HHS Public Access

Author manuscript

J Comput Aided Mol Des. Author manuscript; available in PMC 2018 September 04.

Published in final edited form as:

J Comput Aided Mol Des. 2017 August ; 31(8): 689–699. doi:10.1007/s10822-017-0038-1.

Improving Binding Mode and Binding Affinity Predictions of Docking by Ligand-based Search of Protein Conformations: Evaluation in D3R Grand Challenge 2015

Xianjin Xu[1], **Chengfei Yan**[1,2], and **Xiaoqin Zou**[1,2,3,4,*]

[1]Dalton Cardiovascular Research Center, University of Missouri, Columbia, MO 65211, USA

[2]Department of Physics and Astronomy, University of Missouri, Columbia, MO 65211, USA

[3]Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA

[4]Informatics Institute, University of Missouri, Columbia, MO 65211, USA

Abstract

The growing number of protein-ligand complex structures, particularly the structures of proteins co-bound with different ligands, in the Protein Data Bank helps us tackle two major challenges in molecular docking studies: the protein flexibility and the scoring function. Here, we introduced a systematic strategy by using the information embedded in the known protein-ligand complex structures to improve both binding mode and binding affinity predictions. Specifically, a ligand similarity calculation method was employed to search a receptor structure with a bound ligand sharing high similarity with the query ligand for the docking use. The strategy was applied to the two datasets (HSP90 and MAP4K4) in recent D3R Grand Challenge 2015. In addition, for the HSP90 dataset, a system-specific scoring function (ITScore2_hsp90) was generated by recalibrating our statistical potential-based scoring function (ITScore2) using the known protein-ligand complex structures and the statistical mechanics-based iterative method. For the HSP90 dataset, better performances were achieved for both binding mode and binding affinity predictions comparing with the original ITScore2 and with ensemble docking. For the MAP4K4 dataset, although there were only eight known protein-ligand complex structures, our docking strategy achieved a comparable performance with ensemble docking. Our method for receptor conformational selection and iterative method for the development of system-specific statistical potential-based scoring functions can be easily applied to other protein targets that have a number of protein-ligand complex structures available to improve predictions on binding.

Keywords

molecular docking; scoring function; binding mode; binding affinity; ligand similarity; drug discovery

*Correspondence to: Xiaoqin Zou, zoux@missouri.edu.

1 Introduction

In order to promote the improvement and the development of computational methods for new drug discovery, several valuable benchmarking datasets containing experimentally determined and accurately assessed data (e.g. binding modes and affinities) of a series of protein-ligand complexes were provided by the Community Structure-Activity Resource (CSAR, http://www.csardock.org/, 2010–2014) [1–4] at first and then by the Drug Design Data Resource (D3R, https://www.drugdesigndata.org, starting from 2015). In the Grand Challenge 2015, D3R presented two protein-ligand datasets, one for human heat shock protein 90 (HSP90) [5] and the other one for human mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4) [6]. Two prediction stages were held for each dataset. In stage 1, the D3R organizers provided the ligand SMILES strings and several protein structures (four for HSP90 and two for MAP4K4, respectively) covering two distinct conformational states, “open” and “closed”. Participants were free to use other protein structures for docking, and were required to predict both binding modes and binding affinities (or affinity rankings). In stage 2, the crystal structures of protein-ligand complexes for binding mode prediction in stage 1 were released, and the participants were asked to re-predict the binding affinities (or affinity rankings).

A common way to make such predictions is molecular docking, which predicts the binding mode and binding affinity of a complex formed by a protein and a ligand. Molecular docking is one of the most powerful computational tools in the modern era of structure-based drug discovery [7–11]. Generally, a docking strategy can be divided into two components, sampling and scoring. For sampling, putative binding modes (orientations/conformations) of a ligand near the binding site of a protein are generated through a searching algorithm. Currently, molecular docking programs are able to quite exhaustively search the conformational space for a small molecule such as a ligand, but protein flexibility remains a big challenge due to the huge number of degrees of freedom. The scoring component aims to predict the binding energy for a given binding mode by using either a physical or empirical energy function. The mode with the lowest energy score is predicted as the most favorable binding mode. Moreover, scoring functions are also used to rank the binding affinities of different ligands against a given protein receptor (*i.e.* virtual screening). Therefore, an ideal scoring function should be able to distinguish the native or near-native binding mode from other alternative modes, and to distinguish active ligands from inactive ligands. Unfortunately, such an ideal scoring function is difficult to derive, because of the complicated solvent effect, conformational entropic effect, and so on. A rigorous scoring function would be too computationally expensive to be used for molecular docking studies in which hundreds of putative binding modes are generated for a single ligand. The development of an efficient and reliable scoring function is another challenge for molecular docking.

With the increasing number of crystal structures solved for a specific protein (e.g., an important druggable protein) co-bound with distinct ligands in the Protein Data Bank (PDB) [12], the abundant structural information points to a new way to tackle challenges in molecular docking such as protein flexibility and imperfect scoring functions. For example, for one of the D3R targets, human HSP90, there are over 200 entries in the PDB containing

178 distinct co-bound ligands. A straightforward strategy is to consider all the conformations of the protein available in the PDB during the docking process, which is referred to as the ensemble docking strategy (i.e. docking with multiple receptor conformations) [13]. However, the computational time increases with the number of receptor conformations, and the chance of false positives also increases due to the inaccuracy of the current scoring functions [14]. Therefore, in this study, instead of directly using multiple protein conformations, the information of co-bound ligands was considered to serve as a filter for relevant protein conformational changes. The approach is based on the molecular similarity principle, which states that molecules with similar structures tend to have similar biological activities [15,16]. In this approach, it is assumed that similar protein conformations are induced by the binding of similar ligands [17,18]. Specifically, ligand similarities were calculated for a query ligand against all co-bound ligands in the PDB for the same protein. The PDB entry containing a co-bound ligand that has the best ligand similarity score with the query ligand was further used for docking. Moreover, for the HSP90 target, a system-specific scoring function was recalibrated based on our latest version of statistical potentials, ITScore2 [19–21]. ITScore2 was originally derived using an iterative method based on statistical mechanical principles that circumvents the long-standing reference state problem in the statistical potential-based scoring functions. In this study, ITScore2 was recalibrated using the protein-ligand complex structures available for HSP90 in the PDB to generate an HSP90-specific scoring function, referred to as ITScore2_hsp90. The results of HSP90 showed that our new strategy improves the performance on both binding mode and binding affinity predictions. Regarding the MAP4K4 dataset, encouragingly, although there were only a few number of protein-ligand complex structures available for MAP4K4, our strategy of docking a query ligand into the protein structure that was selected with the ligand-similarity approach alone achieved a comparable performance with ensemble docking for binding mode prediction.

2 Materials and Methods

2.1 D3R datasets

Two datasets, HSP90 and MAP4K4, were held for the D3R Grand Challenge 2015 and are described in detail in the overview article of this special issue [22]. Briefly, the HSP90 dataset consists of 180 ligands that were designed to bind to the ATP binding site of HSP90. The molecular weights (MWs) of all the atoms of these ligands range from 135 to 566. The total numbers of all the atoms range from 18 to 58. The total numbers of the rotatable bonds range from 0 to 10. Four crystal structures of the protein receptor (PDB IDs: 2JJC, 2XDX, 4YKR, and 4YKY) were provided by the D3R organizers. The stage 1 of HSP90 was to predict the binding modes of six specified ligands (HSP90_40, HSP90_44, HSP90_73, HSP90_164, HSP90_175, and HSP90_179) and to predict the binding affinities of all the 180 ligands against the HSP90. The MWs of the six ligands with known binding modes range from 222 to 447, and their total numbers of the rotatable bonds range from 3 to 7. The stage 2 was to re-predict the binding affinities of all the 180 ligands, with the provision of the six crystal structures that were held for the binding mode prediction in stage 1.

On the other hand, the MAP4K4 dataset comprises 30 ligands binding to the active site of the kinase MAP4K4. The MWs of these ligands range from 174 to 439. The total numbers of heavy atoms range from 23 to 45. The total numbers of the rotatable bonds range from 1 to 6. Two crystal structures of MAP4K4 (PDB IDs: 4OBO and 4U44) were provided by the organizers. The stage 1 was to predict the binding modes of all the 30 ligands and the binding affinities of 18 of these ligands against the MAP4K4. The stage 2 was to re-predict the binding affinities of the 18 ligands with the provision of all the co-bound crystal structures available to D3R.

2.2 Available protein-ligand complex structures

All the released protein structures with co-bound ligands for HSP90 and MAP4K4 were collected from the Protein Data Bank (PDB). Only the PDB entries in which the protein receptors were from the same species as the D3R targets (i.e., human) were kept. Each PDB entry was searched for the HET information; if an entry contained only water molecules or ions, it was removed because of the absence of a ligand. The remaining PDB entries were manually examined. The PDB entries containing functional small molecules binding to the active site were kept, and those containing only buffer or detergent ligands were removed from the dataset. If there were more than one PDB entries containing an identical ligand, the structure with a higher resolution was kept. A total of 178 protein-ligand complex structures were obtained for HSP90 and their PDB IDs are listed in the Supplementary Materials. The same strategy was applied to MAP4K4; a total of eight protein-ligand complex structures were found, and their PDB IDs are 4OBO, 4OBP, 4OBQ, 4RVT, 4U43, 4U44, 4U45, and 4ZK5.

2.3 Ligand similarity calculation

The ligand similarity between a query ligand (i.e., a D3R ligand) and a template ligand (i.e., a ligand in one of the released protein-ligand complex structures) was calculated by the program SHAFTS [23, 24]. SHAFTS is a hybrid approach for 3D similarity calculation by combining molecular shape overlay and pharmacophore feature matching between the two ligands. Thus, the hybrid similarity consists of a ShapeScore (shape-densities overlap) and a FeatureScore (pharmacophore feature fit values). Both ShapeScore and FeatureScore are normalized to [0, 1]. The hybrid similarity is the sum of ShapeScore and FeatureScore and scaled to [0, 2], with 0 corresponding to no similarity and 2 corresponding to the same ligand.

Because SHAFTS uses a semi-rigid strategy to characterize molecular flexibility, the conformational ensembles of a query ligand need to be pre-generated. In this study, OMEGA (version 2.5.1.4 OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) [25, 26] was used to generate 3D conformations (with a maximum number of 200) of query ligands using the SMILES strings provided by the D3R organizers.

2.4 Scoring functions

In this round of D3R challenge, an in-house scoring function ITScore was used for both binding mode evaluation and binding affinity prediction. ITScore is an efficient statistical potential-based scoring function derived by using a statistical mechanics-based iterative

method based on a training set of experimentally determined protein-ligand complex structures. The basic idea of the iterative method is to circumvent the reference state problem by improving atom-pair potentials iteratively through the comparison of the calculated ($g_{ij}^{k}(r)$) and the experimental pair distribution functions ($g_{ij}^{*}(r)$) until the native complex structures in the training database can be correctly discriminated from decoy structures. The idea is described by the following equations:

$$u_{ij}^{k+1}(r) = u_{ij}^{k}(r) + \Delta u_{ij}^{k}(r) \quad (1)$$

$$\Delta u_{ij}^{k} = \lambda k_B T\left[g_{ij}^{k}(r) - g_{ij}^{*}(r)\right] \quad (2)$$

Here, i and j represent the types of a pair of atoms, r is the distance between the two atoms, k_B is the Boltzmann constant, and T is the absolute temperature. Without loss of generality, k_BT is set to unit 1. λ is a parameter to control the speed of convergence and is set to 0.5. $u_{ij}^{k}(r)$ is the pairwise statistical potential of atom pair ij in the k-th step, and $u_{ij}^{k+1}(r)$ is the updated interaction potential in the following step. Given a set of initial potentials $\{u_{ij}^{0}(r)\}$, the potentials are updated based on the above equations until all the native binding modes have the lowest energy scores compared with their decoys. A detailed description of the ITScore method is available in the reference [19].

There are two versions of ITScore according to the training sets. The first version, ITScore1 [19], was derived based on a training set of 786 protein-ligand complexes collected from the PDB database. The latest version, ITScore2 [21], was re-trained using the crystal structures in the refined set of PDBbind 2012 [27, 28] containing a total of 2897 protein-ligand complexes; the size of the new training set for ITScore2 was much larger than the old training set for ITScore1. ITScore2 was used in the present study except for HSP90.

For HSP90, because a large number of HSP90-ligand complex structures were available in the PDB, an HSP90-specific scoring function (referred to as ITScore2_hsp90) was re-calibrated using the aforementioned 178 HSP90-ligand complex structures as the training set. Specifically, the potentials in ITScore2 were set as the initial potentials, $\{u_{ij}^{0}(r)\}$, for the derivation of ITScore2_hsp90. Then, decoy binding modes were sampled for each ligand in this training set by using a modified version of AutoDock Vina [21]. In the original AutoDock Vina program [29], the maximum number of output modes is 20. In our modified version [21], this number can be set to any value, and was set to 500 in this study. Namely, up to 500 binding modes were generated for each ligand. Finally, the potentials $\{u_{ij}^{k}(r)\}$ were updated iteratively using Equations (1) and (2) until the native binding modes of all the HSP90-ligand complexes in the training set have the lowest scores compared with their corresponding decoys, respectively. The resulting system-specific scoring function, ITScore2_hsp90, was used in both the binding mode prediction and the binding affinity

prediction for the HSP90 dataset, and was compared with ITScore2. In contrast, for MAP4K4, there exist too few known MAP4K4-ligand complex structures to generate a system-specific scoring function.

2.5 Sampling and scoring

For a query ligand, 3D similarities against all co-bound ligands from the available protein-ligand complex structures were calculated by the program SHAFTS. The PDB entry that had the best hybrid similarity score (cutoff = 0.8) with the query ligand was used for docking.

In the present study, AutoDock Vina [29] was employed for sampling putative binding modes. For a given protein structure, the geometric center of the co-bound ligand was used as the center of the search box. The size of the cubic search box was set to 30 Å, which was large enough to cover the whole binding pocket. In the docking process, the protein was treated as a rigid body and single bonds of ligands were considered to be rotatable. The exhaustiveness value was increased to 30 (default = 8) to ensure exhaustive sampling. The original Vina source code was modified to set the maximum of output modes to 500 (default = 20). The putative binding modes were sampled and ranked by Vina; the output modes were then re-scored with ITScore2_hsp90 and ITScore2 for the HSP90 dataset and with ITScore2 for the MAP4K4 dataset, respectively.

3 Results

To summarize our methods, in the present study, a ligand 3D similarity calculation program named SHAFTS was employed to search for a protein structure with a co-bound ligand sharing high similarity with the query ligand for the docking use. This method is referred to as docking_sel (i.e., docking with a selected receptor conformation based on ligand similarity) in the remaining text. A modified version of AutoDock Vina was used for sampling the putative binding modes. The latest version of an in-house scoring function, ITScore2 and ITScore2_hsp90, was used for both binding mode and binding affinity predictions.

For comparison, the results obtained with the Vina score were also analyzed in this work. Moreover, in order to evaluate the performance of the docking_sel method, the ensemble docking approach (i.e., using multiple receptor conformations) was also used for comparison. In ensemble docking, for a specific protein and a query ligand, the binding modes generated from docking with different protein conformations were merged for ranking. The details of the results on binding mode and binding affinity predictions are presented in the following sections.

3.1 Binding mode prediction

The quality of all the predicted binding modes was assessed by calculating the root-mean-square deviation (RMSD) between a ligand in the predicted mode and its native co-bound mode in the released crystal structure. Specifically, the protein structures were matched using the MatchMaker tool of UCSF Chimera [30], and the RMSDs of the heavy atoms in the ligands were calculated using the maximum common substructure (MCS) functionality of the OEChem Python toolkit (version 2.5.1.4, OpenEye Scientific Software, Santa Fe,

NM. http://www.eyesopen.com) [25, 26]. The MCS functionality was used to correct the ligand atom renumbering and to account for compound symmetries. Then, the top 5 models predicted by each scoring function (ITScore2 and/or ITScore2_hsp90, and Vina score) were analyzed for both docking_sel and ensemble docking.

3.1.1 The HSP90 dataset—For the HSP90 dataset, the putative binding modes were generated either by docking with a selected protein conformation (i.e., docking_sel) or by using two ensemble docking strategies. The two ensemble docking strategies differ in the number of protein conformations: One of them, referred to as ensemble_4, employed the four HSP90 conformations provided by the D3R organizers. The other ensemble docking strategy, referred to as ensemble_32, employed 32 HSP90 conformations (see Table S1). In each sampling method (docking_sel, ensemble_4 or ensemble_32), for a query compound, the binding modes generated from docking with different HSP90 conformations were merged for further ranking. Then, the merged binding modes were ranked by ITScore2, ITScore2_hsp90 and the Vina score, respectively. Therefore, there were a total of nine combinations (i.e., three sampling methods combined with three scoring functions) for the binding mode prediction of a query ligand. The performance of each combination on the six queried HSP90 ligands are shown in Fig. 1.

Fig. 1A shows the RMSDs of these six ligands between the top predicted modes (i.e., associated with the best predicted binding score for each ligand) and their corresponding native binding modes in the crystal structures. In addition, the RMSD values were sorted for each prediction strategy (a combination of a sampling strategy and a scoring strategy), and the mean value and the median value of the RMSDs of the six ligands are reported in the parentheses inside the figure legend. Impressively, docking with the selected protein conformation always achieved a better performance than the ensemble docking strategies for all the three scoring functions. Moreover, for each sampling strategy, the HSP90 system-specific scoring function ITScore2_hsp90 always archived a better performance than the two other scoring functions. In summary, the combination of docking_sel and ITScore2_hsp90 (i.e., ITScore2_hsp90_sel) performed significantly better than other combinations, with mean and median RMSDs of 1.39 Å and 0.76 Å, respectively.

Fig. 1B shows the assessment of binding mode prediction using the best mode (i.e., the mode with the lowest RMSD value in comparison with the crystal structure) among the top 5 predicted binding modes for each ligand case. Both docking_sel and ensemble_4 yielded excellent performances for all the three scoring functions. Notably, small values of mean and median RMSD (≤ 1.10 Å) were achieved. In contrast, ensemble_32, in which a large number (32) of protein conformations were used for docking, yielded disappointing results, with mean or median RMSD ≥ 4.50 Å for all the scoring functions. These results indicate that docking with an inaccurate protein conformation (i.e., a conformation that is significantly different from the bound protein structure) may result in a lower binding score than docking with a protein conformation that is close to the bound structure, due to the inaccuracy of current scoring functions.

As aforementioned, a total of 178 PDB entries containing human HSP90 co-bound with distinct ligands were collected. According to the conformation of a fragment (colored

magenta in Fig. S1) near the binding pocket, the protein conformations were classified into three groups, “closed”, “semi-close”, and “open” states. Notably, the protein structures that were provided by D3R covered only “open” and “closed” states. Therefore, it was examined whether the ligand similarity method correctly selected the right conformation for each ligand. Table S1 lists the PDB ID of the selected protein structure and the corresponding ligand similarity score for each query ligand. Impressively, the protein conformation of each selected structure for docking is consistent with the corresponding protein conformation in the crystal structure for all the six query ligands. For example, for the query ligand HSP90_73, the co-bound ligand in 3RLP shares the highest similarity with HSP90_73, having a hybrid similarity score of 1.58. Fig. 2A shows the complex structure of 3RLP and the superimposed conformation of HSP90_73 taken from the crystal structure (PDB entry: 4YKW) provided by D3R after stage 1. Fig. 2B shows the superimposition of the crystal structure 4YKW and the top binding mode predicted by the strategy ITSore2_hsp90_sel. The RMSD is only 0.6 Å.

For our best binding mode prediction strategy ITScore2_hsp90_sel, the only failed case is HSP90_44 when only the top model was considered, using RMSD ≤ 2.0 Å as the success criterion. However, the large RMSD is contributed by the pyridine-3-sulfonamide group, which is outside the binding pocket, as plotted in Fig. 2C. As reported in the PDB file of this crystal structure (PDB ID: 4YKT), the B factors of the atoms in the pyridine-3-sulfonamide group are much higher than the B factor of other atoms, indicating the intrinsic fluctuations of this group in the binding mode. It is also reported by D3R that this group actually binds to the neighboring monomer of the protein. In fact, our top predicted binding mode is very close to the native binding mode if this fluctuation group of the ligand is not considered, with an RMSD of 0.54 Å.

3.1.2 The MAP4K4 dataset—As mentioned in the Materials and Methods, there were only eight co-bound crystal structures available, insufficient to generate a system-specific scoring function for the MAP4K4 dataset. Thus, only two scoring functions, the Vina score and ITScore2 were used for the MAP4K4 dataset. Similarly, a total of three docking strategies were used for the putative binding mode generation, docking_sel (i.e., docking with a selected protein conformation) and two ensemble docking strategies. One ensemble docking strategy utilized the 2 MAP4K4 conformations that were provided by the D3R organizers, referred to as ensemble_2. The other ensemble docking strategy utilized all the 8 MAP4K4 conformations available in the PDB, referred to as ensemble_8. Therefore, there were a total of six combinations (i.e., three sampling methods combined with two scoring functions) for the binding mode prediction of a query ligand. The performance of each combination on the 30 queried MAP4K4 ligands are shown in Fig. 3.

Fig. 3A and B shows the binding mode prediction results for each strategy considering the top mode and the best in the top 5 modes, respectively. The RMSD values were sorted for each prediction strategy, and the mean value and the median value of RMSDs of the 30 ligands are reported in the figure legend. Because no PDB entry was found that contained a similar ligand (hybrid similarity score ≥ 0.8) for MAP32, only 29 ligands were evaluated in the docking_sel category (i.e., docking with the selected protein structure). The PDB entry

names of the selected protein structures and the corresponding ligand similarity scores are listed in Table S2.

For the Vina score, ensemble dockings performed better than docking_sel, and ensemble_8 (docking using all the 8 MAP4K4 structures) yielded the best performance for both the top 1 and the top 5 modes. For ITScore2, docking_sel and ensemble dockings achieved similar performances when the top mode was considered for each ligand. When top 5 modes were considered, ensemble dockings performed better than docking_sel, and ensemble_8 yielded the best performance, with the mean and median RMSDs of 2.37 Å and 2.06 Å, respectively.

Overall, ensemble dockings performed better than docking_sel for the MAP4K4 dataset. The prediction strategy Vina_score_ensmeble_8 yielded the best performance with a mean RMSD of 4.61 Å when the top mode was considered. ITScore2_ensemble_8 yielded the best performance with a mean RMSD of 2.37 Å when the top 5 modes were considered. It is not surprising that docking_sel did not perform as good as ensemble docking for the MAP4K4 dataset are not as good as the results for the HSP90 dataset. because there were only 8 MAP4K4 crystal structures available for the protein structure selection. A second reason could be the high flexibility of the binding site on MAP4K4 due to a flexible P-loop (see below) [6]. Encouragingly, docking_sel still achieved a comparable performance with the ensemble docking strategies. As shown in Fig. 3A, for ITScore2, the mean/median value of RMSD of docking_sel is 5.44/6.00 Å whereas the corresponding values for ensemble_2 and ensmbel_8 are 5.58/5.57 Å and 5.69/5.82 Å, respectively.

Although all the co-bound crystal structures of the ligands in the MAP4K4 dataset were provided for the D3R community after Stage 1 predictions, so far only the crystal structure of the co-bound ligand MAP15 has been publicly released (PDB entry: 4U41). For this ligand, the PDB entry 4OBP (chain A) was selected for docking_sel. Fig. 4 shows the superimposition of 4OBP and 4U41. The co-bound ligands in these two structures have a good hybrid similarity score (1.47) and share a similar binding mode. However, the binding pocket conformations in the two structures are distinct with each other, especially in the P-loop and in the helix, both colored orange for 4OBP and green for 4U41, respectively. The conformation of the flexible pocket is sensitive to the ligand size, making the binding mode prediction for MAP4K4 much more challenging than for HSP90. In the case of MAP11, the co-bound ligand in the PDB entry 4U44 shares a high similarity (hybrid similarity score = 1.84) with the query ligand, and the binding pocket conformations in the two MAP4K4 structures are very similar. However, a correctly predicted binding mode with an RMSD of 1.02 Å failed to be ranked among the top 5 modes for both the Vina score (#29 in ranking) and ITScore2 (#9 in ranking).

3.2 Binding affinity prediction

Next, the binding affinity prediction results were analyzed for each docking strategy. For each protein, the best (i.e., most negative) score value of a compound was used to compare with the corresponding experimental affinity. Both affinity rankings and free energy predictions were evaluated in this study. Specifically, affinity rankings were evaluated in terms of the Kendall's tau and Spearman's rho rank correlation coefficients. Binding free energy predictions were evaluated in terms of Pearson's correlation coefficients and RMSD.

Different to the RMSD calculated for the binding mode prediction, here, the RMSD is the measure of the average distance between the experimental affinities -pIC_{50} (i.e., log(IC_{50})) and the predicted energy scores of the query ligands. To avoid confusion, RMSD(-pIC_{50}) was used for binding affinity predictions in this section. Considering that the values of Vina score and ITScores have different scales, they were re-scaled by dividing the predicted scores with a factor γ. For each prediction strategy, γ was set to the ratio of the mean value of the predicted scores to the mean value of -pIC_{50}. Therefore, after re-scaling, the predicted free energies have the same mean value as the experimental affinities -pIC_{50}.

Besides the Vina score, ITScore2 and/or ITScore2_hsp90, the change of the solvent-accessible surface area (ΔSASA) of a ligand upon binding and the ligand molecular weight were also used for references in the assessment of affinity rankings and free energy predictions. Naccess2.1.1 [31] was employed to calculate ΔSASA for the HSP90 dataset based on the top mode of the docking strategy ITScore2_hsp90_sel. Naccess2.1.1 were also used for the ΔSASA calculations for the MAP4K4 dataset based on the crystal structures.

3.2.1 HSP90—There were a total of 180 HSP90 ligands for affinity rankings. Regarding the free energy predictions, because D3R did not release the exact IC_{50} values for the ligands with IC_{50} values above 50 μM, only the ligands having $IC_{50} \leq 50$ μM (a total of 150 ligands) were included for the Pearson R and RMSD(-pIC_{50}) calculations. The results of affinity rankings and free energy predictions are reported in Table 1.

For affinities rankings, both the Kendall's tau and Spearman's rho rank correlation coefficients were calculated for each prediction strategy. Both tau and rho range from -1 to 1, where -1 corresponds to a perfectly reversed ranking and 1 corresponds to a perfect ranking. Table 1 shows that all the predicted rankings correlate positively with the experimental results. The tau values for ΔSASA and molecular weight are 0.06 and 0.15, respectively. Interestingly, similar tau values (around 0.19) were observed for the Vina score when different docking strategies (docking_sel or ensemble dockings) were performed. For both ITScore2 and ITScore2_hsp90, docking_sel achieved better performances than ensemble dockings. Moreover, for docking_sel, the HSP90 system-specific scoring function ITScore2_hsp90 yielded a better performance (0.22) than the other two scoring functions, the Vina score and ITScore2. As reported in the overview paper in this special issue [22], the mean and median tau values for the HSP90 affinity rankings are 0.15 and 0.17, respectively, based on the results submitted by nearly 40 participating groups. The performance of our best prediction strategy ITScore2_hsp90_sel was ranked #8 in stage 1 among all the submitted results. In stage 2, the six co-bound crystal structures released by D3R were used for energy score calculations. The energy scores of the other ligands were identical to those in stage 1. Then, the corresponding tau value of ITScore2_hsp90_sel is 0.23, ranked #5 in stage 2. Similar results were observed by using the Spearman's rho as the criterion.

In order to assess the free energy predictions, Pearson's correlation coefficient (R) and RMSD(-pIC_{50}) were calculated between the predicted binding energy scores and -pIC_{50}. Because ΔSASA and molecular weight are positive values, absolute values of -pIC_{50} were used for both R and RMSD(-pIC_{50}) calculations. Interestingly, the molecular weight again performed better than ΔSASA when R was used as the criterion. However, ΔSASA yielded a

smaller value of RMSD(-pIC_{50}) than the molecular weight, indicating that the two parameters complement each other.

As reported in the fourth column of Table 1, for all the three scoring functions, the R value of docking_sel is slightly larger than or equal to the R values of ensemble dockings, indicating that docking with the selected protein conformation also slightly improves the binding free energy predictions. Unfortunately, no improvement was found in RMSD(-pIC_{50}) for docking_sel, as shown in the fifth column of Table 1. Different to the performances in affinity rankings, Vina score yielded a better performance (i.e., larger R and lower RMSD(-pIC_{50})) than ITScore2_hsp90.

3.2.2 MAP4K4—A total of 18 MAP4K4 ligands were provided by D3R for the binding affinity prediction. It is noted that this number is too small to obtain reliable conclusions [32]. The results for both affinity rankings and free energy predictions are reported in Table 2. It is disappointing that only one (vina_score_ensembel_2) out of the six prediction strategies yielded a significantly better performance (for both affinity rankings and free energy predictions) than ΔSASA or molecular weight. The unsatisfactory performances result from the failures in binding mode predictions (see the subsection on binding mode prediction), because the predicted binding scores were based on the corresponding binding modes.

Because all the crystal structures of the MAP4K4 dataset were provided by D3R organizers after stage 1, the binding scores were also calculated based on crystal structures by using both the Vina scoring function and ITScore2. Performances of affinity rankings and free energy predictions for the two scoring functions are reported in Table 2. Both the Vina score and ITScore2 achieved improved performances for affinity rankings, with tau values of 0.24 and 0.32, respectively. ITScore2 performed significantly better than the Vina score for both affinity rankings and free energy predictions (see the last two rows in Table 2). These results suggest that correct binding modes are important to the binding affinity prediction in docking studies.

4 Discussions and Conclusion

Protein flexibility is one of the most challenging problems in molecular docking studies. In the present work, a 3D ligand similarity calculation method was employed to search a protein receptor structure with a bound ligand sharing high similarity with the query ligand for docking. This novel strategy avoids the exhaustively search in the protein conformational space, which is too large to be handled by the existing methods. Our strategy was successfully applied to the HSP90 dataset; significant improvements were achieved especially for the binding mode prediction in comparison to docking with multiple protein conformations (i.e., ensemble docking). For the binding affinity predictions, the strategy yielded better performances than the ensemble docking for ITScore2 and/or ITScore2_hsp90, whereas similar performances were achieved for the Vina score. Obviously, the disadvantage of this strategy is that it cannot be used for a protein target with no or few co-bound crystal structures available. Encouragingly, when this strategy was applied to the MAP4K4 dataset in which there were only eight protein-ligand complex

structures available, its performance was comparable with those obtained by ensemble dockings.

The second challenge in molecular docking is the scoring function. We introduced a system-specific scoring function for HSP90 based on our scoring function ITScore2. Briefly, ITScore2 was recalibrated using the known HSP90-ligand complex structures with the iterative method to generate an HSP90 system-specific scoring function, ITScore2_hsp90. ITScore2_hsp90 always yielded significantly better performance than the original ITScore2 and the Vina score for the binding mode prediction. In addition, improvements for the binding affinity predictions were also observed for docking with a selected protein conformation (docking_sel). However, the method requires a number of co-bound crystal structures for a target protein, which is not suitable for the MAP4K4 dataset. Fortunately, this requirement will be satisfied for more and more protein targets, with the growing number of protein-ligand complex structures deposited in the PDB.

In summary, during the D3R Grand Challenge 2015, we developed a systematic strategy by using the information embedded in the known protein-ligand complex structures to improve both binding mode and affinity predictions. A ligand similarity calculation method was employed to search a receptor structure with a bound ligand sharing high similarity with the query ligand for docking. Our statistical potential-based scoring function, ITScore2, was recalibrated using the known protein-ligand complex structures with the iterative method to generated a system-specific scoring function. This strategy can be easily applied to other protein targets with a number of protein-ligand complex structures available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support to XZ from OpenEye Scientific Software Inc. (Santa Fe, NM, http://www.eyesopen.com) is gratefully acknowledged. This work was supported by the NSF CAREER Award DBI-0953839, the NIH R01GM109980, and the American Heart Association (Midwest Affiliate) 13GRNT16990076 to XZ. The computations were performed on the high performance computing infrastructure supported by NSF CNS-1429294 (PI: Chi-Ren Shyu) and the HPC resources supported by the University of Missouri Bioinformatics Consortium (UMBC).

References

1. Smith RD, Dunbar JB Jr, Ung PM, et al. CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. J Chem Inf Model. 2011; 51:2115–2131. [PubMed: 21809884]
2. Damm-Ganamet KL, Smith RD, Dunbarr JB Jr, et al. CSAR benchmark exercise 2011–2012: evaluation of results from docking and relative ranking of blinded congeneric series. J Chem Inf Model. 2013; 53:1853–1870. [PubMed: 23548044]
3. Smith RD, Damm-Ganamet KL, Dunbar JB Jr, et al. CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. J Chem Inf Model. 2016; 56:1022–1031. [PubMed: 26419257]
4. Carlson HA, Smith RD, Damm-Ganamet KL, et al. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. J Chem Inf Model. 2016; 56:1063–1077. [PubMed: 27149958]

5. Solit DB, Rosen N. Hsp90: a novel target for cancer therapy. Curr Top Med Chem. 2006; 6:1205–1214. [PubMed: 16842157]
6. Crawford TD, Ndubaku CO, Chen H, et al. Discovery of selective 4-Amino-pyridopyrimidine inhibitors of MAP4K4 using fragment-based lead identification and optimization. J Med Chem. 2014; 57:3484–3493. [PubMed: 24673130]
7. Brooijmans N, Kuntz ID. Molecular Recognition and Docking Algorithms. Annu Rev Biophys Biomol Struct. 2003; 32:335–373. [PubMed: 12574069]
8. Kitchen DB, Decornez H, Furr JR, et al. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. Nat Rev Drug Discovery. 2004; 3:935–949. [PubMed: 15520816]
9. Huang S-Y, Zou X. Advances and Challenges in Protein-Ligand Docking. Int J Mol Sci. 2010; 11:3016–3034. [PubMed: 21152288]
10. Huang S-Y, Grinter SZ, Zou X. Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions. Phys Chem Chem Phys. 2010; 12:12899–12908. [PubMed: 20730182]
11. Grinter SZ, Zou X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-based Drug Design. Molecules. 2014; 19:10150–10176. [PubMed: 25019558]
12. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
13. Huang S-Y, Zou X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. Proteins. 2007; 66:399–421. [PubMed: 17096427]
14. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008; 18:178–184. [PubMed: 18302984]
15. Willett P, Barnard JM, Downs GM. Chemical similarity searching. J Chem Inf Model. 1998; 38:983–996.
16. Bender A, Glen RC. Molecular similarity: a key technique in molecular informatics. Org Biomol Chem. 2004; 2:3204–3218. [PubMed: 15534697]
17. Subramanian J, Sharma S, B-Rao C. Modeling and Selection of Flexible Proteins for Structure-Based Drug Design: Backbone and Side Chain Movements in p38 MAPK. ChemMedChem. 2008; 3:336–344. [PubMed: 18081134]
18. Huang SY, Li M, Wang J, et al. HybridDock: A Hybrid Protein–Ligand Docking Protocol Integrating Protein-and Ligand-Based Approaches. J Chem Inf Model. 2016; 56:1078–1087. [PubMed: 26317502]
19. Huang S-Y, Zou X. An Iterative Knowledge-based Scoring Function to Predict Protein-Ligand Interactions: I. Derivation of Interaction Potentials. J Comput Chem. 2006; 27:1866–1875. [PubMed: 16983673]
20. Huang S-Y, Zou X. An Iterative Knowledge-based Scoring Function to Predict Protein−Ligand Interactions: II. Validation of the Scoring Function. J Comput Chem. 2006; 27:1876–1882. [PubMed: 16983671]
21. Yan C, Grinter SZ, Merideth BR, et al. Iterative Knowledge-Based Scoring Functions Derived from Rigid and Flexible Decoy Structures: Evaluation with the 2013 and 2014 CSAR Benchmarks. J Chem Inf Model. 2016; 56:1013–1021. [PubMed: 26389744]
22. Gathiaka S, Liu S, Chiu M, et al. D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. J Comput Aided Mol Des. 2016; 30:651–668. [PubMed: 27696240]
23. Liu X, Jiang H, Li H. SHAFTS: a Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening. J Chem Inf Model. 2011; 51:2372–2385. [PubMed: 21819157]
24. Lu W, Liu X, Cao X, et al. SHAFTS: a Hybrid Approach for 3D Molecular Similarity Calculation. 2. Prospective Case Study in the Discovery of Diverse P90 Ribosomal S6 Protein Kinase 2 Inhibitors to Suppress Cell Migration. J Med Chem. 2011; 54:3564–3574. [PubMed: 21488662]
25. Hawkins PC, Skillman AG, Warren GL, et al. Conformer Generation with Omega: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. J Chem Inf Model. 2010; 50:572–584. [PubMed: 20235588]

26. Hawkins PC, Nicholls A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. J Chem Inf Model. 2012; 52:2919–2936. [PubMed: 23082786]
27. Wang R, Fang X, Lu Y, et al. The PDBbind Database: Methodologies and Updates. J Med Chem. 2005; 48:4111–4119. [PubMed: 15943484]
28. Cheng T, Li X, Li Y, et al. Comparative Assessment of Scoring Functions on a Diverse Test Set. J Chem Inf Model. 2009; 49:1079–1093. [PubMed: 19358517]
29. Trott O, Olson AJ. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J Comput Chem. 2010; 31:455–461. [PubMed: 19499576]
30. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera – A Visualization System for Exploratory Research and Analysis. J Comput Chem. 2004; 25:1605–1612. [PubMed: 15264254]
31. Hubbard SJ, Thornton JM. Naccess. Computer Program. Department of Biochemistry and Molecular Biology, University College London; 1993.
32. Carlson HA. Check your confidence: size really does matter. J Chem Inf Model. 2013; 53:1837–1841. [PubMed: 23909878]

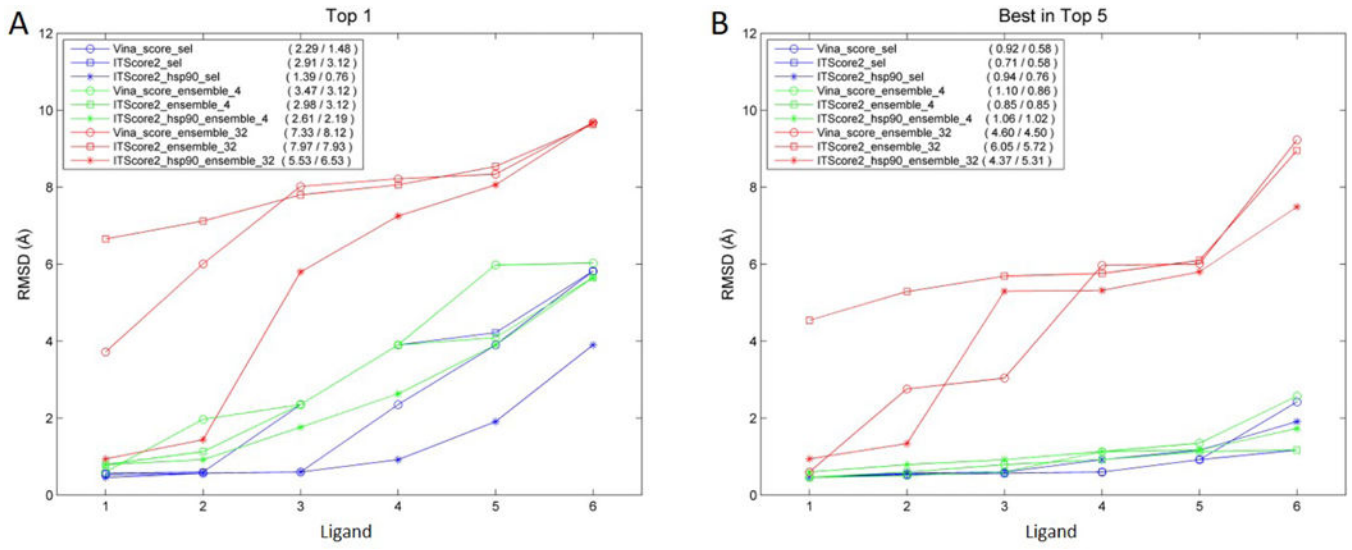

Fig. 1.

Binding mode prediction results of the HSP90 dataset (six queried ligands) based on the top mode (A) or the best (the mode with the lowest RMSD) in the top 5 modes (B) for each prediction strategy. The mean value and the median value of RMSDs of the ligands for each method are reported in the legend.

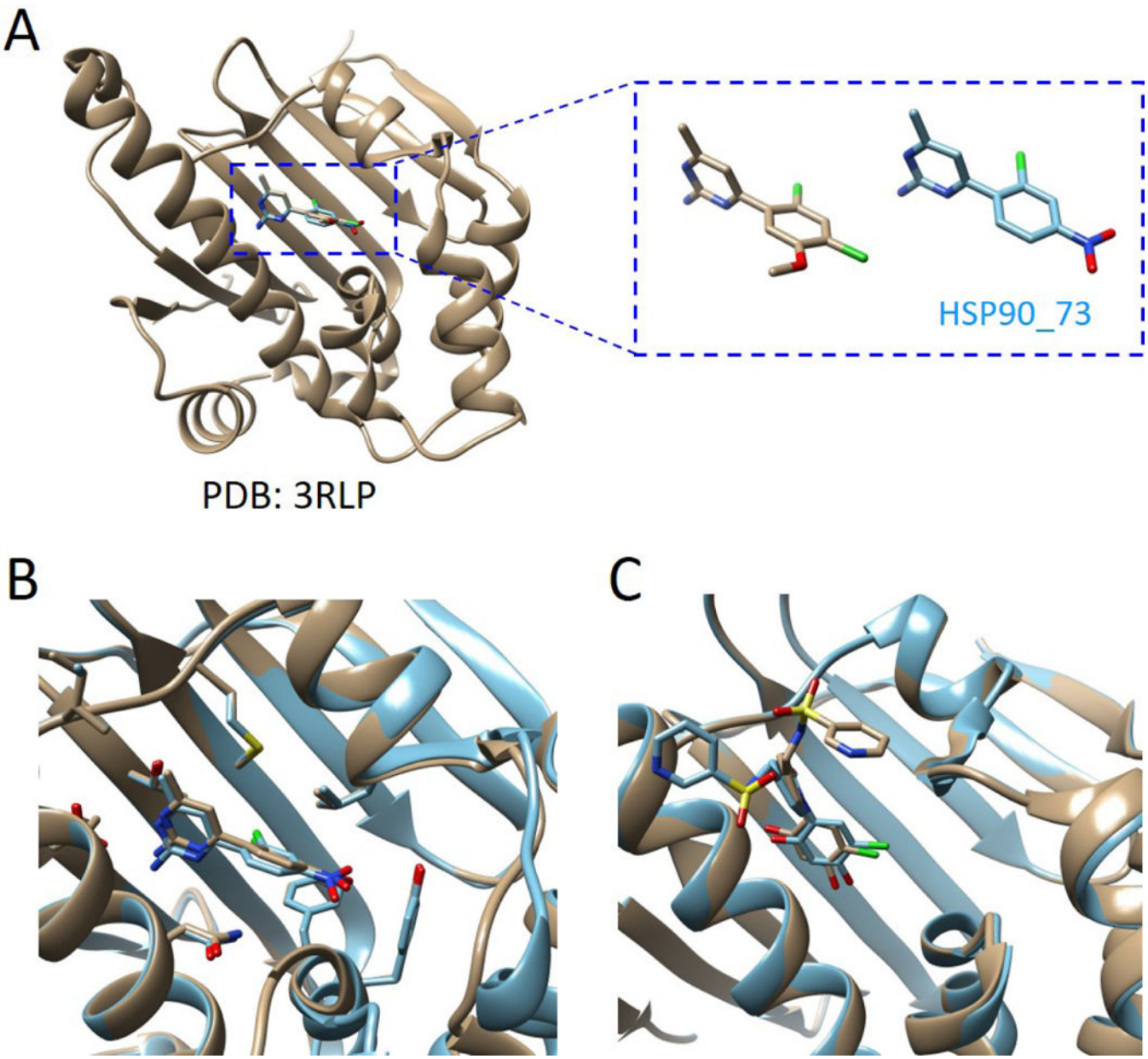

Fig. 2.

A selceted protein structure (A) for docking and two complex structures (B-C) from HSP90 dataset. A. The selected PDB entry (3RLP) containing a co-bound ligand sharing a high 3D similarity (hybrid score 1.58) with the query ligand HSP90_93. B. The top predicted binding mode of HSP90_73 (colored cyan) was superimposed to the crystral structure (PDB ID: 4YKW, colored tan). The proteins are matched by the UCSF Chimera [30]. The ligands and contact residues are represented by the stick model. C. The top predicted binding mode of HSP90_44 (colored cyan) was superimposed to the crystral structure (PDB ID: 4YKT, colored tan).

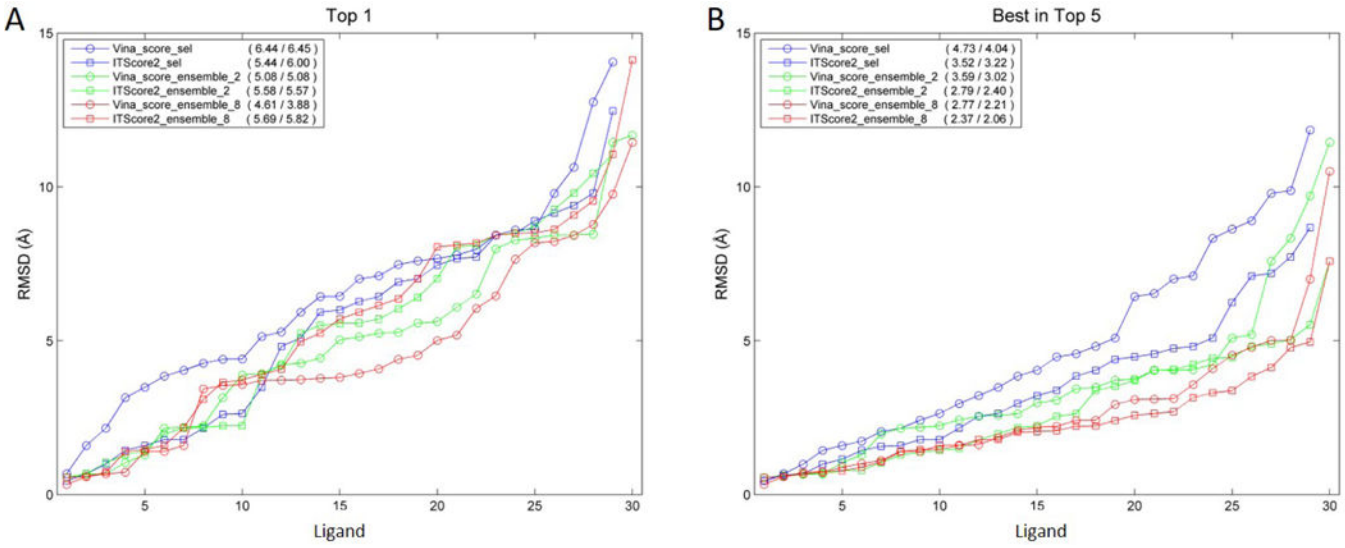

Fig. 3.

Binding mode prediction results of the MAP4K4 dataset (a total of 30 queried ligands) based on the top mode (A) or the best (the mode with the lowest RMSD) in the top 5 modes (B) for each prediction strategy. The mean value and the median value of RMSDs of the ligands for each method are reported in the legend.

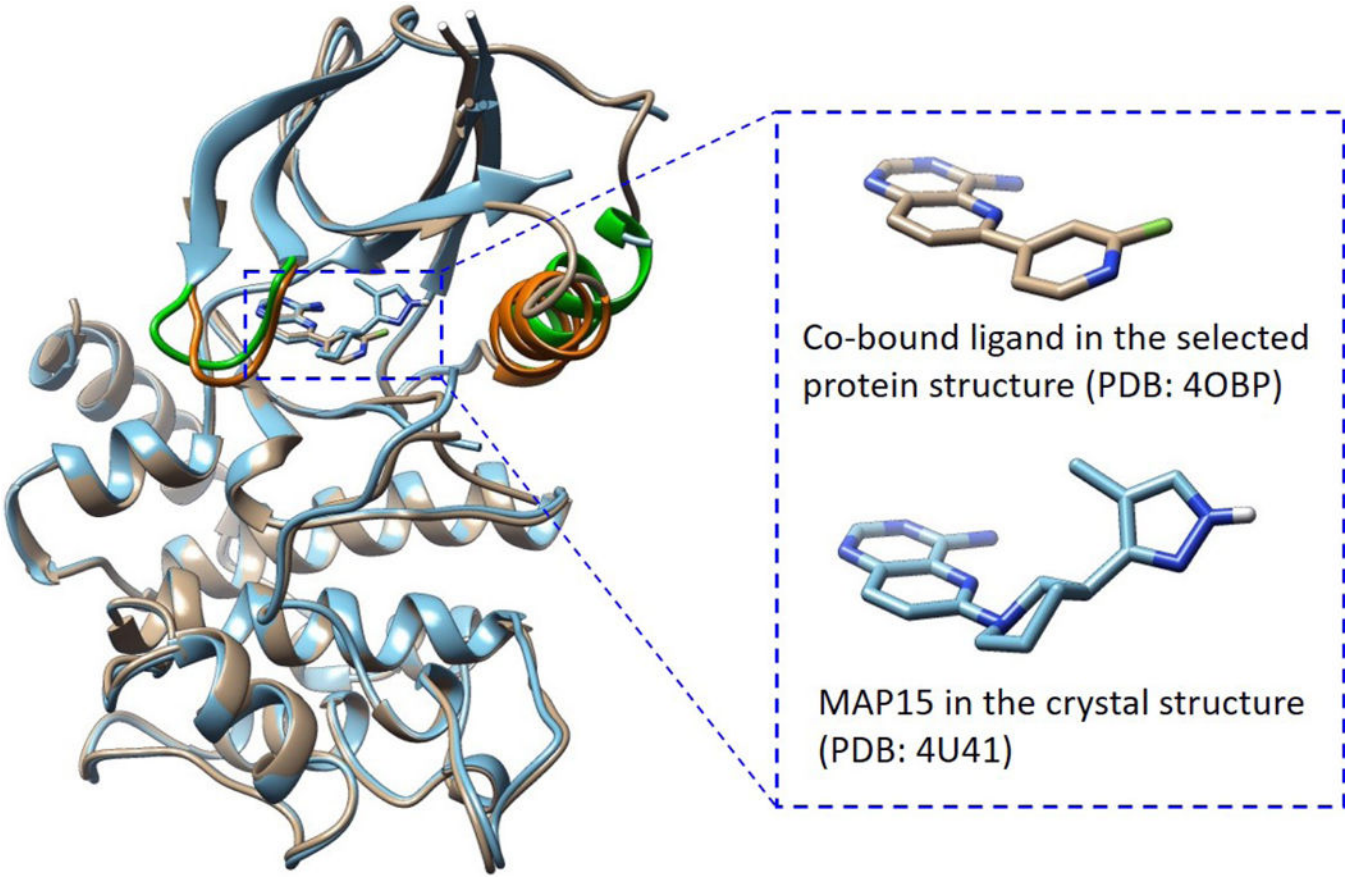

Fig. 4.

Crystal structures of MAP4K4 with co-bound ligands. The cyan-colored structure is the crystal structure of a query ligand MAP15 bound with MAP4K4 (PDB entry: 4U41). The tan-colored structure (PDB entry: 4OBP) was selected by the ligand similarity method and was used for binding mode prediction for MAP15. The flexible regions (the P-loop and the helix) in the binding pocket are colored orange and green for 4OBP and 4U41, respectively.

Table 1

Binding affinity prediction results for the HSP90 dataset.

Prediction strategy	Affinity rankings		Free energy predictions	
	Kendall's tau	Spearman's rho	Pearson R	RMSD (-pIC$_{50}$)
ΔSASA	0.06	0.09	0.19	1.37
Molecular weight	0.15	0.21	0.32	1.62
Vina_score_sel	0.19	0.28	0.39	1.14
ITScore2_sel	0.17	0.24	0.34	1.31
ITScore2_hsp90_sel	0.22	0.31	0.37	1.30
Vina_score_ensemble_4	0.20	0.29	0.39	1.13
ITScore2_ensemble_4	0.14	0.20	0.30	1.30
ITScore2_hsp90_ensemble_4	0.14	0.19	0.26	1.32
Vina_score_ensemble_32	0.18	0.27	0.36	1.18
ITScore2_ensemble_32	0.16	0.23	0.34	1.26
ITScore2_hsp90_ensemble_32	0.16	0.22	0.29	1.24

Table 2

Binding affinity prediction results for the MAP4K4 dataset.

Prediction strategy	Affinity rankings		Free energy predictions	
	Kendall's tau	Spearman's rho	Pearson R	RMSD (-pIC$_{50}$)
ΔSASA	0.12	0.22	0.30	1.19
Molecular weight	0.10	0.17	0.20	1.63
Vina_score_sel	0.11	0.13	0.22	1.32
ITScore2_sel	0.03	0.00	0.03	1.52
Vina_score_ensemble_2	0.22	0.23	0.39	0.97
ITScore2_ensemble_2	0.12	0.22	0.29	1.10
Vina_score_ensemble_8	0.08	0.09	0.23	1.14
ITScore2_ensemble_8	0.11	0.14	0.24	1.25
Vina_score_crystal	0.24	0.23	0.15	1.75
ITScore2_crystal	0.32	0.40	0.30	1.39