the correlation values were 0.39 for both stage of BACE-1 and 0.31 for subchallenge 2 on CatS).

When ligand-based scoring methods were included in the ranking together with structure-based methods, our group kept a similar rank, ranked 3rd out of 36 submissions for BACE-1 stage 1, 8th among 54 submissions for BACE-1 stage 2, and 27th among 55 for CatS in subchallenge 2. To draw a baseline, we computed two metrics: a one-dimensional (1D) compound similarity and ΔSASA. The 1D compound similarity was calculated using MACCS key fingerprint Tanimoto coefficient using RDKit [45] between the compounds of each subchallenge and the cognate ligands of 6EJ2 and 3IEJ, the receptor structures we used. For BACE-1, using the MACCS key showed Kendall's τ and Spearman's ρ correlation coefficients of 0.12 and 0.17, respectively. For CatS, Kendall's τ and Spearman's ρ correlation coefficients were 0.11 and 0.16, respectively. Therefore, PL-PatchSurfer2.0 performed better than such a simple 1D ligand-based method.

To obtain ΔSASA of each ligand, the ligands were docked to the receptor structures (6EJ2 and 3IEJ for BACE-1 and CatS, respectively) using DOCK6 [4], which produced up to 50 poses per ligand. Then, the poses were reordered by ΔSASA and the pose with the largest absolute ΔSASA was selected as prediction using ΔSASA. For BACE-1, ΔSASA showed Kendall's τ and Spearman's ρ correlation coefficients of 0.18 and 0.26, respectively. For CatS, Kendall's τ and Spearman's ρ correlation coefficients were 0.22 and

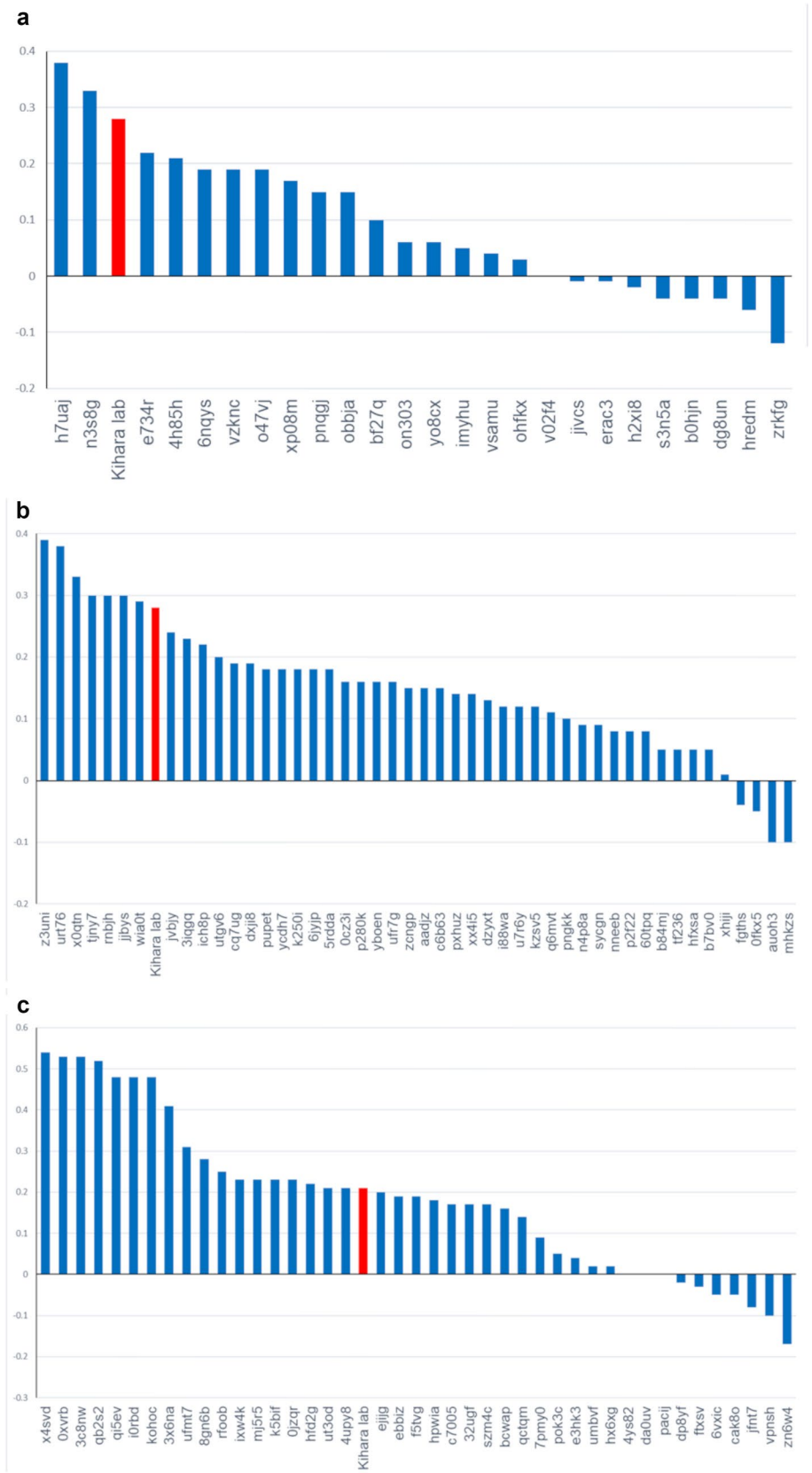


**Fig. 2** Kendall's Tau correlation coefficients of structure-based scoring methods. The x-axis is the group IDs and the Y-axis shows Kendall's Tau correlation coefficients. **a** BACE-1 stage 1, **b** BACE-1 stage 2, **c** CatS. Our group's bars are colored in red in each graph

0.33, respectively. Thus, for BACE-1, PL-PatchSurfer2.0 performed better than ΔSASA, while it was slightly worse for CatS. Indeed, for CatS 25 structure-based submissions out of 43 were performed equal to or worse than ΔSASA. This may indicate that interaction between binding ligands and CatS are mostly dominated by van der Waals interaction.

It caught our attention that our performance in terms of the correlation coefficients to $IC_{50}$ was identical for stage 1 and stage 2 (τ: 0.28 and ρ: 0.39), although we had different sets of receptor structures available for the two stages. For stage 1, we only used 6EJ2 while for stage 2 we were provided with 20 crystal structures of BACE-1, with which we computed the average score against them. The main reason for the same performance would be due to the high similarity of these receptor structures. The average pairwise root mean square deviation (RMSD) among the 20 provided structures in stage 2 was 0.39 Å by TM-align [44], reflecting that their 20 ligands share a large macrocycle (Fig. 3a) and the average RMSD from the 20 structures and 6EJ2 was 1.01 Å. In general, it has been shown that having an ensemble of similar receptor structures does not contribute much in structure-based virtual screening [6], and it is probably particularly true for our PL-PatchSurfer2.0 because of its design of using the surface-patch-based coarse-grained representation.

Next, we discuss why the performance of PL-PatchSurfer2.0 was worse for CatS than the BACE-1 cases in terms of both correlation coefficient values and the relative ranking among other participants. We found the main reason would be the size of the binding pocket in CatS we used for screening. The binding pocket of CatS is formed by three subpockets. Two side-chains, PHE211 and LYS64 vary their side-chain conformations depending on bound ligands [46], and changes the size of the pocket accordingly (Fig. 3b). It turned out that the size of the 459 compounds we needed to screen is larger than the co-crystallized ligand of 3IEJ, which we used as the receptor structure. The molecular weights of the compound library range from 566.5 to 825.2 g/mol, with an average value of 683.2 g/mol. In addition, the ligands in the library have 39–56 heavy atoms with an average of 46.6 heavy atoms. However, the molecular weight and the number of heavy atoms of the co-crystallized ligand of 3IEJ is smaller, at 522.8 g/mol and 38 heavy atoms. Thus, the ligand in 3IEJ is substantially smaller than the ligands we needed to screen. This caused a problem when we defined a binding pocket in the receptor structure using 3IEJ, as we define the pocket by ray-casting from the center position of the cognate ligand, which resulted in a pocket that may be too small for screening the 459 compounds. In addition, the pocket defined in 3IEJ could not form a suitable conformation for the large molecules. As illustrated in Fig. 3c, the two mobile residues change their side-chain conformations depending on which ligand binds. Since the library compounds were larger than the 3IEJ molecule, the two residues might not have the same conformation as in 3IEJ.

To investigate the effect of the binding pocket definition in PL-PatchSurfer2.0, we screened the 459 compounds using another receptor structure, 5QC5. We chose this PDB entry because the cognate ligand has a molecular weight of 663.8 g/mol and 45 heavy atoms, closer to the average values of the library. With this new receptor structure, Kendall's τ correlation coefficient and Spearman's ρ correlation coefficient increased to 0.27 and 0.41 from 0.22 and 0.31, respectively. This result is comparable to the BACE-1 result we obtained, 0.28 and 0.39 for Kendall's and Spearman's coefficients, respectively. We also examined if better receptor structure (i.e. 5QC5) could have been selected by considering the PL-PatchSurfer2.0's scores for the 459 compounds. The average scores were 0.543 and 0.548 for 3IEJ and 5QC5, respectively. Thus, the scores for the two receptor structures were similar but slightly better for 3IEJ, which indicates the proper receptor structure could not be found from the PL-PatchSurfer2.0 scores. This benchmark

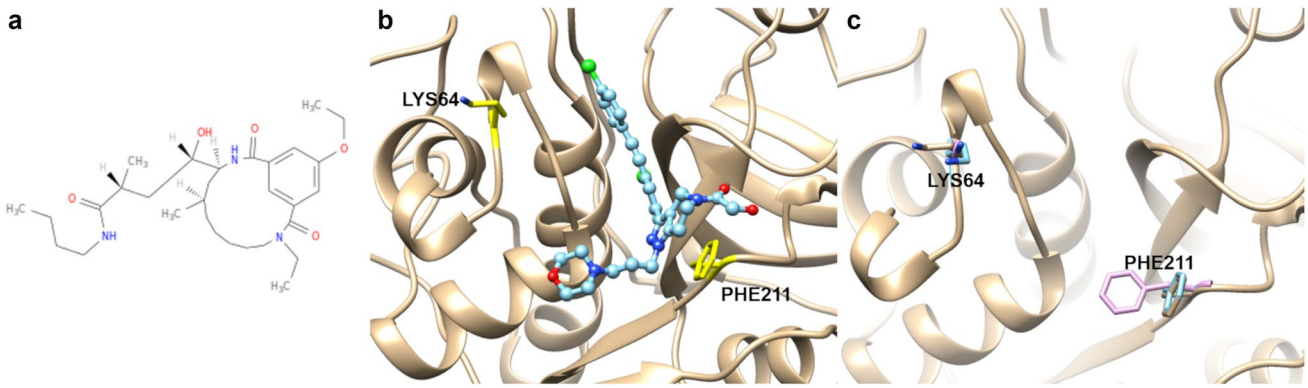


**Fig. 3** **a** The 2D structure of BACE molecule 1 of stage 1a and stage 1b subchallenges, **b** the CatS binding pocket. Flexible residues, PHE211 and LYS64 are labeled and colored in yellow, **c** changes of side-chain conformation of PHE211 and LYS64. Three crystal structures of CatS are superimposed, and the two side-chains are colored; 3IEJ (gold), 5QC5 (blue), and 1NPZ (pink)

implies that selection of the receptor structure based on the compound library characteristics is important for virtual screening.

## Effect of the Boltzmann-weighted scoring

We submitted the order of ligands sorted by the Boltzmann-weighted score (Eq. 1). To examine the effect of the Boltzmann-weighted scoring, we scored a ligand by considering only the lowest conformer score (LCS), where the score of a ligand is determined as the lowest scored conformer. With the LCS, Spearman's ρ correlation coefficients for BACE-1 stage 1 and stage 2 were lower than what we submitted with the Boltzmann-weighted scoring, 0.36 and 0.35 for stage 1 and stage 2, respectively. This result shows the Boltzmann-weighted scoring was effective, which is consistent with our previous work [22]. On the other hand, the result implies that in general Patch-Surfer2.0 was not able to select the best (lowest RMSD) conformer as the top choice. In Table 2, we generated 50 conformers for 20 bound cognate ligands of BACE-1 provided in stage 1 of subchallenge 1 and ranked them using their receptor structures with Patch-Surfer2.0 score (Table 2). Table 2 shows that the lowest RMSD conformations were not selected by Patch-Surfer2.0 and their score ranks were often very low. These results come from the design of the PL-PatchSurfer2.0 algorithm: PL-PatchSurfer2.0 was designed to be insensitive to small changes of conformations so that the method shows stable performance [22–24]. This design of PL-PatchSurfer2.0 shows unique strength when available receptor conformations deviate from ligand-bound form but it can become weakness in cases such as pose selection of bound ligands.

## Binding pose prediction for BACE-1 (stage 1a and 1b of Subchallenge 1)

In subchallenge 1, binding pose prediction for ligands for BACE-1 were the tasks in two different settings, stage 1a and 1b (Table 1). As described in the Methods section, we first generated docking poses of a query ligand in the binding pocket of BACE-1 using AutoDock Vina, which were then ranked by PL-PatchSurfer2.0. We used this procedure

**Table 2** Conformation selection performance of PL-PatchSurfer2.0

| Ligand ID | Minimum RMSD of generated conformations (Å) | PL-PatchSurfer Score Rank of minimum RMSD conformation | RMSD of LCS conformer (Å) | RMSD rank of LCS conformer |
|---|---|---|---|---|
| 1 | 1.339 | 15 | 1.715 | 13 |
| 2 | 1.103 | 33 | 2.817 | 32 |
| 3 | 0.926 | 31 | 2.033 | 15 |
| 4 | 1.624 | 35 | 3.485 | 50 |
| 5 | 1.523 | 7 | 2.359 | 29 |
| 6 | 1.595 | 11 | 2.175 | 14 |
| 7 | 1.958 | 38 | 3.314 | 46 |
| 8 | 1.127 | 25 | 2.246 | 27 |
| 9 | 1.420 | 9 | 1.923 | 12 |
| 10 | 1.118 | 31 | 1.804 | 9 |
| 11 | 1.526 | 28 | 1.842 | 5 |
| 12 | 1.466 | 35 | 2.360 | 24 |
| 13 | 0.797 | 5 | 1.240 | 5 |
| 14 | 1.723 | 42 | 2.481 | 22 |
| 15 | 1.313 | 4 | 2.691 | 38 |
| 16 | 1.670 | 34 | 1.971 | 6 |
| 17 | 0.759 | 11 | 1.445 | 17 |
| 18 | 0.836 | 40 | 1.119 | 10 |
| 19 | 1.109 | 12 | 2.667 | 41 |
| 20 | 1.780 | 31 | 2.573 | 26 |

For the 20 ligands provided for Subchallenge 1, pose prediction category, we generated 50 conformations with OMEGA. For each conformation. RMSD from the bound form of the ligand was measured and PL-PatchSurfer2 score was given using their cognate structure as a receptor

The second column from left shows the minimum RMSD observed in the 50 conformations for each ligand. The score rank of the minimum RMSD conformer is shown in the subsequent column. The column for RMSD of lowest conformer score (LCS) shows the RMSD of the top ligand with the top PL-PatchSurfer2.0 score, and the rank of the RMSD among the 50 conformers is shown on the rightmost column

because it worked well in the Community Structure–Activity Resource (CSAR) 2014 [47], which we have participated in. Submissions were evaluated by the average RMSD of the pose 1 conformations over the 20 ligands and the average RMSD of the best out of 20 poses. In terms of the average RMSD of pose 1, our group was ranked 67th (the RMSD was 7.71 Å) out of 74 submissions and 61st (RMSD: 8.26 Å) out of 70 submissions for stage 1a and stage 1b, respectively. The average RMSD of the best pose of the 20 ligands was 4.06 Å and 8.26 Å, respectively. Thus, our protocol was not very successful in these two stages.

To understand the reason for the failure, we analyzed the performance of binding pose generation by AutoDock Vina and scoring by PL-PatchSurfer2.0 separately by applying the two methods to 5YGX and 6EJ2, BACE-1 structures with a bound ligand conformation. For each of these structures, the cognate ligand was removed from the binding pocket and re-docked by AutoDock Vina. 20 poses were generated and evaluated by the PL-PatchSurfer2.0 score.

Figure 4a, b showed the lowest scoring poses generated by AutoDock Vina for 5YGX and 6EJ2, respectively. Their RMSD values were 8.19 Å and 0.23 Å, respectively. The RMSDs of predicted 20 conformations of 5YGX ranged from 4.41 to 9.07 Å even though it was a cognate ligand docking, where we expected better (lower) RMSD values. Therefore, obviously, the pose generating step failed for 5YGX. On the other hand, for 6EJ2, a near-native binding pose was successfully generated and PL-PatchSurfer2.0 was able to select it. In Fig. 4c we show the distribution of the PL-Patchsurfer2.0 score relative to RMSD values of the poses generated by AutoDock Vina. This result suggests that PL-PatchSurfer2.0 would be able to select a near-native bound pose if any in the ensemble of poses generated by AutoDock Vina.

We further analyzed what actually happened in the Challenge in Table 3, where we examined the quality of ligand conformations generated at each step in the pose prediction. The steps were the initial ligand conformation generation by OpenBabel (the two left columns in Table 3), the docking using Autodock Vina, and the docking pose selection by PL-PatchSurfer2.0. It turned out that generating the initial ligand conformation was not very successful as RMSD values were all around 2 Å or higher. The docking poses generated by Autodock Vina (the columns named "Stage 1a/1b lowest RMSD samples") were also not very accurate. RMSDs of the initial conformations by OpenBabel and the docking poses by Autodock Vina had only a weak correlation (0.16 for stage 1a and 0.33 for stage 1b). However, the quality of the docking poses would have been better if more accurate initial structures had been generated because the cognate docking results that used the bound form of the ligands (the rightmost column) were much better. Finally, there is a substantial correlation between the PL-PachSurfer2.0's choices and the lowest RMSD of conformation ensembles by Autodock Vina (0.67 for stage 1a and 0.52 for stage 1b). Therefore, consistent with the previous analysis with Fig. 4, the pose prediction by PL-PatchSurfer2.0 would have been better with better pose generation by Autodock Vina.

The unsuccessful sampling of AutoDock Vina might have also influenced other participants. For stage 1a, 17 groups were reported to have an average pose 1 RMSD of over 4.0 Å, and 11 groups out of them used AutoDock Vina or its variants as their sampling method.

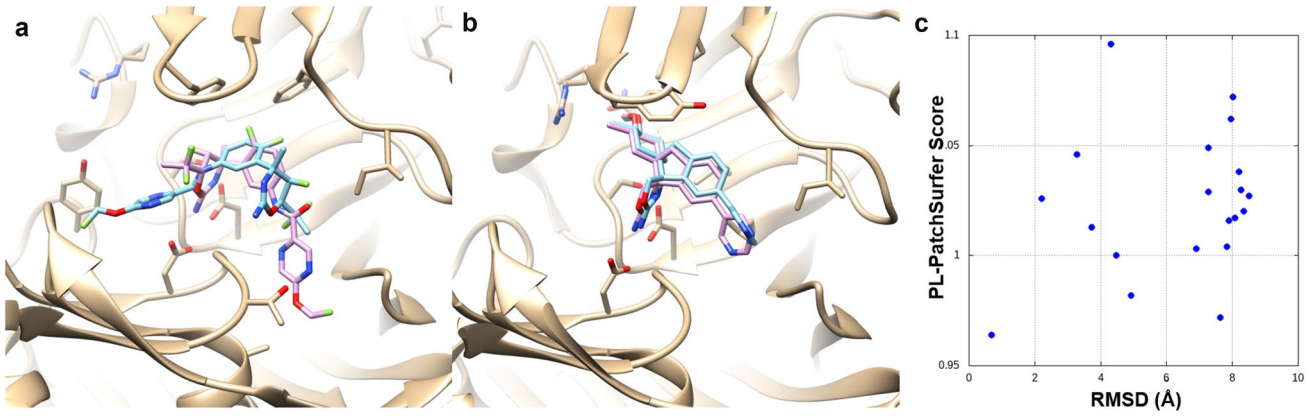


**Fig. 4** The lowest scoring poses (blue) and their native structures (pink) of AutoDock Vina cognate docking. **a** 5YGX. RMSD: 8.19 Å, **b** 6EJ2. RMSD: 0.23 Å, **c** PL-PatchSurfer2.0 score of 20 6EJ2 cognate docking poses relative to RMSD values to the native

**Table 3** RMSD values of ligands at each step of the pose prediction of BACE-1 in stage 1

| Ligand | OpenBabel Conformation | OpenBabel Macrocycle | Stage 1a top scoring pose by PL-PS2.0 | Stage 1a lowest RMSD sampled | Stage 1b top scoring pose by PL-PS2.0 | Stage 1b lowest RMSD sampled | AutoDock Vina cognate docking top pose |
|---|---|---|---|---|---|---|---|
| 1 | 3.10 | 1.86 | 10.74 | 4.11 | 8.39 | 1.90 | 9.49 |
| 2 | 2.90 | 1.79 | 9.42 | 2.96 | 10.99 | 1.96 | 0.45 |
| 3 | 3.38 | 2.79 | 3.24 | 2.80 | 7.88 | 6.29 | 0.65 |
| 4 | 3.38 | 3.06 | 11.03 | 3.11 | 7.89 | 7.92 | 0.33 |
| 5 | 2.91 | 3.50 | 5.42 | 3.50 | 4.58 | 4.58 | 3.06 |
| 6 | 5.00 | 3.17 | 7.26 | 3.31 | 14.07 | 9.43 | 9.76 |
| 7 | 4.10 | 3.42 | 6.52 | 5.24 | 13.21 | 9.56 | 9.21 |
| 8 | 3.67 | 3.57 | 9.89 | 3.64 | 5.51 | 5.51 | 0.36 |
| 9 | 3.12 | 2.59 | 6.12 | 2.79 | 9.62 | 4.53 | 0.72 |
| 10 | 3.57 | 2.79 | 12.74 | 6.46 | 11.87 | 8.06 | 0.93 |
| 11 | 3.34 | 0.80 | 1.66 | 1.66 | 7.17 | 2.13 | 11.24 |
| 12 | 2.53 | 1.43 | 10.45 | 4.20 | 3.02 | 2.13 | 0.62 |
| 13 | 2.58 | 1.95 | 5.33 | 2.75 | 2.51 | 2.51 | 0.38 |
| 14 | 2.94 | 2.13 | 4.97 | 2.98 | 8.46 | 3.04 | 1.07 |
| 15 | 4.14 | 2.71 | 9.43 | 3.86 | 8.10 | 3.59 | 0.35 |
| 16 | 2.15 | 1.33 | 9.12 | 2.54 | 2.72 | 1.58 | 12.18 |
| 17 | 1.93 | 2.18 | 4.79 | 3.10 | 11.75 | 5.89 | 0.46 |
| 18 | 2.59 | 1.80 | 9.55 | 3.82 | 8.58 | 6.12 | 0.47 |
| 19 | 2.97 | 2.75 | 8.75 | 3.84 | 10.62 | 3.84 | 0.60 |
| 20 | 4.99 | N/A[a] | 2.78 | 2.70 | 8.96 | 1.09 | 1.25 |

OpenBabel Conformation and OpenBabel Macrocycle columns (2nd and 3rd column) are RMSDs of the initial ligand conformations that were input for Autodock Vina. The RMSD values were computed relative to the bound form of each ligand (or Macrocycle only) by aligning the ligand only

The fourth and the sixth columns from the left are the RMSD values of the top scoring poses by PL-PatchSurfer2.0. The RMSD values were calculated after superimposition of the receptor structures

The fifth and the seventh columns are the RMSD values of the nearest-native conformation out of 20 conformations sampled by AutoDock Vina,

The last column shows the lowest RMSD values of cognate docking performed with Autodock Vina

[a]Ligand 20 does not have a macrocycle

## Conclusions

We have participated in GC4 of D3R with our software, PL-PatchSurfer2.0. The challenge provided a valuable opportunity for the community to understand the current status of existing computational methods as well as for the developers to reexamine their own methods. Our group was moderately successful in the binding affinity ranking but failed in the binding pose prediction challenges. Overall, the scoring by PL-PatchSurfer2.0 was working throughout the challenges, as we discussed, there are several technical lessons we learned through post-analysis we performed. Particularly, for binding pose prediction, the initial structure generation and the pose-generating step needs to be revisited as OpenBabel AutoDock Vina did not work well in this challenge.

**Acknowledgements** We thank Charles Christoffer for proofreading the manuscript. This work was partly supported by the National Institutes of Health (R01GM123055), the National Science Foundation (DMS1614777, CMMI1825941), and the Purdue Institute of Drug Discovery.

## References

1. Śledź P, Caflisch A (2018) Protein structure-based drug design: from docking to molecular dynamics. Curr Opin Str Biol 48:93–102
2. Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational methods in drug discovery. Pharmacol Rev 66:334–395
3. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461
4. Allen WJ, Balius TE, Mukherjee Sm Brozell SR, Moustakas DT, Lang PT, Case DA, Kuntz ID, Rizzo RC (2015) DOCK6: impact of new features and current docking performance. J Comput Chem 36:1132–1156
5. Sauton M, Lagorce D, Villoutreix BO, Miteva MA (2008) MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinf 9:184

6. Craig IR, Essex JW, Speigel K (2010) Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. J Chem Inf Model 50:511–524
7. Shivakumar D, Williams J, Wu Y, Damm W, Shelly J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput 6:1509–1519
8. Wang L, Deng Y, Wu Y, Kim B, LeBard DN, Wandschneider D, Beachy M, Friesner RA, Abel R (2017) Accurate modeling of scaffold hopping transformations in drug discovery. J Chem Theory Comput 13:42–54
9. Dong X, Ebalunode JO, Yang SY, Zheng W (2011) Receptor-based pharmacophore and pharmacophore key descriptors for virtual screening and QSAR modeling. Curr Comput Aided Drug Des 7:181–189
10. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ (2016) Computational protein–ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc 11:905–919
11. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL key for use in drug discovery. J Chem Inf Model 42:1273–1280
12. Venkatraman V, Chakravarthy PR, Kihara D (2009) Application of 3D Zernike descriptors to shape-based ligand similarity searching. J Cheminform 1:19
13. Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V (2010) Advances in computational methods to predict the biological activity of compounds. Expert Opin Drug Discov 5:633–654
14. Hattori M, Okuno Y, Goto S, Kanehisa M (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. J Am Chem Soc 125:11853–11865
15. Koes DR, Camacho CJ (2011) Pharmer: efficient and exact pharmacophore search. J Chem Inf Model 51:1307–1314
16. Nguyen DD, Cang Z, Wu K, Wang M, Cao Y, Wei G-W (2019) Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. J Comput-Aided Mol Des 33:71–82
17. Li H, Leung K-S, Wong M-H, Ballester PJ (2014) Substituting random forest for multiple linear regression improves binding affinity prediction of scoring function: cyscore as a case study. BMC Bioinf 15:1
18. Jimenez J, Skalic M, Martinez-Rosell G, De Fabritiis G (2018) $K_{DEEP}$: protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Info Model 58:287–296
19. Feinberg EN, Sur D, Wu Z, Husic BE, Mai H, Li Y, Sun S, Yang J, Ramsundar B, Pande VS (2018) PotentialNet for molecular property prediction. ACS Cent Sci 4:1520–1530
20. Cang ZX, Wei G-W (2017) TopologyNet: topology based deep convolutional and multi-task neural networks for biomolecular property predictions. PLoS Comput Biol 13:e1005690
21. Galeb Z, Parks CD, Chiu M, Yang H, Shao C, Walters WP, Lambert MH, Nevins N, Bembenek SD, Ameriks MK, Mirzadegan T, Burley SK, Amaro RE, Gilson MK (2019) D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings. J Comput-Aided Mol Des 33:1–18
22. Shin W-H, Christoffer C, Wang J, Kihara D (2016) PL-PatchSurfer2: improved local surface matching-based virtual screening method that is tolerant to target and ligand structure variation. J Chem Inf Model 56:1676–1691
23. Shin W, Kihara D (2018) Virtual ligand screening using PL-PatchSurfer2, a molecular surface-based protein–ligand docking method. Methods Mol Biol 1762:105–121
24. Hu B, Zhu X, Monroe L, Bures MG, Kihara D (2014) PL-PatchSurfer: a novel molecular local surface-based method for exploring protein–ligand interactions. Int J Mol Sci 15:15122–15145
25. Novotni M, Klein R (2003) 3D Zernike descriptors for content based shape retrieval. In: Proceedings of eighth ACM symposium on solid modeling and applications, Washington, pp 216–225
26. Han X, Sit A, Christoffer C, Chen S, Kihara D (2019) A global map of the protein shape universe. PLoS Comput Biol 15:e1006969
27. Sael L, Li B, La D, Fang Y, Ramani K, Rustamov R, Kihara D (2008) Fast protein tertiary structure retrieval based on global surface shape similarity. Proteins 72:1259–1273
28. Esquivel-Rodríguez J, Xiong Y, Han X, Guang S, Christoffer C, Kihara D (2015) Navigating 3D electron microscopy maps with EM-SURFER. BMC Bioinf 16:181
29. Venkatraman V, Yang YD, Sael L, Kihara D (2009) Protein–protein docking using region-based 3D Zernike descriptors. BMC Bioinf 10:407
30. Esquivel-Rodriguez J, Yang YD, Kihara D (2012) Multi-LZerD: multiple protein docking for asymmetric complexes. Proteins 80:1818–1833
31. Shin W-H, Zhu X, Bures MG, Kihara D (2015) Three-dimensional compound comparison methods and their application in drug discovery. Molecules 20:12841–12862
32. Zhu X, Xiong Y, Kihara D (2015) Large-scale binding ligand prediction by improved patch-based method Patch-Surfer2.0. Bioinformatics 31:707–713
33. Chikhi R, Sael L, Kihara D (2010) Real-time ligand binding pocket database search using local surface descriptors. Proteins 78:2007–2028
34. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nano-systems: application to microtubules and the ribosome. Proc Natl Acad USA 98:10037–10041
35. Heiden W, Moeckel G, Brickmann J (1993) A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. J Comput-Aided Mol Des 7:503–514
36. Cheng T, Zhao Y, Li X, Lin F, Xu Y, Zhang X, Li Y, Wang R (2007) Computation of octanol–water partition coefficients by guiding an additive model with knowledge. J Chem Inf Model 47:2140–2148
37. Prati F, Bottegoni G, Bolognesi ML, Cavalli A (2018) BACE-1 inhibitors: from recent single-target molecules to multitarget compounds for Alzheimer's disease. J Med Chem 61:619–637
38. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo LD, Christie C, Dalenberg K, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranović V, Guzenko D, Hudson DP, Kalro T, Liang Y, Lowe R, Namkoong H, Peisach E, Periskova I, Prlić A, Randle C, Rose A, Rose P, Sala R, Sekharan M, Shao C, Tan L, Tao Y-P, Valasatava Y, Voigt M, Westbrook J, Woo J, Yang H, Young J, Zhuravleva M, Zardeck C (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47:D464–474
39. Ameriks MK, Bembenek SD, Burdett MT, Choong IC, Edwards JP, Gebauer D, Gu Y, Karlsson L, Purkey HE, Staker BL, Sun S, Thurmond RL, Zhu J (2010) Diazinones as P2 replacements for pyrazole-based cathepsin S inhibitors. Bioorg Med Chem Lett 20:4060–4064
40. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminf 3:33
41. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge Structural Database. J Chem Inf Model 50:572–584

42. Boratyn GM, Camacho C, Copper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y, Raytselis Y, Sayers EW, Tao T, Ye J, Zaretskaya I (2013) BLAST: a more efficient report with usability improvements. Nucleic Acids Res 41:W29–W33
43. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35:W522–525
44. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309
45. RDKit: Open-source cheminformatics. https://www.rdkit.org
46. Lee-Dutra A, Wiener DK, Sun S (2011) Cathepsin S inhibitors: 2004–2010. Expert Opin Ther Pat 21:311–337
47. Zhu X, Shin W-H, Kim H, Kihara D (2016) Combined approach of Patch-Surfer and PL-PatchSurfer for protein–ligand binding prediction in CSAR 2013 and 2014. J Chem Inf Model 56:1088–1099