JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility

Sheng Tian,[†,§,#] Huiyong Sun,[†,#] Peichen Pan,[‡] Dan Li,[‡] Xuechu Zhen,[*,§] Youyong Li,[*,†] and Tingjun Hou[*,†,‡]

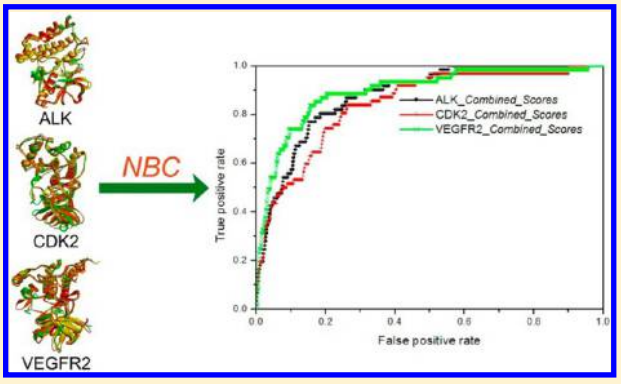
[†]Institute of Functional Nano and Soft Materials (FUNSOM), Soochow University, Suzhou, Jiangsu 215123, P. R. China

[‡]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

[§]College of Pharmaceutical Sciences, Soochow University, Suzhou, Jiangsu 215123, P. R. China

Supporting Information

ABSTRACT: In this study, to accommodate receptor flexibility, based on multiple receptor conformations, a novel ensemble docking protocol was developed by using the naïve Bayesian classification technique, and it was evaluated in terms of the prediction accuracy of docking-based virtual screening (VS) of three important targets in the kinase family: ALK, CDK2, and VEGFR2. First, for each target, the representative crystal structures were selected by structural clustering, and the capability of molecular docking based on each representative structure to discriminate inhibitors from non-inhibitors was examined. Then, for each target, 50 ns molecular dynamics (MD) simulations were carried out to generate an ensemble of the conformations, and multiple representative structures/snapshots were extracted from each MD trajectory by structural clustering. On average, the representative crystal structures outperform the representative structures extracted from MD simulations in terms of the capabilities to separate inhibitors from non-inhibitors. Finally, by using the naïve Bayesian classification technique, an integrated VS strategy was developed to combine the prediction results of molecular docking based on different representative conformations chosen from crystal structures and MD trajectories. It was encouraging to observe that the integrated VS strategy yields better performance than the docking-based VS based on any single rigid conformation. This novel protocol may provide an improvement over existing strategies to search for more diverse and promising active compounds for a target of interest.



INTRODUCTION

With the explosion of protein structures afforded by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, the high-throughput screening (HTS) technique has been used widely to identify promising drug leads for targets of interest. However, HTS is very expensive and can usually be employed at the industrial level only by large pharmaceutical companies. Moreover, the high positive rate of HTS is criticized by medicinal chemists. As an important complementary approach to HTS, virtual screening (VS), especially molecular docking-based VS,[1−4] has received increasing attention.[5−8] By docking small organic molecules into the binding sites of macromolecular targets and estimating the binding potencies in terms of docking scores, a small fraction of screened compounds would be chosen as putative hits for biological testing. For almost all molecular docking programs, the simplistic rigid "lock-and-key" interaction philosophy between rigid receptor and flexible ligand seems to be unreliable to achieve accurate docking results.[1] As pointed by Bottegoni et al., selecting a single receptor conformation may hamper the quality of the obtained potential active compounds.[9] In addition, previous studies showed that the predictions of molecular docking based on different structures of the same target can be quite different, because the binding patterns characterized by these different static complexes deposited in the Protein Data Bank (PDB) may be different.[10−12]

It is widely recognized that protein flexibility plays an important role in protein−ligand recognition and needs to be incorporated into molecular docking calculations.[1,13−15] But due to the unaffordably high computational cost, it is not feasible to screen a massive library of small organic molecules against all possible conformations of a target.[16] For molecular docking, the use of multiple receptor conformations (MRCs) for a specified target is a more realistic and attractive way of restricting the protein conformational space to a more reasonable scope.[1,17−24] According to this consideration, utilizing a set of static conformations of a target, including experimentally derived crystal structures (X-ray and/or NMR) or computationally generated structures, to represent the dynamic nature of a receptor is the most popular strategy in structure-based drug design

Received: July 13, 2014
Published: September 18, 2014

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

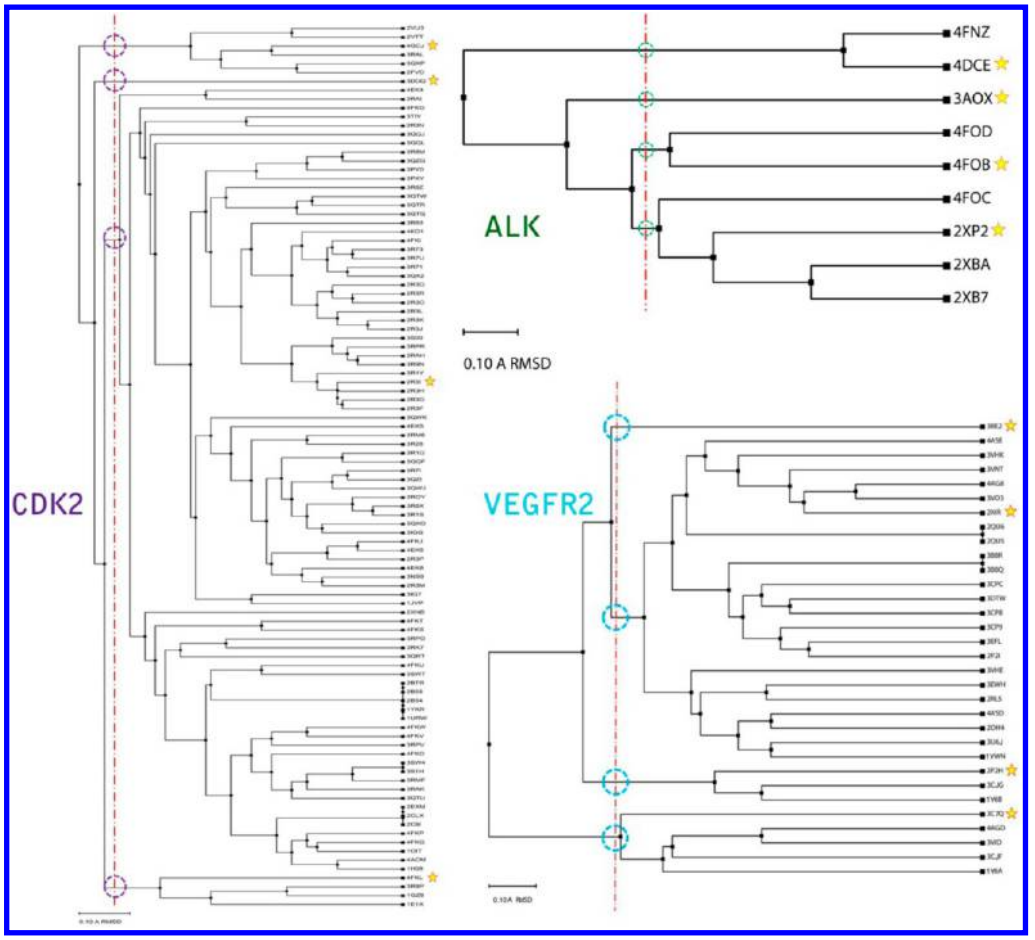


Figure 1. Structural clustering for the crystal structures of the three targets by using the *phylogenetic tree* method in VMD.

(SBDD).[1,25−32] However, MRC docking (also termed as ensemble docking) has some issues to be addressed. For example, Cavasotto and Abagyan raised the following questions: Although the use of various target structures seems to be the best choice to incorporate flexibility, what should be the source of these structures, how many are needed, and how should they be used or results be combined?[33] In order to answer these questions, extensive efforts have been made in the previous studies.[20,31,32] For instance, recent publications illustrate that MRCs can be retrieved from the PDB database or generated by molecular dynamics (MD) simulations. Sufficient and diverse crystal structures obtained from PDB offer a good source for sampling protein flexibilities of some targets.[31,32] In addition, MD simulations provide an alternative way to generate a structurally diverse subset of a protein structure.[34−36] Hagler et al. investigated the performance of ensemble docking based on four representative crystal structures of CDK2 or HIV protease, and they found that, compared with a single crystal structure, ensemble docking yields higher enrichment of known inhibitors.[37] In addition, recent studies reported by Hagler et al. illustrate that the structures generated by MD or replica exchange molecular dynamics (REMD) simulations can increase the enrichment of known inhibitors significantly and enhance the diversity of the hits from molecular docking.[37] Besides, the authors also stated that the docking based on ensemble crystal structures could give higher enrichment in VS and more diverse potential active compounds than the docking based on ensemble MD structures for three studied targets, including androgen receptor, HIV protease, and CDK2.[38]

To balance the computational efficacy and prediction accuracy of MRC docking, a relatively small set of representative structures should be selected. In 2010, Abagyan et al. explored the best way to select experimental protein structures for MRC docking, and they found that molecular docking based on 3−5 structures or a single conformation with the largest cocrystallized ligand for a specified target is more likely to yield satisfactory results.[39] Barril and Morley assessed the performance of MRC docking to CDK2 and HSP90, and they observed that two conformations might be enough to achieve the highest enrichments in the top 1% fraction.[30] It appears that the use of too many receptor conformations may not necessarily improve the docking performance because a large number of more false positives may reduce the enrichment rate in VS.

Another question is also quite challenging for MRC docking: how can the predictions of molecular docking based on multiple conformations be combined for a specified target? A normal way is to merge independent docking runs into one entry and then re-rank or re-score. However, similar to the consensus scoring by using different computational techniques for a single structure, the strategy by simply re-scoring or re-ranking was quite sensitive to each individual docking study.[9] Recently, machine-learning approaches have been applied to the field of VS,[40−44] and they can integrate the predictions of different scoring functions for a single structure or those of a single scoring scheme for multiple structures of a target into more versatile and accurate classification models for VS.[40,43] In 2004, Klon et al. employed naïve Bayesian classification (NBC) to integrate the predictions from five scoring functions, and observed that the enrichment

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

factor of this consensus scoring can be improved.[40] Besides, in our previous studies, we also combined the predictions of molecular docking for multiple complex structures of ROCK1 into a single integrated model by using NBC, and we found that the integrated model achieves the best prediction power.[42,43]

The previous studies illustrate that the prediction capacities of docking-based VS are quite sensitive to the selected structures used in the calculations.[10−12] In addition, the MRC docking results demonstrated that the performance of molecular docking based on crystal structures are better than that based on MD snapshots.[37] However, the MRC docking based on crystal structures and that based on MD snapshots were rarely compared to the same targets.[33,37−39] Moreover, in order to incorporate protein flexibility in docking-based VS, the merging/shrinking procedures, primarily by re-scoring or re-ranking to manage the predictions of molecular docking based on multiple structures for a specific target, is quite sensitive to the individual docking calculations.[33,37−39] Thus, the use of more powerful methodologies to combine the multiple docking predictions is quite necessary. In this study, considering the pharmaceutical importance of the kinase family in drug design/discovery, three important kinase targets—anaplastic lymphoma kinase (ALK), cyclin-dependent kinase 2 (CDK2), and vascular endothelial growth factor receptor 2 (VEGFR2)—were selected to develop and evaluate the integrated VS protocol in this study. To our knowledge, this is the first case to investigate the performances of the naive Bayesian classifiers to integrate the predictions of molecular docking by using multiple structures for kinase targets. Here, for each target, four diverse crystal structures were chosen by using the structural clustering method, and the discrimination capacity of each representative structure to separate the known inhibitors from non-inhibitors was evaluated by using two scoring modes supported by the *Glide* docking. Then, for each target, the crystal complex with best discrimination power was used as the initial structure for 50 ns MD simulations, and four representative snapshots were extracted from each MD trajectory by using structural clustering. The discrimination powers of the MD structures were also evaluated and compared with those based on the crystal structures for three targets. Additionally, the performance difference of molecular docking for each pair of structures was quantitatively evaluated by the correlation between the docking scores. Finally, in order to incorporate protein flexibility into docking-based VS, a novel integrated strategy was built by using the NBC technique based on the docking scores of the representative structures for each target. Moreover, the best Bayesian classifier for each target was also built by removing the docking predictions of redundant structures. The study gives deep insight into the importance of incorporating protein flexibility to molecular docking and provides an improved way over existing strategies to search for more diverse and promising active compounds for kinase targets of interest.

METHODS AND MATERIALS

Preparation of Representative Crystal Structures for Each Target. The crystallographic structures of the complexes of three targets, ALK, VEGFR2, and CDK2, were retrieved from the RCSB Protein Data Bank (PDB).[45] The numbers of the crystal complexes for ALK, VEGFR2, and CDK2 are 9, 32, and >300, respectively. Previous studies demonstrated that the use of 3−5 experimental structures is enough to guarantee the satisfactory performance of MRC docking.[39] Therefore, for each target, four diverse complexes were chosen by using the following structural clustering strategy: (1) the 100 crystal structures of CDK2 with the highest resolutions were selected; (2) the

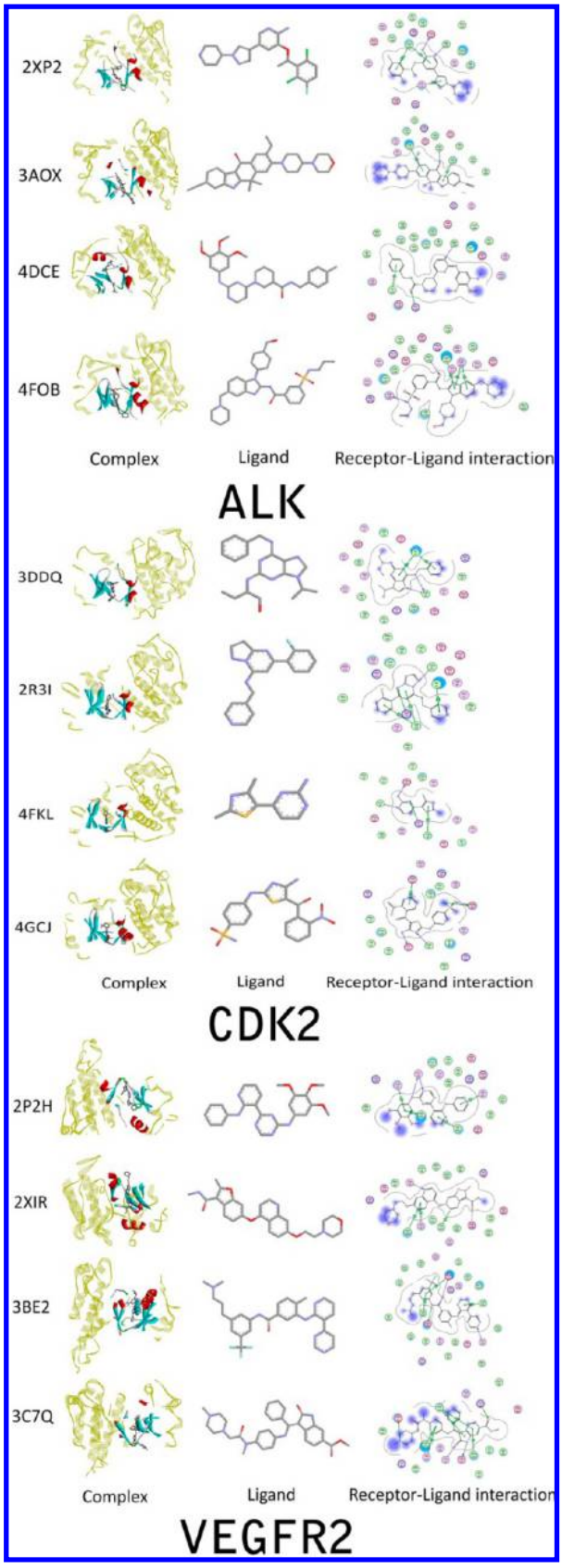


Figure 2. Four representative complexes and the interactions between each cocrystallized ligand and the corresponding receptor for three targets.

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

binding pocket was defined as the residues within 10 Å of the cocrystallized ligand; (3) the residues of the binding pockets were structurally aligned by using the STAMP algorithm in VMD;[46] (4) a phylogenetic tree was created by using the *create phylogenetic tree* protocol in VMD based on the root-mean-square deviation (RMSD) values. According to the phylogenetic tree shown in Figure 1, four representative complexes with the highest resolutions were selected. For each target, the four representative complexes and the interactions between the ligands and corresponding receptors are depicted in Figure 2.

Preparation of Validation Dataset for Molecular Docking. For each target, a validation dataset with known inhibitors and non-inhibitors was prepared for molecular docking. The known inhibitors for each target were obtained from the *BindingDB* database.[47] The inhibitors without quantitative biological activities (IC_{50} or K_i) and the duplicates were removed. Then, the non-inhibitors were randomly selected from the *ChemBridge* database[48] based on the 2D similarity (*Tanimoto Coefficient*) of the FCFP_4 fingerprint by using the *Find Diverse Molecules* module in Discovery Studio 3.1.[49] The use of commercial database is an effective way to represent the chemical space of a decoy set reported in the previous study.[48] For each target, the inhibitors were compared with the non-inhibitors, and the duplicates were removed from the non-inhibitor dataset. Considering the computational accuracy and efficiency, based on the 2D similarity (*Tanimoto Coefficient*) of the FCFP_4 fingerprint, 200 diverse known inhibitors for each target were chose by using the *Find Diverse Molecules* module in Discovery Studio 3.1.[49] Then, to mimic the unbalanced nature between inhibitors and non-inhibitors, the ratio between non-inhibitors and inhibitors was set to 20. At last, for each target, the validation dataset with 200 known inhibitors and 4000 non-inhibitors was prepared for the following analysis.

Molecular Docking Procedure. Molecular docking calculations were carried out by using *Glide*[50] in Schrödinger 9.0.[51] For each protein structure, the *Protein Preparation Wizard* module in Schrödinger 9.0[51] was employed to remove all crystallographic water molecules, add missing side chains and hydrogen atoms, assign protonated states and partial charges with the OPLS-2005 force field, and minimize the structure. The minimization stage was terminated when the RMSD of the non-hydrogen atoms reached a maximum value of 0.3 Å.

All the molecules in each validation dataset were processed by using the *Ligprep* module in Schrödinger 9.0.[51] For each compound, the ionized states and tautomers were created by using *Epik* at pH = 7.0.[51] For the known inhibitors with 3-D structures, the original chiralities were reserved. Because the non-inhibitors chosen from *ChemBridge* database do not have 3-D structural information, the different combinations of chiralities were generated by setting the maximum number of the stereoisomers for each compound to 32. All of the other parameters were kept as the default setting in *Ligprep*. At last, the numbers of the prepared inhibitors for ALK, CDK2, and VEGFR2 are 205, 208, and 202, respectively, and the numbers of the prepared non-inhibitors for ALK, CDK2, and VEGFR2 are 7565, 7565, and 7773, respectively.

Then, the binding pocket with the size of 10 Å × 10 Å × 10 Å that was centered on the mass center of the cocrystallized ligand for each complex was defined by using the *Receptor Grid Generation* module of *Glide*.[51] The default settings in *Glide* were used for grid generation. All the molecules in each validation dataset were docked into the binding pocket of each protein structure and then scored by the standard precision (SP) and extra precision (XP) scoring modes of *Glide*. During the initial phase of the molecular docking calculation, five thousand poses per compound were generated and then the best 400 poses were chosen for energy minimization. A dielectric constant of 2.0 and 100 steps of conjugate gradient minimization were applied in energy minimization phase. The performances of the SP and XP scoring functions of the *Glide* docking for different complexes of three targets were evaluated and compared.

Table 1. Docking Power and Discrimination Power of the *Glide* Docking of the Four Representative Crystal Complexes for ALK, CDK2, and VEGFR2

kinase target		docking power (RMSD/Å)		discrimination power (*P*-value)	
		SP	XP	SP	XP
ALK	4DCE	0.33	0.53	3.48×10^{-47}	8.37×10^{-58}
	3AOX	1.36	0.87	3.52×10^{-49}	1.65×10^{-38}
	4FOB	1.47	3.59	1.25×10^{-71}	1.71×10^{-45}
	2XP2	0.29	0.34	1.06×10^{-48}	1.17×10^{-44}
CDK2	3DDQ	1.44	1.41	2.49×10^{-60}	2.43×10^{-57}
	2R3I	0.23	0.32	3.42×10^{-28}	5.66×10^{-24}
	4FKL	1.99	0.20	8.02×10^{-24}	1.02×10^{-16}
	4GCJ	0.14	0.20	3.23×10^{-45}	6.34×10^{-47}
VEGFR2	2P2H	0.60	0.72	1.96×10^{-55}	2.86×10^{-58}
	2XIR	2.01	1.86	6.24×10^{-98}	2.27×10^{-92}
	3BE2	1.39	1.29	3.69×10^{-82}	5.48×10^{-95}
	3C7Q	4.31	1.28	5.20×10^{-50}	7.15×10^{-51}

Representative Protein Structures Derived from MD Simulations. The crystal complexes (PDB entries: 4FOB[52] for ALK, 3DDQ[53] for CDK2, and 2XIR[45] for VEGFR2) with the best discrimination capacities to distinguish the known inhibitors from non-inhibitors were used as the initial structures for MD simulations.

The MD simulations were carried out by using the Amber11 molecular simulation package.[54] The inhibitor in each complex was optimized and the electrostatic potential was calculated at the HF/6-31G* level by using the Gaussian 09 program,[55] and the atomic partial charges were obtained by the restraint electrostatic potential fit (RESP) approach.[56] The general AMBER force field (*gaff*)[57] and AMBER ff99SB force field[58] were used for the inhibitor and protein, respectively. The partial charges and force field parameters for the inhibitor were generated by using the *antechamber* suite in AMBER 11.[54] As counterions, Na^+ or Cl^- were added to the grids to neutralize the unbalanced charges in the complexes. Then, each system was immersed into a rectangular TIP3P[59] water box that is extended 10 Å from any solute atom in all three dimensions.

The Particle Mesh Ewald method[60] was employed to handle the long-range electrostatic interactions in a periodic boundary box. By using the *sander* program in AMBER 11,[54] each system was subjected to three stage minimizations before the MD simulations. In the first stage, 1000 cycles of minimization (500 cycles of steepest descent and 500 cycles of conjugate gradient minimization) were carried out with the backbone carbons constrained (50 kcal/(mol·Å^2)). Subsequently, 1000 cycles of minimization with relatively weaker constraint (10 kcal/(mol·Å^2)) were conducted. Finally, 5000 cycles (1000 cycles of steepest descent and 4000 cycles of conjugate gradient minimization) of full energy minimization without any constraint were applied. After minimization, each system was gradually

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

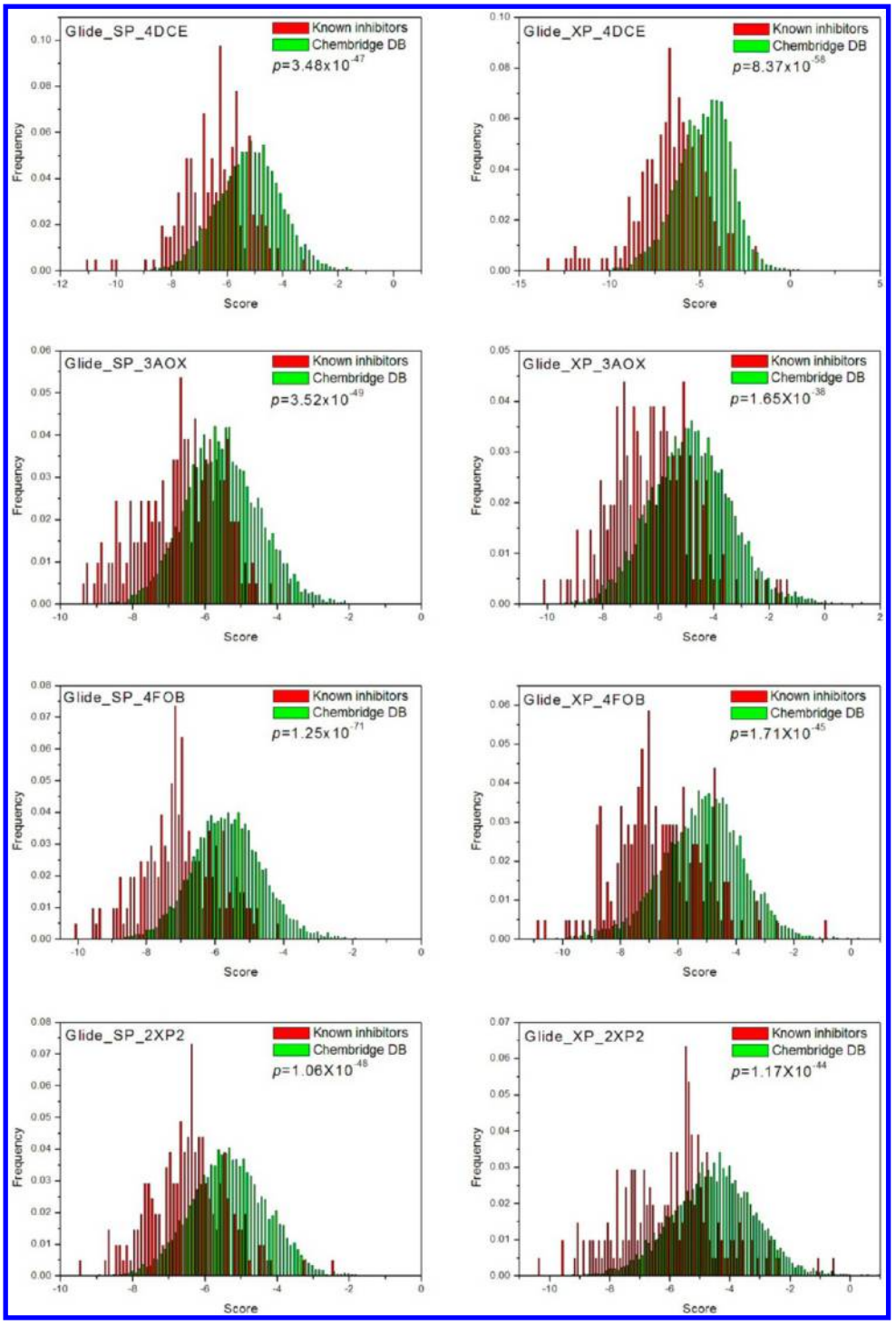


Figure 3. Distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative crystal complexes for ALK.

heated from 0 to 300 K over a period of 50 ps with 2.0 kcal/(mol·Å^2) restraints on the complex, and then another 50 ps NPT (T = 300 K, P = 1 bar) MD simulations were followed. Subsequently, the 50 ns NPT (T = 300 K, P = 1 bar) MD simulations were performed. SHAKE was applied to constrain all bonds involving hydrogen atoms,[61] and the time step was set to 2.0 fs. Coordinate trajectories were saved every 20 ps for the whole MD runs. For each system, 2500 snapshots in each trajectory were named as PDBID_n (n = 1, 2, ..., 2500).

Then, 100 snapshots were evenly extracted from the stable phase (30−50 ns) of each MD trajectory. The residues within 10 Å of the cocrystallized inhibitors for the 100 complexes of each target were defined as the binding pockets. The residues of the binding pockets were structurally aligned by using the STAMP algorithm in VMD.[46] Based on the RMSD structural similarity matrixes, the 100 complexes were clustered into four categories by using the K-means clustering method. For each clustering category, the hypothetical cluster centers were determined and then the complex that has the minimum dissimilarity index with the hypothetical cluster center was selected as the representative structure for three targets.

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

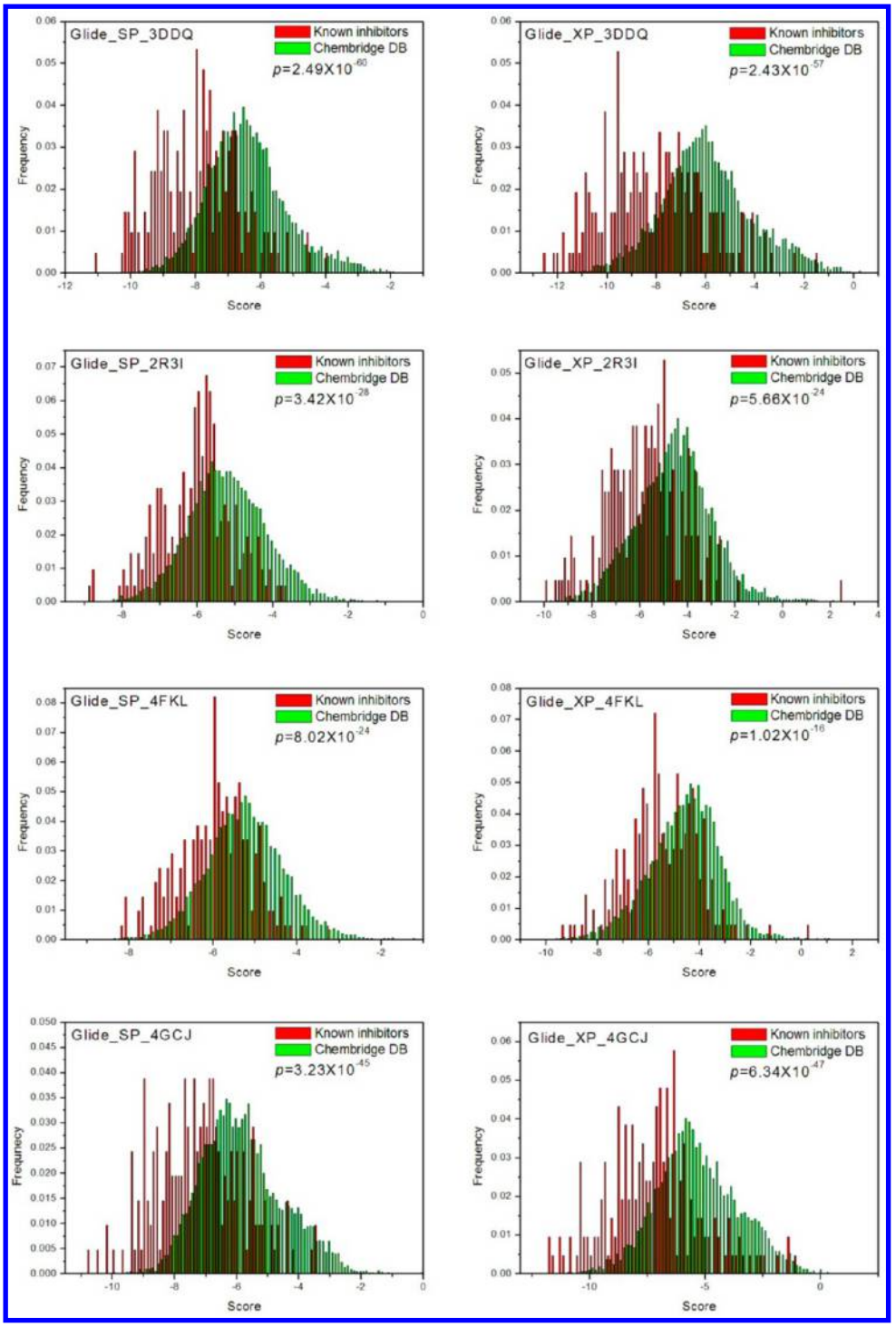


Figure 4. Distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative crystal complexes for CDK2.

Integrating Predictions from Individual Runs. Standard docking-based VS was carried out independently on each structure. Each molecule in the validation dataset was evaluated quantitatively by a *Glide* docking score, and eight docking scores (4 for crystal structures and 4 for MD structures) were obtained for each molecule. In our previous studies, we have employed the machine learning approaches, including naïve Bayesian classification (NBC) and recursive partitioning, to solve binary classification problems.[42,43,62−65] Compared with other machine learning approaches, NBC can handle large amounts of data, learns fast, and is tolerant of random noise during model building. Besides, NBC requires only a small amount of training sample to estimate the parameters (means and variances of the variables) needed for classification.

Here, NBC was employed to integrate the results from individual runs based on multiple protein structures. First, the eight docking scores for each target were served as the independent variables (X), and the class label (1 for an inhibitor and 0 for a non-inhibitor) as the response variable (Y). Then, NBC was employed to develop the classifiers to distinguish the known

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

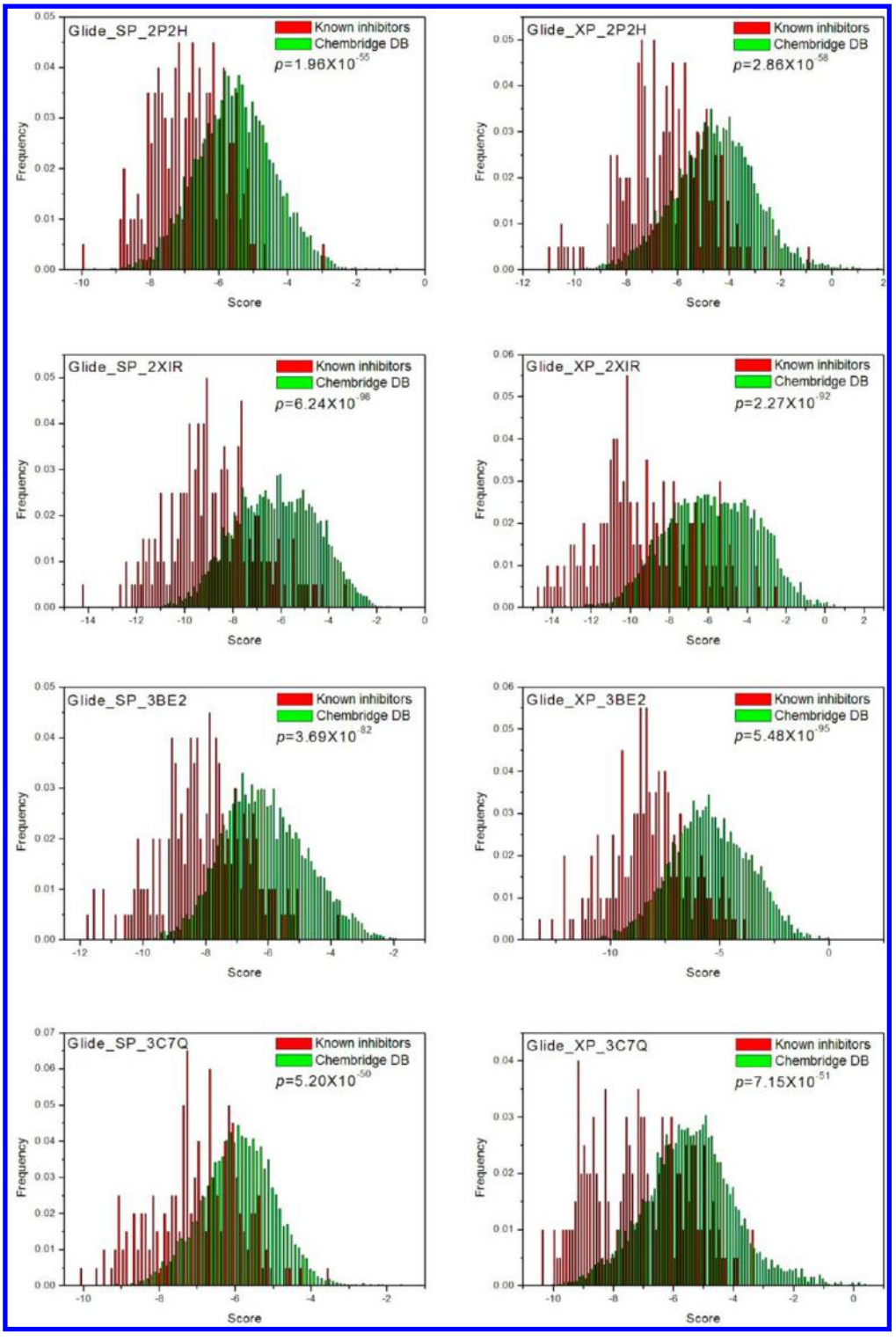


Figure 5. Distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative crystal complexes for VEGFR2.

inhibitors from non-inhibitors. All classification models were developed in Discovery Studio 3.1.[49] Based on five-fold cross validation, we calculated the following measures to evaluate the classification performance of each naïve Bayesian classifier, usinging true positive (TP), false positive (FP), true negative (TN), and false negative (FN): sensitivity,

$$SE = TP/(TP + FN)$$

specificity,

$$SP = TN/(TN + FP)$$

global accuracy (GA), and Matthews correlation coefficient (C),

$$C = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}$$

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

Moreover, the integral of a ROC plot, or the area under the curve (AUC), was used as a quantitative performance assessment for each classifier. Compared with the enrichment factor, which suffers from a number of significant drawbacks, the AUC value is a more global and widely used measure for assessing the performances of docking-based VS.[66,67]

In order to examine the actual prediction power of each classifier, 70% of the inhibitors and non-inhibitors randomly chosen from the validation dataset of each target form the training set, and the remaining compounds form the test set. The tested compounds were validated by the classifier developed based on the training set.

RESULTS AND DISCUSSION

Performance of Molecular Docking Based on a Single Crystal Structure. At first, the "docking power", an important index of the reliability for molecular docking to reproduce the experimental binding modes, was evaluated. For each representative crystal structure, the inhibitor was extracted from the complex and re-docked into the binding pocket. The root-mean-squared-deviation (RMSD) between the predicted binding pose and the respective experimental structure was calculated, and RMSD $\leq$ 2.0 Å is used as the criterion of successful docking. As shown in Table 1, the *Glide* docking can reproduce the experimental binding poses of the inhibitors for most complexes. For ALK and CDK2, all the complexes by using the SP or XP scoring mode of *Glide* can satisfy the requirement of RMSD $\leq$ 2.0 Å. For VEGFR2, only one complex, 3C7Q, cannot achieve successful prediction by using the SP scoring of *Glide* (Table 1).

Then, the "discrimination power", which is defined as the capability to distinguish the known inhibitors from non-inhibitors in a validation dataset, was evaluated. The student's *t*-test was used to assess the significance of the difference between the means of the distributions of docking scores of the known inhibitors and non-inhibitors. As shown in Table 1 and Figures 3, 4, and 5, for the three targets, the known inhibitors and non-inhibitors can be well distinguished in terms of the *P*-values ($<10^{-20}$). However, the prediction capacities of the four representative complexes for each target are quite different. For example, for ALK, the *P*-value for the *Glide* SP scores of 4FOB is up to 1.25×10^{-71}, while that of 3AOX is only 1.65×10^{-38}. Similar phenomena can also be observed for CDK2 and VEGFR2 (Table 1). The results imply that the prediction accuracies of docking are variable across different structures of the same target.

Generally speaking, the XP scoring is believed to be more precise than the SP scoring.[50] However, we observe that the XP scoring does not always perform better than the SP scoring for all the studied systems. For ALK and CDK2, the XP scoring outperforms the SP scoring for only one structure in terms of the *P*-values. For VEGFR2, the predictions of the XP scoring for three of the four structures are better (Table 1). Therefore, it is obvious that the comparison of the performances of the different scoring functions is quite necessary before employing the *Glide* docking-based VS to any target.

As pointed out by Rueda,[39] molecular docking based on the protein structure in complex with the largest cocrystallized ligand is likely to achieve the best discrimination capacity to separate inhibitors from non-inhibitors. Thus, we investigated the influence of the bulkiness, i.e., molecular weight (MW), molecular volume (MV), and solvent-accessible surface area (SASA), of the inhibitors on the performances of molecular docking. As shown in Table S1 in Supporting Information, the sizes (MW, MV, or SASA) of the inhibitors do not show any relationships with the discrimination capacities of molecular docking, which is not consistent with the previous study.[39] For ALK and CDK2, the structures of inhibitors with the largest MVs (4FOB and 3DDQ) achieve the best discrimination capacities of molecular docking, but for VEGFR2, the structure of inhibitor with the largest MV (3C7Q) is not that with the best prediction power (2XIR). Therefore, for a specific target, the most optimal structure for docking-based VS cannot be simply determined by the size of the ligand.

In summary, for the three targets, the *Glide* docking can satisfy the requirement of "docking power" (RMSD $\leq$ 2.0 Å) and "discrimination power" (*P*-value $\leq 10^{-20}$) by using whether SP or XP scoring, suggesting that the *Glide* docking can be used as a reliable VS tool for ALK, CDK2, and VEGFR2.

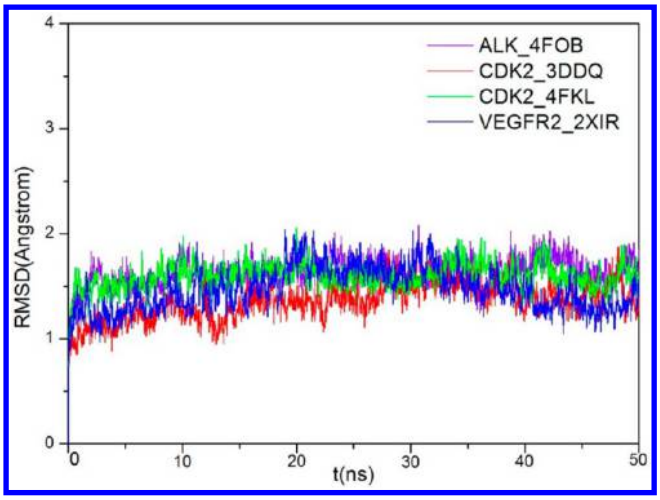


Figure 6. Time evolution of the root-mean-square derivations (RMSD) of the receptor residues within 10 Å around the cocrystallized inhibitor for the representative complex of three targets. Violet, the 4FOB complex of ALK; red, the 3DDQ complex of CDK2; green, the 4FKL complex of CDK2; and blue, the 2XIR complex of VEGFR2.

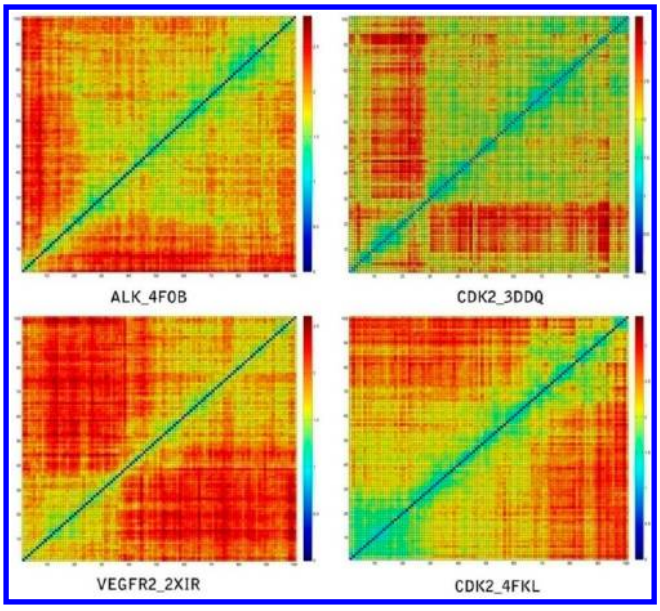


Figure 7. Clustering the 100 conformations/snapshots derived from the MD simulations by serving 4FOB of ALK, 3DDQ and 4FKL of CDK2, and 2XIR of VEGFR2 as the starting structures.

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

Table 2. Docking Power and Discrimination Power of the *Glide* Docking of the Four Representative Complexes Derived from the MD Simulations for ALK, CDK2, and VEGFR2

kinase target		docking power (RMSD/Å)		discrimination power (*P*-value)	
		SP	XP	SP	XP
ALK	4FOB_1510	2.95	2.35	2.75×10^{-56}	4.31×10^{-39}
	4FOB_1610	2.83	4.83	1.96×10^{-56}	4.56×10^{-53}
	4FOB_2010	5.73	6.61	2.76×10^{-30}	6.08×10^{-30}
	4FOB_2400	3.11	2.63	1.42×10^{-34}	5.00×10^{-25}
CDK2	4FKL_1610	5.05	3.94	5.34×10^{-39}	6.14×10^{-25}
	4FKL_1890	3.67	3.65	3.27×10^{-40}	1.20×10^{-26}
	4FKL_2350	1.77	1.77	1.20×10^{-42}	5.07×10^{-44}
	4FKL_2470	1.91	1.82	5.59×10^{-37}	3.53×10^{-27}
	3DDQ_1650	1.30	1.31	1.74×10^{-38}	8.11×10^{-31}
	3DDQ_2040	1.78	1.68	3.34×10^{-53}	1.00×10^{-53}
	3DDQ_2140	1.78	1.48	6.67×10^{-62}	4.64×10^{-61}
	3DDQ_2500	1.35	1.39	3.84×10^{-55}	1.38×10^{-55}
VEGFR2	2XIR_1560	2.25	2.49	9.39×10^{-69}	6.53×10^{-71}
	2XIR_1670	4.12	2.09	1.95×10^{-63}	8.33×10^{-58}
	2XIR_1920	2.55	5.13	8.11×10^{-81}	1.42×10^{-75}
	2XIR_2290	2.92	3.14	1.88×10^{-74}	2.11×10^{-78}

Performance of Molecular Docking Based on a Single MD Structure. For a specific target, the available crystal structures may be not enough to cover some important conformational space that can be recognized by a number of inhibitors. In such case, computational technique, such as MD simulations, can be used to generate a large number of conformations for docking-based VS. According to this consideration, for each target, we carried out 50 ns MD simulations and selected four representative structures from the MD trajectory by structural clustering (Figures 6 and 7).

Based on the four representative structures derived from the MD simulations for each target, the performances of molecular docking by using the SP and XP scoring modes of *Glide* were evaluated. As shown in Table 2, for ALK and VEGFR2, the *Glide* docking based on the MD structures cannot reproduce the binding poses of the inhibitors with high accuracies (RMSD > 2.0 Å). However, in case of CDK2, the *Glide* docking based on the MD structures meets the RMSD requirement. For example, the RMSD between the inhibitor pose of 4FOB_2010 and that predicted by the SP scoring of *Glide* is 5.73 Å (Table 2). However, the inhibitor binding pose of 4FOB_2010 predicted by docking is quite closer to that of the crystal structure (4FOB) than that of 4FOB_2010 (Figure 8). Our finding is not so surprising. It is believed that any crystallographic structure of a protein only presents a single snapshot trapped in the low energy states of the protein, but the protein–ligand recognition is more closely related to the dynamics of target. Therefore, the binding pattern of the ligand to a MD snapshot may be visibly different from that to the crystal structure.

Then, for each target, the discrimination powers of the four MD structures were examined (Table 2). The distributions of the SP or XP scores of the four representative MD snapshots for ALK, CDK2, and VEGFR2 by using 4FOB, 3DDQ and 2XIR as initial structures are depicted in Figures S1, S2, and S3 in Supporting Information. The *P*-values associated with the difference in the mean docking scores of the inhibitors versus those of the non-inhibitors are less than 10^{-20} by using either the SP or XP scoring, suggesting that the MD structures can yield satisfactory predictions in molecular docking.

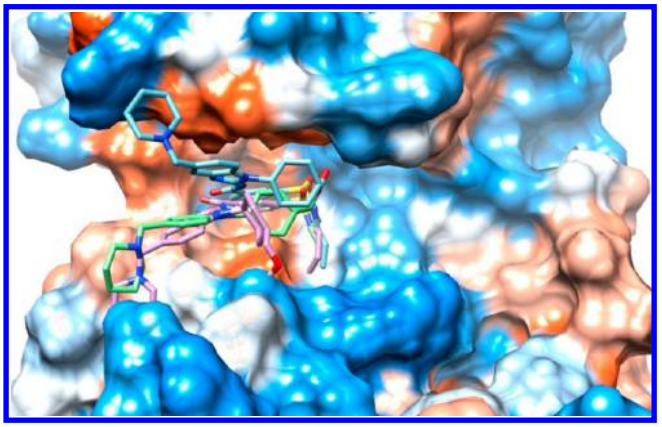


Figure 8. Structural superimposition of the inhibitor pose (green) of 4FOB, the inhibitor pose (blue) of a MD snapshot of 4FOB (4FOB_2010), and the inhibitor pose (pink) of 4FOB_2010 predicted by using the SP scoring of *Glide*.

Furthermore, the discrimination powers of docking based on the representative crystal structures and those based on the MD structures were compared. For two targets, ALK and VEGFR2, the discrimination power of the best crystal structure outperforms that of the best MD structure (Tables 1 and 2). However, for CDK2, the best discrimination power of the MD structure (3DDQ_2140, *P*-value = 6.67×10^{-62}) is slightly better than that of the crystal structure (3DDQ, *P*-value = 2.49×10^{-60}). In addition, in order to investigate whether the MD simulations can improve the performance of docking-based VS, 50 ns MD simulations based on 4FKL with the worst discrimination capacity of CDK2 (Table 1) was also carried out. Interestingly, we observed that the predictions based on a MD snapshot (5.07×10^{-44} for 4FKL_2350) are superior to those based on the initial crystal structure obviously (1.02×10^{-16} for 4FKL). The interesting phenomenon suggests that MD simulations may be used to improve VS results, especially for the targets with limited

dx.doi.org/10.1021/ci500414b | *J. Chem. Inf. Model.* 2014, 54, 2664–2679

Table 3. Pearson Correlation Coefficient Squared (r_p^2) of the Docking Scores of the Known Inhibitors Predicted by Molecular Docking Based on the Four Crystal Structures and Four MD Snapshots of ALK, CDK2, and VEGFR2

	ALK							
	4FOB	4DCE	3AOX	2XP2	4FOB_1510	4FOB_1610	4FOB_2010	4FOB_2400
4FOB	1	0.037	0.124	0.251	0.816	0.053	0.026	0.075
4DCE	0.037	1	0.001	0.032	0.030	0	0.018	0.031
3AOX	0.124	0.001	1	0.068	0.124	0.222	0	0.050
2XP2	0.251	0.032	0.068	1	0.200	0.102	0.067	0.191
4FOB_1510	0.816	0.030	0.124	0.200	1	0.060	0.019	0.073
4FOB_1610	0.053	0	0.222	0.102	0.060	1	0.015	0.156
4FOB_2010	0.026	0.018	0	0.067	0.019	0.015	1	0.116
4FOB_2400	0.075	0.031	0.050	0.191	0.073	0.156	0.116	1
	CDK2							
	4GCJ	4FKL	3PXF	3DDQ	3DDQ_1650	3DDQ_2040	3DDQ_2140	3DDQ_2500
4GCJ	1	0.050	0.075	0.132	0.107	0.080	0.145	0.109
4FKL	0.050	1	0.325	0.381	0.687	0.396	0.650	0.650
3PXF	0.075	0.325	1	0.252	0.350	0.231	0.396	0.462
3DDQ	0.132	0.381	0.252	1	0.546	0.886	0.472	0.462
3DDQ_1650	0.107	0.687	0.350	0.546	1	0.556	0.719	0.718
3DDQ_2040	0.080	0.396	0.231	0.886	0.556	1	0.516	0.515
3DDQ_2140	0.145	0.650	0.396	0.472	0.719	0.516	1	0.808
3DDQ_2500	0.109	0.650	0.462	0.462	0.718	0.515	0.808	1
	VEGFR2							
	3C7Q	3BE2	2XIR	2P2H	2XIR_1560	2XIR_1670	2XIR_1920	2XIR_2290
3C7Q	1	0.610	0.493	0.715	0.539	0.516	0.356	0.506
3BE2	0.610	1	0.598	0.577	0.651	0.671	0.455	0.590
2XIR	0.493	0.598	1	0.500	0.718	0.803	0.691	0.745
2P2H	0.715	0.577	0.500	1	0.464	0.497	0.352	0.464
2XIR_1560	0.539	0.651	0.718	0.464	1	0.762	0.540	0.828
2XIR_1670	0.516	0.671	0.803	0.497	0.762	1	0.736	0.754
2XIR_1920	0.356	0.455	0.691	0.352	0.540	0.736	1	0.533
2XIR_2290	0.506	0.590	0.745	0.464	0.828	0.754	0.533	1

available crystal structures and/or those with poor docking performances. The results illustrate that on average the predictive power of the crystal structures are target-dependent and only slightly better (or comparative to) than the MD structures, which is not completely consistent with the findings observed by Hagler et al. and Nichols et al.[37,68]

Docking Results Based on Crystal Structures versus MD Snapshots. In order to give a quantitative comparison of the docking results for any two structures of the same target, the Pearson coefficients squared (r_p^2) of the docking scores of the known inhibitors for the two compared crystal structures were calculated and are shown in Table 3 and Figure 9a. Ideally, if the predicted binding scores based on the two structures are identical, the Pearson coefficient should be 1.0. The results illustrate that in some cases the predictions based on two different structures are quite different. For example, for the four crystal structures of ALK and CDK2, the *Glide* docking scores of the inhibitors do not show good linear correlations, and the best correlations (r_p^2) are only 0.251 (4FOB and 2XP2 for ALK) and 0.381 (3DDQ and 4FKL for CDK2), respectively. For VEGFR2, the predictions based on two different structures show better correlations than ALK and CDK2, indicated by relatively higher r_p^2 (0.49−0.72, Table 3 and Figure 9a).

The Pearson correlation coefficients squared (r_p^2) of the *Glide* docking scores of the known inhibitors for any two MD snapshots of the same target were also calculated (Table 3 and Figure 9a). For CDK2 and VEGFR2, the highest r_p^2 values are even higher than 0.8, but for ALK, the r_p^2 values are much lower.

Then, for each target, the docking scores of the inhibitors based on a crystal structure and a MD structure were compared. The highest r_p^2 values for ALK, CDK2, and VEGFR2 are 0.816 (4FOB and 4FOB_1510), 0.886 (3DDQ and 3DDQ_2040), and 0.803 (2XIR and 2XIR_1670), respectively. Obviously, the best correlation coefficient can only be obtained between the starting crystal structure and a MD snapshot based on this crystal structure. The reason for this interesting phenomenon was investigated. For eight representative structures of each target, the residues within 10 Å of the cocrystallized ligand were identified as the binding pockets first and then structurally aligned by using the STAMP algorithm in VMD.[46] The RMSD value for each structure pair was calculated and the phylogenetic tree based on the RMSD values was depicted by using the *create phylogenetic tree* protocol in VMD.[46] As shown in Table S2 and Figure S5 in the Supporting Information, the highest r_p^2 can be achieved for the two structures with similar binding pocket (lowest RMSD).

In addition, except for VEGFR2, most correlation coefficients for ALK and CDK2 are not high (Table 3 and Figure 9a). Apparently, MD provides some snapshots that are different from the crystal structures. Furthermore, the Spearman ranking coefficient square (r_s^2) of the *Glide* docking scores of the known inhibitors for each pair of structures were calculated (Table S3 and Figure 9b). The results demonstrate that the predictions by using different structures in docking-based VS are also quite different.

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

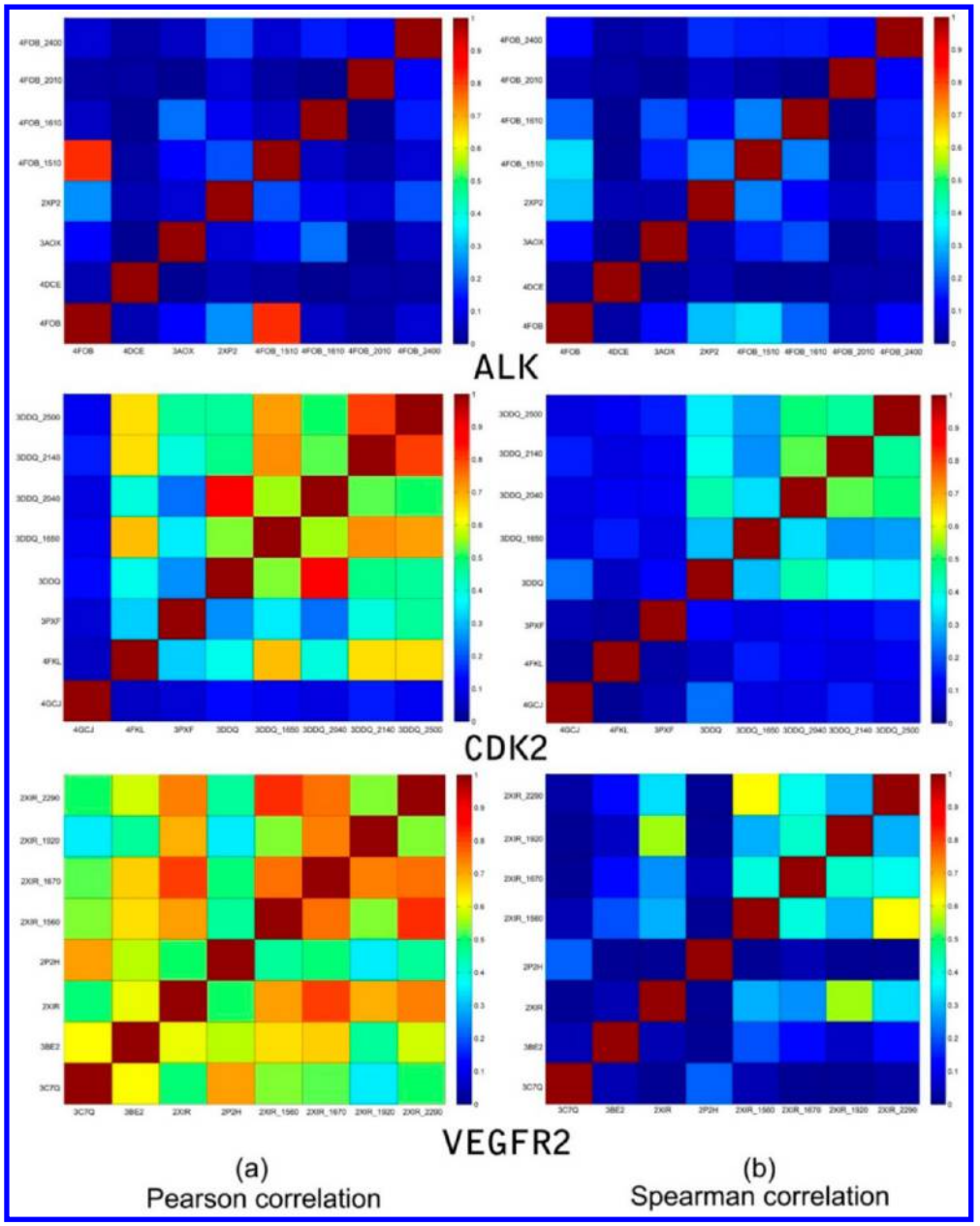


Figure 9. (a) Pearson correlation coefficient squared (r_p^2) and (b) Spearman ranking coefficient squared (r_s^2) of the *Glide* docking scores of the known inhibitors for each pair of eight representative complexes including four crystal structures and four MD snapshots for ALK, CDK2, and VEGFR2.

In summary, for ALK, the *Glide* docking scores of the known inhibitors for any structure pair (except for 4FOB and 4FOB_1510) do not show high linear correlation (r_p^2). For CDK2, low correlation is observed for any two structures, and the correlations for any two MD snapshots are slightly higher than those for any two crystal structures. For VEGFR2, compared with ALK and CDK2, the linear correlations are obviously higher. According to the above observations, we can make the following conclusion: in most cases, the predictions given by two different structures (PDB or MD structures) of the same target are not highly correlated, especially for ALK (Table 3 and Table S3 in the Supporting Information). In other words, if different complexes are used as the templates for docking-based VS, the top hits are possibly quite different. Therefore, the protein structures for molecular docking should be selected cautiously.

An Integrated VS Protocol Based on Multiple Structures. As described previously, if different representative structures are used for docking-based VS, the predicted hits may be considerably different. Moreover, it is worth noting that although a structure yields lower enrichment than others, it still can be used in VS or ensemble docking because it may represent a more unique protein conformation that is able to recognize different types of ligands. Therefore, it is not a good way to use a single structure for molecular docking while discard the others. It is quite possible that the predictions of docking based on different structures of a same target may complement each other, and therefore the integration of these results may minimize the induce-fit effect for ligand binding to the target to some extent.

In order to integrate the predictions of molecular docking based on different structures into a more versatile and intelligent way, the NBC technique was employed, based on the docking scores for the eight representative structures of the same target, including four PDB structures and four MD structures. As we mentioned in the previous studies,[42,43] some compounds in the

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

Table 4. Performance of the Naïve Bayesian Classifiers Based on the Docking Scores of Different Representative Complexes Retrieved from the PDB and MD Simulations in the Training Set[a]

	descriptor	TP	FN	FP	TN	SE	SP	GA	C	AUC
ALK	2XP2_SP	103	41	1543	3753	0.715	0.709	0.709	0.148	0.743
	3AOX_SP	79	65	934	4362	0.549	0.824	0.816	0.154	0.734
	4DCE_XP	96	48	1179	4117	0.667	0.777	0.774	0.168	0.746
	4FOB_SP	139	5	3807	1489	0.965	0.281	0.299	0.089	0.685
	Combined_PDB[b]	103	41	1117	4149	0.715	0.789	0.787	0.194	0.824
	4FOB_1510_SP	28	116	87	5209	0.194	0.984	0.963	0.199	0.632
	4FOB_1610_SP	82	62	991	4305	0.569	0.813	0.806	0.154	0.742
	4FOB_2010_XP	85	59	1350	3946	0.590	0.745	0.741	0.122	0.688
	4FOB_2400_SP	78	66	1221	4075	0.542	0.769	0.763	0.117	0.711
	$Combined\_MD_{4FOB}$[c]	110	34	1655	3641	0.764	0.688	0.690	0.155	0.797
	$Combined\_Scores_{Total}$[d]	92	52	658	4638	0.639	0.876	0.869	0.240	0.836
CDK2	3DDQ_SP	135	11	3728	1568	0.925	0.296	0.313	0.079	0.683
	3PXF_SP	139	7	4329	967	0.952	0.183	0.203	0.057	0.609
	4FKL_SP	86	60	1653	3643	0.589	0.688	0.685	0.096	0.658
	4GCJ_XP	102	44	1570	3726	0.699	0.704	0.703	0.141	0.724
	Combined_PDB[b]	97	49	1325	3971	0.664	0.750	0.748	0.152	0.777
	3DDQ_1650_SP	110	36	1873	3423	0.753	0.646	0.649	0.134	0.731
	3DDQ_2040_SP	96	50	1528	3768	0.658	0.711	0.710	0.130	0.695
	3DDQ_2140_SP	96	50	1070	4226	0.658	0.798	0.794	0.179	0.786
	3DDQ_2500_SP	90	56	1029	4267	0.616	0.806	0.801	0.169	0.768
	$Combined\_MD_{3DDQ}$[c]	99	47	960	4336	0.678	0.819	0.815	0.203	0.811
	$Combined\_Scores_{Total}$[d]	106	40	1147	4149	0.726	0.783	0.782	0.196	0.821
VEGFR2	2P2H_XP	94	47	1145	4151	0.667	0.784	0.781	0.171	0.736
	2XIR_SP	83	58	496	4800	0.589	0.906	0.898	0.255	0.777
	3BE2_XP	115	26	1269	4027	0.816	0.760	0.762	0.210	0.802
	3C7Q_XP	126	15	2919	2377	0.894	0.449	0.460	0.110	0.714
	Combined_PDB[b]	123	18	1130	4166	0.872	0.787	0.789	0.249	0.905
	2XIR_1560	96	45	893	4403	0.681	0.831	0.827	0.211	0.791
	2XIR_1670	114	27	1831	3465	0.809	0.654	0.658	0.153	0.739
	2XIR_1920	57	84	246	5050	0.404	0.954	0.939	0.248	0.725
	2XIR_2290	121	20	2619	2677	0.858	0.505	0.515	0.116	0.759
	$Combined\_MD_{2XIR}$[c]	94	47	788	4508	0.667	0.851	0.846	0.223	0.846
	$Combined\_Scores_{Total}$[d]	114	27	767	4529	0.809	0.855	0.854	0.286	0.910

[a]TP, true positive; FN, false negative; FP, false positive; TN, true negative; SE, sensitivity; SP, specificity; GA, global accuracy; *C*, Matthews correlation coefficient; AUC, area under the curve. [b]Combined_PDB represents the combination of the docking scores of four representative PDB crystal complexes. [c]Combined_MD represents the combination of the docking scores of four representative MD snapshots. [d]Combined_Scores represents the combination of the docking scores of four representative PDB crystal complexes and four MD snapshots.

validation dataset cannot be docked into the binding pockets, and a penalty docking score (PDS) of 20 was artificially assigned to each undocked compound.[43] For each target, the Bayesian classifier was built first by using the docking scores based on any of the four crystal structures. The performances of the Bayesian classifiers for the training set and test set of each target are summarized in Tables 4 and 5, respectively. Almost every Bayesian classifier based on the individual docking scores of each crystal structure can give satisfactory predictions for the test set, except for the structure 3PXF of CDK2 (AUC = 0.644).

Then, for each target, the Bayesian classifier was built by using the docking scores of any of the four MD structures (Tables 4 and 5). As listed in Table 5, almost all of the Bayesian classifiers exhibit acceptable prediction accuracies, with AUC values greater than 0.7. Furthermore, it is interesting to find that the Bayesian classifiers for some MD structures are even more predictive than those for the initial crystal structures. For example, for ALK (4FOB), CDK2 (3DDQ), and VEGFR2 (2XIR), the AUCs for the crystal structures are 0.800, 0.782, and 0.834, respectively, but those for the MD structures are 0.818 (4FOB_1610), 0.787 (3DDQ_2140), and 0.842 (2XIR_1670), respectively.

Furthermore, for each target, a Bayesian classifier by integrating the docking scores based on the four crystal structures, a Bayesian classifier by integrating the docking scores based on the four MD structures and a Bayesian classifier by integrating the docking scores of the all crystal and MD structures were developed. As shown in Table 5, for ALK, the Bayesian classifier by combining the docking scores based on the four crystal structures (AUC = 0.868) is more predictive than that based on the four MD structures (AUC = 0.847); for VEGFR2, the Bayesian classifier by combining the docking scores based on the four crystal structures (AUC = 0.876) is comparative to that based on the four MD structures (AUC = 0.873); and for CDK2, the Bayesian classifier by combining the docking scores based on the four crystal structures (AUC = 0.818) is slightly worse than that based on the four MD structures (AUC = 0.831).

By combining all of the docking scores based on eight structures of each target, the AUC values of the Bayesian classifiers for

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

Table 5. Performance of the Naïve Bayesian Classifiers Based on the Docking Scores of Different Representative Complexes Retrieved from the PDB and MD Simulations in the Test Set[a]

	descriptor	TP	FN	FP	TN	SE	SP	GA	C	AUC
ALK	2XP2_SP	47	14	643	1626	0.770	0.717	0.718	0.170	0.798
	3AOX_SP	34	27	455	1814	0.557	0.799	0.793	0.140	0.769
	4DCE_XP	34	27	414	1855	0.557	0.818	0.811	0.152	0.773
	4FOB_SP	44	17	415	1854	0.721	0.817	0.815	0.216	0.800
	Combined_PDB[b]	48	13	389	1880	0.787	0.829	0.827	0.252	0.868
	4FOB_1510_SP	48	13	896	1373	0.787	0.605	0.610	0.128	0.785
	4FOB_1610_SP	54	7	887	1382	0.885	0.609	0.616	0.161	0.818
	4FOB_2010_XP	39	22	647	1622	0.639	0.715	0.713	0.124	0.696
	4FOB_2400_SP	48	13	864	1405	0.787	0.619	0.624	0.133	0.748
	Combined_MD_{4FOB}[c]	47	14	541	1728	0.770	0.762	0.762	0.196	0.847
	Combined_$Scores_{Total}$[d]	47	14	362	1907	0.770	0.840	0.839	0.256	0.878
CDK2	3DDQ_SP	41	21	435	1834	0.661	0.808	0.804	0.187	0.782
	3PXF_SP	46	16	1110	1159	0.742	0.511	0.517	0.081	0.644
	4FKL_SP	45	17	887	1382	0.726	0.609	0.612	0.110	0.709
	4GCJ_XP	45	17	663	1606	0.726	0.708	0.708	0.152	0.746
	Combined_PDB[b]	44	18	454	1815	0.710	0.800	0.798	0.200	0.818
	3DDQ_1650	49	13	885	1384	0.790	0.610	0.615	0.131	0.763
	3DDQ_2040	39	23	653	1616	0.629	0.712	0.710	0.120	0.758
	3DDQ_2140	34	28	410	1859	0.548	0.819	0.812	0.151	0.787
	3DDQ_2500	45	17	666	1603	0.726	0.706	0.707	0.151	0.770
	Combined_MD_{3DDQ}[c]	44	18	507	1762	0.710	0.777	0.775	0.184	0.831
	Combined_$Scores_{Total}$[d]	43	19	436	1833	0.694	0.808	0.805	0.200	0.843
VEGFR2	2P2H_XP	36	25	454	1815	0.590	0.800	0.794	0.153	0.761
	2XIR_SP	50	11	639	1630	0.820	0.718	0.721	0.188	0.834
	3BE2_XP	46	15	438	1831	0.754	0.807	0.806	0.221	0.829
	3C7Q_XP	31	30	389	1880	0.508	0.829	0.820	0.140	0.711
	Combined_PDB[b]	51	10	507	1762	0.836	0.777	0.778	0.229	0.876
	2XIR_1560_XP	39	22	443	1826	0.639	0.805	0.800	0.175	0.787
	2XIR_1670_SP	53	8	652	1617	0.869	0.713	0.717	0.202	0.842
	2XIR_1920_SP	41	20	424	1845	0.672	0.813	0.809	0.194	0.809
	2XIR_2290_XP	41	20	440	1829	0.672	0.806	0.803	0.189	0.796
	Combined_MD_{2XIR}[c]	46	15	279	1990	0.754	0.877	0.874	0.291	0.873
	Combined_$Scores_{Total}$[d]	54	7	576	1693	0.885	0.746	0.750	0.227	0.894

[a]TP, true positive; FN, false negative; FP, false positive; TN, true negative; SE, sensitivity; SP, specificity; GA, global accuracy; *C*, Matthews correlation coefficient; AUC, area under the curve. [b]Combined_PDB represents the combination of the docking scores of four representative PDB crystal complexes. [c]Combined_MD represents the combination of the docking scores of four representative MD snapshots. [d]Combined_Scores represents the combination of the docking scores of four representative PDB crystal complexes and four MD snapshots.

ALK, CDK2, and VEGFR2 are 0.878, 0.843, and 0.894, respectively (Table 5 and Figure 10). That is to say, the Bayesian classifiers by integrating the predictions based on the eight representative structures are the most predictive, and they afford a practical way to balance the computational demanding and accuracy. Moreover, the results suggest that MD structures can really improve the VS results only based on X-ray structures.

As mentioned above, according to the analyses of r_p^2 of the docking scores of the known inhibitors for each representative structure pair (Table 3), we know that the docking predictions of some representative structure pair of the same target are highly correlated. Here, two structure were regarded as redundant structure if r_p^2 is higher than 0.7. The Bayesian classifiers by removing the redundant structures ($r_p^2 > 0.7$) of each target were developed (Tables S4 and S5 in the Supporting Information). For ALK, after removing a redundant structure (4FOB_1510, Table 3), the AUC value of the Bayesian classifier based on the four PDB crystal structures and three MD snapshots is 0.879, which is even higher than that based on the eight representative structures. For CDK2, four Bayesian classifiers by removing the redundant structures one by one were built. We found that the AUC value of the Bayesian classifier based on the four PDB structures and three MD snapshots was comparative to that based on the eight representative structures. For VEGFR2, by removing the redundant structures, eight Bayesian classifiers based on different representative structure combinations were also generated. The results demonstrated that the prediction capacities of three out of eight Bayesian classifiers were superior or comparative to the Bayesian classifier based on the eight representative structures. The best Bayesian classifier (AUC = 0.898) can be developed by integrating the docking scores of six structures, including 3C7Q, 3BE2, 2XIR, 2XIR_1670, 2XIR_1920, and 2XIR_2290 (Table S5 in the Supporting Information). It is apparent that, by removing the docking results of redundant structures, the performances of the Bayesian classifiers can be improved.

In most cases, if multiple structures are available, none should be discarded a prior because it seems that no system-independent

dx.doi.org/10.1021/ci500414b | J. Chem. Inf. Model. 2014, 54, 2664−2679

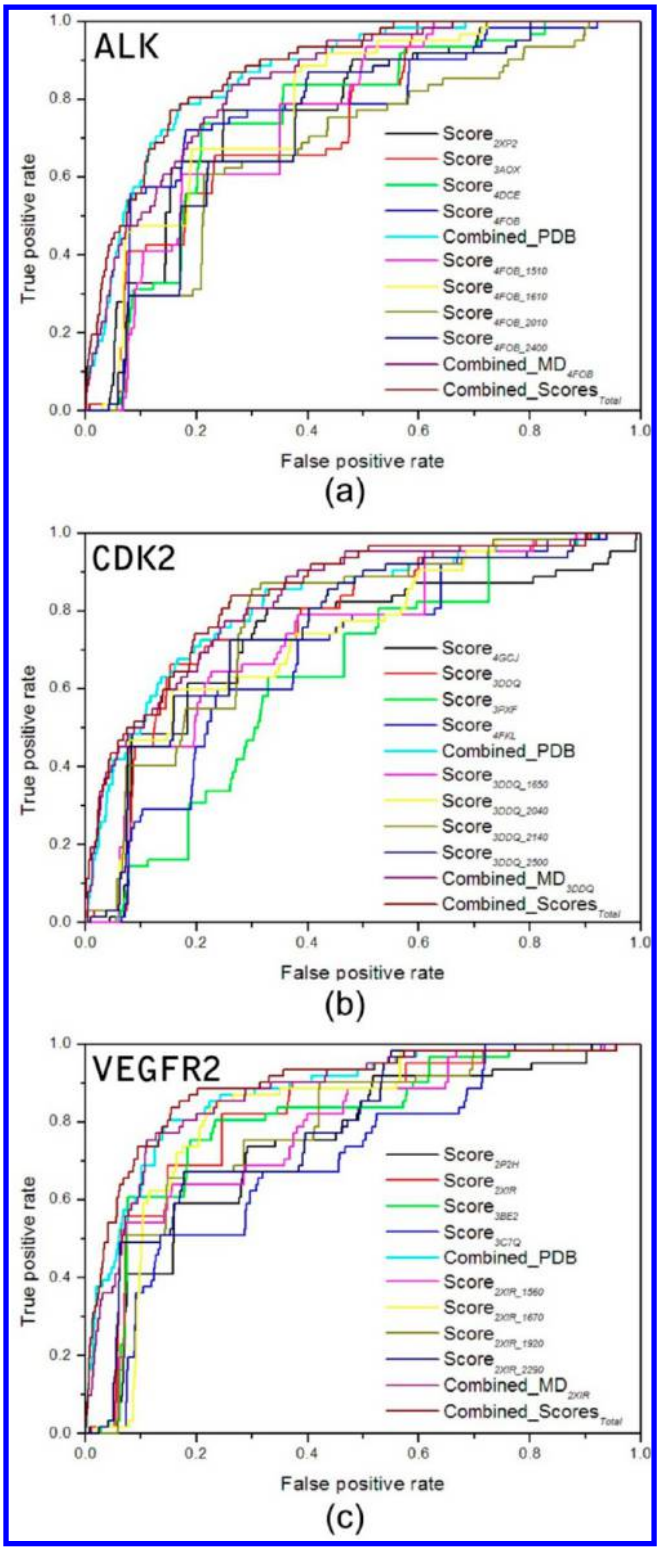


Figure 10. ROC curves of the naïve Bayesian classifiers by using the docking scores as molecular descriptors for ALK, CDK2, and VEGFR2.

method or simple guideline can be used to select the most optimal structure for VS beforehand. Certainly, we can evaluate and compare the VS performances of all available structures (crystal or MD), and identify the best one prior to a more extensive VS calculation. However, developing an integrated model by combining information from multiple structures and evaluated the prediction capacities is absolutely a better and reasonable way.

CONCLUSIONS

In this study, we compared the VS predictive powers of molecular docking based on crystal or/and MD structures for three important targets of the kinase family—ALK, CDK2, and VEGFR2. The results demonstrate that the discrimination power to separate the inhibitors from non-inhibitors is obviously different for different representative structures of each target. On average, the predictive power of the crystal structures is slightly better than (or comparable to) that of the MD structures. However, the discrimination abilities for some MD snapshots show even better prediction accuracies than the respective crystal structures. It is encouraging to observe that, by integrating the prediction results of the representative crystal and MD structures of the same target, the naïve Bayesian classifier is more predictive. Moreover, the best Bayesian classifier for each target can be developed by removing the redundant structures. The training process of NBC can incorporate protein flexibility into docking efficiently and minimize the potential to generate false positives used by most simple integration processes. The integrated classifiers can serve as powerful tools in VS to find more promising and diverse inhibitors of kinases.

ASSOCIATED CONTENT

Supporting Information

Table S1, molecular weights, molecular volumes, and solvent accessible surface areas of the cocrystallized ligands of the four representative PDB crystal structures for ALK, CDK2, and VEGFR2; Table S2, RMSDs of the residues within 10 Å of the ligands of each structure pair of the four PDB crystal structures and four MD snapshots for ALK, CDK2, and VEGFR2; Table S3, Spearman correlation coefficient squared of the docking scores of the known inhibitors predicted by docking based on the four PDB crystal structures and four MD snapshots of ALK, CDK2, or VEGFR2; Table S4, performance of the naïve Bayesian classifiers based on the docking scores of the crystal structures and MD snapshots by removing the redundant structure one by one in the training set; Table S5, performance of the naïve Bayesian classifiers based on the docking scores of the PDB crystal structures and MD snapshots by removing the redundant structure one by one in the test set; Figure S1, distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative MD snapshots by using 4FOB as the starting complex for ALK; Figure S2, distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative MD snapshots by using 3DDQ as the starting complex for CDK2; Figure S3, distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative MD snapshots by using 2XIR as the starting complex for VEGFR2; Figure S4, distributions of the *Glide* docking scores of the known inhibitors and non-inhibitors in the validation dataset for the four representative MD snapshots by using 4FKL as the starting complex for CDK2; and Figure S5, structural clustering for the residues within 10 Å of the ligand of the four crystal structures and four MD snapshots by using the *phylogenetic tree* method in VMD. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail (T.H.): tingjunhou@zju.edu.cn or tingjunhou@hotmail.com.

*E-mail (Y.L.): yyli@suda.edu.cn.
*E-mail (X.Z.): zhenxuechu@suda.edu.cn.

Author Contributions

#S.T. and H.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by the National Science Foundation of China (21173156), the National Basic Research Program of China (973 program, 2012CB932600), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Jiangsu Key Laboratory for Carbon-Based Functional Materials and Devices, and Collaborative Innovation Center of Suzhou Nano Science and Technology.

■ REFERENCES

(1) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178−184.
(2) Kitchen, D. B.; Decornez, 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.
(3) Shoichet, B. K.; McGovern, S. L.; Wei, B. Q.; Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439−446.
(4) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Design* **2002**, *16*, 151−166.
(5) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894.
(6) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580−594.
(7) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening—an overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(8) Hou, T. J.; Xu, X. J. Recent development and application of virtual screening in drug discovery: An overview. *Curr. Pharm. Des.* **2004**, *10*, 1011−1033.
(9) Bottegoni, G.; Rocchia, W.; Rueda, M.; Abagyan, R.; Cavalli, A. Systematic Exploitation of Multiple Receptor Conformations for Virtual Ligand Screening. *PLoS One* **2011**, *6*, No. e18845.
(10) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, *51*, 6237−6255.
(11) B-Rao, C.; Subramanian, J.; Sharma, S. D. Managing protein flexibility in docking and its applications. *Drug Discovery Today* **2009**, *14*, 394−400.
(12) Zhou, S.; Li, Y.; Hou, T. Feasibility of Using Molecular Docking-Based Virtual Screening for Searching Dual Target Kinase Inhibitors. *J. Chem. Inf. Model.* **2013**, *53*, 982−996.
(13) Apostolakis, J.; Pluckthun, A.; Caflisch, A. Docking small ligands in flexible binding sites. *J. Comput. Chem.* **1998**, *19*, 21−37.
(14) Mangoni, M.; Roccatano, D.; Di Nola, A. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins* **1999**, *35*, 153−162.
(15) Zhao, Y.; Sanner, M. F. Protein-ligand docking with multiple flexible side chains. *J. Comput.-Aided Mol. Design* **2008**, *22*, 673−679.
(16) Cosconati, S.; Marinelli, L.; Di Leva, F. S.; La Pietra, V.; De Simone, A.; Mancini, F.; Andrisano, V.; Novellino, E.; Goodsell, D. S.; Olson, A. J. Protein Flexibility in Virtual Screening: The BACE-1 Case Study. *J. Chem. Inf. Model.* **2012**, *52*, 2697−2704.
(17) McCammon, J. A. Target flexibility in molecular recognition. *Biochim. Biophys. Acta (BBA)—Proteins Proteomics* **2005**, *1754*, 221−224.
(18) Carlson, H. A. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(19) Cheng, L. S.; Amaro, R. E.; Xu, D.; Li, W. W.; Arzberger, P. W.; McCammon, J. A. Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J. Med. Chem.* **2008**, *51*, 3878−3894.
(20) Amaro, R. E.; Li, W. W. Emerging Methods for Ensemble-Based Virtual Screening. *Curr. Top. Med. Chem.* **2010**, *10*, 3−13.
(21) Nichols, S. E.; Swift, R. V.; Amaro, R. E. Rational Prediction with Molecular Dynamics for Hit Identification. *Curr. Top. Med. Chem.* **2012**, *12*, 2002−2012.
(22) Sinko, W.; Lindert, S.; McCammon, J. A. Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design. *Chem. Biol. Drug Design* **2013**, *81*, 41−49.
(23) Huang, S.-Y.; Zou, X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins—Struct., Funct. Bioinf.* **2007**, *66*, 399−421.
(24) Lexa, K. W.; Carlson, H. A. Protein flexibility in docking and surface mapping. *Q. Rev. Biophys.* **2012**, *45*, 301−343.
(25) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.
(26) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213−218.
(27) Dixit, A.; Verkhivker, G. M. Integrating Ligand-Based and Protein-Centric Virtual Screening of Kinase Inhibitors Using Ensembles of Multiple Protein Kinase Genes and Conformations. *J. Chem. Inf. Model.* **2012**, *52*, 2501−2515.
(28) Tan, L.; Geppert, H.; Sisay, M. T.; Guetschow, M.; Bajorath, J. Integrating Structure- and Ligand-Based Virtual Screening: Comparison of Individual, Parallel, and Fused Molecular Docking and Similarity Search Calculations on Multiple Targets. *ChemMedChem* **2008**, *3*, 1566−1571.
(29) Damm, K. L.; Carlson, H. A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* **2007**, *129*, 8225−8235.
(30) Barril, X.; Morley, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48*, 4432−4443.
(31) Meagher, K. L.; Carlson, H. A. Incorporating protein flexibility in structure-based drug discovery: Using HIV-1 protease as a test case. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.
(32) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(33) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209−225.
(34) Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632−5633.
(35) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531−568.
(36) Wong, C. F.; Kua, J.; Zhang, Y. K.; Straatsma, T. P.; McCammon, J. A. Molecular docking of balanol to dynamics snapshots of protein kinase A. *Proteins—Struct., Funct. Bioinf.* **2005**, *61*, 850−858.
(37) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Generation of Receptor Structural Ensembles for Virtual Screening Using Binding Site Shape Analysis and Clustering. *Chem. Biol. Drug Design* **2012**, *80*, 182−193.
(38) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols. *J. Phys. Chem. B* **2012**, *116*, 6952−6959.
(39) Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *J. Chem. Inf. Model.* **2010**, *50*, 186−193.

(40) Klon, A. E.; Glick, M.; Davies, J. W. Combination of a naive Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J. Med. Chem.* **2004**, *47*, 4356−4359.
(41) Chen, B.; Harrison, R. F.; Papadatos, G.; Willett, P.; Wood, D. J.; Lewell, X. Q.; Greenidge, P.; Stiefl, N. Evaluation of machine-learning methods for ligand-based virtual screening. *J. Comput.-Aided Mol. Design* **2007**, *21*, 53−62.
(42) Tian, S.; Li, Y.; Li, D.; Xu, X.; Wang, J.; Zhang, Q.; Hou, T. Modeling Compound-Target Interaction Network of Traditional Chinese Medicines for Type II Diabetes Mellitus: Insight for Polypharmacology and Drug Design. *J. Chem. Inf. Model.* **2013**, *53*, 1787−1803.
(43) Tian, S.; Sun, H.; Li, Y.; Pan, P.; Li, D.; Hou, T. Development and Evaluation of an Integrated Virtual Screening Strategy by Combining Molecular Docking and Pharmacophore Searching Based on Multiple Protein Structures. *J. Chem. Inf. Model.* **2013**, *53*, 2743−2756.
(44) Sato, T.; Honma, T.; Yokoyama, S. Combining Machine Learning and Pharmacophore-Based Interaction Fingerprint for in Silico Screening. *J. Chem. Inf. Model.* **2010**, *50*, 170−185.
(45) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(46) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.
(47) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198−D201.
(48) Krueger, D. M.; Evers, A. Comparison of Structure- and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. *ChemMedChem* **2010**, *5*, 148−158.
(49) *Discovery Studio 3.1*; Accelrys Inc.: San Diego, 2012; http://www.accelrys.com.
(50) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.
(51) *Schrödinger 9.0*; Schrödinger, LLC: New York, NY, 2009; http://www.schrodinger.com.
(52) Lewis, R. T.; Bode, C. M.; Choquette, D. M.; Potashman, M.; Romero, K.; Stellwagen, J. C.; Teffera, Y.; Moore, E.; Whittington, D. A.; Chen, H.; Epstein, L. F.; Emkey, R.; Andrews, P. S.; Yu, V. L.; Saffran, D. C.; Xu, M.; Drew, A.; Merkel, P.; Szilvassy, S.; Brake, R. L. The Discovery and Optimization of a Novel Class of Potent, Selective, and Orally Bioavailable Anaplastic Lymphoma Kinase (ALK) Inhibitors with Potential Utility for the Treatment of Cancer. *J. Med. Chem.* **2012**, *55*, 6523−6540.
(53) Bettayeb, K.; Oumata, N.; Echalier, A.; Ferandin, Y.; Endicott, J. A.; Galons, H.; Meijer, L. CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene* **2008**, *27*, 5797−5807.
(54) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(55) Frisch, M. J.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; et al. *Gaussian 09*; Gaussian. Inc.: Wallingford, CT, 2009.
(56) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(57) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(58) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. *AMBER 12*; University of California: San Francisco, 2012.
(59) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(60) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An W log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(61) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(62) Li, D.; Chen, L.; Li, Y.; Tian, S.; Sun, H.; Hou, T. ADMET Evaluation in Drug Discovery. 13. Development of in Silico Prediction Models for P-Glycoprotein Substrates. *Mol. Pharmaceutics* **2014**, *11*, 716−726.
(63) Wang, S.; Li, Y.; Wang, J.; Chen, L.; Zhang, L.; Yu, H.; Hou, T. ADMET Evaluation in Drug Discovery. 12. Development of Binary Classification Models for Prediction of hERG Potassium Channel Blockage. *Mol. Pharmaceutics* **2012**, *9*, 996−1010.
(64) Tian, S.; Wang, J.; Li, Y.; Xu, X.; Hou, T. Drug-likeness Analysis of Traditional Chinese Medicines: Prediction of Drug-likeness Using Machine Learning Approaches. *Mol. Pharmaceutics* **2012**, *9*, 2875−2886.
(65) Chen, L.; Li, Y.; Zhao, Q.; Peng, H.; Hou, T. ADME Evaluation in Drug Discovery. 10. Predictions of P-Glycoprotein Inhibitors Using Recursive Partitioning and Naive Bayesian Classification Techniques. *Mol. Pharmaceutics* **2011**, *8*, 889−900.
(66) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Design* **2008**, *22*, 133−139.
(67) Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: pitfalls and traps. *J. Comput.-Aided Mol. Design* **2008**, *22*, 179−190.
(68) Nichols, S. E.; Baron, R.; Ivetac, A.; McCammon, J. A. Predictive Power of Molecular Dynamics Receptor Structures in Virtual Screening. *J. Chem. Inf. Model.* **2011**, *51*, 1439−1446.