Published in final edited form as:

J Chem Inf Model. 2016 June 27; 56(6): 1013–1021. doi:10.1021/acs.jcim.5b00504.

Iterative Knowledge-based Scoring Functions Derived from Rigid and Flexible Decoy Structures: Evaluation with the 2013 and 2014 CSAR Benchmarks

Chengfei Yan, Sam Z. Grinter, Benjamin Ryan Merideth, Zhiwei Ma, and Xiaoqin Zou*

Department of Physics and Astronomy, Department of Biochemistry, Dalton Cardiovascular Research Center, and Informatics Institute, University of Missouri, Columbia, MO 65211

Abstract

In this study, we developed two sets of iterative knowledge-based scoring functions, ITScore_pdbbind(rigid) and ITScore_pdbbind(flex), using the rigid decoy structures and flexible decoy structures that were generated based on the protein-ligand complexes in the refined set of PDBbind 2012, respectively. These two scoring functions were evaluated using the 2013 and 2014 CSAR benchmarks. The results were compared with the results of two other scoring functions, Vina scoring function and ITScore, the scoring function that we previously developed from the rigid decoy structures of a smaller set of protein-ligand complexes. A graph-based method was developed to evaluate the root-mean-square deviation (RMSD) between two conformations of the same ligand with different atom names and orders due to different file preparations, and the program is freely available. Our study showed that the two new scoring functions that were developed based on the larger training set yielded significantly improved performance on binding mode predictions. For binding affinity prediction, all the four scoring functions showed protein-dependent performances. We suggest to develop protein family-dependent scoring functions for accurate binding affinity prediction.

Graphical abstract

1 INTRODUCTION

Molecular docking is a useful tool to speed up the process of drug design.^{1–5} Most molecular docking tools are composed of two major components, sampling and scoring. The sampling component is responsible for generating possible binding configurations between the protein and the ligand. The scoring component is for distinguishing the native (or near-

*Correspondence to: X. Zou; zoux@missouri.edu, 573-882-6045 (tel.), 573-884-4232 (fax).

native) configurations from other configurations, and for predicting the binding affinities. The Community Structure-Activity Resource (CSAR) provides valuable resources for the development and assessment of docking methods.^{6–8} The 2013 and 2014 CSAR tasks include protein design (2013 CSAR Phase 1), scoring (2013 CSAR Phase 2 and 2014 CSAR Phase 1) and cross docking (2013 CSAR Phase 3 and 2014 CSAR Phase 2). For the protein design, the participants were asked to determine which designed proteins, among over one dozen of designed sequences, were able to bind to a given steroid. In the scoring exercises, the participants were asked to select the near-native binding mode from an ensemble of pre-generated decoys for each given protein-ligand complex. For cross docking, each target was provided with the extracted structure of the protein bound with a different ligand. The unbound ligand conformations in mol2 format or the ligand SMILES strings were also provided. The participants were asked to determine the binding mode between the protein and each ligand and to rank these ligands based on the predicted binding affinities. Because the result of protein design is dependent upon protein structural modeling, which is beyond the scope of molecular docking, we focus on scoring and cross docking in the present study. The 2013 and 2014 CSAR benchmarks consist of a total of four sets of targets (see Figure 1), including the binding affinities and native structures. Among them, the designed protein binders (DIG10.2, DIG18 and DIG20) to the steroid digoxigenin (DIG) (CSAR 2013) were contributed by Baker and colleagues.⁹ The Factor Xa (FXA), Spleen tyrosine kinase (SYK), and tRNA (guanine-N(1)-)-methyltransferase (TRMD) were donated by GlaxoSmithKline (GSK). The 2013 and 2014 CSAR benchmarks offer an unique opportunity for unbiased evaluation of available docking tools and scoring functions by providing novel, systematic and high-quality experimental binding data.

In the present work, we re-trained the iterative knowledge-based scoring functions^{10–13} by using the structures in the refined set of PDBbind 2012,^{14,15} which contains 2897 protein-ligand complexes in total. This nonredundant structural set is much larger than the training set for our original iterative knowledge-based scoring function, ITScore, which contains 786 protein-ligand complexes.¹⁰ Two sets of decoy structures were used for training, based on the same refined PDB-bind set: rigidly sampled decoys and flexibly sampled decoys. The resulting scoring functions are referred to as ITScore_pdbbind(rigid) and ITScore_pdbbind(flex), respectively. Together with the original ITScore^{10,11} and the scoring function of Vina (Vina Score),¹⁶ which were used as references, a total of four scoring functions were evaluated on the 2013 and 2014 CSAR benchmarks. The sampling task of molecular docking was performed by Vina. The details are described as follows.

2 MATERIAL AND METHODS

2.1 The Scoring Functions

Two scoring functions, ITScore_pdbbind(rigid) and ITScore_pdbbind(flex), were re-trained from the refined set of the PDBbind 2012 using the iterative method which circumvents the reference state problem. The main idea of this method is to iteratively adjust the pairwise interaction potentials by comparing the experimentally observed pair distribution functions $(g_{ij}^*(r))$ calculated from the native binding modes and the predicted pair distribution

functions $(g_{ij}^k(r))$ calculated from the sampled decoys (including the native binding modes) for every step; each decoy carries a Boltzmann weight calculated with the interaction potentials of the current step. The iteration stops when all the native binding modes have the lowest energies compared with their decoys, respectively. The idea is described by the following equation:

$$u_{ij}^{k+1}(r) = u_{ij}^k + \Delta u_{ij}^k(r), \quad \Delta u_{ij}^k = \frac{1}{2} K_B T [g_{ij}^k(r) - g_{ij}^*(r)] \quad (1)$$

Here, i and j represent the types of a pair of atoms, respectively, and r is the distance between the atom pair. K_B is the Boltzmann constant, and T is the temperature. Without loss of generality, $K_B T$ is set to unit 1. $u_{ij}^k(r)$ are the pairwise interaction potentials in the k -th step, and u_{ij}^{k+1} are the updated interaction potentials for the next step. Given a set of initial potentials $u_{ij}^0(r)$, the potentials are updated using the above iterative equation, until all the native binding modes are ranked with the lowest energies among the corresponding decoys, respectively. The iterative process generally converges within 100 steps. Detailed description of the iterative method is provided in ref. 10.

To derive the ITScore_pdbbind(rigid) scoring function, the native ligand conformer and the bound protein were extracted from the crystal structure for each protein-ligand complex in the training set. The decoys for each ligand were then sampled by rigidly docking the native ligand conformer against the bound protein structure using our rigid docking software, MDock.^{10,11,17} The top 200 modes ranked by ITScore were kept for each complex. To generate flexible decoys for the derivation of the ITScore_pdbbind(flex) scoring function, first, up to 200 conformers were created for each ligand using the default parameters of Omega 2.4.6 (OpenEye Scientific Software, Santa Fe, NM, USA).^{18,19} Then, these conformers were independently docked to the bound structure of the protein as rigid bodies using MDock. The top 500 modes ranked by ITScore among all the docked modes based on different conformers of the same ligand were kept for each ligand. The remaining procedures were identical to the procedures for the extraction of the ITScore scoring function.¹⁰ The ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) scoring functions were implemented in MDock.

It is noted that 500 flexible modes (*cf.* 200 rigid modes) were kept for each ligand so as to cover larger conformational space. The increase of the number of decoys will not result in overtraining because of the following reason: During the interaction potential derivation, for the calculation of the above pair distribution functions, each decoy carries a Boltzmann weight. Namely, the decoys that have significantly higher energies will be assigned a weight close to zero and therefore have little impact on the iteration process.

In this study, four sets of scoring functions, ITScore_pdbbind(rigid), ITScore_pdbbind(flex), ITScore, and Vina Score, were assessed with the 2013 and 2014 CSAR benchmarks.

2.2 The 2013 and 2014 CSAR benchmarks

The 2013 CSAR exercises focused on the designed binders to DIG. Three targets, DIG10.2, DIG18 and DIG20, with very similar sequences (> 90% sequence identity) and almost identical structures were provided. For DIG18 and DIG20, CSAR provided the 3D protein structures that were separated from the crystal structures of the bound complexes, and 200 poses for each ligand that were sampled by DOCK6.5.²⁰ These poses contained exactly one near-native mode ($RMSD < 1.0 \text{ \AA}$) for each ligand. The participants were asked to select the near-native binding modes from the given decoys. For DIG10.2, CSAR released the 3D structure of the protein and 10 ligands with unbound conformations. Analysis afterwards showed that the 3D structure of the protein was in fact extracted from the crystal structure of DIG10.2 bound with cs337, one of the ten ligands. The participants were asked to rank these ligands according to the predicted binding affinities. All the 3D structures were provided in *mol2* files.

The 2014 CSAR benchmarks include three targets, FXA, SYK and TRMD. In Phase 1, the participants were asked to score three ligands for FXA, five ligands for SYK and fourteen ligands for TRMD, respectively. For each complex, the bound structure of the protein and 200 pre-generated poses sampled by DOCK6.5 with exactly one near-native mode ($RMSD < 1.0 \text{ \AA}$) were provided in *mol2* files. In Phase 2, the participants were asked to predict the binding modes and relative rankings of three sets of ligands for FXA, one set of ligands for SYK and one set of ligands for TRMD, respectively. The ligands were provided with SMILES strings.

In the present study, the 2013 and 2014 CSAR benchmarks were grouped into two sections: The scoring section refers to the identification of the near-native binding modes from the pre-generated decoys. The cross docking section refers to the prediction of the binding modes and ranking of the ligands in terms of binding affinity prediction for the given protein structures bound with different ligands.

2.3 The Method for the Scoring Section

Four sets of scoring functions, ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex), and Vina Score, were employed to identify the near-native binding modes from the pre-generated decoys for targets DIG18, DIG20, FXA, SYK, and TRMD, respectively. The same protocol was used for ITScore, ITScore_pdbbind(rigid), and ITScore_pdbbind(flex). Specifically, each mode was locally optimized and scored with the scoring function in combination with the downhill simplex algorithm.²¹ It is noted that these three scoring functions consider interatomic interactions only. The decoys were then ranked according to the energy scores. The mode with the lowest score was considered as the predicted binding mode, and the score was considered as the relative binding free energy of the ligand.

As for Vina Score, the scoring protocol is slightly more complicated than the previous protocol to keep consistency with the docking and scoring strategy of the Vina program. In the original Vina program, the Vina Score is the sum of the inter-molecular and intra-molecular contributions, and is used for ranking the sampled binding modes. However, the output energy score of the original Vina program for each mode is calculated in the

following way: First, the intra-molecular contributions of the top ranked mode is subtracted. Then the resulting score is transformed with a conformation-independent function.¹⁶ In other words, Vina predicts binding affinities in terms of only the inter-molecular energies, but ranks the binding modes with the sum of the inter-molecular and intra-molecular energies. Therefore, in this study, the original Vina program cannot be applied directly to scoring and ranking CSAR's pre-generated decoys, because the output score for each decoy would exclude the intra-molecular contribution, which would not be Vina's way to rank different modes/decoys. Accordingly, we modified Vina's source code for scoring CSAR's pre-generated decoys. Specifically, we output both the inter-molecular energy score and the intra-molecular energy score for each decoy, and used their sum to rank the decoys. For binding affinity prediction, we used the original Vina program, namely, we used the inter-molecular energy scores for the binding affinities.

2.4 The Method for the Cross Docking Section

2.4.1 Preparing the receptors and the ligands—For DIG10.2, CSAR provided the 3D structure of the protein and the unbound structure of the ligands. However, for FXA, SYK and TRMD, the participants were asked to select the protein structure for each target from the crystal structures released in Phase 1, and to convert the ligands from the SMILES strings into 3D structures. In this study, for the protein structure selection, the released structures of each protein target were superimposed via UCSF Chimera²² and then manually examined. The structure that shared the most similarities among all the structures was used for docking. Each ligand SMILES string was converted to a mol2 file using Omega 2.4.6 by setting “maxconfs”, the maximum number of output conformations, to 1.

2.4.2 Sampling and Scoring—Vina was used for sampling the putative binding modes. For each target, the crystal structures of the protein that were released in the previous phase were superimposed via UCSF Chimera. Then, the geometric center of the bound ligands in the crystal structures was calculated. The geometric center was used as the center of the search box for Vina. The size of the cubic search box was set to 30 Å, which is sufficiently big to cover the whole binding pocket. The protein was set to be rigid, whereas all the rotatable bonds for each ligand were set to be flexible. The “exhaustiveness” parameter was set to 30 (default: 8) to ensure exhaustive sampling. We also modified the original source code of Vina, which allows for a maximum of 20 output modes, to set the maximum number of the output modes to 500.

The top mode ranked by Vina was considered as the predicted binding mode and the corresponding score was considered as the predicted binding free energy for ligand ranking, for the cross docking result of the Vina Score. ITScore, ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) were then used to score the sampled decoys with local simplex minimization via MDock, respectively. For each scoring function, the top ranked mode for each complex was considered as the predicted binding mode, and the corresponding score was used to rank the ligands.

2.5 The Method for Evaluation

For the scoring section, because there was only one near-native mode ($RMSD < 1.0 \text{ \AA}$) among the given decoys for each complex, scoring was defined as a success if the near-native mode was ranked as No. 1.

For the cross docking section, the Pearson correlation coefficient (r) between the predicted binding free energy scores and $\log(Kd)$ values (or $\log(IC50)$ values) was used for binding affinity prediction. To assess the binding mode prediction, the following method was used: First, the crystal structures for each protein target were superimposed using Chimera. Afterwards, the heavy-atom $RMSD$ between the predicted binding mode and the ligand in the crystal structure was calculated. If the $RMSD$ was smaller than $2.0 \text{ \AA}$, the prediction was considered to be successful.

The challenge for calculating the $RMSD$ arose from the fact that the atom names and orders in the predicted binding modes, which were generated by Omega, were different from those in the crystal structures (see Figure 2). The different assignment of the ligand atom names and orders is a common problem for docking due to the necessity of ligand preparation (e.g., charge assignment and conformational sampling). Unfortunately, to the best of our knowledge, there is no free program for academic users to solve this problem. To address the challenge, we developed a general method based on graph theory to build the atom maps between the Omega-generated ligand files and the original ligand files extracted from the crystal structures. Specifically, we built a graph based on each ligand using the graph-tool,²³ setting the atoms as the nodes and the chemical bonds as the edges. We labeled each node by the chemical element of the corresponding atom. We then used graph-tool to build the isomorphic mappings between the two labeled graphs to match both the topologies and the labels (i.e., chemistry). With the isomorphic mappings, we converted the Omega-generated atom names and orders to those used in the crystal structures for $RMSD$ calculations. If a ligand contains structurally symmetric components, there exist more than one isomorphic mappings. For example, Figure 2 shows the atom names of the ligand GTC000223 in the crystal structure (left) and in the Omega-generated structure (right). Both ($O2$, $O31$), ($O1$, $O32$)... and ($O1$, $O32$), ($O2$, $O31$)... are isomorphic mappings. In such cases, the mapping that produced the smallest $RMSD$ was selected.

3 RESULTS

3.1 The Scoring Section

The results of ITScore_pdbbind(rigid), ITScore_pdbbind(flex), ITScore, and Vina Score for the scoring section are shown in Figure 3A. The details are summarized in Table 1. The fact that ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) yielded much higher success rates than ITScore suggests that the knowledge-based scoring functions can be significantly improved by using a larger training set. Surprisingly, ITScore_pdbbind(flex) yielded a slightly lower success rate than ITScore_pdb(rigid), although ITScore_pdbbind(flex) was derived from flexible decoys that were much more diverse than the rigid decoys for ITScore_pdb(rigid) derivation. Vina Score achieved a very similar performance, slightly worse than ITScore_pdbbind(rigid) but slightly better than ITScore_pdbbind(flex).

3.2 The Cross Docking Section

3.2.1 The binding mode prediction—We were able to evaluate binding mode prediction for only those protein-ligand complexes with released structures. For TRMD, all the crystal structures of the protein-ligand complexes were released. For FXA and SYK, however, only the structures of a small subset of the complexes were released, respectively. For DIG10.2, no structures of the complexes were released and therefore the DIG10.2 system is not discussed in this study for binding mode prediction. The results of binding mode prediction are summarized in Table 2 and Figure 3B. First, as shown in the last column of Table 2, the near-native binding modes ($RMSD < 2.0\text{\AA}$) for all the ligands of FXA and SYK modes were successfully sampled using Vina. However, for TRMD, the near-native binding modes for five ligands were not sampled. After examining these failed cases, we found that a flexible loop region in the bound structure is important to the binding of these ligands, but the protein structure that we selected for docking has a different conformation in this region, resulting in severe atomic clashes with the native ligand conformations (see Figure 4). Second, Figure 3B shows that ITScore_pdbbind(rigid) achieved the best performance, which is slightly better than ITScore_pdbbind(flex), but significant better than Vina Score and ITScore. It is encouraging that both ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) failed to rank the near-native mode sampled by Vina within top 3 conformations for only one ligand.

The performance of the scoring functions shown in Figure 3B may be restricted by two possible causes. The first possible cause is the protein flexibility. Our rigid treatment with the protein fails to model the conformational changes induced by ligand binding, which is important in certain cases. The second possible cause is the quality of the initial 3D ligand structures generated by Omega based on the one dimensional SMILES strings. The bond lengths and bond angles of the ligands, which were fixed during the later docking computation as Vina sampled only the dihedral angles, were determined by Omega's default settings, may be different from those of the native structures. These possible deviations may also hinder the docking computation to reach the ligand native binding modes or to top-rank the sampled native modes. To evaluate the impact of these two factors on the performances of the scoring functions, two sets of docking studies were performed separately, using the same cross docking protocol as described in Subsection 2.4.2. The first set of docking studies (Cross docking(Exp)) used the protein structures prepared in Subsection 2.4.1 for previous cross docking studies, however, the 3D structures of the ligands were extracted from the crystal structures for the protein-ligand complexes released later by CSAR. This docking practice was to test whether the initial 3D structure generated by Omega has a significant impact on the docking performances. It is noted that the initial configurations of the ligands, including the dihedral angles, positions and orientations, were randomized by Vina during on-the-fly sampling. Therefore, the docking computations were not biased to the native ligand modes, and the differences in the docking results originated from the slight differences in the bond lengths and bond angles between the Omega-generated ligand structures and the experimental ligand structures. The second set of docking studies (Bound docking) used the structures of the proteins and of the ligands that were all extracted from the crystal structures. The ligand configurations were still sampled by Vina. This bound docking practice was to focus on the impact of protein flexibility on the docking

performances. The results of these two sets of docking studies are summarized in Table 3 and Table 4. Figure 5 shows the comparison of these results with the previous cross docking results (Cross docking(Omega)), considering the top prediction only for each complex.

As seen from Figure 5, for all the scoring functions, the performances were significantly better when the bound protein structures were used. In contrast, the differences in the bond lengths and bond angles between the Omega-generated structures and the bound ligand structures resulted in minimal impact on the performances of the scoring functions. In fact, the success rates for ITScore and ITScore_pdbbind(rigid) were even slightly lower when the bound ligand structures were used as the initial structures for Vina conformational sampling during docking. The slightly lower success rates might be accidental, which might result from Vina's Monte Carlo-based sampling. The results in Figure 5 suggest that protein flexibility is the key factor that limits the performances of binding mode prediction in our study.

3.2.2 The binding affinity prediction—The pK_d ($-\log(K_d)$) values of all the ligands were released for DIG10.2. The pIC_{50} ($-\log(IC_{50})$) values of all the ligands were released for FXA, SYK, and TRMD. The performances were evaluated by calculating the Pearson correlation coefficients (r) between the predicted binding energy scores and the $-pK_d$ s for DIG10.2 (or $-pIC_{50}$ s for FXA, SYK, and TRMD). The results are summarized in Table 5. As seen from this table, the performances of the four scoring functions were target-dependent. For example, ITScore and Vina Score yielded good correlations ($r > 0.5$) with the experimental results for DIG10.2 and TRMD, but performed poorly for FXA and SYK. ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) performed much better on SYK than DIG10.2, FXA and TRMD. In fact, according to CSAR's analysis, no method achieved a correlation coefficient better than 0.5 on any set of ligands for FXA, and only one method in the thirty-five methods yielded $r > 0.5$ for both SYK and TRMD (see the article written by the CSAR organizers in this special issue). Overall, ITScore and Vina Score exhibited similar performances, and ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) achieved similar performances. Although ITScore was trained using the same iterative method as the method that was used for ITScore_pdbbind(rigid) and ITScore_pdbbind(flex), the training set for ITScore which contains 786 protein-ligand complexes is much smaller than the training set for ITScore_pdbbind(rigid) and ITScore_pdbbind(flex). Vina Score was trained using the refined set of PDBbind 2007, which consists of 1300 complexes and is a subset of the training set for ITScore_pdbbind(rigid) and ITScore_pdbbind(flex). To find the underlying reason for the target-dependent performances of the four scoring functions, for each CSAR target, the number of complexes with homologous proteins (defined as the sequence identity higher than 35.0%) in each training set was calculated. The results are listed in Table 6. It can be seen that all the three training sets contain the largest numbers of homologous proteins for FXA but few homologous proteins for DIG10.2 and TRMD. The enrichments of the homologous proteins in the training sets for each target show no correlation with the performances of the binding affinity prediction. Therefore, the enrichments of homologous proteins in the training sets are not the reason for the target-dependent performances of the four scoring functions.

Next, to test whether the ligand conformational sampling in this study has an impact on binding affinity prediction, the ligands for TRMD were scored using the released crystal structures with local minimization. This rescoring process was performed only for TRMD, because for other targets only a tiny portion of the complexes were provided with the crystal structures (i.e., 0 out of 10 for DIG10.2, 3 out of 163 for FXA, and 8 out of 276 for SYK). The results are summarized in Table 7. It can be seen that the correlations increased for ITScore_pdbbind(rigid) and ITScore_pdbbind(flex), but remained lower than 0.5. The correlation barely changed for Vina Score. For ITScore, the correlation even decreased from 0.82 to 0.55, despite the use of the correct binding modes. These results suggest that binding mode prediction is not the bottleneck for the accuracy of binding affinity prediction in this study. More experimental data may help further understanding of the target-dependent results.

4 DISCUSSION AND CONCLUSION

In the present study, two sets of iterative knowledge-based scoring functions, ITScore_pdbbind(rigid) and ITScore_pdbbind(flex), were developed based on the refined set of PDBbind 2012, using rigidly sampled decoys and flexibly sampled decoys, respectively. These two scoring functions, together with ITScore and Vina Score, were evaluated with the 2013 and 2014 CSAR benchmarks for binding mode and binding affinity predictions. The results showed that ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) achieved much better performances than ITScore on binding mode prediction, suggesting that the use of a larger training set can significantly improve the performance of an iterative knowledge-based function on binding mode prediction. However, the decoys generated from flexible sampling of the conformations of the ligands in the training set (for ITScore_pdbbind(flex)) do not offer improvement on performance compared with the decoys generated from rigid ligand sampling (for ITScore_pdbbind(rigid)). The underlying reason may be because our knowledge-based scoring functions do not consider the internal interactions within the ligand. In addition, our results showed that when the bound protein structures were used, Vina Score yielded a performance that was comparable to those of ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) (see Figure 3A and Figure 5). Yet, for cross docking, ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) achieved significantly higher success rates than Vina Score (see Figure 3B). The findings suggest that ITScore_pdbbind(rigid) and ITScore_pdbbind(flex) showed more tolerance to protein conformational changes than Vina Score. All these results suggest that the accuracy and robustness of binding mode prediction can be improved based on a large training set by using the iterative knowledge-based scoring functions.

For binding affinity prediction, however, the four scoring functions performed much less satisfactorily and showed significant target dependence. It would be desirable to develop target-dependant scoring functions for reliable binding affinity predictions. Furthermore, because of the imperfection in the scoring function and the neglect of the protein conformational changes upon binding, errors in the binding mode prediction such as having local minimum rather than global minimum also poses challenges for the binding affinity prediction. Indeed, according to the results of our group and other groups released by CSAR,

the binding affinity prediction exercise is challenging and no method can succeed for all the cases.

In this study, a graph-based method was also developed to calculate the RMSD between two conformations of the same ligand but with different atom names and orders due to different file preparation. The program is freely available to academic users upon request.

5 ACKNOWLEDGEMENT

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

References

1. Brooijmans N, Kuntz ID. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* 2003; 32:335–373. [PubMed: 12574069]
2. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery.* 2004; 3:935–949. [PubMed: 15520816]
3. Huang S-Y, Zou X. Advances and Challenges in Protein–Ligand Docking. *Int. J. Mol. Sci.* 2010; 11:3016–3034. [PubMed: 21152288]
4. 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]
5. 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]
6. Dunbar JB Jr, Smith RD, Yang C-Y, Ung PM-U, Lexa KW, Khazanov NA, Stuckey JA, Wang S, Carlson HA. CSAR Benchmark Exercise of 2010: Selection of the Protein–Ligand Complexes. *J. Chem. Inf. Model.* 2011; 51:2036–2046. [PubMed: 21728306]
7. Damm-Ganamet KL, Smith RD, Dunbar JB Jr, Stuckey JA, Carlson HA. 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]
8. Dunbar JB Jr, Smith RD, Damm-Ganamet KL, Ahmed A, Esposito EX, Del-proposto J, Chinnaswamy K, Kang Y-N, Kubish G, Gestwicki JE, Stuckey JA, Carlson HA. CSAR Data Set Release 2012: Ligands, Affinities, Complexes, and Docking Decoys. *J. Chem. Inf. Model.* 2013; 53:1842–1852. [PubMed: 23617227]
9. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D. Computational Design of Ligand–Binding Proteins with High Affinity and Selectivity. *Nature.* 2013; 501:212–216. [PubMed: 24005320]
10. 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]
11. 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]
12. Huang S-Y, Zou X. Scoring and Lessons Learned with the CSAR Benchmark Using an Improved Iterative Knowledge-based Scoring Function. *J. Chem. Inf. Model.* 2011; 51:2097–2106. [PubMed: 21830787]
13. Grinter SZ, Yan C, Huang S-Y, Jiang L, Zou X. 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.* 2013; 53:1905–1914. [PubMed: 23656179]

14. Wang R, Fang X, Lu Y, Yang C-Y, Wang S. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* 2005; 48:4111–4119. [PubMed: 15943484]
15. Cheng T, Li X, Li Y, Liu Z, Wang R. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* 2009; 49:1079–1093. [PubMed: 19358517]
16. 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]
17. Huang S-Y, Zou X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins: Struct., Funct., Bioinf.* 2007; 66:399–421.
18. Hawkins PC, Skillman AG, Warren GL, Ellingson BA, Stahl MT. 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]
19. 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]
20. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID. DOCK 6: Combining Techniques to Model RNA–Small Molecule Complexes. *RNA.* 2009; 15:1219–1230. [PubMed: 19369428]
21. Nelder JA, Mead R. A Simplex Method for Function Minimization. *Comput. J.* 1965; 7:308–313.
22. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera – A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
23. Peixoto, TP. The Graph-Tool Python Library. figshare. 2014. <http://dx.doi.org/10.6084/m9.figshare.1164194>

Figure 1.

The four targets in the 2013 and 2014 CSAR benchmarks. The proteins surfaces are colored in terms of hydrophobicity, from dodger blue for the most hydrophilic surface to orange red for the most hydrophobic surface. For each protein target, one representative ligand is displayed to show the binding pocket. A. DIG10.2 (DIG18 and DIG20); B. FXA; C. SYK; D. TRMD.

Figure 2.

An example of two conformations of the same ligand with different atoms names and orders. The atoms are colored based on the chemical elements. A. The ligand GTC000223 extracted from the crystal structure of the SYK-GTC000223 complex. B. The initial 3D structure of the ligand GTC000223 generated by Omega.

Figure 3.

A. Scoring: The performances of ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina Score on ranking the near-native binding modes from the pre-generated decoys provided by CSAR within top 1 and top 3, respectively. B. Cross docking: The success rates of ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina score for correctly predicting the binding modes within top 1 and top 3, respectively.

Figure 4.

The conformational change in a loop region resulted in the failure on sampling the near-native binding mode for TRMD. The bound structure of the protein is colored grey. The protein structure used for cross docking is colored cyan. It can be seen that the valine residue in the protein structure for cross docking clashes severely with the native structure of the ligand.

Figure 5.

The performances of ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina score: 1) cross docking with the initial ligand conformations generated by Omega based on the SMILES strings, 2) cross docking with the initial ligand conformations extracted from the released crystal structures, and 3) direct docking with the bound protein structures and with the initial ligand conformations also extracted from the crystal structures. Only the top 1 prediction was considered.

Table I

The performances for ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina Score on identifying the near-native poses (within top 1 or top 3, respectively) from the pre-generated decoys.

Target		ITScore		ITScore_pdbbind(rigid)		ITScore_pdbbind(flex)		Vina Score	
Name	No. of ligands	Top 1	Top 3	Top 1	Top 3	Top 1	Top 3	Top 1	Top 3
DIG(18&20)	2	2	2	2	2	1	2	1	2
FXA	3	2	3	3	3	3	3	2	3
SYK	5	4	4	5	5	4	4	5	5
TRMD	14	7	10	13	14	13	14	14	14
TOTAL	24	15	19	23	24	21	23	22	24

Table II

The performances of ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina Score on cross docking.

Name	Target	No. of ligands ^a	ITScore			ITScore_pdbbind(rigid)			ITScore_pdbbind(flex)			Vina Score		
			Top 1	Top 3	Top 500 ^b	Top 1	Top 3	Top 500 ^b	Top 1	Top 3	Top 500 ^b	Top 1	Top 3	Top 500 ^b
FXA		3	0	0	3	3	3	3	2	2	2	2	2	3
SYK		8	0	1	7	8	8	4	7	1	4	8	8	8
TRMD		31	19	22	24	25	23	26	19	25	26	26	26	26
Total		42	19	23	34	36	29	36	22	31	37	37	37	37

^aEach number in this column stands for the number of the ligands with which the complexes have crystal structures released by CSAR.

^bEach number in this column represents the number of the ligands whose near-native binding modes were successfully sampled ($RMSE < 2.0\text{\AA}$) for each target.

Table III

The performances of ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina Score on cross docking with unbound protein structures and initial ligand conformations extracted from the crystal structures.

Target		ITScore			ITScore_pdbbind(rigid)			ITScore_pdbbind(flex)			Vina Score		
Name	No. of ligands	Top 1	Top 3		Top 1	Top 3		Top 1	Top 3		Top 1	Top 3	Top 500
FXA	3	0	1	3	3	3	3	3	3	2	2	2	3
SYK	8	0	1	5	7	4	6	2	5	8			
TRMD	31	18	26	22	26	24	27	20	24	28			
Total	42	18	28	30	36	31	36	24	31	39			

Table IV

The performances of ITScore, ITScore_pdbbind(rigid), ITScore_pdbbind(flex) and Vina Score on direct docking with bound protein structures and initial ligand conformations extracted from the crystal structures.

Target		ITScore			ITScore_pdbbind(rigid)			ITScore_pdbbind(flex)			Vina Score		
Name	No. of ligands	Top 1	Top 3		Top 1	Top 3		Top 1	Top 3		Top 1	Top 3	Top 500
FXA	3	2	2	2	3	3	2	2	3	2	2	2	3
SYK	8	3	4	7	7	8	6	7	7	5	6	6	8
TRMD	31	19	30	28	28	30	28	30	30	27	31	31	31
Total	42	24	36	38	38	41	36	40	40	34	39	39	42

Table V

The Pearson correlation coefficients (r) between experimental binding affinities and the scores of the top modes for cross docking.

Target	No. of ligands	ITScore	ITScore_pdbbind(rigid)	ITScore_pdbbind(flex)	Vina Score
DIG10.2	10	0.73	0.26	0.11	0.79
FXA_set1	45	0.15	0.44	0.30	0.09
FXA_set2	67	-0.11	0.08	0.14	0.19
FXA_set3	51	-0.18	-0.24	-0.18	0.11
SYK	276	0.31	0.53	0.59	0.33
TRMD	31	0.82	0.25	0.24	0.65

Table VI

The numbers of the complexes in each training set whose proteins are homologous to the given target.

Training Set ^a	No. of Complexes	DIG10.2	FXA	SYK	TRMD
ITScore_pdbset	786	0	30	9	0
PDBbind 2007	1300	0	158	5	0
PDBbind 2012	2897	0	248	34	1

^aITScore_pdbset refers to the training set for ITC. The training set for Vina Score is the refined set of PDBbind 2007. The training set for ITC_pdbbind(rigid) and ITC_pdbbind(flex) is the refined set of PDBbind 2012.

Table VII

The Pearson correlation coefficients (r) between the experimental binding affinities and the scores of the native binding modes

Target	No. of ligands	ITScore	ITScore_pdbbind(rigid)	ITScore_pdbbind(flex)	Vina Score
TRMD	31	0.55	0.45	0.34	0.67