Knowledge-Based Scoring Functions in Drug Design. 1. Developing a Target-Specific Method for Kinase−Ligand Interactions

Mengzhu Xue,† Mingyue Zheng,† Bing Xiong,* Yanlian Li, Hualiang Jiang, and Jingkang Shen*

State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai, China 201203

Received May 10, 2010

Protein kinases are attractive targets for therapeutic interventions in many diseases. Due to their importance in drug discovery, a kinase family-specific potential of mean force (PMF) scoring function, kinase-PMF, was developed to assess the binding of ATP-competitive kinase inhibitors. It is hypothesized that target-specific PMF scoring functions may achieve increased performance in scoring along with the growth of the PDB database. The kinase-PMF inherits the functions and atom types in PMF04 and uses a kinase data set of 872 complexes to derive the potentials. The performance of kinase-PMF was evaluated with an external test set containing 128 kinase crystal structures. We compared it with eight scoring functions commonly used in computer-aided drug design, either in terms of the retrieval rate of retrieving "right" conformations or a virtual screening study. The evaluation results clearly demonstrate that a target-specific scoring function is a promising way to improve prediction power in structure-based drug design compared with other general scoring functions. To provide this rescoring service for researchers, a publicly accessible Web site was established at http://202.127.30.184:8080/scoring/index.jsp.

INTRODUCTION

Computer-aided drug design has become a common tool for drug discovery.[1−4] Recently, structure-based drug design utilizing macromolecular three-dimensional structure information has demonstrated its high efficiency not only in discovering novel compounds but also in a cost-saving manner.[5,6] Among these computational methods, molecular docking is a major player for predicting protein−ligand interactions and is used extensively in lead discovery and optimization.[7] Basically, docking involves a conformational sampling procedure to generate conformations in the binding pocket and a scoring phase to rank these conformations. With the continuing increase in available computational resources, conformational sampling is no longer a significant obstacle.[8−10] Although positioning the important waters and the flexibility of receptors are still hard to contend with, basically, accurate scoring poses a major challenge to the success of molecular docking.

The common scoring methods in molecular docking can be roughly grouped into three types: force-field methods, empirical scoring functions, and knowledge-based potentials. The force-field-based functions aim to describe protein−ligand interactions using elementary physical forces and should be accurate in theory. However, due to the subtle nature of the intermolecular forces, many of them can only approximately obtain the enthalpic interaction energy. For the entropic contribution of changes in solvation and conformational mobility, one needs to adopt a more computationally intensive method to accurately estimate the change in free energy. For example, in the recently released UCSF DOCK software, the Generalized-Born/Surface-Area method (GBSA) was implemented in addition to the original VDW and electrostatics energy term. This method shows improved scoring accuracy at the expense of significantly increased computational time.[11] Empirical scoring functions, such as the Xscore function,[12] are commonly trained against a set of protein−ligand crystal structures and are constructed by linear regression of interaction contributions to binding affinities. Clearly, these methods strongly rely on the accuracy of the binding affinity data,[13] which usually bear large uncertainties. Finally, the knowledge-based potentials are developed from a statistical analysis of protein−ligand complexes,[14−16] rooted from Sippl's reverse Bolztmann's potential construction method.[17] The potential of mean force (PMF) scoring converts structural information into free energies without any knowledge of binding affinities and thus is expected to be more applicable. PMF scoring implicitly balances many opposing contributions to binding including solvation effects, conformational entropy, and interaction enthalpy.[14] This is an ingenious treatment because calculating these terms explicitly often results in large errors.

However, for a given target class, performance of the above-mentioned generic scorings is always unsatisfactory. The development of scoring functions for specific target classes may provide a way to increase accuracy in the prediction of binding interactions. In the present work, we will report a trial of this strategy and demonstrate its use with a protein kinase system.

Protein kinase is one of the largest target families in the human genome. The family's key function is to mediate most of the signal transduction in eukaryotic cells. By doing this, these kinases control many important cellular processes in metabolism, transcription, apoptosis, and differentiation. The family therefore represents a very attractive target class for

* Corresponding authors. E-mail: jkshen@mail.shcnc.ac.cn (J.S.); bxiong@mail.shcnc.ac.cn (B.X.).
† Authors contributed equally to this work.

10.1021/ci100182c
Published on Web 08/03/2010

therapeutic interventions in many disease states such as cancer, diabetes, inflammation, and arthritis.[18]

Although protein kinases differ greatly in their cellular functions, they share a conserved catalytic domain which phosphorylates the substrate by consuming ATP. Due to the small molecular binding nature of the ATP-binding site, it has been the main focus of inhibitor design.[19,20] Most kinase inhibitors that have been developed are ATP-competitive, such as the approved drugs Iressa, Tarceva, and Gleevec. These molecules mimic the binding of the adenine ring of ATP through forming hydrogen bonds with the hinge of the kinases or hydrophobic interactions around the adenine-binding region.[21]

Given the importance of the protein kinase family, much research has been conducted to study them using biochemical, biophysical, structural biological, and genomic sequencing methods. Data available from these investigations has proliferated.[22,23] In 2002, Manning and co-workers[24] identified and catalogued protein kinases in the human genome (the "kinome"). Furthermore, structural biological efforts have resulted in the deposit of thousands of kinase crystal structures in the PDB database, which is a valuable resource for drug development.

Due to the specificity of the structural basis of the kinases' ATP binding sites and the conservation of the catalytic kinase domains, a kinase-specific PMF function was developed to treat the scoring of ATP-competitive inhibitors, following the supposition that specific scoring functions may have better applicability for certain target families. The results presented here clearly demonstrate that this is very promising, especially with a continuing increase in the number of solved crystal structures.

MATERIALS AND METHODS

Construction of Kinase-Specific PMF Scoring Function. *Data Set.* To compile the kinase crystal structure data set from the PDB database, a FASTA format file containing the sequences of 516 human kinase domains was downloaded from the Kinome Web site.[25] A locally implemented BLAST program was then run against the whole PDB sequences (downloaded from PISCES server[26]) using the 516 kinase sequences as the queries. The results from BLAST searching were further analyzed with an in-house python script to get the protein crystal structures with E values lower than 10^{-60}. The resulting PDB structures were considered as representatives of kinase. These structures were then downloaded from the PDB Web site. Finally, only structures containing HETATM ligands in the ATP binding sites and of resolution better than 3.0 Å were retained, resulting in 1000 complexes. These crystal structures were split into two data sets, one for deriving PMF statistical potentials (872 complexes) and the other for evaluating the resulted kinase-PMF scoring function (128 complexes).

Scoring Function. The PMF scoring method was introduced by Muegge and Martin in 1999.[14] After verification[27] and further amendment,[28] it represents a useful scoring function in structure-based drug design. The PMF scoring function is defined as the sum over all protein−ligand atom pair interaction potentials $A_{ij}(r)$ at distance r (eq 1) of protein atom type i and ligand atom type j. r^{ij}_{cutoff} is the distance at which atom pair interactions are truncated.

$$\text{PMF\_score} = \sum_{\substack{kl \\ r<r^{ij}_{\text{cutoff}}}} A_{ij}(r) \tag{1}$$

The $A_{ij}(r)$ values are calculated as

$$A_{ij}(r) = -k_{\text{B}}T \ln\left[f^{j}_{\text{Vol\_corr}}(r)\frac{\rho^{ij}_{\text{seg}}(r)}{\rho^{ij}_{\text{bulk}}}\right] \tag{2}$$

where k_{B} is the Boltzmann factor, T is the absolute temperature, and $f^{j}_{\text{Vol\_corr}}(r)$ is a ligand volume correction factor. $\rho^{ij}_{\text{seg}}(r)$ is the number density of atom pairs of type ij at a certain atom pair distance r. $\rho^{ij}_{\text{bulk}}(r)$ is the number density of a protein−ligand atom pair of type ij in a reference sphere with a radius of 12 Å. For convenience of calculation, a set of spherical shells was defined called segments seg(r) with thickness m consecutively separating the sphere with radius R (12 Å) into R/m segments. m was chosen to be 0.2 Å.

Specially, the ligand volume correction factor was introduced to normalize the PMF according to the available solvent/protein volume by ignoring the ligand−ligand interactions.[27] 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 were defined as follows:

$$\hat{\rho}^{ij}(r) = \rho^{ij}(r)\frac{\rho^{kj}(r)}{\rho^{kj}(r) + \rho^{lj}(r)} \tag{3}$$

$$\hat{\rho}^{ij}_{\text{bulk}}(r) = \rho^{ij}_{\text{bulk}}(r)\frac{\rho^{kj}_{\text{bulk}}(r)}{\rho^{kj}_{\text{bulk}}(r) + \rho^{lj}_{\text{bulk}}(r)} \tag{4}$$

$$f^{j}_{\text{Vol\_corr}}(r) = \frac{\hat{\rho}^{ij}(r)\rho^{ij}_{\text{bulk}}(r)}{\hat{\rho}^{ij}_{\text{bulk}}(r)\rho^{ij}(r)} \tag{5}$$

where $\rho^{kj}(r)$ designates the number densities of protein atoms k of any type around a ligand atom of type j at distance r. $\rho^{lj}(r)$ designates the number densities of ligand atom l of any type around a ligand atom type j at distance r.

Compared to PMF99,[14] PMF04[28] introduces new atom types both for the protein and ligand. So to comparatively evaluate the kinase-specific PMF scoring function, we adopted the same developing method and atom types used in the original PMF04 paper to implement our program. In total, 17 protein atom types and 34 ligand atom types were used to make the atom pairs and calculate the occurrences in each distance shell from 872 kinase structures. For the protein atom typing, a python script was used to identify each atom in 20 common amino acids and then assign a PMF receptor type accordingly. On the basis of the open source C++ programming library OpenBabel,[29] an in-house program was implemented to recognize the ligand atom bonding information and assign the atom types according to the ligand types described in PMF04 (see the Supporting Information for atom typing Python script). Then, a series of programs were developed to perform the following calculations: atom pair occurrences within a certain distance shell, ligand volume correction factor values, and finally atom pair potentials. Importantly, as suggested by Muegge, if the total occurrence of atom pairs in a 12 Å radius was less than 1000, it was discarded for later $A_{ij}(r)$ calculation. For comparison,

PMF04 scoring was also implemented locally by using the above-mentioned atom typing program but with Muegge's calculated atom pair potential data supplied in the original paper.[28]

Evaluation of the Scoring Function. *Test Set.* To better verify the targeted kinase-PMF scoring function, 128 crystal structures were selected on the basis of the kinase phylogenetic trees and ligand diversity. They consist of 65 kinases picked out from different subtrees of the phylogenetic tree and 101 different ligands with rotatable bonds clustered into five groups. To generate more reliable conformations such as those commonly encountered in computational drug design, we chose Autodock4[30] as the conformation generator, one of the most commonly used tools in molecular docking. The parameters for docking are set as follows: npts="75,75,75", tstep=0.5, qstep=10.0, dstep=10.0, ga_pop_size=300, ga_num_evals=8000000, and ga_run=50. The original ligand conformation in the crystal structure and the MMFF force field minimized ligand conformation (refined in the protein binding site) were also included in feasible conformation sets, resulting in a total of 52 ligand conformations for each complex. The RMSDs of every conformation were then calculated against the original ligand conformation in the crystal structures for later analysis.

Comparison with Other Scoring Functions. For comparison, eight additional scoring functions were applied to the same test data set, including the scoring function implemented in the AutoDock4 program, Xscore, five scoring functions from the CScore module in Sybyl (version 7.3; F-Score,[31,32] G-Score,[33,34] D-Score,[35] PMF-Score,[14] and Chem-Score[36]), and a locally implemented PMF04. These scoring functions can be roughly grouped into three categories: (I) force-field-based methods, i.e., AutoDock4, G-Score, and D-Score; (II) empirical scoring functions, i.e., F-Score, Chem-Score, and Xscore; and (III) knowledge-based scoring functions, i.e., kinase-PMF, PMF04, and PMF-Score in Sybyl.

Every scoring function was evaluated with regard to how closely the lowest-energy conformation resembled the one observed in the crystal structure. Different RMSD levels (0.5, 1.0, 1.5, 2.0 Å) were defined to check the sensitivity, and then the occurrence of the lowest-energy conformation predicted by every scoring function was calculated at the 4 RMSD level. It was also examined by correlation between the calculated scores and the experimental binding affinity, by utilizing the commonly used square of Pearson's correlation coefficient (r^2) as the metric.[37]

Virtual Screening. In addition, a case study was conducted to further assess the utility of kinase-PMF in virtual screening. CDK2 was chosen as the target (PDB access number 1CKP[38]), and a ligand database was prepared to consist of 10 known CDK2 inhibitors collected from the literature and 990 compounds randomly selected from the Maybridge database.[39] In order to avoid biasing the virtual screening results, care was taken in selecting the compounds. Initially, the molecular weight and log P of compounds in the Maybridge database were calculated, and only compounds having these two properties similar to known CDK2 inhibitors were saved. Next, 990 compounds were randomly selected from this saved database. With consideration for the consumption of computational time, first, the OMEGA program[40] was chosen to generate ligand conformations. For every ligand, on average 75 conformations were saved; then, fast rigid exhaustive docking software, FRED,[41] was adopted to perform the virtual screening. The nine scoring functions mentioned above were applied to rescore the conformations generated by FRED. For each scoring function, the best scored conformation and the corresponding score were recorded for each compound in the data set, and these compounds were then ranked according to their recorded scores. Next, the number of known inhibitors found in the top 5% of ranked compounds was counted as a measure to evaluate the performance of a scoring function. Finally, the top 100 low-energy compounds rescored by kinase-PMF were extracted and clustered by the Pipeline Pilot Enterprise program[42] to compare their scaffold patterns with those of known CDK2 inhibitors.

Table 1. Averaged Occurrences of Selected Protein−Ligand Pairs in Kinase-PMF, PMF04, and PMF99[a]

ligand atom types	protein atom types			
	CP	ND	OA	CF
cF	636 377 181	283 192 90.5	295 170 90.5	479 265 143
cP	462 448 181	208 229 90.5	215 201 71.9	349 312 114
CP	184 556 360	81 289 181	84 250 181	135 361 227
OA	57 129 90.5	25 66.0 45.4	26 58.0 45.4	42 86.0 90.5
ND	103 35.0 36.0	46 18.0 18.0	48 16 18.0	79 23.0 28.6

[a] The atom pair occurrences in the data set of 872, 7152, and 697 complexes used to derive kinase-PMF, PMF04, and PMF99 have been scaled by the crystal structure number of each data set. The three numbers in each cell are for kinase-PMF, PMF04, and PMF99, respectively. The atom type definitions in the table are as follows: CP, polar aliphatic carbon; ND, nitrogen as hydrogen bond donor; OA, oxygen as hydrogen bond acceptor; CF, nonpolar aliphatic carbon; cF, nonpolar aromatic carbon; cP, polar aromatic carbon.

RESULTS

Comparison of PMF04 with Kinase-PMF. Initial analysis of the training data set shows that it contains about 133 unique kinase proteins in 872 complexes structures and 626 different ligand compounds. Table S1 (see the Supporting Information) shows the occurrences of atom pairs over 1000. The largest number of atom pairs in kinase-PMF is between the polar aliphatic carbon of the protein and the nonpolar aromatic carbon of the ligand (CP-cF), which is different from the most frequent atom pair in PMF99[14] and PMF04.[28] The occurrence of CP-cF is at the fourth position in both PMF04 and PMF99.

For a better comparison, the atom pairs which rank in the top six are listed in Table 1. They are scaled by the crystal structure number of each data set to eliminate the bias resulting from the size of the data set used in PMF construction. It can be inferred that, for ligands of kinases, aromatic structures may be important no matter whether these aromatic structures are polar or nonpolar (ligand atom type cP and cF); for kinases, the higher occurrence of aliphatic carbons (CP and CF) may imply the dominance of hydrophobic residues around ligands. In contrast, for general protein−ligand complexes, the occurrence rate of aromatic atoms is lower. It was also noted that two atom pair types involving hydrogen bonds are different from each other—a hydrogen bond acceptor oxygen in the kinase paired with a hydrogen bond donor nitrogen (OA-ND) in the ligand appears more frequently than the opposite pair (ND-OA).

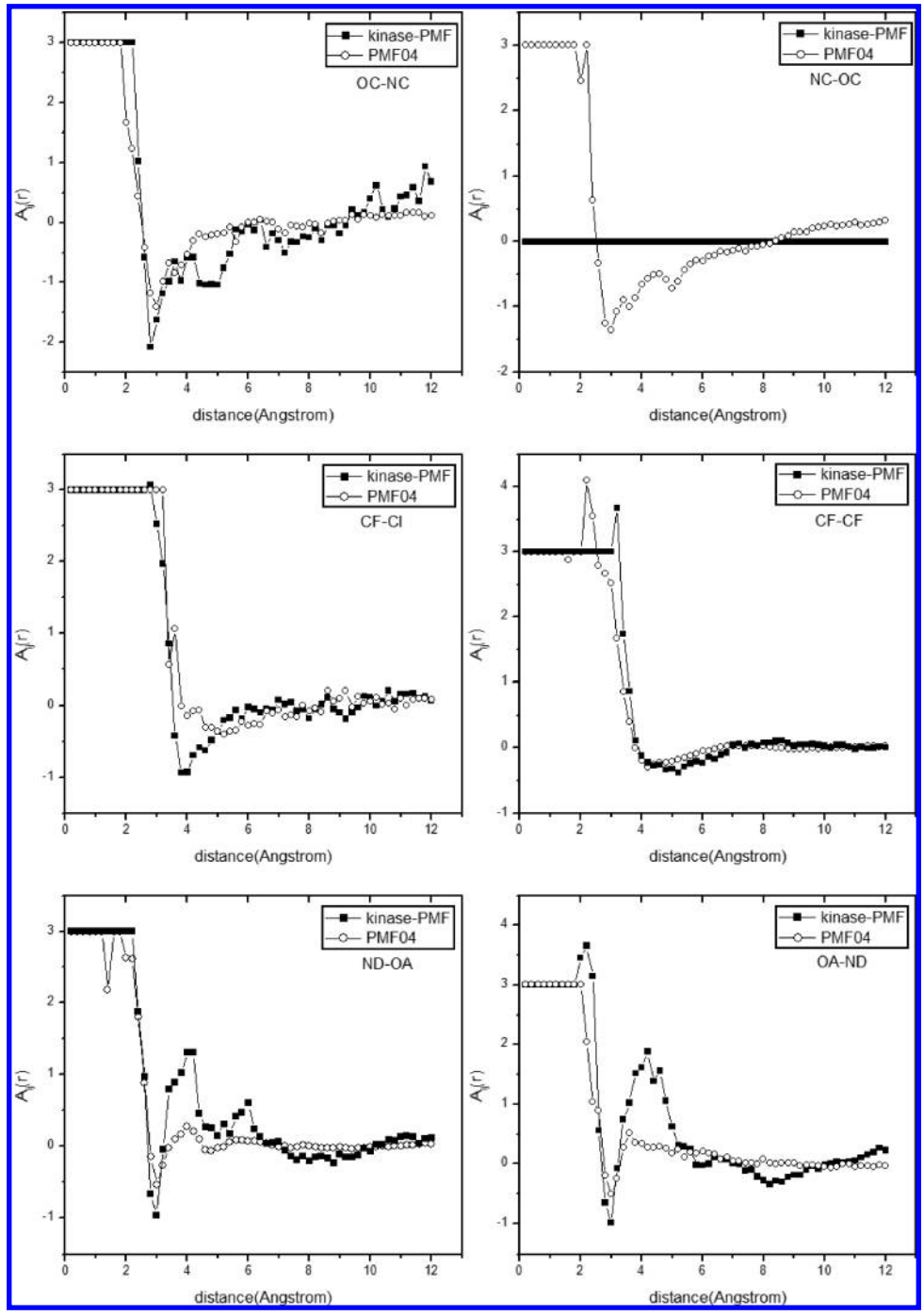

Figure 1. The potentials against atom pair distances of selected atom pairs by kinase-PMF and PMF04. The *x* axis is the atom pair distance, ranging from 0 to 12 Å, while the *y* axis is the potential of kinase-PMF and PMF04. The four letters on each graph denote the types of atom pairs. The first two indicate the protein atom type, and the last two indicate the ligand atom type. The bold squares represent kinase-PMF, and hollow circles represent PMF04.

Figure 1 shows the differences between the PMF04 and kinase-PMF potentials for a select set of protein−ligand atom pairs. The salt bridge between a charged nitrogen and charged oxygen looks very different in kinase-PMF and PMF04. From Figure 1B, the atom pair NC-OC in kinase-PMF occurs less than 1000 times, while this interaction type occurs over 1000 times in PMF04. From Figure 1A, the potential of the atom pair OC-NC of kinase-PMF diverges significantly from the one of PMF04 at 4.2−5.0 Å, which means at this distance range the occurrences are much higher in kinase−ligand systems. As shown in Figure 1C and D, the van der Waals interactions are more consistent in both general protein−ligand interactions and kinase−ligand interactions. For hydrogen bonding interactions between atom pair ND-OA or OA-ND, the curves are similar in shape, but kinase-PMF is more rugged. In these two atom pairs, they all show a repulsive peak around 4.0 Å. This may be due to the specific interactions of the kinase−inhibitor complexes rather than the small size of the data set used in the present study, as the averaged occurrence number of OA-ND is larger in the kinase−ligand data set (see Table 1).

The volume correction factor f_{Vol_corr} for various ligand atom types is illustrated in Figure 2. First, the polar aromatic carbon (cP) possesses the highest volume correction factor,

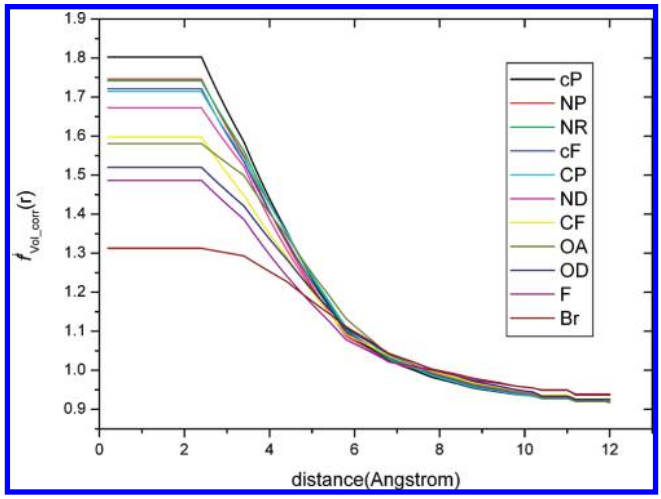

Figure 2. The ligand volume correction factor $f_{Vol\_corr}(r)$, shown as a function of distance (r) for the 12 ligand atom types that occur most frequently in the data set. The ligand atom types are listed in descending order starting with the one that has the highest volume correction factor: cP, NP, NR, cF, CP, ND, CF, OA, OD, F, and Br.

followed by the planar nitrogen (NP) and then the planar nitrogen in a ring structure (NR). The result is consistent with an important structural characteristic of most ATP-competitive kinase inhibitors, in which an aromatic nitrogen heterocycle commonly presents to mimic the ATP's adenine ring to form hydrogen bonds with the hinge of the kinase.[43] This is very different from the result in general protein−ligand interactions[27] in which the planar nitrogen (NP) has the highest volume correction factor. Second, the value of the hydrogen bond donor nitrogen (ND) is far higher than the hydrogen bond donor oxygen (OD), which is in common with the average occurrences analyzed above. Together, these results indicate that the ligand generally contains a hydrogen bond donor nitrogen, while the kinase provides the hydrogen bond acceptor, usually oxygen. Finally, the low volume correction factor of atom types F and Br may result from their low occurrences in the structures of kinase inhibitors compared to PMF,[27] where F has a high volume correction factor.

To investigate the overall scoring function, the 1000 crystal structures used in the present study were also used to calculate the energy scores by kinase-PMF and PMF04. As illustrated in Figure 3, they also differ greatly. The mean and SD values for kinase-PMF are −211.825 and 54.166 and for PMF04 are −42.826 and 22.819, revealing that the distribution of scores by kinase-PMF is widely dispersed. From the graph, it can be seen that scores by PMF04 mostly concentrate in the range of −125 to 50, while those by kinase-PMF are dispersed in a wide range from −400 to 50. One can infer that kinase-PMF is more sensitive for the identification of distinct complexes, which may imply a better prediction for active conformation selection.

Ranking Conformations. A data set of 128 complexes was selected to generate decoy conformations. A total of 65 kinases in the test set were picked from different groups in the phylogenetic tree, and 101 unique ligands could be clustered into five groups (2, 4, 6, 8, 10) on the basis of their rotatable bond number. There are about 54 kinases also contained in the training set, although the complex structures are different. The introduction of diversity in both kinases

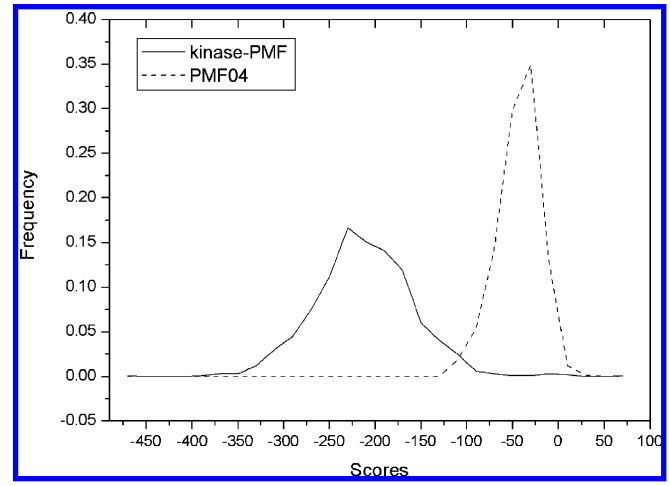

Figure 3. The distribution of scores in kinase-PMF and PMF04.

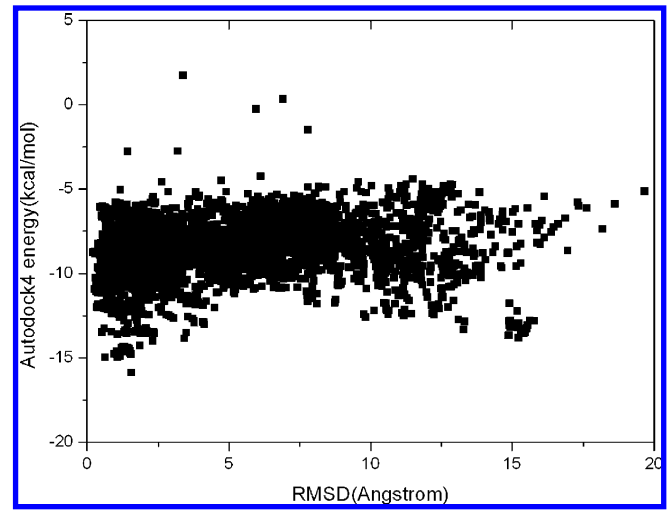

Figure 4. AutoDock4 binding free energy versus RMSD of the generated ligand conformations.

and ligands is expected to better represent the characteristics of the kinase−ligand interactions.

Generating reasonable conformations is a very important issue in testing the method. If the generated conformation is energetically unfavorable, it will be much easier to select the right one as expected. The AutoDock4 program is a commonly used tool in the field of structure-based drug design for predicting binding conformations. Here, we utilize AutoDock4 to perform docking studies on these complexes in order to generate more realistic ligand conformations in the kinase binding sites. As shown in Figure 4, the generated ligand conformations have a wide distribution with respect to RMSD and energy. Although for each ligand only 50 conformations were generated, these conformations are selected after a large-scale docking energy evaluation (800 0000 energy evaluations in each GA run; see the docking parameter in Materials and Methods section), most of which are of high quality in terms of predicted binding energy. It was shown in Figure 4 that there is no correlation between RMSD and AutoDock4 energy, and there are many conformations that bind very well to the kinases but have large RMSD values compared to the native conformations. This raised the challenge of picking the "right" conformation from these conformational decoys. As described in the Materials and Methods section, if the conformation with the best score has a lower RMSD value, the prediction is thought

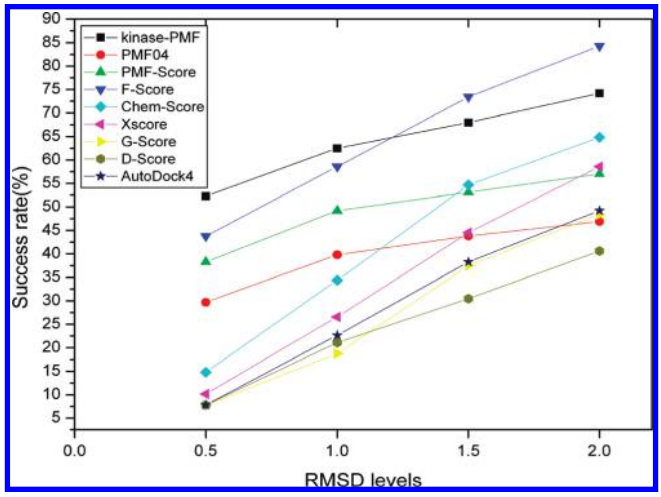

Figure 5. The occurrence of the best scored conformations at different RMSD levels (0.5, 1.0, 1.5, 2.0 Å), with the crystal conformations included in the statistics.

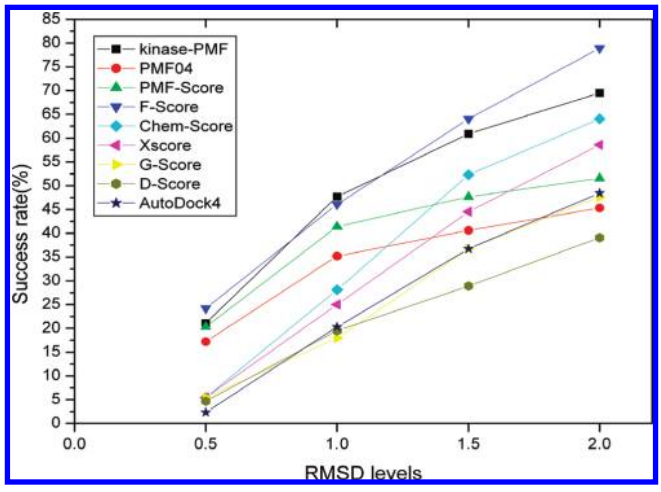

Figure 6. The occurrence of the best scored conformations at different RMSD levels (0.5, 1.0, 1.5, 2.0 Å), without the crystal conformations included in the statistics.

to be successful (at different levels: 0.5, 1.0, 1.5, 2.0 Å). The success rates for all of the nine scoring functions at different RMSD levels are illustrated by Figures 5 and 6, either including the ligand conformations of experimental complex structures or not.

From Figures 5 and 6, it was found that kinase-PMF and F-Score outperform the others at all four RMSD levels. However, different test sets may give different results. For example, Wang et al.[44] had reported the success rate of some of the scoring functions mentioned here. As listed in their paper, at RMSD level 2.0, the performance of test scoring functions was as follows: F-Score (74%), Xscore (66%), AutoDock3 (62%), PMF implemented in cerius2 (52%), G-Score (42%), Chem-Score (35%), and D-Score (26%).

In more detail, by including conformations from the 128 crystal structures in the data set, kinase-PMF can identify 67 conformations in which the RMSD values are less than or equal to 0.5 Å, followed by F-Score with a success number of 56. The other scoring functions behave more or less poorly at the level of 0.5 Å, meaning that they have an inferior ability to pick out the ligand conformation close to the one observed in experimental complex structures. At the 1.0 Å RMSD level, kinase-PMF is moderately better than the F-Score function. Beyond the RMSD 1.0 Å level, the F-Score function shows a good ability to identify the near native conformations. Nevertheless, it is clearly shown that the kinase-PMF function surpasses the PMF-Score implemented in SYBYL as well as the locally implemented PMF04.

When excluding crystal ligand conformations from the data set, F-Score, kinase-PMF, or PMF-Score show similar performances at the 0.5 Å level, and kinase-PMF is still superior to PMF04 or PMF-Score at all the RMSD levels. Comparing the PMF type scoring function with other scoring functions in Figures 5 and 6 at the 0.5 Å or 1.0 Å level, it can be seen that such structure statistical potential functions all have the ability to identify the near native conformations. This may imply that PMF scoring is more tolerant of close contact than current force fields and empirical-based scoring functions. Taken together, this clearly shows that kinase-PMF can identify more near-native conformations, even after excluding conformations observed in experimental complex structures.

To further evaluate the kinase-PMF function, another test of the correlation between experimental binding affinity data and calculated energy scores of the crystal conformations has been conducted. The K_i or K_d data of 55 kinase−inhibitor complexes were collected from several databases and literature sources[45] (Table S2 in the Supporting Information). The calculation shows that the correlations are all very low (see Supporting Information Table S3). Although this is consistent with results described in Muegge's work[28] and the work of Englebienne et al.,[37] it clearly points toward improving the scoring function (advances will be reported elsewhere).

In summary, scoring functions of the PMF type have better capability of differentiating the native conformations from decoys, while the accuracy of empirical scoring functions depends on the concrete forms of functions, and the tested force-field based methods generally perform poorly for this kinase data set.

Virtual Screening Study. As described in the Materials and Methods section, a total of 1000 molecules, including 10 known CDK2 inhibitors (the structures of known CDK2 inhibitors are listed in Table S4, Supporting Information), were docked into the CDK2 binding site to assess the enrichment of nine scoring functions. First, the OMEGA program was used to generate conformations for 1000 molecules, and 100 low-energy conformations were saved for each molecule. These conformations were then considered as different molecules and subjected to a FRED rigid docking study. A total of 35 000 low-energy conformations were output from FRED and saved for later rescoring. After rescoring by each scoring function, the best score of each compound was collected and ranked for analysis. In Figure 7, one can see that at the level of 1% (top 10 molecules), kinase-PMF can find 50% of known inhibitors, followed by PMF in Sybyl and FScore, both with 30% enrichment. At the level of 4−5%, kinase-PMF also identifies more known inhibitors and outperforms the other eight scoring functions.

A detailed analysis of the structural pattern of the top 100 low-energy compounds rescored by kinase-PMF was performed. The essential structural pattern of kinase ligands is that they usually consist of an aromatic ring which forms a conservative hydrophobic and hydrogen bond interaction with the hinge of CDK2.[38] By utilizing an in-house scaffold fragmentation program,[46] the scaffolds located at the hinge

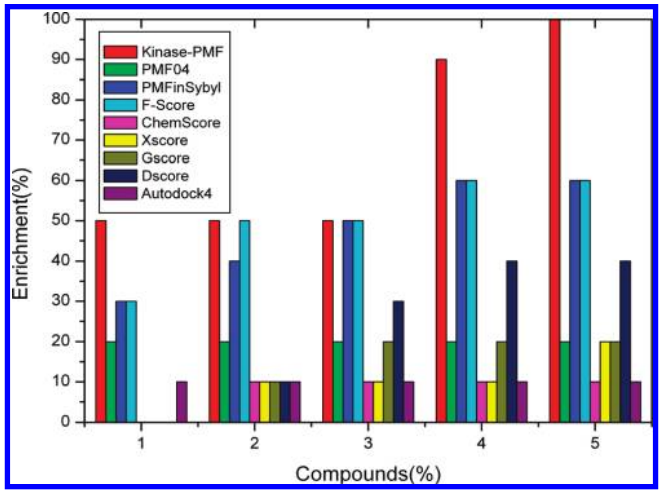

Figure 7. The number of known inhibitors found from the top scored compounds.

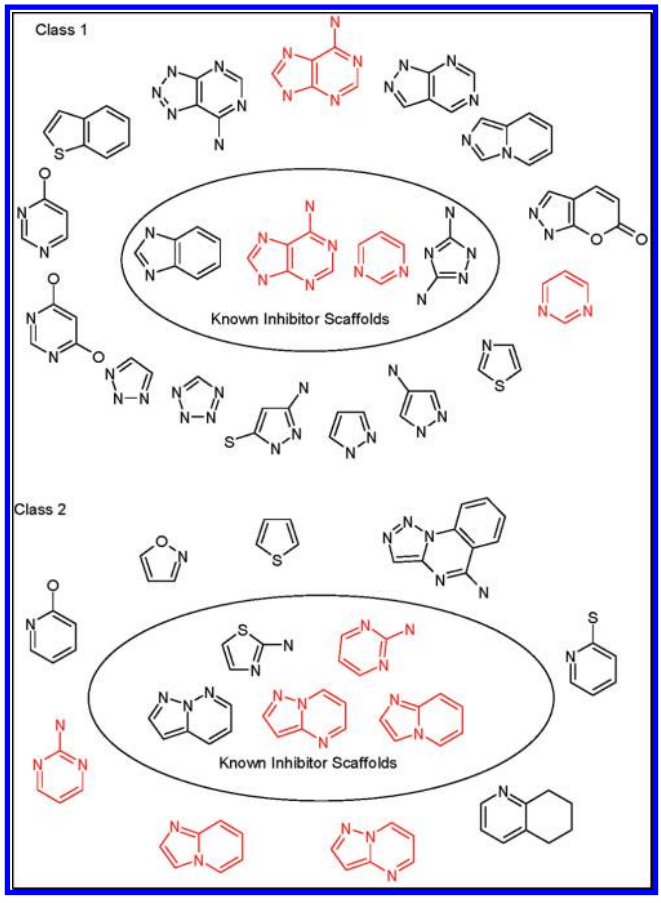

Figure 8. The scaffolds of known inhibitors and Maybridge compounds belonging to clusters 1 and 2. Scaffolds in the circle are from 10 known CDK2 inhibitors, while the scaffolds out of the circle are from 90 compounds in the Maybridge database. The scaffolds in red are the same as the scaffolds from known CDK2 inhibitors.

part were clustered into seven classes. From Figure 8, one can see that the scaffolds of known inhibitors can be clustered into two classes, all of which are structures of aromatic nitrogen-including heterocycles such as pyrimidine. The scaffolds from the top rescored Maybridge compounds belonging to these two classes are also similar, or even identical. Though scaffolds in classes 3 and 5 (the structures of other classes are available in Figure S1, Supporting Information) are almost all pyrimidinone or pyridazinone and their derivatives, it is likely that their carbonyl oxygen atom can form a hydrogen bond with CDK2. This structural characteristic is consistent with the ligand volume factor results and the characteristics of the interactions of kinase−inhibitor systems.

DISCUSSION

Target-Specific Scoring Function. The scoring function is an essential component in the computational prediction of protein−ligand interactions and in virtual screening. However, proteins belonging to different categories of structure or physiological function may be very different from each other. Kinases, for example, share a catalytic domain conserved in structure and sequence. Target-specific scoring methods have been developed with the hope of improving the predictability of certain target classes. Seifert[47] did a pioneer study on two kinases. By incorporating specific filtering rules into the scoring function, they found that the DUD data sets for FGFR1 and VEGFR2 had improved significantly (AUC of ROC from 0.46 to 0.86 and 0.66 to 0.77). The result clearly proves that target-specific scoring can improve the results of virtual screening on unique target families.

Target-specific scoring functions are commonly based on established scoring functions and either extended, recalibrated, or accompanied by special filters to achieve better performance. Breu et al.[48] applied partial-least-squares regression considering multimode binding and a variable influence as an extension to the current AFMoC approach and then validated it with a data set of 79 thrombin inhibitors. Compared to the conventional AFMoC, the correlation coefficient of the newly developed approach was improved from 0.57 to 0.61. In addition to this, Martin and Sullivan[49,50] developed AutoShim, an empirically parametrized target-specific scoring function created by adding pharmacophore-like interaction terms to other general scoring functions in which Flo+ in QXP was chosen. In this method, they used the program Magnet to add point-pharmacophore shims for filtering and training sets on activity data to weight the shims by partial least-squares regression and then scored with the Flo+ component of QXP. In test sets of seven different kinases, the predictive R^2 had been improved, with the best value for CHK1 from 0.12 to 0.57 and the worst for PIM1 from 0.01 to 0.21. In our work, PMF was chosen as the original scoring function, and a data set of 872 kinase−ligand complexes was used to reconstruct it. By using the nonoverlap kinase complexes as the test set, the success rate dramatically improved from 46.9% of PMF04 to 74.2%. This also demonstrates the potential of target-class approaches as a particularly promising method for selecting bioactive-like conformations.

One may be concerned that there is a limitation with the availability of large-scale crystal structures for the development of target-specific scoring functions. Because of the statistical nature of knowledge-based scoring functions, it is necessary to have a large sample data set to characterize the population. From the present study and PMF99, it was found that generally around 1000 crystal structures are sufficient for the construction of meaningful scoring functions with PMF-type methods.

Bioactive Conformation. When a ligand binds to a protein, it transits to a special conformation (induced-fit

mechanism) or shifts the conformational population to a bound form (population-shift mechanism). These are considered as bioactive conformations. The conformation in a crystal complex is recognized as a representative of the bioactive conformation. However, most docking software can only generate bioactive-like conformations in which the RMSD values are less than or equal to 2.0 Å. In the present study, the kinase-PMF scoring function was evaluated with two methods. One of these included the crystal structures in the ligand conformation set for each complex, while the other excluded them. By doing this, it was found that the PMF type scoring function is better at differentiating the right conformations with low RMSD values. This may imply that the crystal structures usually contain some very close interactions, which may result in a large positive energy in force field and empirical type scoring functions. Compared with other scoring functions, kinase-PMF demonstrates a superior ability to pick out near-native conformations. By including the ligand crystal structure conformations in conformational sets, the success rate of kinase-PMF at the 0.5 Å RMSD level can improve about 20−30% (from 21.1% to 52.3%) compared with PMF04 (from 17.2% to 29.7%) or PMF_Score in SYBYL (from 20.3% to 38.3%).

Correlation between Scores and Experimental Values. From the results of the binding affinity correlation test, one can see that the correlation of scores by each method with experimental values is generally low. This may result from the following causes: (I) The accuracy of the experimental binding affinity data is limited. (II) Data of different sources should not be directly integrated. To address this problem, we tried to collect data from the same literature,[45] but there remain some data from others' work. (III) The PMF method does not discriminate contributions of ligands with different binding affinities. Instead, it takes all atom pairs *ij* in the same shell as equally important in statistical analysis, which may cause problems in later binding affinity prediction. Currently, we are refining the PMF scoring function by taking the experimental binding affinity data into account and will report it elsewhere.

Virtual Screening. Kinases represent an important drug target class, which has spurred many researchers to seek novel compounds for specific kinase inhibitor development. To further investigate the ability of kinase-PMF in virtual screening, a rescoring test was conducted for nine scoring functions. Comparing them, it was found that F_Score and PMF_Score in Sybyl and kinase-PMF performed well in the CDK2 virtual screening study. The enrichment of F_Score and PMF_Score is about 60% in the top 5% compounds of the data set, while kinase-PMF is superior in identifying all known compounds. Although the overall correlation of binding affinity is generally low in all nine score functions, the ability to select the active compounds from the database is not directly dependent on the binding affinity ranking. The structural characteristics of kinase ligands and interaction patterns between kinase and inhibitors may be implicitly incorporated into the kinase-PMF scoring function. This may provide kinase-PMF with the ability to select compounds containing such structural patterns. Scaffold analysis also demonstrates that top-ranked compounds usually contain similar scaffolds found in known CDK2 inhibitors. These compounds will be suitable for further experimental investigation to see whether these are actually active in CDK2 inhibition assays.

CONCLUSION

A target-specific scoring function called kinase-PMF has been developed with a kinase data set of 872 complexes from the PDB database. The scoring function inherits the functional form and atom types in PMF04. Compared to PMF04, PMF99, and PMF_Score in Sybyl, kinase-PMF is greatly different in several respects: first, it has different occurrences of some atom pairs, reflecting the special characteristics of kinase−ligand interactions. Second, the ligand volume correction factor is also different in some atom types, in accordance with the difference in occurrences of atom pairs. Finally, it has a wider range of scores, which implies a better sensitivity for binding conformation prediction at different levels. Compared with eight other methods in rescoring evaluation, kinase-PMF has the highest success rate not only in identifying good conformations from decoys but also in picking out crystal conformations, proving its excellent sensitivity. In the virtual screening test, it also behaves well. Although the correlations between the scores of the nine methods and experimental affinity data are generally low, kinase-PMF can still identify the known CDK2 inhibitors contained in random selected Maybridge molecules.

Together, the results above testify that the predictability of conformational ranking and virtual screening can be improved by a target-oriented scoring scheme and that kinase-PMF is a target-specific scoring function for kinase study. This will empower researchers to search and optimize hit compounds in kinase inhibitor development.

ACKNOWLEDGMENT

We thank Dr. Albert M. Berghuis and Dr. David L. Burk (McGill University, Canada) for providing critical comments and suggestions. This work was financially supported by National Science and Technology Major Project (2009ZX09501-010 to J.S.), State Key Program of Basic Research of China (Grant 009CB918502 to B.X.), and Science and Technology Commission of Shanghai Municipality(08DZ1980200 to B.X.).

Supporting Information Available: A table of logarithms of proten−ligand atom pair occurrences in the kinase database. A table of correlation tests with kinases' PDB codes, affinity data, and scores of the nine methods. The 872 PDB codes used in kinase-PMF scoring function construction and the other 128 PDB codes used in the test and 10 CDK2 inhibitors for the virtual screening study. We also provide the python codes for assigning the protein and ligand atom types in the PMF scoring function. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Marshall, G. R. Computer-aided drug design. *Annu. Rev. Pharmacol. Toxicol.* **1987**, *27*, 193–213.
(2) Dixon, J. S. Computer-aided drug design: getting the best results. *Trends Biotechnol.* **1992**, *10*, 357–363.
(3) Jackson, R. C. Update on computer-aided drug design. *Curr. Opin. Biotechnol.* **1995**, *6*, 646–651.

(4) Tang, Y.; Zhu, W. L.; Chen, K. X.; Jiang, H. L. New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. *Drug Discovery Today* **2006**, *3*, 307–313.
(5) Gane, P. J.; Dean, P. M. Recent advances in structure-based rational drug design. *Curr. Opin. Struct. Biol.* **2000**, *10*, 401–404.
(6) Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **2003**, *10*, 787–797.
(7) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
(8) Good, A. C.; Cheney, D. L. Analysis and optimization of structure-based virtual screening protocols (1): exploration of ligand conformational sampling techniques. *J. Mol. Graphics Modell.* **2003**, *22*, 23–30.
(9) Liwo, A.; Czaplewski, C.; Oldziej, S.; Scheraga, H. A. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134–139.
(10) Kallblad, P.; Dean, P. M. Efficient conformational sampling of local side-chain flexibility. *J. Mol. Biol.* **2003**, *326*, 1651–1665.
(11) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK6: Combining techniques to model RNA-small molecule complexes. *RNA* **2009**, *15*, 1–12.
(12) Wang, R. X.; Lai, L. H.; Wang, S. M. Further development and validation of empirical scoring functions of structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.
(13) Pham, T. A.; Jain, A. N. Customizing scoring functions for docking. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 1–18.
(14) 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.
(15) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
(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) Sippl, M. J. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 473–501.
(18) Levitzki, A. Protein kinase inhibitors as a therapeutic modality. *Acc. Chem. Res.* **2003**, *36*, 462–469.
(19) Noble, M. E.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science.* **2004**, *303*, 1800–1805.
(20) Liao, L.; Jie, J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50*, 409–424.
(21) Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.
(22) Williams, D. H.; Mitchell, T. Latest developments in crystallography and structure-based destgn of protein kinase inhibitors as drug candidates. *Curr. Opin. Pharmacol.* **2002**, *2*, 567–573.
(23) Zhang, J. W.; Aizawa, M.; Amari, S.; Iwasawa, Y.; Nakano, T.; Nakata, K. Development of KiBank, a database supporting structure-based drug design. *Comput. Biol. Chem.* **2004**, *28*, 401–407.
(24) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912–1934.
(25) Manning, G. KinBase. http://kinase.com/kinbase/FastaFiles/ (accessed September 25, 2008).
(26) Wang, G.; Dunbrack, R. L. PISCES: a protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589–1591.
(27) Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem.* **2001**, *22*, 418–425.
(28) Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895–5902.
(29) Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray, R. P.; Rzepa, H.; et al. The blue obelist-interoperability in chemical informatics. *J. Chem. Inf. Model.* **2006**, *46*, 991–998.
(30) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. Software news and update a semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.
(31) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(32) Rarey, M.; Kramer, B.; Lengauer, T. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics.* **1999**, *15*, 243–250.
(33) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.
(34) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(35) Meng, C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
(36) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolinine, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.
(37) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins. *J. Chem. Inf. Model.* **2009**, *49*, 1568–1580.
(38) Gray, N. S.; Wodicka, L.; Schultz, P. G.; et al. Exploiting Chemical Libraries, Structure, and Genomics in the Search for Kinase Inhibitors. *Science* **1998**, *281*, 533–538.
(39) Irwin, J. J.; Shoichet, B. K. ZINC-a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(40) Bostrom, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449–462.
(41) Mcgann, M. R.; Almond, H. R.; Nicholls, A.; et al. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.
(42) Vellay, S. G. P.; Miller Latimer, N. E.; Paillard, G. Interactive text mining with pipeline pilot: a bibliographic web-based tool for pubmed. *Infect. Disord. Drug Targets.* **2009**, *9*, 366–374.
(43) Workman, P.; van Montfort, R. L. M. Structure-based design of molecular cancer therapeutics. *Trends. Biotechnol.* **2009**, *27*, 315–328.
(44) 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.
(45) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; et al. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.
(46) Yan, B. B.; Xue, M. Z.; Xiong, B.; Liu, K.; Hu, D. Y.; Shen, J. K. ScafBank: a public comprehensive Scaffold database to support molecular hopping. *Acta Pharmacol. Sin.* **2009**, *30*, 251–258.
(47) Seifert, M. H. J. Targeted scoring functions for virtual screening. *Drug Discovery Today* **2009**, *14*, 562–569.
(48) Breu, B.; Katrin, S.; Holger, G. Consensus adaptation of fields for molecular comparison (AFMoC) models incorporate ligand and receptor conformational variability into tailor-made scoring functions. *J. Chem. Inf. Model.* **2007**, *47*, 2383–2400.
(49) Martin, E. J.; Sullivan, D. C. AutoShim: Empirically Corrected Scoring Functions for Quantitative Docking with a Crystal Structure and IC_{50} Training Data. *J. Chem. Inf. Model.* **2008**, *48*, 861–872.
(50) Martin, E. J.; Sullivan, D. C. Surrogate AutoShim: Predocking into a Universial Ensemble Kinase Receptor for Three Dimensinal Activity Prediction, Very Quickly, without a Crystal Structure. *J. Chem. Inf. Model.* **2008**, *48*, 873–881.

CI100182C