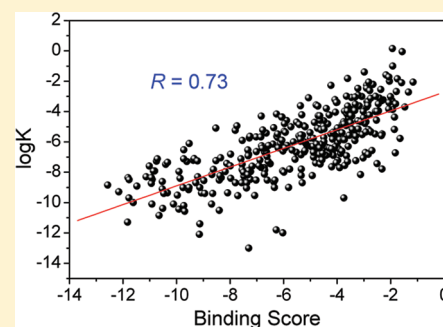


Scoring and Lessons Learned with the CSAR Benchmark Using an Improved Iterative Knowledge-Based Scoring Function

Sheng-You Huang[†] and Xiaoqin Zou^{*,†}

[†]Department of Physics and Astronomy, Department of Biochemistry, Dalton Cardiovascular Research Center, and Informatics Institute, University of Missouri, Columbia, Missouri 65211, United States

ABSTRACT: Based on a statistical mechanics-based iterative method, we have extracted a set of distance-dependent, all-atom pairwise potentials for protein–ligand interactions from the crystal structures of 1300 protein–ligand complexes. The iterative method circumvents the long-standing reference state problem in knowledge-based scoring functions. The resulted scoring function, referred to as ITScore 2.0, has been tested with the CSAR (Community Structure–Activity Resource, 2009 release) benchmark of 345 diverse protein–ligand complexes. ITScore 2.0 achieved a Pearson correlation of $R^2 = 0.54$ in binding affinity prediction. A comparative analysis has been done on the scoring performances of ITScore 2.0, the van der Waals (VDW) scoring function, the VDW with heavy atoms only, and the force field (FF) scoring function of DOCK which consists of a VDW term and an electrostatic term. The results reveal several important factors that affect the scoring performances, which could be helpful for the improvement of scoring functions.



1. INTRODUCTION

An accurate scoring function is crucial for molecular docking and lead discovery/optimization in structure-based drug design.^{1–7} Despite significant progress in the past two decades, the scoring problem remains challenging. There are two important aspects in scoring function development, methodology, and validation.⁷

Regarding the aspect of methodology development, typical scoring functions can be classified into three broad categories: force-field scoring functions, empirical scoring functions, and knowledge-based scoring functions. Force field scoring functions are based on the potential energy functions and parameters derived from quantum mechanical calculations and experimental data.^{8–11} The potential energy function usually consists of an electrostatic term, a van der Waals 6–12 potential term, and terms accounting for conformation changes (i.e., bond stretching, angle bending and bond torsions). Explicit treatment of water molecules is widely used to calculate relative affinities for protein–ligand binding (see ref 12 for review). For absolute affinity calculations, restricted by today's computing power, water is usually treated implicitly, using Poisson–Boltzmann/solvent accessible surface area (PB/SA) approaches,^{13–19} generalized-Born/SA (GB/SA) approaches,^{20–32} or crude approximations such as the distance-dependent dielectric.^{33,34} Force field scoring functions that employ intermediate-level approximations such as implicit solvent treatment (PB/SA and GB/SA) and local conformational sampling (e.g., LIE)³⁵ often introduce empirical, system-dependent weighting parameters to combine different energy components such as the electrostatic interaction term, van der Waals term, and hydrophobic term. The second type of scoring functions – empirical scoring functions

decompose the free energy of binding into multiple empirical energy terms, in which the weighting factors are calibrated by fitting the affinity data of a training set of protein–ligand complexes with known three-dimensional structures via linear regression.^{36–41} The transferability of empirical scoring functions can be improved by increasing the number, diversity, and accuracies (in the affinity data and structures) of the complexes in the training set. Finally, knowledge-based scoring functions (also referred to as statistical potential-based scoring functions) are derived by converting the frequencies of interacting atomic pairs observed in protein–ligand complexes into atomic interaction energy parameters (see ref 42 for a recent review).^{43–58} Most knowledge-based scoring functions use the Boltzmann relationship^{59–61} to make the conversion, see below for more discussions. This approach of extracting microscopic energy parameters from macroscopic structural data circumvents the difficulty of resolving the contributions of individual energy components by lumping them together, leading to a good transferability of this type of scoring functions for affinity predictions on a variety of sets of protein–ligand complexes that have a wide range of binding affinities. In addition, the pairwise nature of the knowledge-based scoring functions make them computationally efficient. As a trade-off, it is difficult for knowledge-based scoring functions to explicitly analyze the individual contributions to the free energy of binding, such as

Special Issue: CSAR 2010 Scoring Exercise

Received: February 14, 2011

Published: August 10, 2011

the total electrostatic energies, van der Waals energies, and hydrophobic energies.

Second, regarding the aspect of scoring validation and comparison, standardized, accurate, easy-to-use, and freely available benchmarks are highly valuable for the docking community.⁷ For example, the benchmark should be accurate in its collected binding constant data and three-dimensional structures so as to avoid any introduction of additional experimental errors. The benchmark is also expected to be large in number and diverse in geometry and chemistry so as to test the general applicability of a scoring function. A benchmark for therapeutic purposes should use drug-like and noncovalently bound ligands. It is highly desirable for the whole docking community to use the same standardized benchmarks for easy comparisons on different scoring functions, which can provide helpful insights into future scoring development. Recently, Dr. Carlson and her team at the University of Michigan are leading the efforts on constructing such valuable benchmarks, referred to as CSAR (Community Structure–Activity Resource, <http://www.csardock.org/>). The first CSAR benchmark was released in 2009, which consists of two test sets of diverse protein–ligand complexes of known K_d 's and high-resolution crystal structures. Every structure was carefully examined, and any original atomic assignments that are inconsistent with the corresponding electron density map were fixed.

In the present study, we have used the CSAR benchmark to test our scoring function, ITScore.^{56,57} ITScore is a knowledge-based scoring function, in which the distance-dependent, atom-based interaction energy parameters were derived using a unique iterative method.^{56,62} The advantage of this iterative method over the traditional use of the Boltzmann relationship is that the former circumvents the long-standing challenging reference state problem (see the next section).^{63,64} Consequently, ITScore achieved a significant improvement in ligand binding mode prediction and virtual screening compared with other knowledge-based scoring functions.⁵⁷ With more and more high-quality complex structures available, we have recently developed an improved set of energy potential parameters for ITScore (referred to as ITScore 2.0) by using a much larger training set of protein–ligand complexes and a more physical convergence criterion. ITScore 2.0 was then evaluated using the recently released CSAR benchmark.

2. MATERIALS AND METHODS

2.1. An Iterative Method for Extracting Effective Potentials. Traditionally, the energy parameters in knowledge-based scoring functions or potentials of mean force (PMF) $w_{ij}(r)$ are derived according to the following inverse Boltzmann relation⁶⁵

$$w_{ij}(r) = -k_B T \ln \left[\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right] \quad (1)$$

Here, $\rho_{ij}(r)$ is the intermolecular pair frequencies of the atom types i and j observed at distance r in the high-resolution crystal structures of protein–ligand complexes, and ρ_{ij}^* is the corresponding pair frequencies in a “reference state” in which the atomic interactions are zero. Despite the success achieved in binding affinity prediction,⁴² a well-known challenge in the traditional knowledge-based scoring functions is the inaccessibility of the accurate reference state and thus ρ_{ij}^* in eq 1.⁶³

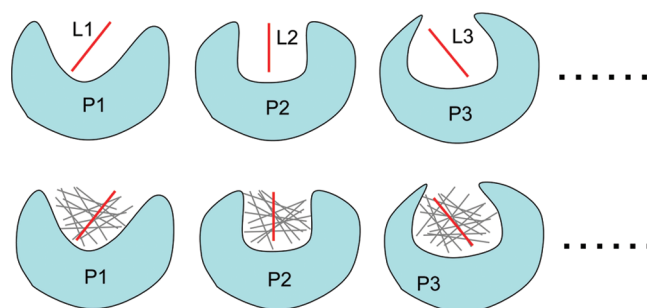


Figure 1. A cartoon illustration of the training set used in our iterative method. The upper row are the experimentally determined complex structures. The red lines stand for the native ligand binding modes, and the gray lines in the lower row represent the decoys.

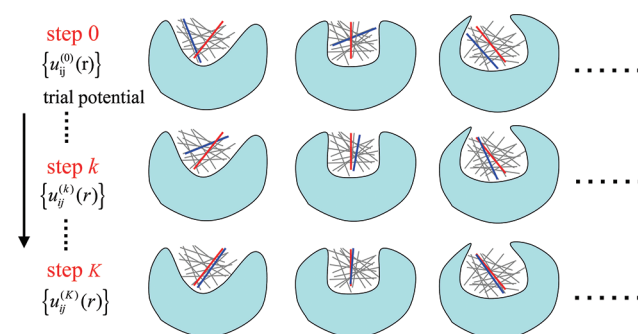


Figure 2. An illustration of the iterative procedure to extract the effective potentials of ITScore. The red, blue, and gray lines stand for the native, predicted, and decoy binding modes, respectively.

Accordingly, we have developed a statistical mechanics-based iterative method to derive the atom-based pairwise interaction potentials for protein–ligand interactions (ITScore), instead of using eq 1 which requires the accurate calculation of ρ_{ij}^* of the reference state. Specifically, for a given set of N native protein–ligand complexes (illustrated in Figure 1, upper row), we generated many decoy structures using UCSF DOCK 4.0.1⁶⁶ (Figure 1, lower row). During this process, only VDW interactions were considered so as to minimize the effect of other energies on the decoy generation. It should be noted that the DOCK program was used only to construct ligand binding mode decoys to train ITScore. In other words, the accuracy of DOCK/VDW on ranking these decoys is irrelevant to the derivation of ITScore energy parameters (as shown below), as long as DOCK/VDW can sample a set of diverse binding decoys that cover the whole binding site. Based on the native structures and the ligand decoys, we improved the initial guess of the pairwise interaction potentials step by step using the following expression

$$u_{ij}^{(k+1)}(r) = u_{ij}^{(k)}(r) + \Delta u_{ij}^{(k)}(r) \quad (2)$$

where

$$\Delta u_{ij}^{(k)}(r) = \frac{1}{2} k_B T [g_{ij}^{(k)}(r) - g_{ij}^{\text{obs}}(r)] \quad (3)$$

Here, $u_{ij}^{(k)}(r)$ is the interaction potential of atom pair ij at distance r in the k -th iterative step, $u_{ij}^{(k+1)}(r)$ is the improved potential by iteration, k_B is the Boltzmann constant, and T is the temperature. $g_{ij}^{\text{obs}}(r)$ is the pair distribution function for the experimentally

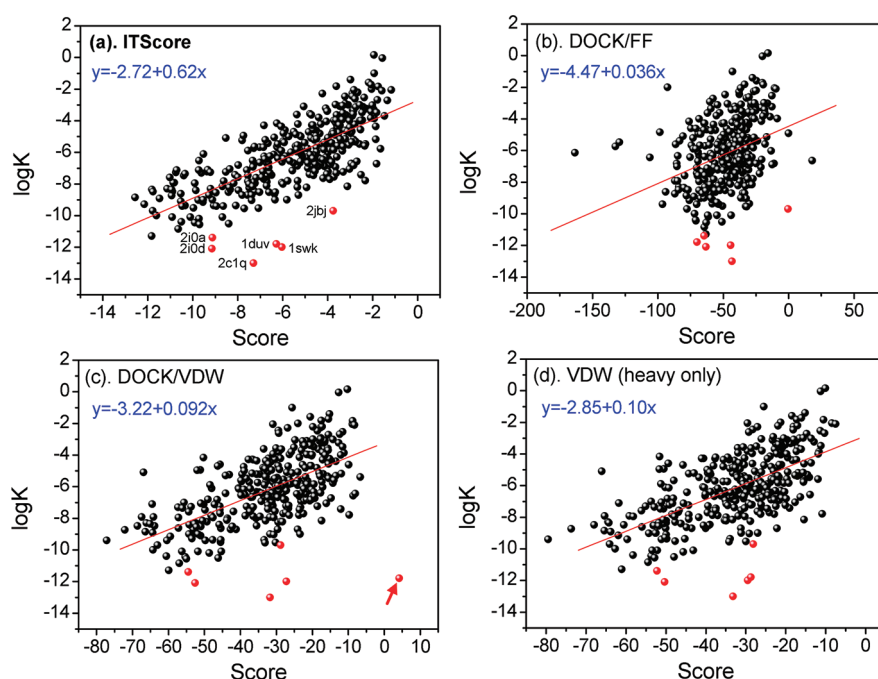


Figure 3. The affinity-score plots for ITScore, DOCK/FF, DOCK/VDW, and VDW(Heavy) with the CSAR benchmark of 345 protein–ligand complexes. Each red line represents the linear fit of the data with the fitting equation shown in blue. The red spheres in each panel show examples of the common outliers for all the four scoring functions. The outlier in panel c with the red arrow is the complex #18 (PDB entry: 1DUV) of set 2.

observed (i.e., native) complex structures. $g_{ij}^{(k)}(r)$ is the average pair distribution function calculated for the whole ensemble consisting of the native structures and the decoys following the Boltzmann distribution. The iterative process is illustrated in Figure 2.

Usually, the initial potentials $u_{ij}^{(0)}(r)$ were inaccurate, and therefore the binding modes predicted by $u_{ij}^{(0)}(r)$ (i.e., the mode with the lowest predicted energy) were significantly different from the native modes (Figure 2, top row). Through the iterative process, the potentials $u_{ij}^{(k)}(r)$ became more and more accurate, and the binding modes predicted by $u_{ij}^{(k)}(r)$ were getting closer and closer to the native structures (Figure 2, middle row). At the end of the iterations, the potential corrections $\Delta u_{ij}^{(k)}(r) \rightarrow 0$, and the potentials $u_{ij}^{(k)}(r)$ converged to a set of effective pairwise interaction potentials $u_{ij}(r)$. Thus, the binding energy score can be calculated by summing up all the atomic pairs between protein atom i and ligand atom j in a complex as

$$E_{\text{ITScore}} = \sum_{i,j} u_{ij}(r) \quad (4)$$

According to the theory of statistical mechanics, the effective potentials extracted from the iterative procedure are expected to be able to reproduce the native structures⁶⁷ (Figure 2, bottom row). The details on the calculations of the pair distribution functions, the choice of the initial potentials, and the statistical mechanics basis of the iteration method are described in our previous studies.^{42,56}

2.2. Training Set for ITScore. In the present study we used a much larger training set of protein–ligand complexes to derive the potentials for ITScore compared to our previous study.⁵⁶ The new training set consists of 1300 protein–ligand complexes from the “refined set” in the 2007 PDBBind database prepared by Wang’s group.^{68,69} In the training set, water molecules and

hydrogen atoms were removed from the proteins. The definitions of the atom types were given in ref 56.

Furthermore, to examine the effect of the training set on ITScore, we also used a smaller training set of 1152 protein–ligand complexes by removing the overlapped entries between the CSAR benchmark and the PDBBind database. The results showed almost no difference as shown in Section 3.4, indicating the robustness of our iterative method. The observation is also consistent with our previous study, in which the removal of a small portion from a large training set will not affect the extracted interaction potentials and thereby the prediction results.⁷⁰

3. RESULTS AND DISCUSSION

3.1. Overall Performance. Using the extracted effective potentials through the iteration, we examined our ITScore 2.0 for its ability of predicting binding affinities with the CSAR benchmark (2009 release). The benchmark consists of set 1 and set 2 with a total of 345 diverse protein–ligand complexes. The performance is measured by the correlation between the calculated binding energy scores and the experimentally determined log K . For reference purpose only, we also calculated the binding energy scores for the same protein–ligand complexes by using the force field (FF) scoring function of DOCK 4.0.1 and its VDW component, respectively (<http://dock.compbio.ucsf.edu/>).⁶⁶ The all-atom FF scoring function is composed of a Coulombic interaction term with a distance-dependent dielectric constant and a 6–12 Lennard-Jones potential for VDW interactions as follows³³

$$E_{\text{FF}} = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) \quad (5)$$

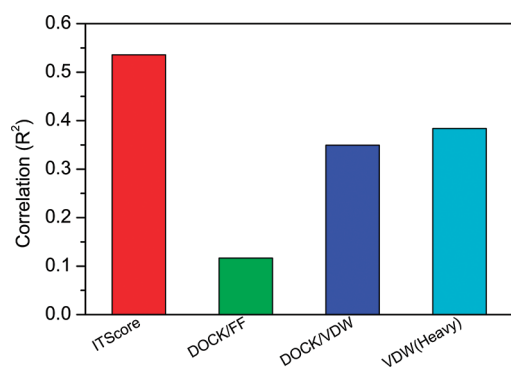


Figure 4. The histograms of the correlations for ITScore, DOCK/FF, DOCK/VDW, and VDW(Heavy) with the CSAR benchmark of 345 protein–ligand complexes.

where r_{ij} stands for the distance of protein atom i and ligand atom j , A_{ij} and B_{ij} are the VDW parameters, and q_i and q_j are the atomic partial charges. The effect of solvents is crudely accounted for by using a distance-dependent dielectric constant $\epsilon(r_{ij})$. Finally, to investigate the effect of hydrogen assignments on the force field scoring function, we also calculated the VDW binding scores for the heavy atoms only as

$$E_{VDW} = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (6)$$

These reference scoring functions could provide crude estimates on the contributions of individual energy components such as electrostatic energies and van der Waals energies to the binding affinities.

Some file preparations were done for the above calculations. In the CSAR benchmark (2009 release), the protein files were provided in the pdb format, and the ligand files were in Sybyl (Tripos, Inc.) mol2-format with AM1-BCC atom type and charge assigned.^{71,72} Because both ITScore and DOCK require mol2-format files for input, we converted the protein files into mol2-format files via UCSF Chimera,⁷³ in which the atom types and charges were assigned with the Amber force fields.⁸ These assignments were used for the DOCK related calculations. For the ITScore calculations, however, the protein and ligand atom types were automatically reassigned from the mol2 files by our MDock software package (<http://zoulab.dalton.missouri.edu/software.htm>).^{56,74,75} MDock requires no charge assignments and considers heavy atoms only. To be comparable with the experimentally determined affinities, the calculated ITScore energy scores were simply scaled by a factor of 10 for all the evaluations. Notice that the scaled ITScore yielded just scores rather than approximate affinities, because no regression calculations were performed to reproduce the experimentally determined binding affinities. Rigid-body optimizations were performed within DOCK and MDock runs to minimize the impact of possible atomic clashes in the crystal structures on scoring.

Figures 3 and 4 show the correlation coefficients of ITScore, DOCK/FF, DOCK/VDW, and VDW (Heavy) for the CSAR benchmark of 345 protein–ligand complexes. The corresponding values of the correlations are listed in Table 1. By default, the correlation refers to the square of the Pearson correlation coefficient, i.e. (R^2). In addition, we also calculated the Spearman's rank correlation coefficients (ρ^2) and Kendall tau

Table 1. Correlations of the Four Scoring Functions on Binding Affinity Prediction for Set 1, Set 2, and All (Set 1 + Set 2) Ligand-Protein Complexes in the CSAR Benchmark (2009 Release)

set	correlations ^a			
	ITScore	DOCK/FF	DOCK/VDW	VDW(Heavy)
set 1	0.527	0.057	0.317	0.330
set 2	0.563	0.253	0.386	0.450
all	0.536	0.117	0.349	0.384
all ^b	0.543	0.155	0.351	0.390
all (ρ^2) ^c	0.559	0.156	0.364	0.386
all (τ^2) ^d	0.301	0.073	0.182	0.193

^aThe correlations refer to the square of the Pearson correlation coefficients (R^2) unless otherwise specified. ^bThe row considers the effect of ligand conformational entropy loss by counting the number of rotatable bonds in each ligand. ^cThe row lists the Spearman rank correlation coefficients (ρ^2). ^dThe row shows the Kendall tau rank correlation coefficients (τ^2).

rank correlation coefficients (τ^2), as shown in Table 1. The value of each correlation is between +1 and −1, and a positive correlation is preferred. A negative correlation means anticorrelation, for which the predictions are opposite to measurements. A correlation of +1 corresponds to perfect prediction of binding affinities, and zero correlation represents a random prediction and thus a complete failure. It can be seen from Table 1 that the relative performances of the four scoring functions were similar regardless of which correlation was used. However, the Kendall tau correlation seems to be the most stringent measurement in the present study, as the values were significantly smaller than the values of the Pearson correlation and the Spearman rank correlation.

Several notable features can be found from the figures. First, ITScore yielded the highest correlation of $R^2 = 0.54$ compared with 0.12 for DOCK/FF, 0.35 for DOCK/VDW, and 0.38 for VDW(Heavy), suggesting the efficacy of our iterative procedure in extracting effective potentials. Second, it is a bit surprising to notice that the DOCK/FF scoring function, which consists of both the electrostatic and VDW terms yielded a lower correlation than the DOCK/VDW scoring function for the CSAR benchmark, implying the importance of rigorous treatment of the electrostatic interactions in the force field scoring functions. Third, the DOCK/VDW for all atoms performed slightly worse than the VDW scoring for heavy atoms only. Detailed examinations on all the scores by DOCK/VDW and VDW(Heavy) revealed that DOCK/VDW yielded a positive score of 4.2 to the complex No. 18 (PDB code: 1DUV) of set 2 ($\log K = -11.8$) as shown in Figure 3(c), a major contribution to the lower correlation of DOCK/VDW. In this protein–ligand complex, the hydrogen atoms assigned for the ligand cause severe atomic clashes even after rigid-ligand optimization, yielding a high score penalty. The observation suggests the importance of correct hydrogen assignments for ligands to the force field scoring functions. Fourth, it is noted that although the DOCK/FF scores range from −165 to 20, most of the data points are within −100 to −1 (Figure 3). Therefore, the correlation for DOCK/FF was recalculated by removing the data that have scores beyond (−100, −1), leading to a significant improvement to $R^2 = 0.17$, compared to 0.11 for the complete set.

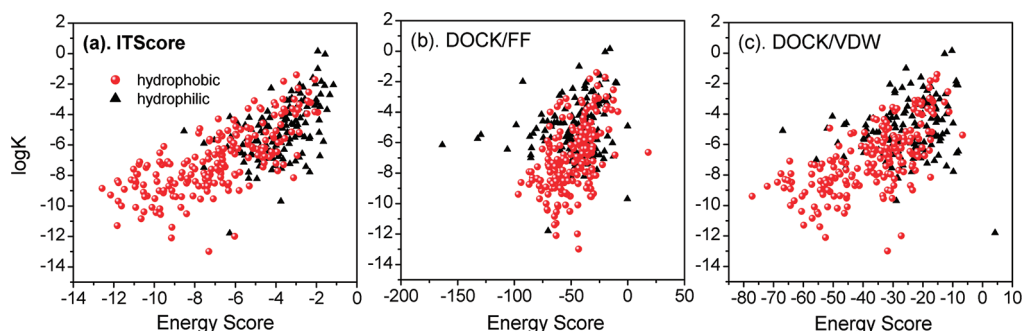


Figure 5. The replots of the affinity-score correlations for ITScore, DOCK/FF, and DOCK/VDW by classifying the whole CSAR benchmark into hydrophobic and hydrophilic complexes. The figure legend in panel a applies to all three panels.

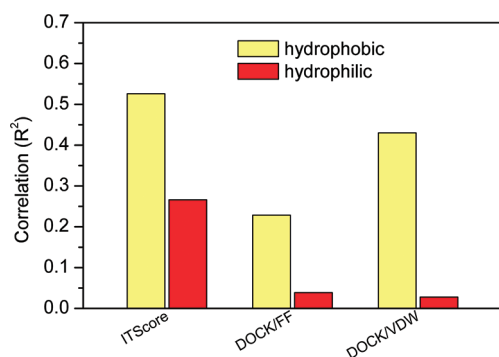


Figure 6. The histograms of the correlations for ITScore, DOCK/FF, and DOCK/VDW for the hydrophobic and hydrophilic complexes in the whole CSAR benchmark, respectively.

Finally, it was also noted that there are common outliers for the four scoring functions. Six of them are 2JBJ, 1SWK, 2C1Q, 1DUV, 2I0A, and 2I0D (see Figure 3). Their binding energy scores were underestimated by all the four scoring functions. These six complexes are also among the 20 entries that were poorly scored by all the 17 participating groups in the CSAR exercise. Examination of 2JBJ, a complex of glutamate carboxypeptidase II (GCP II) and 2-phosphonomethyl-pentanedioic acid (2-PMPA), revealed that two zinc ions in the active site coordinate the phosphonate group of the ligand and three carboxyl groups (D301, D311, and E345) of the protein.⁷⁶ Both oxygen atoms in the charged phosphonate and carboxylate groups would form strong salt bridges with the zinc ions, which is attributed to the tight binding of 2-PMPA. In contrast, neglect of the zinc ions would not only break the salt bridges but also result in possible unfavorable repulsions between negatively charged groups and thus significantly underestimates the binding energy score, as shown in our present results. The ligands in both 1SWK and 2C1Q are biotin, of which the binding involves electron polarization effects. Therefore, accurate calculation of biotin-binding affinities may require quantum mechanical methods.⁷⁷ 1DUV is a complex between the Ornithine transcarbamoylase (OTCase) and a transition-state (TS)-like inhibitor PSOrn.⁷⁸ Its strong binding attributes to not only hydrogen bonds but also unique N–P bonds that are not considered in the present scoring functions. Both 2I0D and 2I0A are HIV-1 protease complexes.⁷⁹ Despite their high experimental affinities, examination on their ligand binding modes showed that part of each ligand stretches outside of the binding

site compared to some less potent HIV-1 protease inhibitors such as 1D4I. It is challenging to understand why the off-pocket ligand fragments result in the high affinities measured by the experiments. The two complexes may deserve further theoretical and experimental studies for their peculiar features.

3.2. Performance Analysis Based on Hydrophobicity/Hydrophilicity. To further evaluate our scoring function, we classified the 345 complexes in the CSAR benchmark into two categories according to the chemical nature of the protein–ligand interactions. We used a classification criterion for hydrophobic/hydrophilic interactions that is similar to the criterion given in ref 80. Namely, we defined the protein–ligand interaction “hydrophobic” if the contribution of the hydrophobic term is larger than that of the hydrogen bond term in HPScore of X-Score³⁸ for a given complex; otherwise we classified the interaction as being “hydrophilic”. The CSAR benchmark was thus divided into a total of 205 hydrophobic complexes and a total of 140 hydrophilic complexes.

Figures 5 and 6 show the correlation of ITScore on the two subsets of hydrophobic and hydrophilic complexes. As a reference, the results of DOCK/FF and DOCK/VDW are also listed in the figures. VDW(Heavy) was not included because of its close performance to DOCK/VDW. It can be seen that all the three scoring functions performed significantly better for the hydrophobic complexes than the hydrophilic complexes. This behavior is particularly prominent for DOCK/FF and DOCK/VDW. The better performance of DOCK/VDW for the hydrophobic complexes can be understood by the fact that VDW interactions make the major contribution and electrostatic interactions make much less contribution to the binding scores for the hydrophobic complexes than for the hydrophilic complexes. Regarding ITScore, although a few outliers may contribute to the lower correlation for the hydrophilic complexes, the difference between the two correlation values may mainly result from a statistical reason, i.e., the majority of the hydrophilic complexes in the data set have a smaller affinity range ($\log K \sim [-8, -2]$) than the hydrophobic complexes ($\log K \sim [-11, -2]$) (see Figure 5). For DOCK/FF, the dramatic difference in the correlations may be due to the overestimation of the electrostatic interactions that will have a much more impact on hydrophilic complexes than on hydrophobic complexes because the atoms of hydrophilic complexes tend to carry more charges than the atoms of hydrophobic complexes. This can also be indicated from the more outliers for the hydrophilic complexes in the score-affinity diagram (Figure 5).

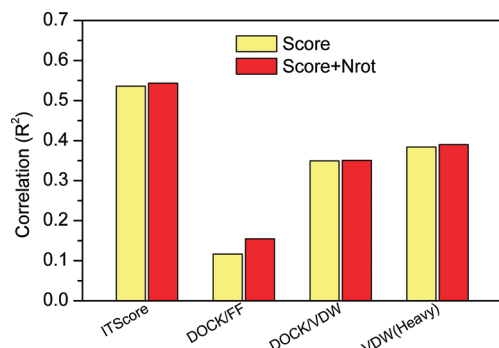


Figure 7. The correlations for ITScore, DOCK/FF, DOCK/VDW, and VDW(Heavy) with or without considering the ligand conformational entropy in terms of the number of rotatable bonds (N_{rot}) in the ligands for the CSAR benchmark of 345 complexes. See the main text for explanations.

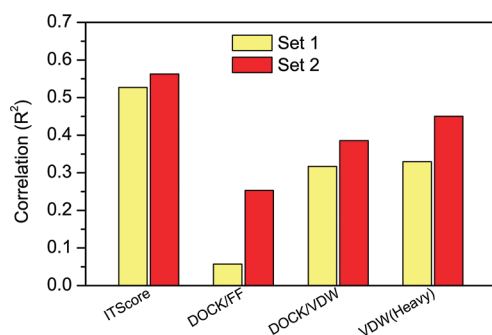


Figure 8. The correlations for ITScore, DOCK/FF, DOCK/VDW, and VDW(Heavy) on the set 1 (169 complexes) and set 2 (176 complexes) of the CSAR benchmark.

3.3. Considering Ligand Conformational Entropy. Another factor on scoring performance is the impact of ligand flexibility. Ligands may change their conformations upon binding. Binding leads to a loss of ligand conformational entropy by constraining a ligand in the binding pocket. As none of the four scoring functions investigated in the present study explicitly account for the conformational entropy, we adopted the empirical, crude approximation that the loss of the ligand conformational entropy is proportional to the number of rotatable bonds N_{rot} in the ligand. Specifically, we added an entropic energy term to each scoring function as follows

$$E_{\text{tot}} = W_x \times E_x + W_{\text{rot}} \times N_{\text{rot}} \quad (7)$$

where E_x stands for E_{ITScore} , $E_{\text{DOCK/FF}}$, $E_{\text{DOCK/VDW}}$, or $E_{\text{VDW(Heavy)}}$. The weighting coefficients W_x and W_{rot} were obtained by a regression method to maximize the correlation between the experimentally measured binding affinities and the calculated binding energy scores for the complexes in the CSAR benchmark. Without loss of generality, we set $W_x = 1.0$ during the regression, yielding the coefficients of $W_{\text{rot}} = 0.10, -6.66, 0.30$, and 0.58 for ITScore, DOCK/FF, DOCK/VDW, and VDW(Heavy), respectively.

Figure 7 shows the correlation coefficients for ITScore, DOCK/FF, DOCK/VDW, and VDW(Heavy) with or without the ligand entropic term. It can be seen from the figure that the inclusion of the empirical entropic term does not significantly



Figure 9. The correlations of ITScore on the CSAR benchmark for the potentials derived from the training set of 1300 protein–ligand complexes and from the nonoverlap training set of 1152 complexes, respectively.

improve the performances of the scoring functions. Interestingly, the coefficient W_{rot} was even negative for DOCK/FF. The finding implies that the number of rotatable bonds may not an effective measurement of the loss of ligand conformational entropy in ligand binding. The first reason is because the number of rotatable bonds, a measure of the ligand flexibility, is an oversimplification of the ligand conformational entropy. The second reason is that a ligand may lose only partial conformational entropy upon binding, but the empirical approximation assumes a complete loss in ligand conformational entropy.

3.4. The Set-Dependence of the Scoring Performance. We next evaluated the dependence of the scoring performance on different data sets. There are two sets in the original CSAR benchmark—released in 2009. Set 1 consists of 176 protein–ligand complexes, most of which were deposited in the Protein Data Bank (PDB)⁸¹ in 2007 and 2008. Set 2 contains 169 complexes deposited in 2006 or earlier. Figure 8 shows the correlation coefficients (R^2) of the four tested scoring functions on set 1 and set 2, respectively. It can be seen that all four scoring functions yielded a higher correlation for set 2 than for set 1. The difference is least significant for ITScore and most prominent for DOCK/FF. The differences in the correlations may result from different physical/chemical properties of the ligands in the two data sets, such as molecular weight, polarity, and flexibility. For example, set 1 contains a larger number of charged ligands than set 2, which may explain the largest difference in the performance of DOCK/FF, a scoring function that involves simple electrostatic calculations.

We further investigated the effect of the training set selection on the performance of ITScore. Figure 9 shows a comparison of the correlations for the ITScore potentials derived from two training sets: one is the complete set of 1300 complexes (default), and the other is the nonoverlap set of 1152 complexes after excluding the entries that are in the CSAR benchmark. It can be seen from the figure that the two training sets yielded very similar correlations, indicating the general applicability of the derived effective potentials.

3.5. The Effect of Protonation and Minimization on Scoring Performance. In addition to the above CSAR benchmark of 345 protein–ligand complexes released in 2009, an updated CSAR benchmark of 343 complexes (referred to as CSAR-NRC HiQ set) was released in 2010 release. Two complexes in Set 2 of the 09' benchmark were removed (#242 with PDB entry 2BB7 and #267 with PDB entry 1NLS), which resulted in 167 complexes in Set 2 of the 10' benchmark. There are two major

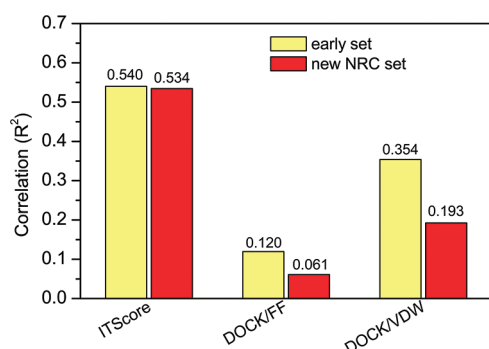


Figure 10. The correlations for ITScore, DOCK/FF, and DOCK/VDW on the early CSAR benchmark (2009 release) and the new CSAR-NRC HiQ benchmark (2010 release) with protonation curation. To be comparable, two complexes were removed from the early released CSAR benchmark so that both benchmarks have the same number of 343 complexes.

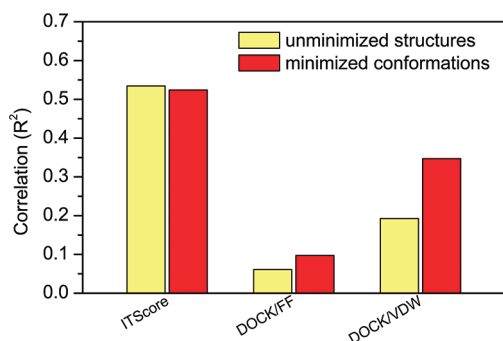


Figure 11. The correlations for ITScore, DOCK/FF, and DOCK/VDW on the unminimized structures and minimized conformations of the new CSAR-NRC HiQ benchmark.

changes in the 10' CSAR-NRC HiQ benchmark compared to the 09' CSAR benchmark. First, the protonation states of some of the ligands were curated manually. Second, unlike the 09' benchmark which gave the PDB-format files only for the crystal structures of the proteins, the 10' benchmark provided the mol2 files for both the crystal structures and the minimized conformations of the protein–ligand complexes. Compared with the pdb-format files, the mol2-format files add the assignments for atom types and charges.

To investigate the effects of protonation and minimization on scoring, we calculated the binding energy scores of ITScore, DOCK/FF, and DOCK/VDW for the CSAR-NRC HiQ benchmark. As VDW(Heavy) was highly correlated with DOCK/VDW in scoring performance, VDW(Heavy) was omitted in this evaluation. It should be noted that the results of DOCK/FF for the 09' CSAR and the 10' CSAR-NRC HiQ benchmarks are not comparable because of the different force fields used for the protein charge assignments for the two benchmarks.

Figure 10 shows the correlations for ITScore, DOCK/FF, and DOCK4/VDW on binding affinity prediction with the original CSAR and new CSAR-NRC HiQ benchmarks. It can be seen that interestingly, the protonation corrections in the new benchmark did not improve the correlation for ITScore but even worsened the correlation for DOCK/FF and DOCK4/VDW. The results can be understood as follows. For ITScore, the interaction

potentials are extracted from the experimentally determined protein–ligand complex structures and thereby implicitly include the corrections for the protonation states upon binding. Consequently, it is expected that the protonation corrections in the new CSAR-NRC HiQ benchmark would not significantly affect the results of ITScore, indicating the robustness of ITScore. The differences in correlation for DOCK/FF and DOCK4/VDW may result from the different atom type and charge assignments in the new CSAR-NRC HiQ set.

To investigate the effect of conformational minimization on scoring, Figure 11 shows a comparison of the correlations before and after the CSAR minimization for the ITScore, DOCK/FF, and DOCK/VDW. Notice that the default parameters were always used for DOCK scoring calculations (i.e., orientational “Minimization” was set to “Yes”) to warrant the best DOCK/FF and DOCK/VDW scores. It can be seen that ITScore yielded a slightly better correlation for the (unminimized) crystal structures than for the minimized structures. The results can be understood by the fact that the potentials of ITScore are derived from the experimentally determined native structures and thus are expected to perform better with the unminimized native structures. In contrast to ITScore, DOCK/FF and DOCK/VDW achieved a better correlation for the minimized conformations than for the unminimized native structures, because the VDW interactions in these scoring functions are sensitive to the ligand positions. Even a few atomic clashes in the native structures can significantly worsen the calculated binding scores, which can be corrected by the conformational minimization procedures that remove these atomic clashes.

4. CONCLUSION

We have developed an improved iterative knowledge-based scoring function (ITScore 2.0) based on the crystal structures of 1300 protein–ligand complexes. ITScore achieved a correlation of $R^2 = 0.54$ on binding affinity prediction with the CSAR benchmark of 345 diverse protein–ligand complexes. For references, we compared the scoring results with those of DOCK/FF, DOCK/VDW, and VDW(Heavy). We also used these simple scoring functions to crudely estimate the contributions of polar and nonpolar interactions to the binding affinities. We found that the scoring performances of DOCK/VDW and VDW(Heavy) were better than expected for the CSAR benchmark, with $R^2 = 0.35$ and 0.38 , respectively, which suggests that VDW alone catches certain general chemical properties in a diverse set of protein–ligand complexes. Although the VDW is not a good scoring function, it may serve as a low-end threshold/reference for validation of scoring functions. The comparison between DOCK/VDW and VDW(Heavy) suggests that correct hydrogen assignments can help improving scoring performance. It was also found that despite the successes reported in the literature, the empirical method using the number of rotatable bonds to consider the ligand conformational entropy does not always improve the performances of scoring functions. Due to its sensitivity to electrostatic interactions, DOCK/FF showed significant dependence on the test set and assignment of charges/protonation states, indicating the importance of force field scoring functions in appropriately treating electrostatics. Because of the sensitivity of the force field scoring functions (including the VDW function) to the atomic positions, a minimization that removes possible atomic clashes in the complexes would significantly improve their performances.

As aforementioned, the electrostatic interactions may be overestimated by the force-field scoring function in DOCK 4.0.1 due to the neglect of the desolvation effect. The later versions of DOCK have included an improved scoring function, SDOCK,^{24–26} to account for the desolvation effect on protein–ligand binding. It will be interesting for future studies to investigate how such desolvation models perform with the CSAR test sets.

AUTHOR INFORMATION

Corresponding Author

*Phone: 573-882-6045. Fax: 573-884-4232. E-mail: zoux@missouri.edu.

ACKNOWLEDGMENT

Support to X.Z. from OpenEye Scientific Software Inc. (Santa Fe, NM) is gratefully acknowledged. X.Z. is supported by NSF CAREER Award 0953839 and NIH grant R21GM088517. The computations were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).

REFERENCES

- (1) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
- (2) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.
- (3) Gilson, M. K.; Zhou, H.-X. Calculation of protein–ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (4) Rajamani, R.; Good, A. C. Ranking poses in structure-based lead discovery and optimization: Current trends in scoring function development. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 308–315.
- (5) Huang, N.; Jacobson, M. P. Physics-based methods for studying protein–ligand interactions. *Curr. Opin. Drug Discovery Dev.* **2007**, *30*, 325–331.
- (6) Huang, S.-Y.; Zou, X. Advances and challenges in protein–ligand docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016–3034.
- (7) Huang, S.-Y.; Grinter, S. Z.; 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.
- (8) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (9) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (10) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Gerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
- (11) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (12) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein–ligand, protein–protein, and protein–nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.
- (13) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128–137.
- (14) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson–Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
- (15) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (16) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339–355.
- (17) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.
- (18) Naim, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. Solvated Interaction Energy (SIE) for Scoring Protein–Ligand Binding Affinities. 1. Exploring the Parameter Space. *J. Chem. Inf. Model.* **2007**, *47*, 122–133.
- (19) Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *48*, 1081–1091.
- (20) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (21) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (22) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- (23) Tjong, H.; Zhou, H. GBr⁶: A parameterization-free, accurate, analytical generalized Born method. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.
- (24) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.
- (25) Liu, H.-Y.; Kuntz, I. D.; Zou, X. Pairwise GB/SA scoring function for structure-based drug design. *J. Phys. Chem. B* **2004**, *108*, 5453–5462.
- (26) Liu, H.-Y.; Zou, X. Electrostatics of ligand binding: Parameterization of the generalized born model and comparison with the Poisson–Boltzmann approach. *J. Phys. Chem. B* **2006**, *110*, 9304–9313.
- (27) Liu, H.-Y.; Grinter, S. Z.; Zou, X. Multiscale generalized born modeling of ligand binding energies for virtual database screening. *J. Phys. Chem. B* **2009**, *113*, 11793–11799.
- (28) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caffisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999**, *37*, 88–105.
- (29) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.
- (30) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein–ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253.
- (31) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.

- (32) Guimaraes, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.
- (33) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy approach to macromolecule-ligand interactions. *J. Comput. Chem.* **1992**, *13*, 505–524.
- (34) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (35) Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.
- (36) Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243–256.
- (37) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. Validate a new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959–3969.
- (38) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.
- (39) Zhang, S.; Golbraikh, A.; Oloff, S.; Kohn, H.; Tropsha, A. A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Model.* **2006**, *46*, 1984–1995.
- (40) Jain, A. N. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.
- (41) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. eHiTS: A new fast, exhaustive flexible ligand docking system. *J. Mol. Graphics Modell.* **2007**, *26*, 198–212.
- (42) Huang, S.-Y.; Zou, X. Mean-force scoring functions for protein-ligand binding. *Annu. Rep. Comput. Chem.* **2010**, *6*, 280–296.
- (43) Verkhrivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng.* **1995**, *8*, 677–691.
- (44) Wallqvist, A.; Jernigan, R. L.; Covell, D. G. A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci.* **1995**, *4*, 1881–1903.
- (45) DeWitte, R. S.; Shakhnovich, E. I. SMOG: de Novo design method based on simple, fast, and accurate free energy estimate. I. Methodology and supporting evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733–11744.
- (46) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.
- (47) Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895–5902.
- (48) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- (49) Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScore^{CSD} – Knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.
- (50) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP – Potential of mean force describing protein-ligand interactions: I. Generating potential. *J. Comput. Chem.* **1999**, *20*, 1165–1176.
- (51) Ishchenko, A. V.; Shakhnovich, E. I. Small molecule growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J. Med. Chem.* **2002**, *45*, 2770–2780.
- (52) Ozrin, V. D.; Subbotin, M. V.; Nikitin, S. M. PLASS: protein-ligand affinity statistical score – A knowledge-based force-field model of interaction derived from the PDB. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 261–270.
- (53) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J. Med. Chem.* **2005**, *48*, 2325–2335.
- (54) Mooij, W. T. M.; Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins* **2005**, *61*, 272–287.
- (55) Yang, C. Y.; Wang, R. X.; Wang, S. M. M-score: a knowledge-based potential scoring function accounting for protein atom mobility. *J. Med. Chem.* **2006**, *49*, 5903–5911.
- (56) 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.
- (57) 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.
- (58) Huang, S.-Y.; Zou, X. Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *J. Chem. Inf. Model.* **2010**, *50*, 262–273.
- (59) Tanaka, S.; Scheraga, H. A. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* **1976**, *9*, 945–950.
- (60) Miyazawa, S.; Jernigan, R. L. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **1985**, *18*, 534–552.
- (61) Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. *J. Mol. Biol.* **1990**, *213*, 859–883.
- (62) Thomas, P. D.; Dill, K. A. An iterative method for extracting energy-like quantities from protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11628–11633.
- (63) Thomas, P. D.; Dill, K. A. Statistical potentials extracted from protein structures: How accurate are they? *J. Mol. Biol.* **1996**, *257*, 457–469.
- (64) Koppensteiner, W. A.; Sippl, M. J. Knowledge-based potentials – Back to the roots. *Biochem. (Moscow)* **1998**, *63*, 247–252.
- (65) McQuarrie, D. A. *Statistical Mechanics*; Harper Collins Publishers: New York, 1976.
- (66) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175–1189.
- (67) Huang, S.-Y.; Zou, X. A statistical mechanics-based method to extract atomic distance-dependent potentials from protein structures. *Proteins* **2011**, DOI: 10.1002/prot.23086, (in press).
- (68) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.
- (69) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977–2980.
- (70) Huang, S.-Y.; Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* **2008**, *72*, 557–579.
- (71) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (72) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (73) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera – A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (74) Huang, S.-Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.

(75) Huang, S.-Y.; Zou, X. Efficient molecular docking of NMR structures: Application to HIV-1 protease. *Protein Sci.* **2007**, *16*, 43–51.

(76) Mesters, J. R.; Henning, K.; Hilgenfeld, R. Human glutamate carboxypeptidase II inhibition: structures of GCPII in complex with two potent inhibitors, quisqualate and 2-PMPA. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 508–513.

(77) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New advance in computational chemistry: full quantum mechanical ab initio computation of streptavidin-biotin interaction energy. *J. Phys. Chem. B* **2003**, *107*, 12039–12041.

(78) Langley, D. B.; Templeton, M. D.; Fields, B. A.; Mitchell, R. E.; Collyer, C. A. Mechanism of inactivation of ornithine transcarbamoylase by N^{δ} -(N' -sulfodiaminophosphinyl)-l-ornithine, a true transition state analogue? Crystal structure and implications for catalytic mechanism. *J. Biol. Chem.* **2000**, *275*, 20012–20019.

(79) Ali, A.; Reddy, G. S.; Cao, H.; Anjum, S. G.; Nalam, M. N.; Schiffer, C. A.; Rana, T. M. Discovery of HIV-1 protease inhibitors with picomolar affinities incorporating N-aryl-oxazolidinone-5-carboxamides as novel P2 ligands. *J. Med. Chem.* **2006**, *49*, 7342–7356.

(80) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.

(81) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.