


Knowledge-Based Scoring Functions in Drug Design: 2. Can the Knowledge Base Be Enriched?

Qiancheng Shen,^{†,||} Bing Xiong,^{‡,||} Mingyue Zheng,^{*,†} Xiaomin Luo,[†] Cheng Luo,[†] Xian Liu,[†] Yun Du,[†] Jing Li,[†] Weiliang Zhu,[†] Jingkang Shen,[‡] and Hualiang Jiang^{*,†,§}

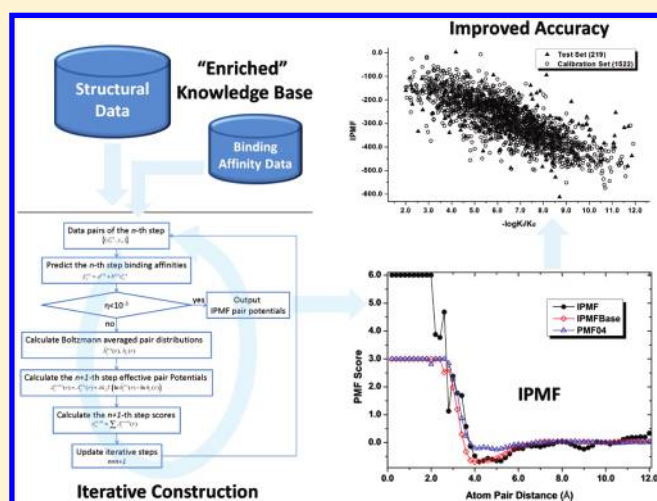
[†]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

[‡]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

[§]School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

 Supporting Information

ABSTRACT: Fast and accurate predicting of the binding affinities of large sets of diverse protein–ligand complexes is an important, yet extremely challenging, task in drug discovery. The development of knowledge-based scoring functions exploiting structural information of known protein–ligand complexes represents a valuable contribution to such a computational prediction. In this study, we report a scoring function named IPMF that integrates additional experimental binding affinity information into the extracted potentials, on the assumption that a scoring function with the “enriched” knowledge base may achieve increased accuracy in binding affinity prediction. In our approach, the functions and atom types of PMF04 were inherited to implicitly capture binding effects that are hard to model explicitly, and a novel iteration device was designed to gradually tailor the initial potentials. We evaluated the performance of the resultant IPMF with a diverse set of 219 protein–ligand complexes and compared it with seven scoring functions commonly used in computer-aided drug design, including GLIDE, AutoDock4, VINA, PLP, LUDI, PMF, and PMF04. While the IPMF is only moderately successful in ranking native or near native conformations, it yields the lowest mean error of 1.41 log K_i/K_d units from measured inhibition affinities and the highest Pearson's correlation coefficient of R_p^2 0.40 for the test set. These results corroborate our initial supposition about the role of “enriched” knowledge base. With the rapid growing volume of high-quality structural and interaction data in the public domain, this work marks a positive step toward improving the accuracy of knowledge-based scoring functions in binding affinity prediction.



INTRODUCTION

Molecular docking has become an essential component of many drug discovery projects. In addition to its extensive usages in virtual screening and subsequent lead optimization, molecular docking was recently employed in *in silico* target identification.^{1–3} Selecting the appropriate drug target can be as important as optimizing the chemical entity that binds to that target.⁴ Therefore, in drug discovery process, an accurate and fast solution to the docking problem is of practical and fundamental importance. Basically, molecular docking involves two steps: a conformational sampling step positioning a ligand in a protein binding site and a scoring step assessing binding affinities of generated poses. While the conformational search is often regarded resolvable to a certain

extent,⁵ the scoring still remains a challenging field of research and poses a major obstacle for the reliability of docking.^{6–8}

Scoring functions are typically classified into three groups: force field based, empirical, and knowledge based. Comprehensive analyses of the strengths and shortcomings of these scoring function types can be found elsewhere.^{6–9} In this study, we will focus on the knowledge-based scoring functions, which generally offer a good balance between accuracy and calculation efficiency. The most common approach for developing this type of function is to extract structural information from experimentally determined

Received: September 2, 2010

Published: December 30, 2010

protein–ligand complexes, as stored in Protein Data Bank (PDB) and Cambridge Structural Database (CSD), and employ the Boltzmann law to transform the atom pair preferences into distance-dependent pairwise potentials. Many different aspects and properties of these potentials have been explored during the last few decades, including the different data sources to derive them,^{10,11} different atom typing schemes,^{12,13} and various models for defining the reference state.^{14–20}

An ideal knowledge-based scoring function is expected to identify the near-native binding poses and at the same time provide a score that is close (or proportional) to the experimentally derived value. The latter goal is also the most useful outcome of a scoring function. However, because of the subtle enthalpic and entropic effects involved in binding and static views of crystal structure coordinates (i.e., the knowledge base), the accurate prediction of binding affinity is essentially beyond the reach of all types of current scoring functions.^{21,22} As pointed out by Bissantz et al., for the molecular interactions observed in PDB or CSD, a large part of the thermodynamic cycle represented by a binding free energy is missing, such as solvation processes, long-range interactions, and conformational changes.²³ Careful approximations or considerations of these contributions have to be made to develop more reliable knowledge-based scoring functions. To this end, Yang et al. addressed the mobility issue of protein atoms in their M-Score. Instead of a fixed position, the location of each protein atom is described by a Gaussian distribution based upon the isotropic B-factors.¹² Huang and Zou explicitly included the contributions from solvation and configurational entropy in their knowledge-based scoring function ITScore/SE.²⁴ In this model, a solvent-accessible surface area-based energy term and an empirical term proportional to the number of rotatable bonds in the molecule were introduced. Compared to the previous version of ITScore without added terms,²⁵ the revised score shows improved accuracy for identifying near-native binding modes and equal performance for binding affinity prediction for a benchmark of 100 diverse protein–ligand complexes. Generally, adding explicit terms to the existing statistical potentials is difficult because it is not clear to what extent these contributions have been captured and whether the additional terms cause an overestimation. In ITScore/SE, the problem was addressed by using an iteration method, where the effective potentials and solvation parameters are derived by iteratively comparing the experimental and predicted structures until the potentials can reproduce the experimentally observed pair distribution functions. In contrast, Muegge and Martin considered the solvation and entropic contributions in their PMF potentials implicitly.¹⁵ In their work, a large cutoff of 12 Å for the reference state and a correction term accounting for the volume taken by the ligand were introduced for the treatment of these opposing contributions. This is an ingenious idea because the double counting problem can be avoided and the function form is simple. Many successful adaptations to the PMF score were made on the basis of this idea.^{11,26,27} Despite all these progresses, the resultant scoring functions show improvements in binding affinity prediction for some ligand series for a particular protein but not for large sets of diverse protein–ligand complexes.

In this study, we describe a novel method for improving knowledge-based scoring functions. As we know, a significant feature of this type of function is that they are solely based on the structural information. With the fast growth of available 3-D structures of protein–ligand complexes and their associated binding affinity data, a question arises as to why the binding

affinity information cannot be used as a part of the knowledge base. In the following sections, we concentrate on addressing this question. Thus, the aim of the study is to investigate how to include experimental binding affinity data in knowledge-based scoring function development and how the additional information affects the performances of derived scoring function, especially in terms of binding affinity prediction performance.

To this aim, we propose a novel iterative approach for extracting improved statistical potentials from 1522 protein–ligand complexes and associated experimental binding data. The procedures are as follows: The function form of PMF originally presented by Muegge and Martin is adopted for deriving initial potentials. Then, the derived potentials are iteratively adjusted by introducing a Boltzmann weighted pair distribution function that utilizes experimental binding data. Meanwhile, the correlation coefficient between the observed binding affinities and calculated scores of the calibration set complexes was monitored. The final potentials are obtained when the change of the correlation coefficient converges. In the end, the potentials derived at the final iterative step, referred to as IPMF, are used for further assessment and validations. The resulting IPMF score is also compared to seven different scoring functions for binding affinity prediction and binding poses reranking.

METHODS

1. Preparation of Data Sets. Protein–ligand complexes from the “refined set” of the publicly available database PDBbind (version 2009) were used in this study, which provides a comprehensive collection of binding affinities for the complexes in the Protein Data Bank (PDB).^{28,29} In the release, the “refined set” consists of a total of 1741 entries. Here, we choose the set because it comprises high-quality complexes with known K_i or K_d values that were carefully checked. The refined set of PDBbind complexes also has a subset called the “core set” of 219 complexes, which is selected through a systematic nonredundant sampling of the refined set with special focus on the diversity of structures and binding data. Detailed procedures to prepare these two sets can be found elsewhere.^{29,30} Briefly, the refined set is clustered by protein sequence similarity using a cutoff of 90%; for each cluster, at least four representatives were selected as members of the core set. The calibration set of this study is compiled from the “refined set” by removing the structures overlapped with the “core set”, yielding 1522 protein–ligand complexes and associated experimental binding data for extracting effective pair potentials. Then, the removed complexes (i.e., the whole “core set”) are used for model validation, constituting the test set. Following the definitions provided in PMF04¹¹ and our previous implementation,²⁷ a set of 17 protein and 34 ligand atom types were used. An in-house python scripts utilizing OpenBabel^{31,32} for the atom typing can be obtained from the Supporting Information.

2. Calculation of Initial Pair Distribution Function. Muegge's method was used here to derive initial values of statistical potentials,^{11,15} which are briefly described as follows: The score of the m -th protein–ligand complex x_m is calculated as the sum over all protein–ligand atom pair potentials $A_{ij}(r)$ at distance r

$$x_m = \sum_{\substack{kl, m \\ r < r_{\text{cut-off}}^{ij}}} A_{ij}(r) \quad (1)$$

where kl is a protein–ligand atom pair of type ij , $r_{\text{cut-off}}^{ij}$ is the distance at which atom pair interactions are truncated. The potential $A_{ij}(r)$ is obtained by

$$A_{ij}(r) = -k_B T \ln g_{ij}(r) \quad (2)$$

where k_B is the Boltzmann constant, T is the absolute temperature, $g_{ij}(r)$ represents the pair distribution function with form

$$g_{ij}(r) = f_{\text{Vol}_{\text{corr},j}}(r) \frac{\rho_{ij}(r)}{\rho_{ij,\text{bulk}}} \quad (3)$$

where $f_{\text{Vol}_{\text{corr},j}}(r)$ is the ligand volume correction factor for a ligand atom type j ; $\rho_{ij}(r)$ is the number density of a protein–ligand atom pair of type ij at a certain atom pair distance r ; $\rho_{ij,\text{bulk}}$ is the number density in a reference sphere with a radius of R ($R = 12 \text{ \AA}$). $\rho_{ij}(r)$ and $\rho_{ij,\text{bulk}}$ can be calculated by

$$\rho_{ij}(r) = \sum_{m=1}^M \frac{n_{ij,m}(r)}{4\pi r^2 dr} \quad (4)$$

$$\rho_{ij,\text{bulk}} = \sum_{m=1}^M \frac{N_{ij,m}}{4\pi R^3/3} \quad (5)$$

where $n_{ij,m}(r)$ and $N_{ij,m}$ are the numbers of atom pair ij in the spherical shell and the reference sphere for the m -th complex, respectively.

Specially, the ligand volume correction factor $f_{\text{Vol}_{\text{corr},j}}(r)$ was introduced to normalize the PMF according to the available solvent/protein volume by ignoring the ligand–ligand interactions.^{15,33} It can be understood as the quotient of effective volumes taken by the ligand in the reference sphere and in a certain spherical shell of thickness Δr . For calculation, a series of formulas are defined as follows

$$\hat{\rho}_{ij}(r) = \rho_{ij}(r) \frac{\rho_{kj}(r)}{\rho_{kj}(r) + \rho_{lj}(r)} \quad (6)$$

$$\hat{\rho}_{ij,\text{bulk}} = \rho_{ij,\text{bulk}} \frac{\rho_{kj,\text{bulk}}}{\rho_{kj,\text{bulk}} + \rho_{lj,\text{bulk}}} \quad (7)$$

$$f_{\text{Vol}_{\text{corr},j}}(r) = \frac{\hat{\rho}_{ij}(r) \rho_{ij,\text{bulk}}}{\hat{\rho}_{ij,\text{bulk}} \rho_{ij}(r)} \quad (8)$$

where $\hat{\rho}_{ij}(r)$ and $\hat{\rho}_{ij,\text{bulk}}$ mean the ligand volume corrected number density and bulk number density, respectively; $\rho_{kj}(r)$ designates the number densities of protein atoms k of any type around a ligand atom of type j at distance r ; $\rho_{lj}(r)$ is the number densities of ligand atom l of any type around a ligand atom type j at distance r .

3. Extraction of Potentials Using an Iterative Method.

Scoring functions assume that a calculated score can be used as a predictor for observed protein–ligand binding constants. Furthermore, if we suppose that the relationship between observed binding affinities (y_m) in terms of $-\log K_i$ or $-\log K_d$ and scores x_m can be described by a linear regression model, we may obtain the predicted binding affinities (\hat{y}_m) from a linear fitted equation:

$$\hat{y}_m = a + bx_m \quad (9)$$

where the model parameters a and b were fitted by minimizing the value of chi-square statistics. In general, the initial scores

x_m are unlikely to generate accurate prediction to y_m . Thus, we constructed an iteration device to improve the potentials. Because variables x_m and $A_{ij}(r)$ were also used and updated during the following iteration procedures, for clarity, hereinafter, we add a superscript (n) to these variables to indicate the current iteration number. So eq 1 can be written as

$$x_m^{(n)} = \sum_{\substack{kl,m \\ r < r_{\text{cut-off}}^{ij}}} A_{ij}^{(n)}(r) \quad (10)$$

We used the superscript (0) to designate a variable used at the start of the iteration. The zeroth potentials were obtained from eq 2, which becomes

$$A_{ij}^{(0)}(r) = -k_B T \ln g_{ij}(r) \quad (11)$$

Detailed iteration procedures are described as follows:

Suppose the calibration set contains M ligand–receptor complexes, and we have M data pairs $\{x_m^{(n)}, y_m\} : m = 1, \dots, M; n = 0, \dots, N\}$ at n -th iteration. By linear regression, we calculate the predicted binding affinity $\hat{y}_m^{(n)}$ based on the PMF score of $x_m^{(n)}$, with the linear fitted equation

$$\hat{y}_m^{(n)} = a^{(n)} + b^{(n)} x_m^{(n)} \quad (12)$$

The basic idea of our method is to correct the trial potentials iteratively by comparing two pair distributions weighted by a Boltzmann factor, $\hat{h}_{ij}^{(n)}(r)$ and $h_{ij}(r)$, until the change of correlation coefficient between $\hat{y}_m^{(n)}$ and y_m converges. The effective pair potentials of the $n + 1$ -th iterative step $A_{ij}^{(n+1)}(r)$ in eq 10 were obtained by modifying the potentials of the proceeding (n) step as

$$\begin{aligned} A_{ij}^{(n+1)}(r) &= A_{ij}^{(n)}(r) + \Delta A_{ij}^{(n)}(r) \\ &= A_{ij}^{(n)}(r) + \lambda k_B T (\ln \hat{h}_{ij}^{(n)}(r) - \ln h_{ij}(r)) \end{aligned} \quad (13)$$

where $\Delta A_{ij}^{(n)}(r)$ is a correction term and λ is a scaling parameter to control the quantity of the correction, and hence, the speed of convergence.³⁴ The larger the λ is, the more quickly an iteration run terminates. Here a small λ value of 0.2 was set to keep the change of potentials slow and steady.

Given a certain protein–ligand complex m and its associated binding affinity y_m , we weighted the counts of atom pairs extracted from m by the Boltzmann factor $e^{y_m/(k_B T)}$ and then computed the Boltzmann average (over the different protein–ligand complexes in the calibration set) of the pairing frequencies using the following weighting function

$$\begin{aligned} h_{ij}(r) &= f_{\text{Vol}_{\text{corr},j}}(r) \frac{\sigma_{ij}(r)}{\sigma_{ij,\text{bulk}}} \\ &= f_{\text{Vol}_{\text{corr},j}} \frac{\sum_{m=1}^M \frac{n_{ij,m}(r) e^{y_m/(k_B T)}}{4\pi r^2 dr}}{\sum_{m=1}^M \frac{N_{ij,m} e^{y_m/(k_B T)}}{4\pi R^3/3}} \end{aligned} \quad (14)$$

where $\sigma_{ij}(r)$ is the weighted number density of a protein–ligand atom pair of type ij at a distance r , and $\sigma_{ij,\text{bulk}}$ is the weighted number density in a reference sphere with a radius of R . These two variables are the counterparts of $\rho_{ij}(r)$ and $\rho_{ij,\text{bulk}}$ in the unweighted pair distribution function $g_{ij}(r)$ in eq 3, respectively.

In addition to y_m , we also have the zeroth approximation of the binding affinities $\hat{y}_m^{(0)}$ from eq 12. Following the same weighting

approach, we obtained another weighted pair distribution function $\hat{h}_{ij}^{(0)}(r)$ utilizing the Boltzmann factor $e^{\hat{y}_m^{(0)}/(k_B T)}$ with the following general function

$$\hat{h}_{ij}^{(n)}(r) = f_{\text{Vol,corr}} \frac{\sum_{m=1}^M \frac{n_{ij,m}(r) e^{\hat{y}_m^{(n)}/(k_B T)}}{4\pi r^2 dr}}{\sum_{m=1}^M \frac{N_{ij,m} e^{\hat{y}_m^{(n)}/(k_B T)}}{4\pi R^3/3}} \quad (15)$$

Accordingly, in the first iterative step, $A_{ij}^{(1)}(r)$ was calculated by the initial potential $A_{ij}^{(0)}(r)$ plus the scaled difference between the two weighted pair distribution functions $\hat{h}_{ij}^{(0)}(r)$ and $h_{ij}(r)$

$$A_{ij}^{(1)}(r) = A_{ij}^{(0)}(r) + \lambda k_B T (\ln \hat{h}_{ij}^{(0)}(r) - \ln h_{ij}(r)) \quad (16)$$

Having $A_{ij}^{(1)}(r)$, the corrected scores of the first iterative step $x_m^{(1)}$ can be obtained from eq 10. As a result, we obtained the M data pairs $\{(x_m^{(1)}, y_m): m = 1, \dots, M\}$ required for the next iteration step. In this way, the effective pair potentials $A_{ij}^{(n)}(r)$ and the calculated scores $x_m^{(n)}$ were updated each iterative step. The iteration terminates when the n -th approximation of the binding affinities $\hat{y}_m^{(n)}$ are close to the experimental values y_m .

To account for inherent uncertainties in experimental data, the smoothing strategy described by Mitchell et al. was applied to potentials of each iteration step, i.e. the value of a raw potential in the i -th bin was set to the weighted average of 1:2:4:2:1 of the potential from bins $(i-2)$ to $(i+2)$.¹⁶ In the present work, the bin size was set to 0.2 Å. The Pearson's correlation coefficient (R_p^2 , defined in the next section) between observed binding affinities and calculated scores from the smoothed potentials was monitored during the iteration progress. We defined the increment quantity of R_p^2 in an interval of 20 steps as the convergence parameter (η) of the iteration

$$\eta^{(n)} = [R_p^{(n)}]^2 - [R_p^{(n-20)}]^2 \quad (17)$$

The convergence criteria was set as $\eta = 10^{-3}$. Thus, the potentials derived at the step, termed as IPMF, were then used for further tests and validation. A flowchart of our iterative method is shown in Figure 1.

It should be noted that R_p^2 is obtained directly from X-ray structures. If a scoring function is overfit, one would expect its results to be sensitive to the exact coordinates used in input complexes. To investigate the effect of trivial geometric differences on predicted binding affinities, we also monitored the correlation coefficient calculated from energy-minimized structures, denoted as $R_p'^2$. The minimization details are as follows: The program SZYBKJ³⁵ with MMFF94 force field and conjugate gradient method was used for in situ optimization. For each complex, the ligand is subjected to up to 500 steps of minimization in the presence of a fixed protein atom environment. The convergence criterion was set to 0.1 kcal/(mol Å).

4. Validation of Derived Scoring Function. As mentioned above, we used the "core set" of 219 protein–ligand complexes to evaluate the performance of the derived potentials on binding affinity prediction and binding mode identification. The results were compared with seven popular scoring functions, including four scoring functions, i.e. PLP,³⁶ PMF,¹⁵ PMF04¹¹, and LUDI,³⁷ implemented in Discovery Studio (version 2.5, Accelrys, Inc.), Glide score³⁸ (GLIDE) of Glide program (version 5.5) in the Schrödinger Suite (version 2009, Schrödinger LLC), AutoDock4 scoring function³⁹ (AD4) in AutoDock⁴⁰ (version 4.2.3), and

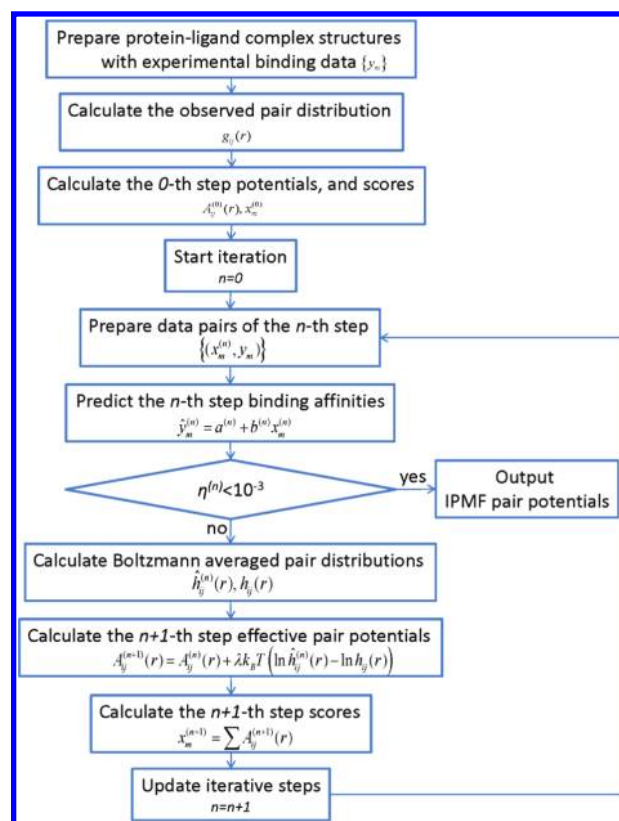


Figure 1. Flowchart showing the iterative approach.

AutoDock Vina (VINA) score⁴¹ (version 1.1.1). These scoring functions can be roughly classified into three groups: (i) empirical scoring functions, including PLP, LUDI, and VINA, (ii) knowledge-based potentials, including PMF and PMF04, and (iii) force-field based approaches, including GLIDE and AD4 scores. These scoring functions were applied to compute the binding scores of the 219 protein–ligand complexes in the test set with default parameter settings.

As defined in eq 9, the performance of each scoring function set can be measured by a linear correlation between its calculated score (x) and the experimental binding affinity (y) of the protein–ligand complexes in the test set. The fitted $-\log(K_b/K_d)$ value \hat{y} is calculated as

$$\hat{y} = a + bx \quad (18)$$

While a number of metrics are available for comparing the performance of binding affinity prediction, the widely used Pearson's correlation coefficient (R_p), standard deviation (SD), and unsigned mean error (ME) in regression were calculated in this study. For a test set containing N complexes, they are computed as

$$R_p = \frac{\sum_m (x_m - \bar{x})(y_m - \bar{y})}{\sqrt{\sum_m (x_m - \bar{x}) \times \sum_m (y_m - \bar{y})}} \quad (19)$$

$$SD = \sqrt{\sum_m (y_m - \hat{y}_m)^2 / (N - 2)} \quad (20)$$

$$ME = \sum_m |y_m - \hat{y}_m| / N \quad (21)$$

where \bar{x} and \bar{y} denote the arithmetic averages of x_m and y_m , respectively. The Spearman's rank order correlation coefficient (R_s) provides a complementary measurement of ranking ability of scoring functions

$$R_s = 1 - \frac{6 \times \sum_i (r_i - s_i)^2}{N^3 - N} \quad (22)$$

in which r_i is the rank of complex i determined by its experimental binding affinity, while s_i is the rank determined by a scoring function. For better comparison, we changed the sign of some scoring functions to ensure that a lower score always indicates a higher binding affinity.

Performance on reproducing near-native binding modes was also tested. For each ligand–protein complex in the test set, up to 10 discrete putative ligand binding poses were first generated using the docking program AutoDock Vina, with the maximum number of binding modes to produce (num_modes) set to 10, and the maximum energy difference between the best binding mode and the worst one set to default (energy_range = 3 kcal/mol). These decoy poses plus the native one in a cocrystallized structure, making a binding pose ensemble for each complex in the test set, were separately assessed by the above-mentioned scoring functions. Next, for each scoring function the best docked modes and their root-mean-square deviation (rmsd) values from the respective crystal structures were collected for the whole test set

$$\text{rmsd} = \sqrt{\frac{\sum_i^n (x_{\text{docked}}(i) - x_{\text{crystal}}(i))^2}{n}} \quad (23)$$

where the sum runs over all n non-hydrogen atoms of a given ligand. We plotted cumulative success rates of different scores to show the distribution of the best predicted poses with respect to the rmsd values.

RESULTS AND DISCUSSION

1. Effects of Iteration. The theory behind knowledge-based scoring functions is Boltzmann's principle, i.e., it assumes that protein–ligand atom pairs observed in a database of structures correlate with the frequencies expected from thermodynamic behavior. However, the structures deposited in PDB or CSD do not provide a thermodynamic ensemble at equilibrium, and it is not clear whether each complex of a heterogeneous structure set is in its own global free energy minimum.¹⁸ Therefore, the atom pair distribution function $g_{ij}(r)$ used in eq 2 is in fact a rough approximation of the frequency that atom pairs would have in thermodynamic equilibrium, so the physical ground of the resulted potentials appears less solid.

Iterative strategy provides a practical means to circumvent the unphysical aspects of the potentials. Previously, Thomas and Dill reported an iterative approach for extracting residue–residue contact potentials to address protein threading problems.⁴² The potentials were improved by iteration until they correctly discriminated a set of known protein folds from decoy conformations. A similar method was also used by Huang and Zou to extract protein–ligand atom pair potentials.²⁵ One of the major differences between these approaches and ours lies in the correction term $\Delta A_{ij}^{(n)}(r)$. In our proposed approach, the first-step

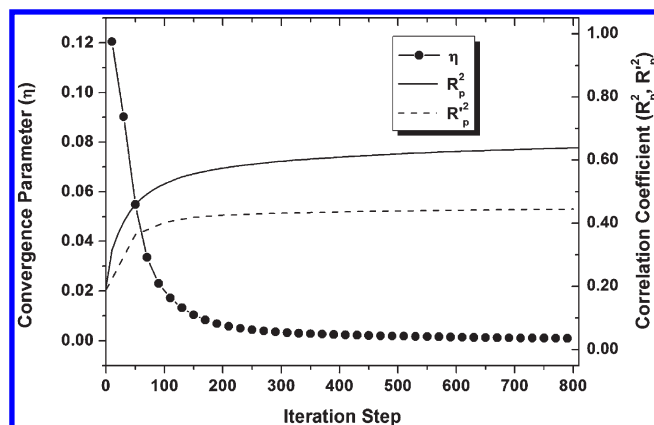


Figure 2. Plots showing the change of the convergence parameter η and Pearson's correlation coefficients R_p^2 and $R_p'^2$ along the iteration.

potentials (eq 16) are equal to $A_{ij}^{(0)}(r)$ plus a correction $\Delta A_{ij}^{(0)}(r)$. Here, the first term $A_{ij}^{(0)}(r)$ is a set of typical knowledge-based potentials that utilizes the atom pair distribution function $g_{ij}(r)$ and only extracts the 3-D structural information. The correction term $\Delta A_{ij}^{(0)}(r)$, in contrast, uses the weighted atom pair distribution functions $h_{ij}(r)$ and $\hat{h}_{ij}^{(0)}(r)$ and exploits extra information of experimental binding data. As defined in eq 14, these two functions assign a weighting factor to atom pairs observed in each complex, which depends exponentially on the observed or predicted binding affinity of the complex. In statistical procedures to derive potentials, atom pairs observed from protein–ligand complexes with tight or loose binding can be discriminately treated through this function. In turn, the resultant potentials are expected to show improvements in binding affinity prediction. If the inference is valid (i.e., $x_m^{(1)}$ is a better predictor to y_m than $x_m^{(0)}$), the correlation coefficient between $\hat{y}_m^{(1)}$ and y_m will be higher than that between $\hat{y}_m^{(0)}$ and y_m , and the information increment $\Delta A_{ij}^{(1)}(r)$ will be smaller than $\Delta A_{ij}^{(0)}(r)$. It means that, with the progressing of the iteration, the correlation coefficient will be enhanced, while the correction to potentials will be gradually reduced. On that ground, the iteration can eventually converge, and it will arrive at a final set of potentials that incorporates both the structural and binding affinity information and more importantly may improve the binding affinity prediction ability.

Figure 2 shows the change of R_p^2 and η as function of iterative steps. This figure supports the above inference: the correlation coefficient R_p^2 between the predicted and observed binding affinities was significantly enhanced from about 0.2 to 0.5 in the first 100 steps and slowly increased to around 0.6 in the following 600 steps. The shapes of the plots reflect that the correction term imposed on the potentials of each ongoing iterative step is slowly decreased. By setting the convergence criterion to 10^{-3} ($\eta = 10^{-3}$), the iteration terminated at the 750th step. Thus, the potentials derived at the step, termed as IPMF, were used for further analyses and validation.

Figure 2 also shows the change of $R_p'^2$, the correlation coefficient for the minimized structures over iteration. The plot follows a similar trend observed for R_p^2 but is consistently lower. For the step deriving IPMF potentials, a $R_p'^2$ value of 0.44 was obtained, which is 0.20 less than the value obtained from the X-ray structures. This result suggests that IPMF potentials may have the risk of fitting noise and also suggests that $R_p'^2$ is a more conservative parameter for measuring the fitting and predicting power of our model.

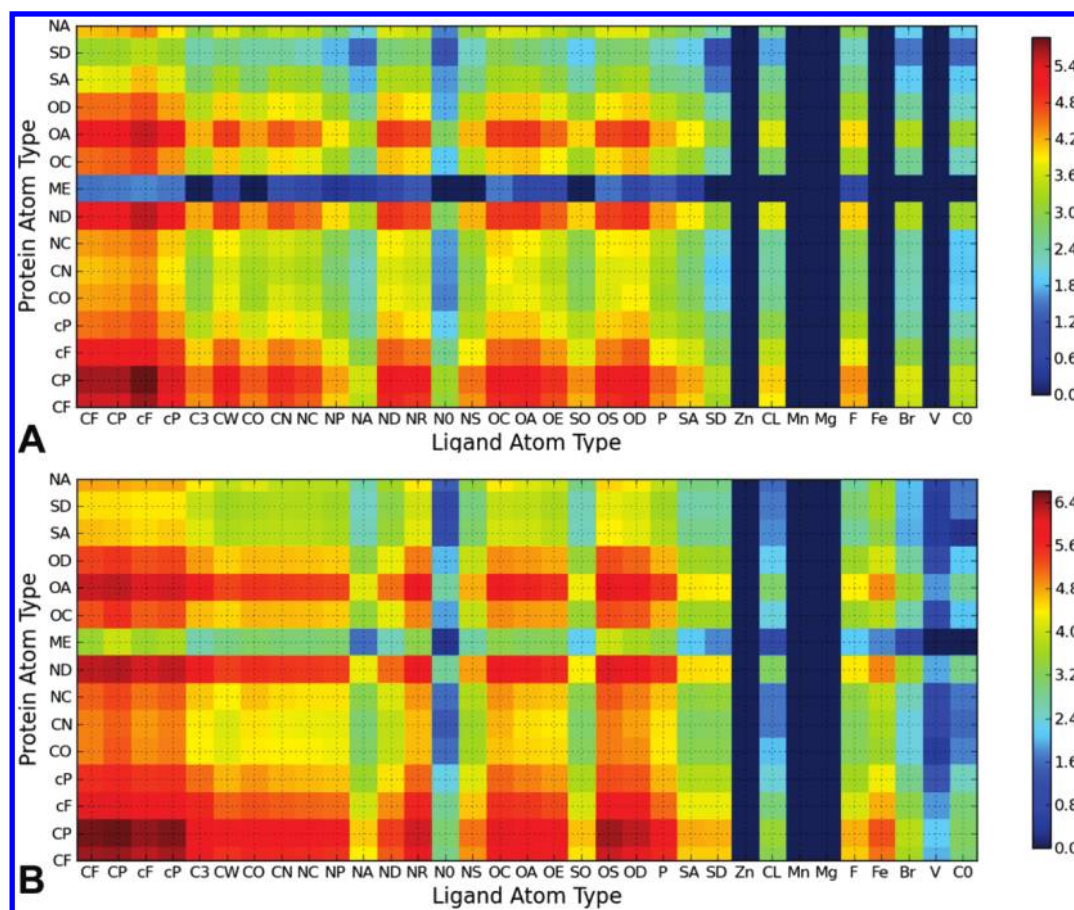


Figure 3. Heat maps showing logarithms (\log_{10}) of occurrence frequencies of atom pairs for (A) the data set of IPMF and (B) the data set of PMF04 (hydrogens are not shown).

2. Properties of IPMF Potentials. Compared to the calibration data set of PMF04, including 7152 protein–ligand complexes, the data set used for extracting IPMF potentials consists of only 1522 samples. A question arises whether the amount of samples in our data set ensures the statistical significance of our derived potentials. We compared occurrence frequencies of atom pairs for the data sets of IPMF and PMF04 with heat maps (Figure 3). After rescaling the color spectrum with the lowest and the largest occurrence values of each data set, we found that the color distributions of these two maps closely resemble to each other. This suggests that these two data sets share a similar atom pair composition.

For each map, different cells show significant color variations, indicating some protein and ligand atom pair types are adequately sampled, while others are not. To derive statistical significant potentials, we retained only the interaction potentials of the atom type pairs whose occurrences were above 1000, corresponding to the warm color cells in Figure 3A. Thus, there are 64% of atom pairs in IPMF database that have effective interaction potentials. For PMF04, this ratio is 77%. Further investigation of the two maps in Figure 3 shows that the ratio difference is mainly caused by the fewer occurrences of atom pairs involving a ferrum atom of ligand (FE) and a metal atom of protein (ME) in the data set of IPMF. It is not clear why FE-related atom pairs have such a high occurrence rate in PMF04. Most iron atoms in PDB are tightly bound ions (such as in heme), which should be better considered as part of a prosthetic group in protein instead of as part of a ligand. Nevertheless, in

medicinal chemistry research, few ligands contain an iron atom, and these interactions are seldom required for scoring. For ME, all the related atom pairs have an occurrence frequency below 1000 in our calibration set. This means that for metal-containing systems, interactions involving metal–ligand pairs are not considered in IPMF. Although it is a rough approximation to use one atom type to characterize various metal types, PMF04 has some effective interaction potentials for this pair type. In this regard, IPMF is not fit for handling protein–ligand complexes where metal–ligand interactions are a part of dominating factors. We defer inclusion of such interactions to future work.

The IPMF potentials between aliphatic carbons (CFCF) and between aliphatic and aromatic carbons (CFcF) are shown in Figure 4 as examples. For comparison, the potentials before iteration, referred to as IPMF-Base, and PMF04 were also plotted. As described in the Method section, IPMF-Base is identical to PMF04 in all aspects, except the database of structures from which they are derived. Therefore, the differences between PMF04 and IPMF-Base can be deemed as the dependence of statistical potentials on calibration data sets, while the differences between IPMF and IPMF-Base reflect the net effects of iteration. Clearly, a similar overall shape is shared by these three types of potentials, where the potential minimum around 4 Å corresponds to hydrophobic interactions between these atom pairs. IPMF, in particular, shows a higher platform near the hard repulsive core region. For a R_p^2 -driven iteration implemented here, this feature means that enhancement of short-distant repulsion between these atom pairs is beneficial

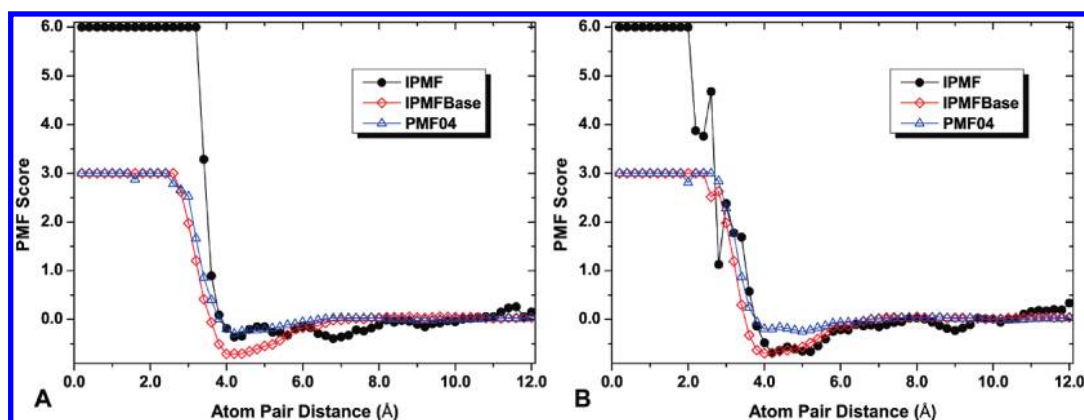


Figure 4. Distance-dependent pair potentials derived from the IPMF, IPMF-Base, and PMF04 for the hydrophobic interactions between CFCF (A) and CFcF (B), in which the first atom-type label refers to a protein atom and the second to a ligand atom.

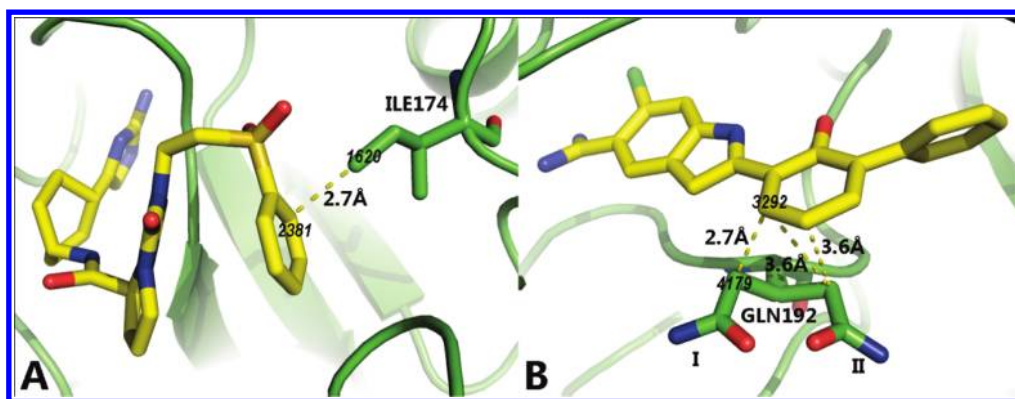


Figure 5. Complex structures with short CFCF atom pair distance: (A) 1D6W and (B) 1GJ7.

to data fitting, hence a better correlation between the calculated and observed binding affinities. Another point of note is that the CFCF potential of IPMF exhibits a local minimum at around 5.0 Å. PMF04 also shows a detectable minimum at the same distance, while IPMF-Base does not. These observations indicate that through our iteration approach IPMF can reproduce some features that can only be captured by extensive data sampling.

It should also be noted, however, that not all potential shapes of IPMF depict meaningful interactions. Take the abrupt peak and valley at 2.6–2.8 Å of the CFCF potential as an example. The distance is much shorter than the favorable distance range for typical interactions involving aromatic carbon and aliphatic carbon ($-\text{CH}_2-$), which is 3.6–4.4 Å.²³ As compared to that of PMF04 and IPMF-Base, these variations apparently exaggerate the minor fluctuation in the data and should reflect the noise instead of the underlying interactions. A detailed structural examination of our calibration set reveals that there are indeed two CFCF pairs with their distance falling in this range. One is in 1D6W (PDB entry, Figure 5A), the atom pair between atom 1620 (CD1 of ILE174 of protein) and 2381 (C2 of the phenyl group of ligand), with the distance of 2.7 Å. The other is in 1GJ7 (Figure 5B), the atom pair between atom 3292 (CG of GLN192 of protein) and 4179 (C4 of the biphenyl-3-yl group of ligand), of which the distance is also 2.7 Å. For 1D6W, because R and R_{free} factors (in crystallography) were not reported, the short distance calls into question the quality of the structure. For 1GJ7, there are two possible positions of the residue GLN192 reported in the

PDB file, denoted as “I” and “II” in Figure 5B. In PDBbind database only the “I” position was kept. While, for the “II” position, all CFCF atom pairs are found in a reasonable distance range. More importantly, the occupancy values of atoms in the “II” location are also higher. These results indicate that the “correct” structure of PDB 1GJ7 should be in the “II” location, instead of the current position deposited in our calibration set. Nevertheless, both CFCF pairs in Figure 5 should be considered as structural noise of the calibration set, which corresponds to our expectation regarding the unphysical shape of CFCF potential at 2.6–2.8 Å.

3. Assessment of Binding Affinities. Table 1 gives the statistical details of using IPMF on a calibration and test set, and the corresponding scatter plots are shown in Figure 6. Clearly, IPMF presents high quality results for data fitting. For a calibration set, significant correlation was observed between the calculated scores and the experimentally determined binding affinities with $R_p^2 = 0.64$. For a test set, IPMF gives a fairly respectable correlation coefficient of 0.40. This means that IPMF can reproduce the known binding affinities of the entire test set with a standard deviation (SD) of 1.76 log units, corresponding to 2.42 kcal/mol in terms of binding free energy at room temperature. From Figure 6, it is also evident that the quality of the prediction results is acceptable. No obvious systematic tendency to under- or overpredict was found for data points of a test set, and outliers are uniformly distributed.

It would be interesting to investigate the prediction performance of IPMF on different protein classes. To yield meaningful

Table 1. Correlations between Binding Scores and Experimentally Determined Binding Affinities^a

data set	scoring function	R_p^2	SD (log K_i/K_d)	ME (log K_i/K_d)	a	b	R_s^2
calibration set	IPMF	0.64	1.17	0.91	1.9472	−0.0159	0.65
test set	IPMF	0.40	1.76	1.41	2.5763	−0.0139	0.43
	VINA	0.29	1.91	1.54	3.1712	−0.4478	0.34
	PLP	0.26	1.95	1.55	3.6870	−0.0268	0.30
	AD4	0.23	1.99	1.60	4.4135	−0.3017	0.27
	LUDI	0.22	2.01	1.60	3.9500	−0.0037	0.24
	GLIDE	0.16	2.13	1.71	4.0634	−0.3467	0.19
	MW	0.15	2.10	1.69	4.7046	0.0048	0.20
	PMF	0.15	2.09	1.68	4.9323	−0.0128	0.15
	PMF04	0.03	2.24	1.81	6.0600	−0.0073	0.02

^aFor the test set, IPMF is compared with different scoring functions.

correlation statistics, we first clustered the complexes according to their sequence similarity, with the threshold set to 30%. Then, the resultant correlation coefficients of clusters with at least 6 members were calculated. As summarized in Table 2, the R_p^2 values for the 11 protein clusters vary significantly, ranging from 0.16 to 0.89. Scatter plots showing the relationship between IPMF scores and observed binding affinities are in the Supporting Information. On one side, these results may suggest that IPMF potentials have different performance on different protein families, which is similar to the situation of PMF04. On the other hand, we must treat these values with caution because that correlation calculated from a small subset may not completely characterize its relationship. Among these subsets, it is not surprising to find that cluster 5 with 9 examples only gave a low R_p^2 value of 0.17. The inferior performance is ascribed to metal-containing proteins of this subset, which confirms our supposition on IPMF potentials from Figure 3. However, the exact reason behind the variable prediction performance of IPMF or PMF04 is still not clear. In Table 2, we also listed the number of homologous proteins presented in the calibration set for each subset. There is no relationship observed between the number of homologous examples used for training and R_p^2 for testing, indicating that the prediction performance of IPMF is not biased by the sampling size of protein family. Recently, Huang and Zou reported a similar finding with a different experiment, in which the training set complexes homologous to their test set complexes were excluded, and potentials were rederived. No significant difference was found between the correlations of the two versions of potentials.²⁵

For the test set, the correlation coefficient for minimized complexes, $R_p'^2$, is also calculated, following the same procedure applied to the calibration set. Satisfactorily, IPMF potentials give an even higher correlation coefficient ($R_p'^2 = 0.41$) to the minimized complexes. It suggests that IPMF does not rely on the exact coordinates of input structures to yield accurate prediction results. Meanwhile, it is interesting to investigate why the drop from R_p^2 to $R_p'^2$ for the training set is not observed for the test set. For this question, we again focus on the case of CFcF potential and its abrupt jumps at 2.6–2.8 Å. Because of the exceptional atom distance, it is expected that this variation in potential would have little effect on the rescoring of binding poses. In our test set, there are no CFcF atom pairs in the distance range, and thus, no complexes in the test set benefited from the variations in the potential values. Given the consistent results of R_p^2 and $R_p'^2$ for testing, we can see from the example of CFcF potential that

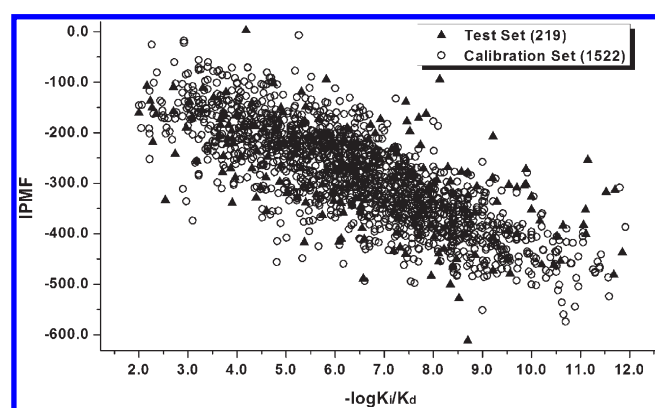


Figure 6. Scatter plots showing the calculated IPMF score as a function of the observed binding affinity. (○) Data points of calibration set ($R_p^2 = 0.64$). (▲) Data points of test set ($R_p^2 = 0.40$).

the predictive performance of IPMF is not arising from overfitting to the training set. Instead, it bears general information and describes underlying molecular interactions.

For comparison, Table 1 summarizes the statistics of other scoring functions on the test set. Detailed prediction results are provided in the Supporting Information. It is noteworthy that there are three different versions of LUDI and two versions of PLP implemented in DS 2.5, i.e., LUDI1, LUDI2, LUDI3, PLP1, and PLP2. We have tested all of these version scores in our study and have found that LUDI3 and PLP1 generate higher R_p^2 values than other versions of these two scores. To avoid confusion, in this paper, we only report the results of LUDI3 and PLP1 and use them to represent LUDI and PLP throughout this paper. Another note is that GLIDE failed to yield reasonable binding scores for some protein–ligand complexes. There are 38 complexes that did not pass the Coulomb–vdW, metal, or H-bond energy filters, which were disregarded on measuring the performance of the GLIDE score.

The empirical- or force field-based scoring functions, i.e., VINA, PLP, AD4, LUDI, and GLIDE, demonstrate observable correlations ($R_p^2 = 0.16–0.29$) between their binding scores and the experimentally determined binding affinities for the entire 219 protein–ligand complexes. The knowledge-based scores PMF04 and PMF, however, show weak or basically no correlation between their scores and the experimental binding affinities with very low R_p^2 values (0.03–0.15). Interestingly, if we use the

Table 2. Correlation between Binding Scores and Experimentally Determined Binding Affinities for Subsets of the Test Set

cluster	no. of complexes	protein family	No. of homologous proteins in calibration set		R_p^2
			ST ^a = 60%	ST = 90%	
1	6	penicillopepsin endothiapepsin	15	15	0.24
2	6	beta-glucosidase A beta-galactosidase	39	32	0.89
3	15	HIV-1 protease	171	171	0.47
4	18	trypsinogen coagulation factor X coagulation factor VII coagulation factor VIIA trypsin beta urokinase elastase thrombin	191	182	0.50
5	9	macrophage metalloelastase stromelysin-1 neutrophil collagenase	16	11	0.17
6	6	heat shock protein 82 heat shock protein 90 heat shock protein HSP90- α	12	10	0.43
7	6	antibody FAB fragment	29	10	0.16
8	6	N-methyl-D-aspartate receptor subunit 1 glutamate receptor 2	12	10	0.63
9	6	thymidylate synthase	7	7	0.47
10	6	acetylcholinesterase	6	6	0.67
11	6	casein kinase II CDK2	51	46	0.36

^a ST is the abbreviation for similarity threshold.

molecular weight (MW) of ligands as descriptors to predict the binding affinities, we may also observe a weak inherent correlation ($R_p^2 = 0.15$) between them. It has been reported that there exists a moderate correlation between MW and experimental binding data.⁴³ For Wang et al.'s data set, a data set containing 100 complexes that has been extensively used for scoring function comparison, MW outperforms many well-known scoring functions.⁴⁴ In the current study, a similar situation was encountered, suggesting that the composition of the evaluated data set deserves our attention. As pointed out by Velec et al., data sets showing correlation with respect to MW is not a trivial issue because scoring functions based on such data may prefer large ligands.⁴⁴ Having the same binding affinity, a larger ligand usually means less room for structure optimization. Therefore, those size-dependent scoring functions will fail at many applications such as fragment-based drug design or virtual screening. To relieve such a concern for IPMF, we calculated the MW of ligands in our calibration set and found that no correlation exists between MW and binding affinities ($R_p^2 = 0.046$).

As mentioned above, molecular docking and scoring are used in different drug discovery stages. For virtual screening, the correlation between their calculated scores and the experimental binding affinities need not necessarily be linear. A scoring function providing the correct rankings of candidate molecules is sufficient. In this regard, the Spearman's rank order correlation

coefficient is a more appropriate measure for the predictive power of scores. As given in Table 1, R_s^2 of IPMF is also significantly higher than other scoring functions, indicating that IPMF has a reasonable level of competence in ranking ligands.

It is evident that IPMF potentials show superior results for the current test set. However, it could be argued that IPMF's performance is artificially enhanced by its training set being related to the test set by the nonredundant sampling procedure, as explained in the first part of the Method section. To address this concern, two more external test sets were used to test the success of our adaptation to PMF04: One is Wang's set⁴⁵, and the other is the test set of ITScore.²⁵ These two sets are relatively large and contain diverse protein–ligand complexes. For both data sets, all complexes overlapped with our training set structures are removed. The resulting complex number and R_p^2 of IPMF and PMF04 are summarized in Table 3. Clearly, IPMF potentials show improved performance in binding affinity prediction.

4. Reranking of Binding Poses. As described above, for each protein–ligand complex in the test set, a conformation ensemble, including up to 11 discrete binding poses, was generated to measure the reranking capability of scoring functions. Using IPMF, the experimentally determined ligand geometries of 42% of all cases, i.e., 92 of the 219 evaluated complexes, are ranked best out of the decoy set. Overall, for 78.1% of all cases, the crystal structures are found in the first three ranks, in which 25.6% are in

Table 3. Correlation Coefficients of IPMF and PMF04 for Two External Data Sets

data set	no. of complexes	scoring function	R_p^2
ITScore set	100 (140) ^a	IPMF	0.37
		PMF04	0.24
Wang's set	79 (100) ^a	IPMF	0.34
		PMF04	0.21

^a The number in parentheses is the original size of the data set without excluding overlapping structures with our calibration set.

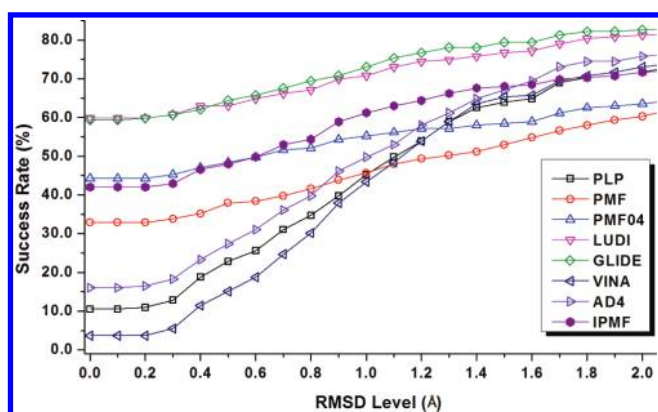


Figure 7. Cumulative plots showing the change of success rates at different rmsd levels for IPMF and other seven other scoring functions on 219 complexes of the test set.

the second rank and 10.5% are in the third. Among the eight scoring functions evaluated here, this recognition rate ranks third, behind LUDI (87.7%) and GLIDE (80.5%) and before AD4 (72.1%).

A similar analysis has been reported previously. Velec et al. evaluated 13 scoring functions on their success rates of identifying the crystal structures.⁴⁴ By using their derived potentials, DrugScore^{CSD}, the success rate at 0.0 Å rmsd level reaches 77%, much higher than those of IPMF and other scores evaluated in our study. However, as pointed out by the authors, we should note that the performance of a scoring function sometimes depends on the composition of the evaluated data set. In their work, a simple Lennard–Jones 12–6 potential yields a high success rate of 57% on recognizing the native ligand poses, suggesting that steric complementarity of complexes in their data set may be significant. Similarly, Ferrara et al. found that with respect to their data set, the steric complementarity between the receptor and its natural ligand is crucial for recognizing near-native structures.⁴³ To examine whether our data set has this issue, we followed Velec's approach to rescore all ligand poses with only short-range interaction terms of AD4. Interestingly, the success rate at 0.0 Å rmsd level is just 11%, significantly inferior to the rate of most scoring functions. This result indicates that binding pose reranking performances obtained in this study are less biased by the steric complementarity issue of data set. Another support for this argument is that PLP score in Velec's work gives a high success rate of 52%, which drops to 10.5% in our study.

Our reranking protocol also allows for the test of retrieving the near native pose out of the same pool of decoys. As depicted in Figure 7, setting the threshold for a valid structure to a rmsd value of 2.0 Å, we obtain that the success rates of IPMF, PMF04, and PMF are 71.7%, 63.5%, and 60.3%, respectively. This result

suggests that the application of iteration yields an improvement of 8.2–11.4% with respect to the identification of native or “well-docked” ligand pose compared to the original versions of PMF. We ascribe such an improvement to a better characterizing of atom pair interactions between protein and ligands. Meanwhile, we must also be aware that the binding pose reranking ability of IPMF is not the best among the scoring functions evaluated here. As shown in Figure 7, the success rates of GLIDE and LUDI are consistently higher than that of IPMF at every rmsd level. In our point of view, the superiority of GLIDE and LUDI is due to their adequate representation of hydrogen bond contributions in molecular interaction. As we know, the hydrogen bond is one of the most important specific interactions in biological recognition processes. The geometry of the hydrogen bond follows strict rules, either in terms of distances or angles between neighboring donor and acceptor atoms. However, the current atom pair potentials neglect the angular preference of hydrogen bonds. Additionally, recent research has suggested that a better description of the hydrophobic effect and solvation differences is crucial to scoring function improvements, which also requires consideration of directionality that cannot be accounted for in the current model.⁴⁶ To address the issue, we find that future endeavor should focus on the incorporation of directional effects of atom interaction pairs.

CONCLUSIONS

On the basis of the concept and formalism of PMF and a novel iteration method, a knowledge-based scoring function, named IPMF, has been developed. The iteration method designed here is to extract the 3-D structural information as well as the binding affinity information to yield an “enriched” knowledge-based model. Compared to a third-party test set containing 219 diverse protein–ligand complexes to seven commonly used scoring functions, the IPMF score performs best in the activity prediction test. In the test of reranking binding poses, IPMF demonstrated minor improvements over other evaluated knowledge-based functions and showed an inferior performance with respect to the other scoring functions. These results suggest that the additional binding affinity information cannot only be used as a part of the knowledge base for developing scoring functions, but also can be of help in improving their binding affinity prediction. In addition to potential usages in drug design, our IPMF approach also provides a well-defined scheme to gradually add binding information into a set of typical statistical potentials, which is applicable to other knowledge-based scoring functions.

ASSOCIATED CONTENT

S Supporting Information. A table listing PDB codes of all test set structures, experimental affinity data, and scores of IPMF and seven scoring functions under test. Scatter plots to show relationships between the calculated IPMF score and observed binding affinity for 11 subsets listed in Table 2. Python codes for assigning the protein and ligand atom types in the IPMF scoring function and calculating IPMF scores. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Telephone: 86-21-50271399; fax: +86-21-50807088; e-mail: myzheng@mail.shcnc.ac.cn (M.Z.) or hljiang@mail.shcnc.ac.cn (H.J.).

Author Contributions

[†]These authors contributed equally to this work.

ACKNOWLEDGMENT

The authors thank Dr. Renxiao Wang and Dr. Xiaoqin Zou for their help in data set compilation. The authors gratefully acknowledge financial support from the State Key Program of Basic Research of China (Grant 2009CB918502), Hi-tech Research and Development Program of China (Grants 2006AA020404), and the National Natural Science Foundation of China (Grants 81001399).

REFERENCES

- (1) Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. TarFisDock: A web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **2006**, *34*, W219–224.
- (2) Muller, P.; Lena, G.; Boilard, E.; Bezzine, S.; Lambeau, G.; Guichard, G.; Rognan, D. In silico-guided target identification of a scaffold-focused library: 1,3,5-Triazepan-2,6-diones as novel phospholipase A2 inhibitors. *J. Med. Chem.* **2006**, *49*, 6768–6778.
- (3) Kellenberger, E.; Foata, N.; Rognan, D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: Methods and problems. *J. Chem. Inf. Model.* **2008**, *48*, 1014–1025.
- (4) León, D.; Markel, S. In Silico Technologies in Drug Target Identification and Validation. In *Drug Discovery Series*; Carmen, A., Ed.; CRC Press: Boca Raton, FL, 2006; Vol. 6, pp 1–480.
- (5) Lipinski, C. A. Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discovery Today* **2003**, *8*, 12–16.
- (6) Kitchen, D. B.; Decornet, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
- (7) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
- (8) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153* (Supplement 1), S7–26.
- (9) Jain, A. N. Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.* **2006**, *7*, 407–420.
- (10) Velec, H. F.; 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.
- (11) Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895–5902.
- (12) Yang, C. Y.; Wang, R.; Wang, S. M-score: A knowledge-based potential scoring function accounting for protein atom mobility. *J. Med. Chem.* **2006**, *49*, 5903–5911.
- (13) Zhao, X.; Liu, X.; Wang, Y.; Chen, Z.; Kang, L.; Zhang, H.; Luo, X.; Zhu, W.; Chen, K.; Li, H.; Wang, X.; Jiang, H. An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA, and RNA. *J. Chem. Inf. Model.* **2008**, *48*, 1438–1447.
- (14) DeWitte, R. S.; Shakhnovich, E. I. SMOG: de Novo design method based on simple, fast, and accurate free energy estimates 0.1. Methodology and supporting evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733–11744.
- (15) 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.
- (16) 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.
- (17) Muegge, I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Perspect. Drug Discovery* **2000**, *20*, 99–114.
- (18) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- (19) Ishchenko, A. V.; Shakhnovich, E. I. SMOG2001: An improved knowledge-based scoring function for protein–ligand interactions. *J. Med. Chem.* **2002**, *45*, 2770–2780.
- (20) 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.
- (21) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (22) Cheng, T. J.; Li, X.; Li, Y.; Liu, Z. H.; Wang, R. X. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.
- (23) Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.
- (24) Huang, S. Y.; Zou, X. Q. Inclusion of solvation and entropy in the knowledge-based scoring function for protein–ligand interactions. *J. Chem. Inf. Model.* **2010**, *50*, 262–273.
- (25) Huang, S. Y.; Zou, X. Q. An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials. *J. Comput. Chem.* **2006**, *27*, 1866–1875.
- (26) Zhao, X. Y.; Liu, X. F.; Wang, Y. Y.; Chen, Z.; Kang, L.; Zhang, H. L.; Luo, X. M.; Zhu, W. L.; Chen, K. X.; Li, H. L.; Wang, X. C.; Jiang, H. L. An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA, and RNA. *J. Chem. Inf. Model.* **2008**, *48*, 1438–1447.
- (27) Xue, M.; Zheng, M.; Xiong, B.; Li, Y.; Jiang, H.; Shen, J. Knowledge-based scoring functions in drug design. I. Developing a target-specific method for kinase–ligand interactions. *J. Chem. Inf. Model.* **2010**, *50*, 1378–1386.
- (28) Wang, R. X.; Lu, Y. P.; Fang, X. L.; Wang, S. M. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein–ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114–2125.
- (29) Wang, R. X.; Fang, X. L.; Lu, Y. P.; Yang, C. Y.; Wang, S. M. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.
- (30) Wang, R. X. A Brief Introduction to the PDBbind Database, version 2009. <http://www.pdbbind.org.cn> (accessed October 2010).
- (31) Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L. The Blue Obelisk: Interoperability in chemical informatics. *J. Chem. Inf. Model.* **2006**, *46*, 991–998.
- (32) The Open Babel Package, version 2.2.3. <http://openbabel.sourceforge.net> (accessed October 2010).
- (33) Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem.* **2001**, *22*, 418–425.
- (34) Almaraz, N. G.; Lomba, E. Determination of the interaction potential from the pair distribution function: an inverse Monte Carlo technique. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2003**, *68*, 011202.
- (35) SZYBKI, version 1.3; OpenEye Scientific Software: Santa Fe, NM, 1998.
- (36) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: Conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317–324.
- (37) Bohm, 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.

- (38) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (39) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.
- (40) 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.
- (41) Trott, O.; Olson, A. J. Software news and update 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.
- (42) 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.
- (43) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.
- (44) 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.
- (45) Wang, R. X.; Lu, Y. P.; Wang, S. M. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
- (46) Abel, R.; Wang, L.; Friesner, R. A.; Berne, B. J. A displaced-solvent functional analysis of model hydrophobic enclosures. *J. Chem. Theory Comput.* **2010**, *6*, 2924–2934.