Published in final edited form as:

J Comput Aided Mol Des. 2019 March ; 33(3): 367–374. doi:10.1007/s10822-019-00185-0.

Predicting Protein-Ligand Binding Modes for CELPP and GC3: Workflows and Insight

Xianjin Xu^{1,2,3,4,†}, Zhiwei Ma^{1,2,3,4,†}, Rui Duan^{1,2,3,4,†}, and Xiaoqin Zou^{1,2,3,4,*}

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

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

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

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

Abstract

Drug Design Data Resource (D3R) continues to release valuable benchmarking datasets to promote improvement and development of computational methods for new drug discovery. We have developed several methods for protein-ligand binding mode prediction during the participation in the D3R challenges. In the present study, these methods were integrated, automated, and systematically tested using the large-scale data from Continuous Evaluation of Ligand Pose Prediction (CELPP) and a subset of Grand challenge 3 (GC3). The results show that current molecular docking methods benefit from the increasing number of protein-ligand complex structures deposited in Protein Data Bank. Using an appropriate protein structure for docking significantly improves the success rate of the binding mode prediction. The results of our template-based method and docking method are compared and discussed. Our future direction include the combination of these two methods for binding mode prediction.

Keywords

Protein-ligand interaction; Molecular docking; Template-based; Molecular similarity; Drug discovery

Introduction

In silico methods for protein-ligand complex structure prediction (e.g. molecular docking) play an increasingly important role in modern drug discovery and development [1–3]. Although great successes have been achieved, current methods are facing several challenges such as the protein flexibility problem and the scoring function problem [4,5]. To promote the improvement of the existing methods and the development of new approaches, Drug Design Data Resource (D3R, <https://www.drugdesigndata.org>, starting from 2015) continues to release valuable benchmarking datasets containing experimentally determined binding structures and affinity data [6–7]. Recently, the D3R Team has developed Continuous

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

†XX, ZM, and RD contributed equally to this work.

Evaluation of Ligand Pose Prediction (CELPP), an automated workflow to process and evaluate the challenge of protein-ligand binding mode prediction. CELPP is held weekly, in which the targets are prepared based on pre-released data from the Protein Data Bank (PDB) [8] (including the ligands and the sequence of their target proteins). In addition to CELPP, D3R organizes one or two rounds of Grand Challenge (GC) every year for both binding mode and binding affinity predictions. In the latest Grand Challenge 3, D3R released a subset of 24 compounds targeting the human Cathepsin S (CatS) for binding mode prediction.

Through the participation in the D3R and its previous predecessor, Community Structure-Activity Resource (CSAR, <http://www.csardock.org>, 2010–2014) [9–12], we developed several novel strategies for predicting protein-ligand complex structures [13–19]. During the CSAR challenges, we improved our in-house scoring function, ITScore2, 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 [15]. During the GC 2015, we improved both binding mode and binding affinity predictions of molecular docking by using the protein structures that contain the co-bound ligands sharing high molecular similarities with the query ligands [13]. A system-specific scoring function was also introduced in this round of GC. Then, in GC2, we developed a template-based approach using known protein-ligand complex structures as templates [14]. These methods use known protein-ligand complex structures in different ways to improve the performance of binding mode prediction, and benefit from the increasing number of experimentally determined protein-ligand complex structures deposited in the PDB.

In this study, we thoroughly analyzed the binding mode prediction results from our recent participation in the D3R challenges, CELPP and GC3. For CELPP, we predicted more than one thousand targets in 42 CELPP weeks (from week22_2017 to week11_2018). For each target, a molecular docking approach was performed using up to 6 protein structures. Five of these protein structures were provided by CELPP. The 6th protein structure was selected from all the known experimental structures of the same protein by using a 3D molecular similarity calculation method. The docking results showed that using a proper protein structure (e.g., containing a co-bound ligand that shares a high molecular similarity with the query ligand) significantly improved the performance of binding mode prediction for the docking method. In addition to molecular docking, a template-based method was also used and the results were compared with the docking results. The analysis showed that the performance of the template-based method highly depends on the molecular similarities between the co-bound ligands in the templates and the query ligands. Similar results were observed for the GC3 targets (CatS subset). Through systematic validation on the large-scale data from D3R (CELPP and GC), we continue to improve our protein-ligand complex structure prediction methods (both docking-based and template-based), which would become a useful, user-friendly tool for computer-aided drug design (CADD) and for biologists' use.

Materials and Methods

The query protein-ligand complexes

The CELPP targets—CELPP releases 20–100 upcoming protein-ligand complex targets every Saturday (12:01 am), and the participants are required to submit predicted binding modes by the following Tuesday (2:59 pm). Experimental structures will be released in the Protein Data Bank on Wednesday (i.e., the day after the submission deadline). For each target, one query ligand and up to five candidate protein structures are provided. These five protein structures are named as LMCSS, SMCSS, hiResHolo, hiResApo, and hiTanimoto. LMCSS stands for the structure that contains the ligand with the largest maximum common substructure to the query ligand. In contrast, SMCSS represents the structure that contains the ligand with the smallest maximum common substructure to the query ligand. hiResHolo denotes the highest-resolution, ligand-bound protein structure, and hiResApo marks the highest-resolution, apo protein structure. Similar to LMCSS, hiTanimoto corresponds to the structure that contains a ligand sharing the highest similarity (calculated by Tanimoto coefficient) with the query ligand. In addition, the participants are allowed to submit predictions based on other protein structures. For each protein structure, participants are asked to submit only one predicted binding mode. More details regarding the CELPP targets are available at the website <https://drugdesigndata.org/about/celpp>.

In this study, we analyzed our predictions based on the 1472 targets that we submitted from week22_2017 to week11_2018. A total of 1636 targets were released during these 42 weeks. Failed submissions were mainly due to the uncommon atoms (such as As) in the query ligands. In addition, for 295 cases out of these 1636 targets, their LMCSS or hiTanimoto protein structures contain co-bound ligands that were identical to the query ligands. These 295 targets were therefore discarded in the analysis of this study.

The GC3 targets—Grand Challenge 3 released a subset for protein-ligand binding mode prediction. This subset contains 24 query ligands targeting the human Cathepsin S (CatS). The organizers provided the SMILES strings of the query ligands, and the participants were free to use any structures of CatS. For each query ligand, the participants were allowed to submit up to 5 predicted binding modes.

Overview of our binding mode prediction methods

For CELPP—For the protein structures provided by CELPP, we used a docking strategy that was developed during previous D3R challenges. Briefly, a modified version of AutoDock Vina [15, 20] was used to generate up to 500 ligand binding modes on a target protein. Then, these putative binding modes were re-ranked using our in-house scoring function, ITScore2 [15]. Finally, the top binding mode based on each protein structure was submitted for each target.

ITScore2 was a statistical-potential-based scoring function. It was derived using an iterative method that circumvents 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. A detailed description of the ITScore2 method is available in the references [15, 21, 22].

Moreover, in addition to the predictions based on the protein structures provided by CELPP, we also submitted a prediction based on our own selected protein structure for each target. This protein structure was selected based on our calculated 3D molecular similarity between the co-bound ligand and the query ligand. Namely, the protein structure that contains a co-bound ligand sharing the highest 3D molecular similarity with the query ligand according to our calculation was used for binding mode prediction.

For GC3—Two approaches were used in binding mode prediction for GC3. The first approach was the docking approach that was used for CELPP. The protein structure that contains a co-bound ligand sharing the highest 3D molecular similarity with the query ligand according to our calculation was used for docking. The second approach was a template-based method. Briefly, the query ligand was superimposed onto the ligands in the known protein-ligand complexes, and the protein-query ligand complexes were ranked using the scoring function ITScore2. The top 5 models were further refined by the minimization step in molecular dynamics (MD) simulation.

Protein preparation

For CELPP—In addition to the protein structures provided by CELPP, we also selected a protein structure from the PDB for each target, referred to as the user-specified protein structure. Specifically, the program Protein-Protein BLAST (v2.6.0+) [23,24] was employed to search protein structures deposited in the PDB using the protein sequence provided by CELPP. The cutoff was set to 90% for both the sequence identity and the sequence coverage. The apo structures (i.e., the structures without co-bound ligands) were discarded. If there were more than one PDB structure that contain the same ligand, the structure with the lowest R-free value was kept. Then, 3D molecular similarities between the co-bound ligands and the query ligand were calculated by the program SHAFTS [25]. The PDB structure that contained the ligand sharing the highest similarity with the query ligand was selected for docking. The whole process was automatically performed for a given target.

For GC3—For the Grand Challenge, we manually checked the CatS structures in PDB, and found 12 PDB entries of human CatS containing different co-bound ligands. These PDB entries (2F1G, 2G7Y, 2H7J, 2R9M, 2R9O, 3IEJ, 3KWN, 3MPE, 3MPF, 3N3G, 4P6E, and 4P6G) were used for both docking method and template-based method.

Ligand preparation and similarity calculation

Up to 500 3D conformers were generated for each query ligand, using the program OMEGA (version 2.5.1.4, OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>) [26, 27] based on the SMILES string provided by CELPP or GC3. Only one conformer for each query ligand was necessary for molecular docking, because the ligand was treated as a flexible molecule in docking. The multiple ligand conformers were for the calculation of the 3D molecular similarities between a query ligand and the co-bound ligands in the known protein-ligand complex structures, using the program SHAFTS [25]. The similarity score

calculated by SHAFTS consists of a ShapeScore (i.e., shape-densities overlap) and a FeatureScore (i.e., pharmacophore feature fit values); both scores are normalized to a range of [0, 1], and the resulting HybridScore (i.e., the sum of ShapeScore and FeatureScore) is scaled to [0, 2], with 0 representing no similarity and 2 corresponding to the same ligand.

Binding mode prediction

The Docking approach—AutoDock Vina [20] was employed for sampling putative binding modes of a query ligand on a target protein. For the CELPP targets, we used the binding site center provided by CELPP as the center of the search box. The size of the cubic search box depended on the size of the query ligand. Specifically, the edge length of the cubic box was set to $\max[2 \times \text{maxL}, 22.5 \text{ \AA}]$, in which maxL was the maximum length of the query ligand structure in all dimensions. For the GC3 targets, the geometric center of the co-bound ligand was used as the center of the search box. The size of the search box was set to 30 Å. For both CELPP targets and GC3 targets, the protein structures were treated as rigid bodies and single bonds of the ligands were considered to be flexible. 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 output modes were then re-ranked using an in-house scoring function ITScore2. ITScore2 is a knowledge-based atomic pairwise scoring function that was derived based on the known protein–ligand complex structures (a refined set of PDBbind 2012 containing 2897 protein–ligand complexes [28, 29]) and the statistical mechanics-based iterative method [15, 21, 22]. Finally, the top mode was submitted for each CELPP target protein structure, and the top 5 modes were submitted for each GC3 target.

The template-based approach—During Grand Challenge 2, we developed a template-based approach for protein–ligand complex structure prediction. The details of the approach are described in Reference [14]. Briefly, molecular similarities of a query ligand with the co-bound ligands in the known complex structures were calculated using the program SHAFTS. Then, the top 5 PDB entries that had the best similarity scores with the query ligand were chosen as the templates. For each template, the query ligand was superimposed onto the ligand using the program SHAFTS. Finally, the protein–query ligand complex was refined using the minimization step of the MD simulation.

A drawback of this template-based approach is that severe clashes/overlaps between protein atoms and ligand atoms were found in the superimposition step (i.e., before minimization) in some cases, particularly in the cases in which only low-quality templates were available. Even MD minimization were unable to remove the clashes. To avoid or reduce these clashes/overlaps, we introduced a re-scoring step after the superimposition step in GC3. Specifically, instead of using the top 5 best PDB entries as the templates, we superimposed each conformer of the query ligand onto all the co-bound ligands in the known complex structures. Then, all the generated protein–query ligand complex structures were ranked using their scores calculated by the scoring function ITScore2. If the ligand RMSD of the two models was less than 2.0 Å, the model with a more negative score was kept. Finally, the top 5 models were refined using the minimization step of the MD simulation.

Results and Discussion

To evaluate the quality of predicted binding modes, we calculated root-mean-square deviations (RMSDs) between the predicted ligand binding modes and the modes in the released experimental complex structures. 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>) [26,27]. The MCS functionality can handle ligand atom renumbering and account for compound symmetries that are often involved in ligand superimposition.

CELPP

For each target, CELPP provided up to 5 protein structures, which were named as candidate protein structures. In addition to the predictions based on these candidate protein structures, we also submitted a prediction based on our own selected protein structure (referred to as the user-specified protein structure) for each target. The prediction results for using the CELPP-provided protein structures and the user-specified protein structures are reported as follows.

The CELPP-provided protein structures—CELPP provided the binding site location for each target. However, for a number of cases, the provided binding site locations were far from those in the released experimental complex structures. Here, we analyzed only the cases in which the binding site information was correctly provided. Namely, the targets were discarded if the distance between the geometry centers of the CELPP binding site and the real binding site were larger than 10 Å. The number of the targets that were kept for each type of protein structures are reported in Table 1.

Fig. 1 shows the results of binding mode prediction using different RMSD values as the thresholds to define successful predictions. The vertical dashed black lines correspond to a commonly-used threshold value of 2.0 Å. In addition to the docking results using CELPP-provided protein structures, we also presented the results of docking using bound protein structures (i.e. the protein structures extracted from the experimental complex structures) for references. In bound dockings, ligand structures were generated from the SMILES strings. Fig. 1A shows the results when the top model was considered for each protein structure. Using the threshold of 2.0 Å as an example, bound docking yielded the highest success rate 30.5%, as expected. The LMCSS protein structures (which contain similar ligands with the query ligands) resulted in a significantly higher success rate (23.2%) than the SMCSS protein structures (which contain dissimilar ligands with the query ligands) for which the success rate was 13.6%. The highest resolution ligand-bound candidate protein structures (hiResHolo) yielded slightly better docking performance (a success rate of 14.8%) than SMCSS, but resulted in significantly worse performance than LMCSS. When the hiResApo protein structures (i.e., the highest-resolution unbound protein structures) were used, the success rate was 8.1%, which was lower than all the other cases. Similar to LMCSS, the hiTanimoto protein structures contained similar ligands with the query ligands and achieved

comparable performances with LMCSS. The same conclusion can be obtained using different threshold values (see Fig. 1).

Fig. 1B presents the results when top 5 models were considered for each protein structure. Dramatically higher success rates were observed for all types of protein structures, in comparison with those when only the top model was considered. Specifically, when 2.0 Å was used as the threshold, the success rates of dockings with the bound, LMCSS, SMCSS, hiResHolo, hiResApo, and hiTanimoto protein structures increased to 46.3%, 40.4%, 28.5%, 30.4%, 17.7%, and 40.6%, respectively. Lower success rates observed in Fig. 1A (considering only the top model) mainly due to the inaccuracy of the current scoring function (ITScore2).

In addition to using the in-house scoring function, ITScore2, we also analyzed the performance of the Vina Score. The results are presented in the Supplementary Materials (Fig. S1). Vina Score achieved similar success rates as ITScore2 for each type of protein structures when the top model was considered. However, ITScore2 yielded significantly better performance than Vina Score when top 5 models were considered. In addition, similar to what was observed for ITScore2, the bound docking yielded the highest success rate. Dockings with the protein structures containing similar ligands with the query ligands (LMCSS or hiTanimoto) achieved significantly higher success rates than dockings with other types of protein structures (SMCSS, hiResHolo, or hiResApo).

Fig. 1C shows the sampling results for the use of each type of protein structures for docking. Using 2.0 Å as the threshold, bound docking yielded the highest success rate (80.4%), similar to the results in Fig. 1A and 1B. Possible reasons for the failures of sampling in some targets could be the non-exhaustive sampling strategy used in Vina and the ignorance of the effects of crystal waters or co-factors during docking. For the cases using the CELPP-provided protein structures for docking, LMCSS or hiTanimoto protein structures resulted in significantly higher success rates than other CELPP-provided protein structures. Specifically, when 2.0 Å was used as the threshold, the success rates of dockings with the LMCSS, SMCSS, hiResHolo, hiResApo, and hiTanimoto protein structures were 71.0%, 53.0%, 57.7%, 45.3%, and 69.8%, respectively.

In addition to the success rates shown in Fig. 1, the median and mean RMSD values for the use of each type of protein structures for docking are reported in Table 1. Similar to the results presented in Fig. 1, docking with the LMCSS and hiTanimoto protein structures achieved similar performances, and their median/mean RMSD values were significantly lower than dockings with the other three types of CELPP-provided protein structures. For instance, as shown in Columns 3 and 4, when the top predicted model was considered for each protein structure, the LMCSS protein structures resulted in significantly lower median/mean RMSD values (4.6 ± 0.2 / 5.1 ± 0.1 Å) than the SMCSS protein structures (5.7 ± 0.2 / 6.0 ± 0.1 Å). The highest resolution ligand-bound candidate protein structures (hiResHolo) yielded slightly better docking performance (5.6 ± 0.2 / 5.8 ± 0.1 Å) than SMCSS, but resulted in significantly worse performance than LMCSS. When the hiResApo protein structures were used, the median/mean RMSD values were 6.4 ± 0.2 / 6.5 ± 0.1 Å, which were larger than all the other cases. Similar to LMCSS, the hiTanimoto protein structures, which

contained similar ligands with the query ligands, achieved comparable performances with LMCSS. The results based on the Vina Score are also reported in Table S1.

In summary, dockings using the LMCSS or hiTanimoto protein structures achieved significantly better performances than dockings using other CELPP-provided protein structures. Similar conclusions were also obtained in the previous Grand Challenges (GC1 and GC2), in which we found that using protein structures containing a co-bound ligand sharing a high molecular similarity with the query ligand would significantly improve the binding mode prediction [13,14]. The above findings are consistent with the assumption that similar protein conformations, particularly similar conformations of the binding pockets, would be induced by binding of similar ligands, and that similar ligands tend to manifest similar binding modes. Next, we will present the results of prediction based on our own selected protein structures (i.e., user-specified protein structures), which were chosen with a 3D molecular similarity calculation method.

The user-specified protein structures—Fig. 2A shows the performances using different RMSD values as the thresholds for predictions based on our own selected protein structures (referred to as hiSHAFTS). For each target, docking was performed on the protein structure that contains a co-bound ligand sharing the highest 3D molecular similarity with the query ligand. The molecular similarity was calculated using the program SHAFTS with the default cutoff of 0.8. Based on the 875 evaluated targets, using 2.0 Å as the threshold for a correct prediction, the success rates were 22.6%, 39.0%, and 67.7% for the top model, the best in top 5 models, and the best in all sampled models, respectively. The median and mean RMSD values of binding mode prediction using hiSHAFTS are also reported in Table 1. The performances were similar to the performances using the best candidate protein structures (i.e., LMCSS or hiTanimoto structures provided by CELPP).

Fig. 2A also shows the results of the template-based method. Here, the hiSHAFTS complex structure was used as the template for each target. The query ligand was superimposed onto the co-bound ligand in the hiSHAFTS complex structure. It should be noted that the MD-based refinement step for the generated protein-query ligand complexes was not performed in the CELPP study, mainly due to the large number and complexity of the CELPP targets. It is surprising that a simple superimposition of the query ligand onto the template achieved a much better performance than the docking-based method. When 2.0 Å was used as the threshold, the success rate of the template-based method was 40.0%, about twice of the success rate achieved by the docking method (considering only the top model).

Furthermore, we analyzed the correlation between the performance of the template-based method and the quality of the templates used for prediction. Fig. 2B plots the RMSD values of each target vs the corresponding similarity scores between the query ligands and the corresponding ligands in the templates, showing that targets with higher similarity values tend to have lower RMSD values ($R=-0.44$). The value of 1.2 is the default cutoff of SHAFTS for a distinction of similar and dissimilar ligands. In this study, at least one good-quality template (SHAFTS score ≥ 1.2) were found for about 63.4% of the 875 evaluated cases. If only the cases in which SHAFTS score ≥ 1.2 were considered, the success rate

increased to 53.1% using 2.0 Å as the threshold. The success rate increased to as high as 71.2% if the SHAFTS score cutoff was set to 1.5.

Although the template-based method achieved higher success rate than the docking method for binding mode prediction, the performance of the template-based method was highly dependent on the quality of the templates. The method usually failed if no good-quality template (e.g. SHAFTS score < 1.2) was found. Furthermore, sometimes large RMSD values were also observed even if high-quality templates (e.g. SHAFTS score > 1.5) were (see Fig. 2B). The failures suggest the need for the improvement of our current template-based method, in which the query ligands were simply superimposed onto the template ligands and the protein induced-fit effects were ignored.

GC3

The template-based method also achieved good performance for the CatS subset, which contains 24 compounds targeting the human Cathepsin S. The median/mean RMSD values are 3.5/4.3 (2.4/3.4) Å when the top model (top 5 models) was (were) considered for each case. Unfortunately, poor performance was observed with the docking method. The median/mean RMSD values were as high as 8.3/9.9 (5.9/6.5) Å when the top model (top 5 models) was (were) considered for each case. The predictions did not improve when the bound protein structures were used for docking. Analyzing the targets in this subset showed that the targets contain large-size flexible compounds and large-size binding pocket and that relatively few interactions were found in the released crystal structures (see Fig. 3 for an example), indicating that entropy may play an important role in binding for this subset. The failure of the current docking method is probably due to the inaccuracy of the scoring function for the entropy calculation for large-size ligand molecules.

Conclusion

In this study, we thoroughly analyzed our prediction results on recent D3R challenges, CELPP and GC3. Our binding mode prediction methods have been systematically tested using the large-scale benchmark (more than one thousand targets) provided by CELPP. The results show that current molecular docking methods can benefit from the rapid increase of the number of protein-ligand complex structures in the PDB. The use of an appropriate protein structure for docking will significantly improve the success rate of the binding mode prediction. In addition to the docking method, the results of a template-based approach have been assessed. The template-based method achieved better performance than the docking method. However, a drawback of the template-based method is its reliance on the templates, and its performance was poor when there was no good-quality template available. On the other hand, the docking-based method does not have this issue. Future combination of these two different methods is on the way to improve the binding mode prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Support to XZ from OpenEye Scientific Software Inc. (Santa Fe, NM, <http://www.eyesopen.com>) is gratefully acknowledged. This work was supported by NIH R01GM109980 (PI: XZ), NIH R01HL126774 and NIH R01HL142301 (PI: Cui) 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. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3:935–947. [PubMed: 15520816]
2. Grinter SZ, Zou X (2014) Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules*. 19:10150–10176. [PubMed: 25019558]
3. Xu X, Huang M, Zou X (2018) Docking-based inverse virtual screening: methods, applications, and challenges. *Biophys Rep* 4:1–16. [PubMed: 29577065]
4. Brooijmans N, Kuntz ID. (2003) Molecular recognition and docking algorithms. *Annual review of biophysics and biomolecular structure*. 2003 6;32(1):335–373.
5. Huang SY, Grinter SZ, Zou X (2014) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 12:12899–12908.
6. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Carlson HA, Burley SK (2016) D3R grand challenge 2015: evaluation of protein–ligand pose and affinity predictions. *J Comput Aided Mol Des* 30:651–668. [PubMed: 27696240]
7. Gaieb Z, Liu S, Gathiaka S, Chiu M, Yang H, Shao C, Feher VA, Walters WP, Kuhn B, Rudolph MG, Burley SK (2018) D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J Comput Aided Mol Des* 32:1–20. [PubMed: 29204945]
8. Berman HM, Westbrook J, Feng Z, et al. (2000) The protein data bank. *Nucleic Acids Res* 28:235–242. [PubMed: 10592235]
9. Smith RD, Dunbar JB Jr., Ung PM, et al. (2011) CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J Chem Inf Model* 51:2115–2131. [PubMed: 21809884]
10. Damm-Ganamet KL, Smith RD, Dunbar JB Jr., et al. (2013) CSAR benchmark exercise 2011–2012: evaluation of results from docking and relative ranking of blinded congeneric series. *J Chem Inf Model* 53:1853–1870. [PubMed: 23548044]
11. Smith RD, Damm-Ganamet KL, Dunbar JB Jr., et al. (2016) CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. *J Chem Inf Model* 56:1022–1031. [PubMed: 26419257]
12. Carlson HA, Smith RD, Damm-Ganamet KL, et al. (2016) CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J Chem Inf Model* 56:1063–1077. [PubMed: 27149958]
13. Xu X, Yan C, Zou X (2017) Improving Binding Mode and Binding Affinity Predictions of Docking by Ligand-based Search of Protein Conformations: Evaluation in D3R Grand Challenge 2015. *J Comput Aided Mol Des* 31: 689–699. [PubMed: 28668990]
14. Duan R, Xu X, Zou X (2018) Lessons learned from participating in D3R 2016 Grand Challenge 2: compounds targeting the farnesoid X receptor. *J Comput Aided Mol Des* 32:103–111. [PubMed: 29127582]
15. Yan C, Grinter SZ, Merideth BR, Ma Z, Zou X (2016) Iterative Knowledge-Based Scoring Functions Derived from Rigid and Flexible Decoy Structures: Evaluation with the 2013 and 2014 CSAR Benchmarks. *J Chem Inf Model* 56:1013–1021. [PubMed: 26389744]
16. Grinter SZ, Yan C, Huang SY, Jiang L, Zou X (2013) Automated Large-Scale File Preparation, Docking, and Scoring: Evaluation of ITScore and STScore Using the 2012 Community Structure–Activity Resource Benchmark. *J Chem Inf Model* 53:1905–1914. [PubMed: 23656179]

17. Huang SY, Zou X (2011) Scoring and lessons learned with the CSAR benchmark using an improved iterative knowledge-based scoring function. *J Chem Inf Model* 51:2097–2106. [PubMed: 21830787]
18. Huang SY, Zou X (2011) Construction and test of ligand decoy sets using MDock: community structure–activity resource benchmarks for binding mode prediction. *J Chem Inf Model* 51:2107–2114. [PubMed: 21755952]
19. Huang SY, Zou X (2007). Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* 66:399–421. [PubMed: 17096427]
20. 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. [PubMed: 19499576]
21. Huang S, Zou X (2006) An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem* 27:1866–1875. [PubMed: 16983673]
22. Huang S, Zou X (2006) An Iterative Knowledge-based Scoring Function to Predict Protein–Ligand Interactions: II. Validation of the Scoring Function. *J Comput Chem* 27:1876–1882. [PubMed: 16983671]
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [PubMed: 2231712]
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC bioinformatics* 10:421. [PubMed: 20003500]
25. Liu X, Jiang H, Li H (2011) SHAFTS: a Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening. *J Chem Inf Model* 51:2372–2385. [PubMed: 21819157]
26. Hawkins PC, 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. [PubMed: 20235588]
27. Hawkins PC, Nicholls A (2012) Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J Chem Inf Model* 52:2919–2936. [PubMed: 23082786]
28. Cheng T, Li X, Li Y, Liu Z, Wang R (2009) Comparative Assessment of Scoring Functions on a Diverse Test Set. *J Chem Inf Model* 49:1079–1093. [PubMed: 19358517]
29. Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) The PDBbind Database: Methodologies and Updates. *J Med Chem* 48:4111–4119. [PubMed: 15943484]
30. Pettersen EF, Goddard TD, Huang CC, et al. (2004) UCSF Chimera – A Visualization System for Exploratory Research and Analysis. *J Comput Chem* 25:1605–1612. [PubMed: 15264254]
31. Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7: 1–26.
32. Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A (2010) How many bootstrap replicates are necessary? *J Comput Biol* 17: 337–354. [PubMed: 20377449]

Figure 1.

The results of binding mode prediction for CELPP using the bound protein structures and the candidate protein structures provided by CELPP for docking. Different RMSD values were used as the thresholds that defined successful predictions. (A) The success rate when only the top predicted model was considered for each protein structure. (B) The success rate when top 5 models were considered for each protein structure. (C) The success rate when all the models were considered for each protein structure.

Figure 2.

(A) The results of binding mode prediction for CELPP using the user-specified protein structures (i.e. hiSHAFTS structures in this study) for both docking-based and template-based methods. Different RMSD values were used as the thresholds for the definition of successful predictions. (B) The correlation between the RMSD values of the binding modes predicted by the template-based method and the corresponding similarity scores between the query ligands and the corresponding template ligands.

Figure 3.

The top predicted binding mode (colored red) of CatS_1 in comparison with the binding mode given by the crystal structure (colored green) released by D3R. The protein is shown in surface representation. The ligand (CatS_1) is plotted in stick representation. The predicted binding mode was obtained from the docking method using the bound protein structure.

Table 1

The results of binding mode prediction for the CELPP targets. The first row (“Bound”) lists the results of using the bound protein structures for docking. The last row (hiSHAFTS) gives the results from the use of the user-specified protein structures (i.e., the protein structure that contains the ligand with the largest HybridScore to the query ligand) for docking, and the other five rows show the results from the use of the candidate protein structures provided by CELPP. The error of each value was estimated using the bootstrap method in which 1000 replicates were used, and was reported in parentheses [31,32].

Protein Structure	Number of Targets	Top 1		Best in Top 5		Best in All	
		Median RMSD (Å)	Mean RMSD (Å)	Median RMSD (Å)	Mean RMSD (Å)	Median RMSD (Å)	Mean RMSD (Å)
Bound	969	4.3 (0.3)	4.9 (0.1)	2.3 (0.1)	3.3 (0.1)	1.2 (0.0)	1.5 (0.0)
LMCSS	969	4.6 (0.2)	5.1 (0.1)	2.6 (0.1)	3.5 (0.1)	1.5 (0.0)	1.9 (0.0)
SMCSS	969	5.7 (0.2)	6.0 (0.1)	3.5 (0.1)	4.2 (0.1)	1.9 (0.1)	2.4 (0.1)
hiResApo	608	6.4 (0.2)	6.5 (0.1)	4.4 (0.2)	4.8 (0.1)	2.2 (0.1)	2.8 (0.1)
hiResHolo	969	5.6 (0.2)	5.8 (0.1)	3.5 (0.1)	4.2 (0.1)	1.8 (0.1)	2.4 (0.1)
hiTanimoto	969	4.9 (0.2)	5.2 (0.1)	2.6 (0.1)	3.5 (0.1)	1.5 (0.0)	1.9 (0.0)
hiSHAFTS	875	4.7 (0.2)	5.3 (0.1)	2.7 (0.1)	3.6 (0.1)	1.6 (0.1)	1.9 (0.0)