Optimization of High Throughput Virtual Screening by Combining Shape-Matching and Docking Methods

Hui Sun Lee,[†] Jiwon Choi,[†] Irina Kufareva,[‡] Ruben Abagyan,[‡] Anton Filikov,[§] Young Yang,[†] and Sukjoon Yoon*,[†]

Department of Biological Sciences, Research Center for Women's Diseases (RCWD), Sookmyung Women's University, Hyochangwongil 52, Yongsan-gu, Seoul, Republic of Korea 140-742, Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and ArQule, Inc., 19 Presidential Way, Woburn, Massachusetts 01801

Received October 20, 2007

Receptor flexibility is a critical issue in structure-based virtual screening methods. Although a multiple-receptor conformation docking is an efficient way to account for receptor flexibility, it is still too slow for large molecular libraries. It was reported that a fast ligand-centric, shape-based virtual screening was more consistent for hit enrichment than a typical single-receptor conformation docking. Thus, we designed a "distributed docking" method that improves virtual high throughput screening by combining a shape-matching method with a multiple-receptor conformation docking. Database compounds are classified in advance based on shape similarities to one of the crystal ligands complexed with the target protein. This classification enables us to pick the appropriate receptor conformation for a single-receptor conformation docking of a given compound, thereby avoiding time-consuming multiple docking. In particular, this approach utilizes cross-docking scores of known ligands to all available receptor structures in order to optimize the algorithm. The present virtual screening method was tested for reidentification of known PPARγ and p38 MAP kinase active compounds. We demonstrate that this method improves the enrichment while maintaining the computation speed of a typical single-receptor conformation docking.

INTRODUCTION

In recent years, virtual high throughput screening (VHTS) has become an essential technique for the discovery of new lead compounds, and it has served as an alternative to experimental high throughput screening in drug discovery. The importance of VHTS in drug discovery is increasing simultaneously with the rapidly growing number of small molecules available in corporate and public libraries.[1] A plethora of available target proteins with high-resolution crystal structures has also accelerated the development of structure-based VHTS methods. Despite recent theoretical and technical improvements in the field,[2] the performance of VHTS methods is still sometimes unsatisfactory in part due to the flexible nature of receptor conformation.[3]

VHTS methods can be classified into two categories: ligand-centric and receptor-centric virtual screening. Ligand-centric methods essentially focus on comparative analysis of the structural shape and chemical or pharmacophore similarity between compounds and known ligands. Therefore, the knowledge of experimentally selected active compounds is a prerequisite for applying ligand-centric methods.[4] On the other hand, receptor-centric methods predict interaction of a given compound with a target receptor. This does not necessarily require experimental data on active compounds. Molecular docking, which is a key method in receptor-centric virtual screening, is a technique that uses computers to predict a binding mode and affinity of a given compound for a target receptor.[5] Docking is a central component in many lead discovery strategies.[6]

A critical issue in receptor-centric virtual screening is to incorporate a dynamic nature of receptor structures. Commonly in molecular docking algorithms, the target protein is kept rigid in a single low-energy conformation, and only conformational and positional flexibility of a ligand is considered. Proteins, however, can have different conformational states with similar energies. In many cases binding site conformation of a receptor exhibits significant motion including rearrangements of side chains and backbone upon ligand binding. This is called 'an induced fit'.[7] Even small local motions of side chains may significantly impact docking results.[8] Therefore, using a single receptor conformation in docking experiments can lead to errors in identification of binding modes and errors in prediction of binding affinities. This can significantly reduce the chances of finding new ligands.[9] In such a flexible system no clear relationship between docking and ranking was found.[10]

There were various attempts to include protein flexibility in the virtual molecular docking procedure. A simple approach is to reduce the van der Waals radii of the receptor and/or ligand atoms or delete some of the side chains in order to eliminate possible close contacts due to rigidity of the receptor conformation.[11] Another approach is to use an ensemble of experimental receptor conformations in ligand docking. Knegtel et al. used crystal and solution structures to generate combined interaction grids by averaging with

* Corresponding author phone: 82-2-710-9415; fax: 82-2-2077-7322; e-mail: yoonsj@sookmyung.ac.kr.
[†] Sookmyung Women's University.
[‡] The Scripps Research Institute.
[§] ArQule, Inc.

10.1021/ci700376c CCC: $40.75
Published on Web 02/27/2008

Table 1. Comparison of Representative Run Times of the Ligand-Centric and the Receptor-Centric Method for p38 Compounds (Actives + Decoys) Used in This Study[a]

	ROCS (min)	docking (min)
1A9U	1.4	362.3
1BL6	1.3	294.6
1BL7	1.5	329.4
1BMK	1.5	284.7
1DI9	1.5	299.7
1KV1	1.4	281.9
1KV2	1.7	320.7
1M7Q	1.5	359.8
av	**1.5**	**316.6**

[a] The measured values of the ligand-centric and the receptor-centric method were obtained from ROCS and GLIDE docking runs, respectively (system information: Red Hat Linux 3.4.4, 3.2 GHz Xeon CPU).

respect to energy and geometry.[12] Shrödinger Induced Fit docking uses an iterative docking protocol which accounts for both ligand and receptor flexibility.[13] In this method, a ligand is docked into a rigid receptor using reduced van der Waals radii, and then the receptor side-chain conformations are sampled for each receptor−ligand binding pose, followed by an energy minimization and a second round of docking. The ICM-flexible receptor-docking algorithm (IFREDA) generates a discrete set of receptor conformations through seeding, soft van der Waals structure relaxation and energy optimization processes, thus sampling the conformational space of the receptor. The de novo generated structures and the native structures are used for the virtual screening, and then the docking results are condensed with a merging and shrinking step.[14] In our previous studies, attempts were made to use a large ensemble of receptor conformations obtained from molecular dynamics simulation (MD) to select a minimal subset of receptor conformations for virtual screening.[15] The subset is selected based on a correlation between the experimental binding affinities and the docking scores obtained from a preliminary docking experiment on a small number of known active compounds.

Despite previous efforts, the receptor-centric virtual screening methods that accommodate structural flexibility of both ligand and receptor are still too computationally expensive for a large library of compounds. Although the merge-and-shrink method using multiple receptor conformations generated de novo or sampled from a set of ensembles is useful in improving the virtual screening effectiveness, the receptor conformations might be not reliable. Thus, it is concerned that the additional receptor conformations may increase the "false positive" rate in enrichment studies. If a number of cocrystal structures bound with various ligands are available for a target protein, it may be an alternative and/or more convenient approach to incorporate conformational flexibility of the receptor in docking. The cognate ligands in the cocrystal structures can be used for ligand-centric virtual screening as well.

Ligand-centric methods are much faster than receptor-centric ones (Table 1). Recently, Hawkins et al. reported that a ligand-centric shape-based virtual screening approach was often superior to a typical single-receptor conformation docking.[16] In this study, single-receptor conformation docking provided inconsistent performance, although the GLIDE docking engines showed an equivalent consistent performance when compared to the ligand-centric method. Generally, docking using multiple conformations of a receptor improves the rank-order performance significantly.[15]

Table 2. List of Crystal Structures of (A) PPARγ and (B) p38 MAP Kinase (p38) and Their Cognate Ligands[a]

PDB	resolution (Å)	ligand	ligand class
		(A)PPARγ	
1FM6	2.10	rosiglitazone (BRL)	1
1I7I	2.35	tesaglitazar (AZ2)	2
1KNU	2.50	carbozol ethoxy phenyl propionic acid (YPA)	3
1NYX	2.65	ragaglitazar (DRF)	4
2ATH	2.28	BPR1H036 (3EA)	5
2G0H	2.30	pyrazol-5-ylbenzenesulfonamide derivative (SP3)	6
2GTK	2.10	indole propionic acid derivative (2O8)	7
2HFP	2.00	*N*-sulfonyl-2-indole carboxamide derivative (NSI)	8
		(B) p38	
1A9U	2.50	SB203580 (SB2)	1
1BL6	2.50	SB216995 (SB6)	2
1BL7	2.50	SB220025 (SB4)	3
1BMK	2.40	SB218655 (SB5)	4
1DI9	2.60	4-anilinoquinazoline derivative (MSQ)	5
1KV1	2.50	4-naphthalen-urea derivative (BMU)	6
1KV2	2.80	4-phenyl-urea derivative (B96)	7
1M7Q	2.40	dihydroquinazolinone derivative (DQ0)	8

[a] The ligand class was used for the classification of active and decoy compounds based on the shape similarity to crystal ligands (see the Methods section for details).

Ligand- and receptor-centric virtual screening methods can be complementary to each other. Thus, a well-tuned combination of these methods may improve the virtual screening effectiveness. In this study, we present a novel docking method that integrates ligand- and receptor-centric virtual screening methods to incorporate receptor flexibility while maintaining the time efficiency of rigid receptor docking. Our method focuses on optimizing the virtual screening effectiveness for a system where multiple experimental receptor−ligand cocrystal structures are available. We developed a "distributed docking" method that uses the advantages of shape matching and multiple-receptor conformation docking methods, with the goal of improving virtual high throughput screening.

MATERIALS AND METHODS

The coordinates for all protein−ligand complexes were obtained from the RCSB Protein Data Bank (PDB).[17] Eight PPARγ ligand-binding domains (PPARγ) and eight p38 MAP kinase (p38) crystal structures bound with different ligands were selected for the present study (Table 2).

The two test sets include collections of ligands known to have binding activities on PPARγ and p38, respectively. We retrieved a total of 119 active compounds for PPARγ and 134 active compounds for p38 from the MDL Drug Data Report (MDDR) database, respectively.[18] In addition, two sets of decoys, which have no known activity on the protein targets, were prepared from the NCI compound library.[19] Recently a directory of useful decoys has been reported, and the procedure for preparing unbiased decoy sets was well described.[20] In the present study, a total of 126 705 compounds in the SD format were downloaded from the NCI Web site. Compounds containing any atoms other than H, C, N, O, F, S, Cl, and Br and improper functional groups

were first removed using FILTER.[21] We then generated 3D energy-minimized conformations for the retained compounds using LigPrep.[22] In the LigPrep operation, the information on chiralities was kept in the input files, and the ionization states were calculated at pH 7.0 without additional generation of the tautomer forms. These compounds were filtered to satisfy Lipinski's rules and additional criteria such as molecular weight (Mw) and solvent-accessible volume (Vs). Specifically, the filters were as follows:

$$\text{Mw: } \geq 300 \text{ and } \leq 600 \text{ for PPAR}\gamma; \quad \geq 200 \text{ and } \leq 650 \text{ for p38}$$

$$\text{Vs: } \geq 1000 \text{ and } \leq 2100 \text{ for PPAR}\gamma; \quad \geq 600 \text{ and } \leq 1800 \text{ for p38}$$

$$\text{number of Lipinski violations: } \leq 2$$

The above properties were calculated from two sets of actives using QikProp,[23] and the threshold values were set up to make the property distribution similar between the active and decoy sets.

To make test compounds unique in comparison to ligands cocrystallized with their target protein in PDB structures (crystal ligands), those with a Tanimoto similarity of >0.5 to any of known crystal ligands were removed from both the active and decoy sets. In addition, to make both active and decoy compounds diverse and not biased to a specific part of chemical space, we eliminated the neighboring compounds with a Tanimoto similarity of >0.7 within the sets. The Tanimoto similarity was calculated using a DayLight fingerprint descriptor.[24] We constructed a 2048-bit vector representation for each compound and calculated the Tanimoto similarity between all the possible compound pairs using the 0/7 path (defaults in the DayLight toolkit). As a result, 77 compounds were prepared as actives and 3317 compounds as decoys for PPARγ. For p38, 111 compounds were prepared as actives and 6450 compounds as decoys.

For ligand-centric shape comparison between crystal ligands and the prepared test set compounds, multiple conformers of the actives and decoy compounds were generated using the OMEGA program (version 2.1.0).[25] The maximum number of rotatable bonds was set large enough to accept all the compounds. A maximum of 200 conformers were allowed for each compound based on default rmsd cutoff (0.8 Å) and energy window (25 kcal/mol). McGaughey et al. examined the enrichment of conformer libraries taken from the in-house flexibases (using maximum conformers of 25 and 100) and the native-OMEGA (conformers of 2000), suggesting that the number of conformations do not systemically affect the enrichment.[26] It suggests that the maximum number of conformers used in our study is a reasonable selection for the ligand-centric shape comparison. In the case of the crystal ligands, their crystal structures were used for shape comparison without further conformation generation. Pairwise 3D shape similarities of the generated conformers against the crystal ligands were quantitatively calculated using the ROCS program (version 2.2).[27] ROCS (Rapid Overlay of Chemical Structures)[27] is a program designed to perform large scale structural comparison by using a superposition method. ROCS compares a Gaussian-based overlap parametrized to reproduce hard-sphere volumes between two molecules. As both the shape and electrostatic components of a molecule are critical for the observed biological activity, an overlap of functional groups was employed with properties such as hydrogen bond donors and hydrophobic groups as well as shape-based matching. The degree of structural similarity between two structures was calculated using the sum of the shape Tanimoto value and the scaled color value, ranging from 0 to 2, where 2.0 represents an exact match of both shape and functional groups between two molecules. To measure the chemical complementarity, we used the ImplicitMillsDean color force field, which is a simple pK_a model defining cations, anions, donors, and acceptors. In our test compounds, the lowest ROCS similarity score of an active compound against a crystal ligand was 0.85. Thus, we applied this cutoff (0.85) for filtering the decoy sets and removed decoy compounds with a ROCS similarity score of "<0.85" against a crystal ligand from the two test sets. As a result, 883 and 978 compounds were finally prepared as the decoy sets for PPARγ and p38, respectively.

GLIDE (version 4.0) and ICM (version 3.3) were used for the docking studies. Prior to the molecular docking, PDB structures of protein targets were prepared using the Schrödinger software package. All water molecules were removed, and multimeric complexes were simplified from the PDB structure. If a PDB structure had any missing side-chain atoms, Prime[28] was used to predict their location.

For GLIDE docking, receptor structures were preprocessed using protein preparation and refinement components in the GLIDE docking package.[29] Hydrogen atoms were added by using the all-atom force field. Side chains that were not close to the ligand-binding site and did not participate in salt bridges were neutralized. A restrained minimization using the OPLS-AA force field was performed for the refinement of the complex structure. This procedure reorients side-chain hydroxyl groups and alleviates potential steric clashes. This minimization continued until an average rms deviation of the non-hydrogen atoms reached the specified limit of 0.3 Å. Then a set of grids for each receptors was generated. The GLIDE docking algorithm performed a series of hierarchical searches for possible locations of the ligand in the binding site region of the receptor. The details of the GLIDE docking methods are described elsewhere.[30] All docking calculations were run in the "Standard Precision" (SP) mode of GLIDE. All the procedures including protein preparation, refinement, grid generation, and docking were performed using default GLIDE parameters.

For ICM docking[31] the initial structures were converted to ICM objects, and grid maps were calculated with a grid spacing of 0.5 Å. Docking was performed with default docking parameters. During the docking, either one of the torsional angles of the ligand was randomly changed or a pseudo-Brownian move was performed. Each random change was followed by 100 steps of local conjugate-gradient minimization. The new conformation was either accepted or rejected according to the Metropolis rule using a temperature of 600 K. The number of Monte Carlo steps in the docking run as well as the length of local minimization was determined automatically by an adaptive algorithm depending on the size and number of flexible torsions in a ligand. Further details of the ICM docking method are described elsewhere.[32]

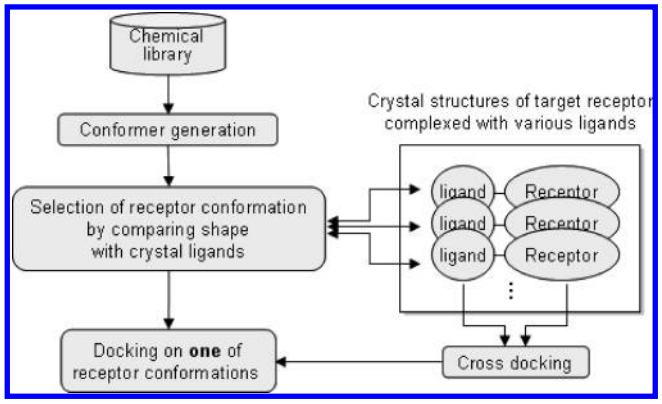

Figure 1. Flowchart of distributed docking method.

To quantitatively compare the performance of various virtual screening methods, our study analyzed the prediction results using receiver operating characteristic (ROC) curves.[33] A ROC curve describes the effectiveness of a given virtual screening method as its ability to avoid "false negatives" (i.e., where a molecule classified as "inactive" is really "active") and "false positives" (i.e., where a molecules classified as "active" is really "inactive"). The effectiveness of the virtual screening method can be quantified by the area under the ROC curve (AUC). An AUC value of 1.0 represents a theoretically perfect performance, while a 0.5 AUC value implies a random performance.

RESULTS AND DISCUSSION

The procedure of the proposed distributed docking method is illustrated in Figure 1. First, multiple 3D conformations were generated for each of the library compound sets (actives and decoys), and a shape comparison was performed with the crystal structure of known ligands. Each library compound was then assigned to a ligand class, predefined in Table 2, based on the shape similarities between the library compound and the class-representing crystal ligands. The chemical structure of the class-representing crystal ligands is displayed in Figures 2 and 3. The receptor structure for docking of a library compound was then selected based on the assigned ligand class. In order to optimally assign a receptor structure to each ligand class, preliminary cross docking of all class-representing crystal ligands to all crystal receptor structures was carried out, and the best scoring pairs were identified in advance.

For the comparisons of performance of the proposed distributed docking method with other virtual screening methods, we carried out the following procedures in parallel using the same test compound sets.

1) Ligand-centric shape comparison − library compounds were sorted by their shape similarity to the crystal ligands. The AUC of sorted compounds was calculated based on their ROCS similarity score.

2) Single-receptor conformation docking − library compounds were docked to a single receptor structure and sorted using the docking score. Since the experiment had eight crystal structures for each test receptor, single-receptor conformation docking was repeated eight times using all structures, and the average performance (i.e., average AUC) was calculated for comparison with other methods.

3) Multiple-receptor conformation docking − each library compound was docked to all available receptor conformations− the same as the single-receptor conformation docking. In this method, however, only the best docking score among all the single-receptor conformation docking results was selected for each compound and used for sorting the compounds.

4) Structure-based distributed docking − each library compound was assigned to a ligand class where the compound showed a higher shape similarity with the class representing crystal ligand than other crystal ligands. Then, the compound was docked to the cognate receptor structure of the class-representing ligand.

5) Score-based distributed docking − each library compound was assigned to a ligand class in the same way as the above structure-based distributed docking method. Then, the compound was docked to the receptor structure that produced the best docking score with the class-representing ligand among all crystal receptor structures. Thus the selected receptor structure may be noncognate to the class-representing ligand.

To demonstrate the present distributed docking method, 16 crystal structures of PPARγ and p38 cocrystallized with diverse ligands (Table 2) were used for the preliminary cross docking. The ligand class was defined for each crystal ligand. The chemical structure of the class-representing crystal ligands is displayed in Figures 2 and 3. Pairwise shape similarity among the crystal ligands is summarized in the hierarchical trees (Figure 4).

In the preliminary cross docking, each crystal ligand was docked to all crystal receptor structures, and the best scoring ligand−receptor pair was identified (Table 3). In the case of PPARγ, only three crystal ligands (1, 6, and 8) generated the best docking score with cognate receptor structures (1FM6, 2G0H, and 2HFP), while the rest of the eight ligands found the best scoring match from noncognate structures. In those latter cases ligands with similar 3D shapes generated the best docking score with a common receptor structure. Ligands 2, 3, 4, and 7, which are found in two neighboring clusters (Figure 4A), showed the best docking results with a common receptor structure (i.e., 1I7I by GLIDE docking and 1KNU by ICM docking). These results imply that the pairwise 3D shape similarity between crystal ligands and the library compounds measured by ROCS can provide a useful tool for assigning an appropriate receptor structure for the docking of library compounds. In the case of p38 and its ligands, five ligands found the cognate receptor structure to be the best scoring match, while three ligands found the best scoring match from the noncognate structures (Table 3B). Although the GLIDE and ICM results were not consistent in identifying the best docking matches for PPARγ test set, the two docking methods were relatively consistent in identifying the best-scoring receptor structure for each ligand. In general, results in Table 3 show that the cognate receptor structure does not always provide the best docking score to the crystal ligand in docking procedures. In addition, the reproduction of crystal poses by docking generally did not correlate with docking scores. Cognate receptor structures, which reproduce accurate crystal poses of the ligand in the docking procedure, may not generate better docking scores than noncognate receptor structures (Table 4). The calculated RMSDs did not show any meaningful correlation with the resolution of the crystallographic structures or the atom occupancy of crystal ligands (Tables 2 and 4). Many docked poses with a rmsd of less than 1.0 Å in the cognate ligand−

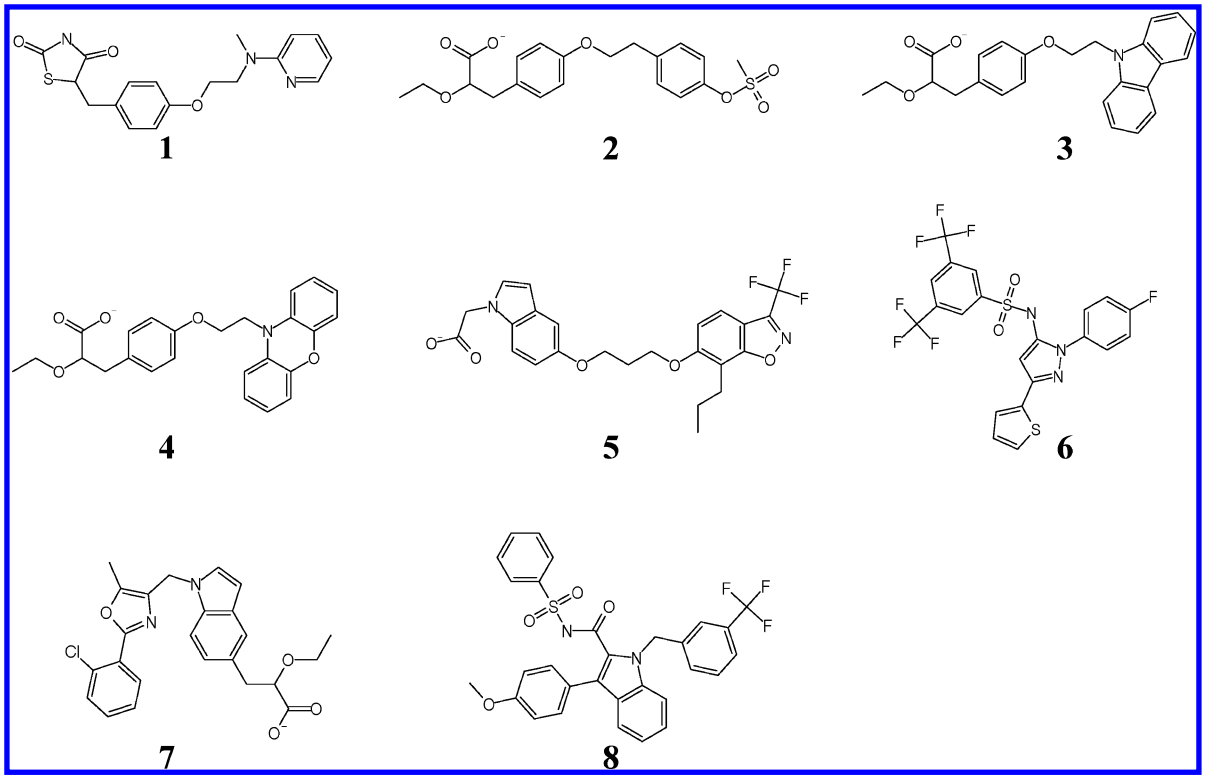

Figure 2. The chemical structure of crystal ligands found in the PPARγ structures. The number represents the ligand class defined in Table 1A.

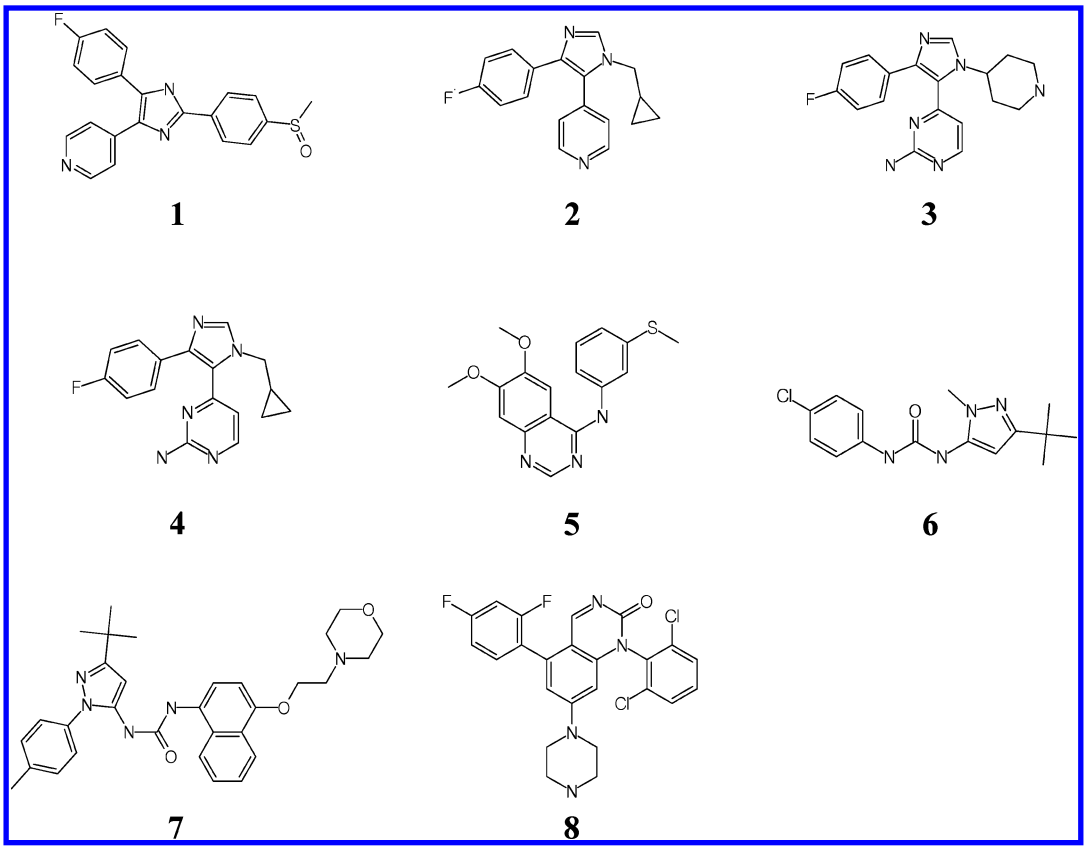

Figure 3. The chemical structure of crystal ligands in the p38 structures. The number represents the ligand class defined in Table 1B.

receptor pairs did not show the best docking score in the cross-docking procedure. The worst result found in the crystal ligand 1 of p38 (rmsd=8.29 Å) was due to an inappropriate tautomeric state of the ligand in docking. Once the conformer was changed to the alternative tautomeric state (Figure 5), the crystal ligand was properly docked to the same receptor structure with an rmsd of 0.83 (data not shown). Thus, the present preliminary cross-docking results provide a useful guideline for selecting appropriate receptor structures for docking and for the scoring of library compounds which are

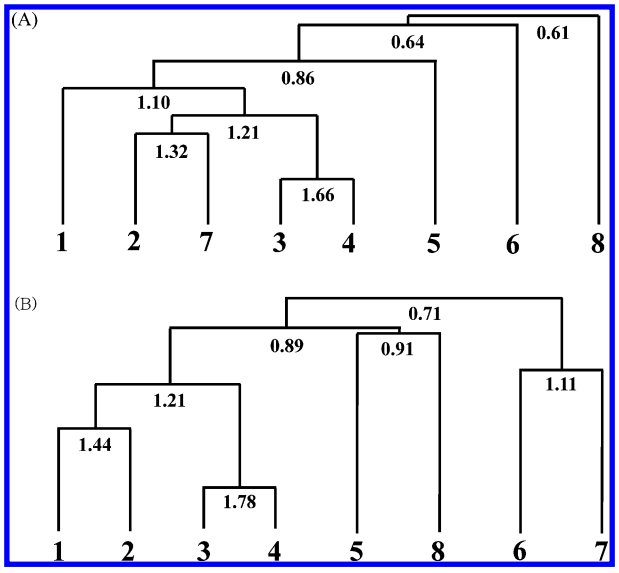

Figure 4. 3D shape similarity of the crystal structure among selected (A) PPARγ and (B) p38 ligands. The hierarchical tree was built using pairwise ROCS similarity scores (indicated in each branching point) in the range of 0−2 (see the Methods section for details). The length of vertical lines is proportional to "2-ROCS score". Bottom numbers represent ligand classes defined in Table 1. The trees are constructed using the UPGMA (Unweighted Pair Group Method Using Arithmetic Average) method, which builds a tree by sequential clustering of the smallest pairwise distance in the distance matrix.

Table 3. Identification of the Best-Scoring Receptor Structure for Each Crystal Ligand by Cross-Docking: A Total of Eight PDB Structures of (A) PPARγ and (B) p38 MAP Kinase Were Cross-Docked with Their Crystal Ligands[a]

(A)

PPARγ ligands	1FM6	1I7I	1KNU	1NYX	2ATH	2G0H	2GTK	2HFP
1	○●							
2		○	●					
3		○	●					
4		○	●					
5	●				○			
6						○●		
7		○	●					
8								○●

(B)

p38 ligands	1A9U	1BL6	1BL7	1BMK	1DI9	1KV1	1KV2	1M7Q
1			●				○	
2			○●					
3			○●					
4				○●				
5					○●			
6							○●	
7							○●	
8								○●

[a] A gray box represents the docking match between a PDB structure and its cognate ligand. The best-scoring match for each ligand is indicated with white (GLIDE docking) and black (ICM docking) circles. The ligand numbers of each target protein are the same as shown in Table 1.

classified in advance based upon their shape similarities to the crystal ligands.

Vigers and Rizzi showed in their study that while the docking programs can reproduce the crystal conformations of ligands bound to their cognate receptors with a high probability of success, they tend not to provide the highest docking scores when comparing multiple potential ligands. They attribute this problem to ligand-dependent biases in the scoring function of docking programs.[34] The biases of the docking scores in the cross-docking matrices in this study are shown in Table 3. In addition, our cross-docking results and shape similarity analysis showed that a set of ligands with high shape similarity may identify a common noncognate receptor structure as the best fit, although a crystal ligand generally identified the cognate receptor structure as the best fit in the docking and scoring practice. This implies that the pairwise 3D shape similarity between crystal ligands and library compounds can provide a useful tool for assigning an appropriate receptor structure for each of the library compounds and for improving the docking results. Thus, in the present study we classified each compound in the test sets into predefined classes based on shape similarities to crystal ligands. We confirmed that test compounds were not significantly biased to any specific classes of crystal ligands (data not shown). As described in the Methods section, all test compounds had a less than 0.5 Tanimoto similarity in chemical topology against any of the crystal ligands. Although known actives are assumed to have no significant chemical similarity to any of the crystal ligands, in general their shape similarity score measured by ROCS to crystal ligands was still higher than the decoy compounds. For the PPARγ test compounds, after shape comparison and classification, the average shape similarity score of the actives against the crystal ligands was 1.10, while the similarity score of the decoys was 0.95. For p38, the scores were 1.28 and 1.06 for actives and decoys, respectively. These results indicate that the measure of the shape similarity itself has a discriminating ability to determine active from decoy compounds.

Rank-ordering of library compounds is a major application of virtual screening methods. The virtual screening performance measured by the AUC score is compared for various docking and shape-based methods (Table 5 and Figure 6). When the compound library was sorted by the shape similarity score (ligand-centric shape comparison method), the AUC score was 0.83 and 0.82 for PPARγ and p38, respectively. For comparison with a typical docking method,

Table 4. RMSD (Å) between Docked and Crystal Poses of PPARγ and p38 Ligands[a]

	PDB ligand							
	1	2	3	4	5	6	7	8
PPARγ	0.23	1.70	0.86	0.41	1.52	0.24	2.89	0.40
p38	8.29	1.36	0.82	0.36	0.22	0.44	0.21	0.48

[a] GLIDE docking was carried out using the cognate receptor structure for each crystal ligand. A gray box indicates that the cognate receptor structure produced the best docking score for a given ligand compared with noncognate structures.

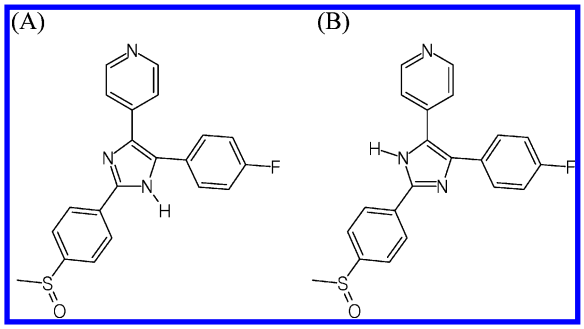

Figure 5. Two tautomeric states of the p38 crystal ligand 1. The RMSDs of the docked poses for tautomeric states (A) and (B) against the crystal pose were 8.29 and 0.83, respectively.

Table 5. Comparison of AUC Values among Various Virtual Screening Methods[a]

		SC-docking	Dist-docking A	Dist-docking B	MC docking	ROCS-shape
GLIDE docking	PPARγ	0.79	0.80	0.87	0.91	0.83
	p38	0.64	0.65	0.74	0.81	0.82
ICM docking	PPARγ	0.61	0.62	0.71	0.77	0.83
	p38	0.70	0.78	0.77	0.88	0.82

[a] The area under curve (AUC) values calculated from the ROC plots of the docking results are used for comparison of virtual screening performance. SC-docking represents single-receptor conformation docking. The AUC of SC-docking is the average of eight repeats of single-receptor conformation docking. Dist-docking A and B represent structure-based distributed docking and score-based distributed docking, respectively. MC-docking represents multiple-receptor conformation docking.

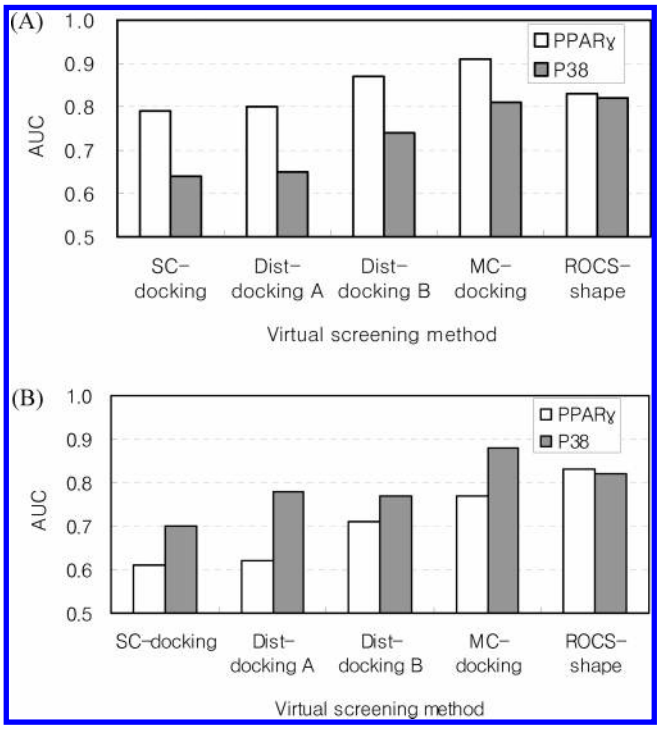

Figure 6. Comparison of various virtual screening methods. The performance of various virtual screening methods was compared on prioritizing ligands of PPARγ and p38 kinase, respectively. The area under curve (AUC) values calculated from ROC plots of the docking results are used for comparison of virtual screening performance. (A) Various docking procedures were carried out using GLIDE. (B) Various docking procedures were carried out using ICM. SC-docking represents single-receptor conformation docking. The AUC of SC-docking is the average of eight repeats of single-receptor conformation docking. Dist-docking A and B represent structure-based distributed docking and score-based distributed docking, respectively. MC-docking represents multiple-receptor conformation docking.

all active and decoy compounds were docked into each of the receptor conformations. As a result, GLIDE produced an average AUC of 0.79 and 0.64 in the single-receptor conformation dockings for PPARγ and p38, respectively. The performance of the ICM program was 0.61 for PPARγ and 0.70 for p38. Although a few cases of single-receptor conformation docking showed better AUC scores than the shape comparison method, the average performance was commonly inferior to that of the shape comparison method. Hawkins et al. reported that the ligand-centric shape comparison was more consistent than, and often superior to, typical single-receptor conformation docking in the rank-ordering of a compound library.[16] The large variation in the AUC scores among individual single-receptor conformation dockings also indicated that the appropriate receptor conformation was a critical factor for the success of virtual screening by the docking and scoring method. It is well-known that rigid single-receptor conformation does not accommodate the docking of compounds with various shapes which requires a conformational shift in the binding pocket. In the case of estrogen receptor-alpha (ERα), it has been reported that the knowledge-based selection of ERα conformation using known active compounds significantly improves the docking and scoring results.[15] However, so far no simple and clear guidelines have been suggested for optimal selection of receptor conformation for docking and scoring.

The performance of multiple-receptor conformation docking method is comparatively analyzed in Figure 6. The best docking score of each test compound was determined from all the parallel runs using different receptor structures, followed by merging and picking the best scores. This method showed an AUC value of 0.91 and 0.81 for PPARγ and p38, respectively, from GLIDE docking (Table 5). In the ICM docking, the AUC was 0.77 for PPARγ and 0.88 for p38, respectively. The multiple-receptor conformation docking methods significantly improved the AUC values in comparison to single-receptor conformation docking. In only one out of four test cases, was the performance of the multiple-receptor conformation docking worse than the shape-comparison method. Multiple-receptor conformation docking for PPARγ and p38 compounds by using the GLIDE and ICM methods, respectively, shows significantly higher AUCs than those from the ROCS method (Figure 6A,B). GLIDE docking of p38 compounds showed a similar performance to the ROCS method.

In the flexible system single-receptor conformation dockings did not provide a consistent virtual screening performance. The ligand-centric shape-comparison method using multiple crystal ligands showed a significantly enhanced hit enrichment than the average performance of the single-receptor conformation dockings (Figure 6). In addition, the shape comparison was generally much faster than the docking methods (for instance, the average computing times were 0.5 min for ROCS shape matching per crystal ligand and 32.9 min for GLIDE single-receptor conformation docking for the active compound set of p38 in our test machine). In multiple-receptor conformation docking the computing time increases linearly with the number of receptor structures. However, it significantly improved the virtual screening performance in comparison not only to the single-receptor conformation docking but also, in many cases, to the ligand-centric shape comparison methods. Thus, in order to achieve a compromise between time and performance in virtual screening our study attempted to combine these two methods into a distributed docking method.

In the structure-based distributed-docking method a receptor structure that was the crystal counterpart of the class representing ligands was assigned to all compounds in the same ligand class. However, this method did not significantly improve the AUC score in comparison with single-receptor conformation docking except for the case of ICM docking for p38 (Figure 6). On the other hand, in the score-based

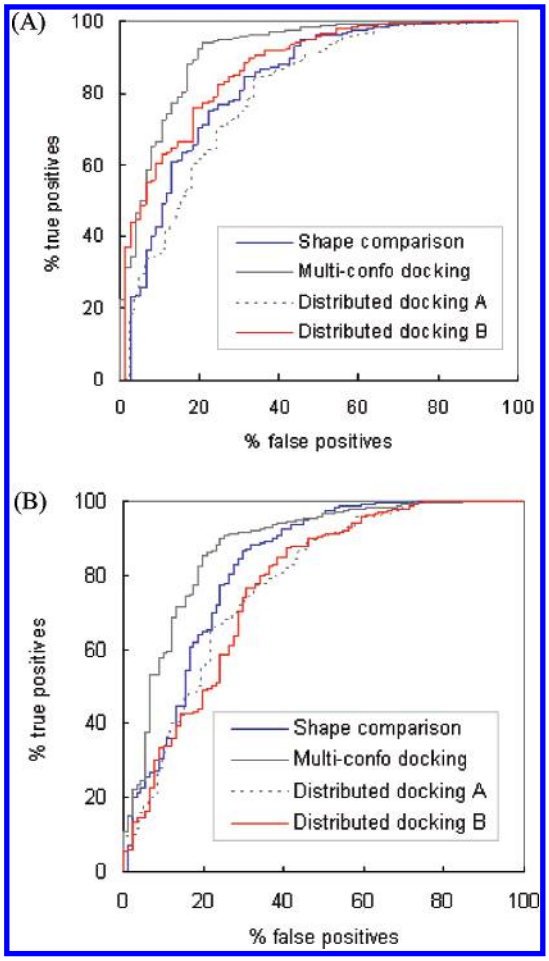

Figure 7. ROCS plot validation of various virtual screening methods on (A) PPARγ and (B) p38. The docking was carried out using (A) GLIDE and (B) ICM. ROC plots represent the enrichment of active compounds against decoys in the screening library. "Distributed docking A" represents the structure-based distributed docking, and "Distributed docking B" represents the score-based distributed docking.

distributed docking a receptor structure that showed the best docking score with the class representing crystal ligand was assigned to compounds in the same ligand class. In this case the virtual screening performance measured by the AUC was somewhere between the single-receptor conformation docking and the multiple-receptor conformation docking methods (Figure 6). Notably, the AUC of score-based distributed docking was higher than (similar to) that of the ligand-centric method in GLIDE docking of PPARγ compounds (ICM docking for p38 compounds). This observation can be interpreted as follows: when multiple-receptor conformation docking showed a better performance than the shape-comparison method, the score-based distributed docking was also superior (or similar) to the shape-comparison method. In our test cases GLIDE docking worked relatively well with PPARγ, while ICM docking worked well with p38. However, the shape-comparison method ROCS showed a consistent AUC ($\sim$0.8) score over all test cases. Thus, it would be critical to select an appropriate docking engine for the optimization of structure-based virtual screening.

For these two cases with relatively good docking results (i.e., GLIDE docking of PPARγ compounds and ICM docking for p38 compounds), an ROC plot of AUCs for various methods was drawn and comparatively analyzed. In general, time-consuming multiple-receptor conformation docking consistently showed excellent performance in the hit enrichment in all ranges of the sampling size (Figure 7A,B). When the sampling size was relatively small including less than 10% of all inactive compounds, the score-based distributed docking showed similar performance in the hit enrichment to multiple-receptor conformation docking and to shape-comparison methods. In a practical application of high throughput virtual screening in a drug discovery project the sampling size for experimental screening is relatively small in comparison to the size of entire libraries. In this sense the present distributed-docking method, which requires a computational time similar to single-receptor conformation docking, has a great advantage over multiple-receptor conformation docking in accelerating virtual screening.

CONCLUSION

When a crystal structure of a target receptor is available, molecular docking is typically used as a virtual screening method. Commonly, however, rigid receptor docking is limited in enriching databases with true inhibitors since this method does not include receptor flexibility in ligand docking. If several receptor cocrystal structures bound with different ligands are available, multiple-receptor conformation docking can be a useful strategy to improve hit enrichment. However, multiple-receptor conformation docking is very time-consuming for large screening libraries. The major challenge in docking is how to efficiently compromise speed with ability to accurately prioritize compounds. In this study we present a novel docking method termed distributed docking. It is a receptor-centric virtual screening optimized by combining docking with ligand-centric shape comparison methods. The method consistently improves screening performance, while maintaining the high speed of a rigid receptor docking. We also show that scoring and ranking obtained from the compound classification procedure using shape comparison only often yields a better virtual screening performance than rigid receptor docking.

ACKNOWLEDGMENT

This work was supported by the SRC/ERC program of MOST/KOSEF (R11-2005-017-01003-0) and by grant No. R01-2006-000-10515-0 from the Basic Research Program of the Korea Science & Engineering Foundation. This work was also supported by a Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2006-311-C00582).

Supporting Information Available: PPARγ and p38 active and decoy sets. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Yoon, S.; Smellie, A.; Hartsough, D.; Filikov, A. Surrogate docking: structure-based virtual screening at high throughput speed. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 483−497.

(2) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580−594.

(3) Carlson, H. A. Protein flexibility is an important component of structure-based drug discovery. *Curr. Pharm. Des.* **2002**, *8*, 1571−1578.

(4) Stahura, F. L.; Bajorath, J. Virtual screening methods that complement HTS. *Comb. Chem. High Throughput Screening* **2004**, *7*, 259−269.
(5) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006**, *65*, 15−26.
(6) 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.
(7) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* **2003**, *2*, 527−541.
(8) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.
(9) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547−562.
(10) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(11) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213−218.
(12) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.
(13) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534−553.
(14) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209−225.
(15) Yoon, S.; Welsh, W. J. Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 88−96.
(16) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74−82.
(17) Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ. http://www.rcsb.org (accessed Jan 2007).
(18) *MDL Drug Data Report*; licensed by Molecular Design Ltd., San Leandro, CA. https://www.discoverygate.com (accessed Nov 2006).
(19) Developmental Therapeutics Program NCI/NIH. http://dtp.nci.nih.gov (accessed Nov 2006).
(20) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(21) *FILTER*; OpenEye Scientific Software, Inc.: Santa Fe, NM.
(22) *LigPrep, version 2.0*; Schrödinger, LLC: New York, NY.
(23) *QikProp, version 2.5*; Schrödinger, LLC: New York, NY.
(24) *Daylight Toolkit*; Daylight Chemical Information Systems, Inc.: Aliso Viejo, CA.
(25) *OMEGA*; OpenEye Scientific Software, Inc.: Santa Fe, NM.
(26) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504−1519.
(27) *ROCS*; OpenEye Scientific Software, Inc.: Santa Fe, NM.
(28) *Prime, version 1.5*; Schrödinger, LLC: New York, NY.
(29) *Glide, version 4.0*; Schrödinger, LLC: New York, NY.
(30) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(31) *ICM*; Molsoft, LLC: San Diego, CA.
(32) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488−506.
(33) Hand, D.; Mannila, H.; Smyth, P. *Principles of Data Mining*; The MIT Press: Cambridge, MA, 2001.
(34) Vigers, G. P.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47*, 80−89.

CI700376C