# Identification of a Minimal Subset of Receptor Conformations for Improved Multiple Conformation Docking and Two-Step Scoring

Sukjoon Yoon† and William J. Welsh*

Department of Pharmacology, University of Medicine & Dentistry of New Jersey−Robert Wood Johnson Medical School (UMDNJ−RWJMS), 675 Hoes Lane, Piscataway, New Jersey 08854

Received July 30, 2003

Docking and scoring are critical issues in virtual drug screening methods. Fast and reliable methods are required for the prediction of binding affinity especially when applied to a large library of compounds. The implementation of receptor flexibility and refinement of scoring functions for this purpose are extremely challenging in terms of computational speed. Here we propose a knowledge-based multiple-conformation docking method that efficiently accommodates receptor flexibility thus permitting reliable virtual screening of large compound libraries. Starting with a small number of active compounds, a preliminary docking operation is conducted on a large ensemble of receptor conformations to select the minimal subset of receptor conformations that provides a strong correlation between the experimental binding affinity (e.g., $K_i$, $IC_{50}$) and the docking score. Only this subset is used for subsequent multiple-conformation docking of the entire data set of library (test) compounds. In conjunction with the multiple-conformation docking procedure, a two-step scoring scheme is employed by which the optimal scoring geometries obtained from the multiple-conformation docking are re-scored by a molecular mechanics energy function including desolvation terms. To demonstrate the feasibility of this approach, we applied this integrated approach to the estrogen receptor α (ERα) system for which published binding affinity data were available for a series of structurally diverse chemicals. The statistical correlation between docking scores and experimental values was significantly improved from those of single-conformation dockings. This approach led to substantial enrichment of the virtual screening conducted on mixtures of active and inactive ERα compounds.

## INTRODUCTION

Since experimental determination of the binding constants between a large number of pairs of molecules is costly and time-consuming, virtual docking and scoring methods have been developed for numerous applications.[1] With the advent of combinatorial chemistry and genomics approaches, drug discovery has changed dramatically.[2] Modern pharmaceutical research with its trend toward high-throughput systems demands fast and accurate characterization of large numbers of compounds for many potential target proteins. Frequently, the chemicals or collection of compounds must be scored to prioritize their suitability for chemical synthesis and biological evaluation. However, despite the rapid spread of computational approaches, stimulated by the availability of high-resolution structural data on proteins, basic problems in computational approaches to docking and scoring still remain unresolved. Three major outstanding challenges are (1) the inclusion of the receptor's conformational flexibility in the docking step, (2) appropriate scoring functions to rank the compounds, and (3) the inclusion of solvent effects.

The energy landscape of most proteins is frequently described in terms of a folding funnel in which there are multiple favorable folded states.[3−5] The folded state of the flexible system has high entropy and, thus, can and will sample a large subensemble of occupied states. The subensemble is a collection of structurally similar and nearly energetically equivalent conformations of the protein that, together, comprise the folded state.[6] Thus, a single structure, even the weighted average provided by a crystal structure, may not describe these substates adequately. It is also important to note that introducing a ligand into the system changes the environment. It may affect the most populated state of the protein, such as would apply to the "induced-fit" model of enzyme action.[6]

The existing docking and scoring algorithms that allow some degree of protein flexibility include methods based on single, semiflexible structures or multiple structures of a protein. Semiflexible single structure docking is computationally efficient and relatively easy to implement in existing programs. However, this approach still restricts the successful ligands in that they must conform to a certain size and conformation. In multiple-conformation docking, the flexibility of the system is represented through the use of multiple structures rather than through use of a single, semiflexible structure. An ensemble of structures for multiple-conformation docking is generated either through molecular dynamics (MD) or Monte Carlo (MC) simulation.[7,8] An efficient compromise between semiflexible and multiple-conformation approaches has also been reported.[9]

To improve the docking result by including multiple conformations that occur less frequently in protein dynamics, generally a large number of receptor conformations (>100) are required. This is the critical disadvantage of multiple-conformation docking when applied to virtual screening

* Corresponding author phone: 732-235-3234; fax: 732-235-3475; e-mail: welshwj@umdnj.edu.
† E-mail address: yoonsu@umdnj.edu.

10.1021/ci0341619 CCC: $27.50 
Published on Web 11/08/2003

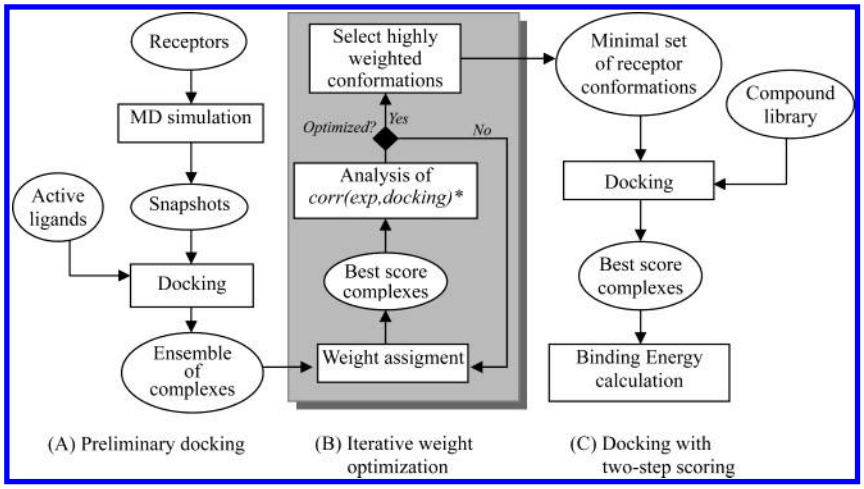


**Figure 1.** Schematic overview of the proposed docking and scoring method. Step A shows a preliminary docking procedure for the preselected active compounds, B includes the knowledge-based weight optimization step to identify the minimal number of receptor conformations that provide the best correlation between experimental binding affinity and docking score, and C shows the two-step scoring scheme combined with flexible docking to improve the prediction of rank-order activity. The asterisk (∗) denotes the correlation of experimental binding affinity (exp) and best docking score (docking) of an active compound.

efforts in pharmaceutical drug discovery, where chemical libraries often contain millions of small-molecule compounds. If instead it would be possible to select the minimal subset of receptor conformations that effectively capture the dynamic nature of the receptor, the computational efficiency of the multiple-conformation docking procedure would improve. Here we describe a knowledge-based approach to identify the minimal subset from a large ensemble of MD-generated receptor conformations. Starting with a small number of active compounds, a preliminary docking operation is conducted on the ensemble of receptor conformations (Figure 1A). Next, the population weight assigned to the ensemble of conformations is optimized to select the minimum subset that yields a strong correlation between experimental binding affinities and docking scores (Figure 1B). This subset is used for subsequent multiple-conformation docking of the entire collection of library compounds (Figure 1C). If the selected subset represents the dynamics of the receptor sufficiently well to overcome the pharmacophoric constraints encountered in single-receptor (rigid) dockings, it would be expected to provide an improvement in results from virtual screening.

Besides the matter of receptor flexibility during the docking operation, another salient issue in docking and scoring involves the scoring function employed for ranking the ligands. The role of the scoring function is critical in two key respects: (1) to distinguish the correct ligand-binding mode as the lowest energy structure from a multitude of alternative solutions and (2) to discriminate bound "active" compounds from a large pool of less active or inactive compounds. To enable rapid identification of the correct pose of a ligand in the binding pocket, scoring functions generally focus only on short-range (e.g., van der Waals, hydrogen bonding) interactions between the ligand and receptor. If, instead, one's primary interest is the prediction of binding affinity for a set of compounds, more reliable estimation of the binding potential or free energy is required. Unfortunately, calculation of electrostatic and desolvation effects, critical for evaluating correct binding affinities, is notoriously difficult and tedious.[1]

A more rational computational approach would be to employ a tiered or hierarchical scoring scheme. Such a scheme may consist of two steps. In step 1, a simple scoring function is deployed for searching the optimal binding geometry in the multiple-conformation docking. In step 2, a more sophisticated molecular mechanics energy function (including desolvation terms) is implemented to predict and rank the binding affinity of ligands for which binding geometries were optimized in step 1 using our multiple-conformation docking approach (Figure 1C). This two-tiered scoring scheme, in conjunction with knowledge-based multiple-conformation docking, will resolve some common limitations in current ligand−protein docking algorithms and provide an improved rank-order activity in a compound library that is not possible with simple or refined scores alone.

To demonstrate and evaluate the utility of our approach, we tested this new docking and scoring procedure on the estrogen receptor α (ERα) system for which several experimental structures with different ligands are available in the Protein Data Bank (PDB). ERα is also a good choice by virtue of the wealth of published experimental binding affinity data for many compounds to this receptor.[10−12] ERα is associated with numerous diseases, including breast cancer, osteoporosis, endometrial cancer, and prostate hypertrophy.[13] In addition to the biochemical and physiological importance of the natural ER ligands, in recent years considerable concern has emerged about the possible deleterious effects of xenoestrogens, particularly exogenous ER agonists. It appears that the ER shows affinity for a remarkably wide range of structurally diverse compounds, such as polyaromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), phytoestrogens, phthalates, and pesticides.[10,14−17] Everyday foodstuffs are another source of estrogen-like compounds. Given the diverse range of compounds that may bind to the ER and exert an effect on human and animal health, there is considerable medical interest in understanding the details of ligand−ER affinity and developing techniques to predict the affinity of compounds for ERα. Previous computational efforts in this area have concentrated mainly on empirical

**Table 1.** Structural Diversity of Test Compounds

| class[a] | entries | subset for preliminary docking | log(RBA) |
|---|---|---|---|
| steroids | 39 | 3-methylestriol | −1.65 |
| | | ethylnylestradiol | 2.28 |
| | | moxestrol | 1.14 |
| | | 5α-antrostane-3β,17β-diol | −0.92 |
| alkylphenolic | 21 | nonylphenol | −1.53 |
| | | 4-ethylphenol | −4.17 |
| diphenyl derivatives | 16 | bisphenol A | −2.11 |
| | | 3-phenylphenol | −3.44 |
| organochlorines | 22 | 2,3,4,5-tetrachloro-4-biphenylol | −0.64 |
| | | dihydroxymethoxychlor olefin | 0.42 |
| other pesticides | 20 | | |
| alkyl hydroxy benzoate | 7 | ethyl 4-hydroxybenzoate | −3.22 |
| phthalates | 8 | | |
| benzophenone | 5 | 4,4-dihydroxybenzophenone | −2.46 |
| others | 37 | aurin | −1.49 |
| | | zearalenol | 1.63 |
| | | 6-hydroxyflavanone | -3.05 |
| total | 175 | 15 | |

[a] Chemical classes were taken from the study by Tong and co-workers.[10] log(RBA) of a compound is the relative binding affinity to 17β-estrogen (See Methods for the detail).

regression-based (quantitative structure−activity relationship (QSAR)) approaches.[18−20] In the present study, we explore the ligand−ERα interactions using the knowledge-based multiple-conformation docking method with an improved scoring scheme.

## METHODS

**Preparation of Receptor Conformations.** Seven crystal structures of the ERα ligand binding domain (LBD) complexed with different ligands were available from the Protein Data Bank (http://www.rcsb.org/pdb). Three ER-LBD structures (PDB ID: 3ERD, 1QKU, and 1L2I) complexed with agonist ligands (diethylstilbestrol, estradiol, and diethyl tetrahydrochrysenediol) were used as starting structures for the MD simulations. Three independent MD simulations were carried out by using the SANDER module of the AMBER7.0 program.[21] AMBER7.0 contains the new generalized atom force field (GAFF) for small molecules in addition to the popular AMBER force field for proteins. Thus, we could conveniently perform dynamics simulations and energy calculations for protein−ligand complexes with AMBER7.0. The MD simulations were conducted in vacuo, since our interest is to explore the conformational diversity within the binding site rather than to gain a precise knowledge of the global backbone and/or side-chain movements. The SHAKE option was implemented to constrain bond-stretching motions, thereby permitting extension of the time step to 2 fs. After initial X-ray structures were minimized for 5000 steps, MD simulations were carried out for 1 ns at 300 K. The distance cutoff for the nonbonded (van der Waals, Coulombic) interactions was set at 300 Å for energy minimization and then reduced to 8 Å for the MD simulations. In each simulation, snapshots were taken at intervals of 10 ps. Consequently, a total of 300 snapshots were collected from three independent simulations (100 snapshots from each simulation).

**Preparation of Ligands.** For the test compounds, we obtained the SMILES representation of 232 compounds for which experimental binding affinities were taken from the published studies by Tong and co-workers.[10,11] The experimental binding affinity was represented by log(RBA). log(RBA) refers to the relative binding affinity, defined as the logarithm of the percent ratio of the $IC_{50}$ between 17β-estradiol and a test compound. Thus the RBA of 17β-estradiol is 100, and log(RBA) of 17β-estradiol is 2. These workers measured log(RBA) as low as −5, defining compounds with log(RBA) < −5 as "nonbinders".[10] Suitable initial 3D structures were generated for these compounds using the 2D→3D conversion algorithm OMEGA[22] for molecules containing rotatable (single) bonds and CORINA[23] for rigid compounds devoid of rotatable bonds. The ERα LDB crystal structures exhibit variability in the C-terminal activation domain (Helix 12)[24] depending on the size of the bound ligands. The three ERα structures used in our MD simulations had well-defined binding pockets with Helix 12 in a "closed" conformation, which effectively filters out large-size compounds that are unable to fit into the binding pockets (Figure 2). Compounds with atom types not recognized by GOLD were also eliminated from the test set. As a result of this preliminary culling process, 175 of the original 232 compounds were retained for this study. Of these 175 test compounds, 15 compounds (Table 1) were selected from structurally diverse classes for the preliminary docking (Figure 1A), while the remaining 160 compounds were used as the test set for the subsequent docking (Figure 1C).

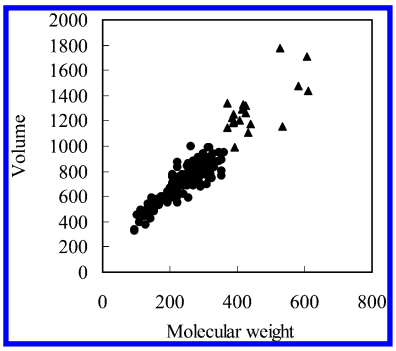


**Figure 2.** Distribution of original 232 compounds from Tong et al.[10,11] in "volume vs molecular weight" space. ● indicates a compound included in this study, and ▲ indicates a compound that was discarded due to the large size.

**Docking Protocol.** The GOLD program[25] (version 1.1) was used for both preliminary docking (Figure 1A) and later minimal subset-based multiple-conformation docking (Figure 1C). The active site was defined to encompass all atoms within a 10 Å radius sphere whose origin was located at the center of the bound ligand. The "library screening mode" was selected for fast dockings, and 10 individual dockings were conducted for each compound. To allow poor nonbonded contacts at the start of each genetic algorithm (GA) run, the maximum distance between hydrogen-bond donors and fitting points was set to 5 Å, and nonbonded van der Waals energies were cut off at a value equal to $k_{ij}$ (i.e., the well depth of the van der Waals energy for the atom pair $i,j$). The "early-termination" option was applied when the top three solutions are within 0.5 Å root-mean-square deviation (RMSD).

**Minimal Subset of Receptor Conformations.** An iterative method for optimizing population weights was used to find the minimal subset of MD snapshots that provides a strong correlation between the experimental binding affinities and docking scores in the preliminary docking. In this study, the Metropolis MC procedure was employed to optimize the population weights assigned to the snapshots. The details of the procedure are described below.

The total number of ERα snapshots from the MD simulations is $N = 300$. A population weight $w_i$ is assigned to each snapshot $i$ either uniformly or randomly such that

$$\sum_{i=1}^{N} w_i = 1$$

The weighted docking score of an active compound $j$ against $i$ snapshot $S_{i,j}$ is

$$S_{i,j} = w_i s_{ij}$$

where $s_{ij}$ is the original GOLD docking score for the compound $j$ and snapshot $i$. The best docking score of an active compound $j$ against $N$ snapshots is thus calculated as

$$S_{\max,j} = \max(S_{1,j}, S_{2,j}, \ldots, S_{N,j})$$

The correlation between $S_{\max}$ and experimental binding affinity Ex can be evaluated by calculating Pearson's correlation coefficient, $r$:

$$r = \frac{\sum(\mathrm{Ex}_j - \langle \mathrm{Ex} \rangle)(S_{\max,j} - \langle S_{\max} \rangle)}{\sqrt{\sum(\mathrm{Ex}_j - \langle Ex \rangle)^2 \sum(S_{\max,j} - \langle S_{\max} \rangle)^2}}$$

A MC procedure was constructed to optimize the population weights. At each iteration of the $r$ calculation, a new population weight set is generated by making random changes to the previous population weight set by a random number generator. After generating a unique random number ($\xi$) for each weight, $\delta x_{\max}$ represents the maximum possible displacement in the ± direction in the weight. A new weight is calculated as follows:

$$w_{\mathrm{new}} = w_{\mathrm{old}} + (2\xi - 1)\,\delta x_{\max}$$

A new $S_{\max}$ is then calculated with $w_{\mathrm{new}}$, and the $r$ for the new set is calculated from $S_{\max}$ and experimental values. If the value of $r$ of the new population weights is higher than its predecessor, then the new set is retained as the starting point for the next iteration. If the new set is lower in $r$ than its predecessor, then the Metropolis algorithm is applied. That is, a Metropolis weight function, $\exp(\Delta r/C)$ is compared to a random number between 0 and 1 where $C$ is a constant. If the weight function is greater than the random number, then the new set is accepted; if not, then it is rejected and the initial set is retained for the next move. This acceptance condition can be written in the following concise fashion:

$$\mathrm{rand}(0,1) \le \exp(\Delta r/C)$$

After the weight optimization, the snapshots with relatively high weights were selected and used for subsequent docking of the 160 test compounds.

**Binding Energy Calculations.** A molecular mechanics energy calculation was employed to predict and rank the binding affinity of a ligand for which the binding geometry was optimized by the knowledge-based multiple-conformation docking and GOLD scoring. The standard molecular mechanics force field is less amenable to searching but, in principle, it should describe more adequately the energetics of ligand−protein interactions. The binding energy $E_{\mathrm{binding}}$ was calculated as follows:

$$E_{\mathrm{binding}} = E_{\mathrm{complex}} - (E_{\mathrm{receptor}} + E_{\mathrm{ligand}})$$

The AMBER7.0 program was used for the energy minimization and binding energy calculations. A 100-step minimization was conducted in vacuo for each ligand−receptor complex, and the final energy was calculated including implicit solvent terms ($E_{\mathrm{GB}} + E_{\mathrm{surf.}}$) represented by the generalized Born/surface area (GB/SA) algorithm.[26] The ratio $E_{\mathrm{GB}}/E_{\mathrm{surf.}}$ estimates the "electrostatic/nonelectrostatic contribution" to solvation. The cutoff for the nonbonded van der Waals and electrostatic energy terms was set to 200 Å to include long-range electrostatic effects. Since bonding and intramolecular nonbonding terms cancel out in the $E_{\mathrm{binding}}$ calculation, only the following four terms including intermolecular interactions are considered.

$$E_{\mathrm{pot.}} = E_{\mathrm{vdw}} + E_{\mathrm{elec}} + E_{\mathrm{GB}} + E_{\mathrm{surf.}}$$

## RESULTS AND DISCUSSION

**Identification of the Minimal Subset of ERα Conformations.** MD simulations of three ERα X-ray structures generated 1 ns trajectories showing about 0.5 Å RMSD fluctuation in the active site (Figure 3). Using 300 snapshots taken from the trajectories at 10 ps intervals, we carried out the preliminary docking of the 15 ERα active compounds (Table 1). A novel aspect in our method was to conduct a preliminary docking using a set of structurally diverse active compounds in order to identify the minimal subset of ERα conformations for the later multiple-conformation docking of the whole test compounds. Although docking results are generally quite sensitive to the conformation of a receptor, the topology of an active site might not be so dependent on the nature of bound ligand.[27] We therefore made the assumption that if a set of ERα conformations provides a strong correlation between the predicted and experimental binding affinities in a subset of compounds, they might

**Table 2.** Minimal Subset of ERα MD Conformations Selected by Iterative Population Weight Optimization

| selected conformations | ER1 | ER2 | ER3 | ER4 | ER5 | ER6 |
|---|---|---|---|---|---|---|
| sampling position | 3erd | 3erd | 1l2i | 1l2i | 1qku | 1qku |
| | 490 ps | 840 ps | 410 ps | 830 ps | 700 ps | 940 ps |
| correlation[a] ($r$) | | | max(ER1,..,ER6) = 0.94 | | | |
| solvent accessible atoms[b] | 147 | 120 | 162 | 165 | 146 | 164 |
| H-bond donors[b] | 1 | 1 | 1 | 2 | 2 | 0 |
| H-bond acceptors[b] | 5 | 3 | 7 | 6 | 4 | 5 |
| rotatable bonds[b] | 1 | 1 | 1 | 1 | 0 | 0 |

[a] max(ER1,..,ER6) was calculated from the preliminary docking using 15 compounds. For a given ligand, max(ER1,..ER6) provides the highest GOLD score among those of ER1−ER6. The $r$ is based on Pearson's correlation coefficient between the docking score (max(ER1,..,ER6)) and experimental binding affinity. [b] Analysis of the binding site (10 Å radius sphere whose origin is located at the center of the bound ligand).

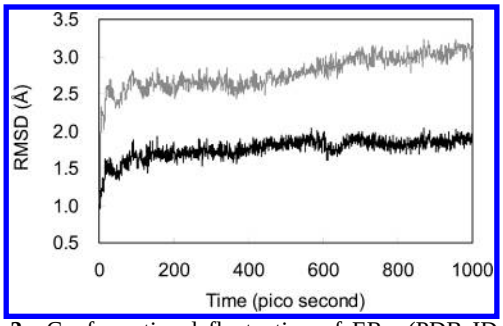


**Figure 3.** Conformational fluctuation of ERα (PDB ID: 3erd) during MD simulation. Gray color represents all-atom RMSD for the whole protein, while black is for only active site residues (5 Å within the ligand).

generally provide good docking results for a larger library of compounds. We repeated the MC procedure to select the minimum number of snapshots. After several repeats of 100 000 steps of weight optimization, six snapshots (ER1−ER6) with relatively high weights were selected from diverse positions in the MD trajectories (Table 2). Two conformations were taken from each of three different MD trajectories. Among different subsets selected from the several repeats of the MC procedure, these six snapshots generated the best correlation coefficient ($r = 0.94$) between the experimental and predicted binding affinity for the 15 compound set. In these six conformations, numbers of solvent accessible atoms and available H-bonding donor/acceptor varied significantly (Table 2, Figure 4). Considering that GOLD calculates the van der Waals and H-bond energies only for solvent accessible atoms in the binding site, these variations will have differential effects on the docking result. The average RMSD of the binding-site heavy atoms among three conformations (ER1, ER3, and ER5) was 2.54 Å (Figure 4). We used these six snapshots for the docking of 160 diverse test compounds.

**Binding Affinity Prediction.** A data set of 89 ERα active compounds (log(RBA) > −5.0)[11] were docked to the six selected conformations (Table 3). The correlation coefficient ($r$) between experimental binding affinity and the docking score ranged from 0.036 to 0.30 for the six individual dockings. The best result ($r = 0.30$) was obtained from the docking to ER3 (GOLD-ER3). When the highest score among six dockings (i.e., max(ER1,..,ER6)) was selected for each compound, and then correlated to the experimental affinity, the $r$ for 89 compounds was significantly improved to 0.50 (GOLD-max; Table 3). This result demonstrates the advantage of the multiple-conformation docking in predicting the binding affinity.

To further improve the correlation, a molecular mechanics energy calculation (AMBER) was employed to predict and rank the binding affinity of a ligand for which the optimal binding geometry was determined by six subset dockings and GOLD scorings. Long-range electrostatic contributions and desolvation effects are not included in GOLD scoring. Thus, we expected that re-scoring by AMBER (including GB/SA) would significantly improve the prediction of binding affinity. Indeed, the correlation between binding energy and experimental affinity improved dramatically ($r = -0.75$; Table 3).

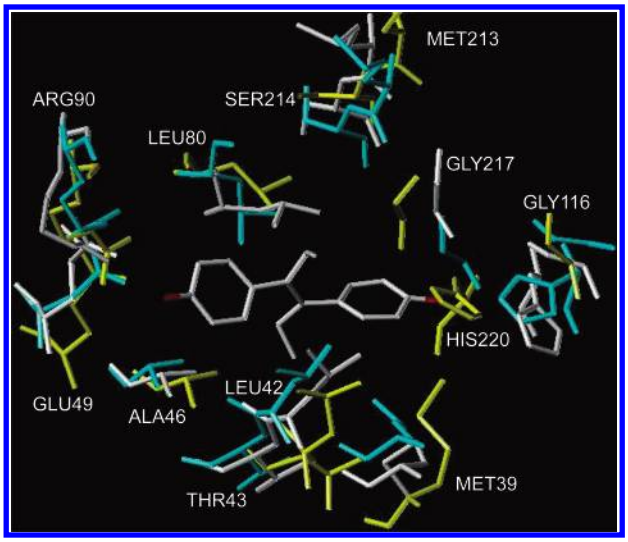


**Figure 4.** Comparison of active site residues in three different ERα conformations from MD simulations. Gray, cyan, and yellow refer to ER1, ER3, and ER5, respectively (see Table 2). The ligand presented is diethylstilbestrol that was complexed with ER1 during the MD simulation.

**Table 3.** Correlation between Predicted and Experimental Binding Affinity in 89 Active Ligands (logRBA > −5.0) Using the Minimal Subset of ERα MD Conformations[a]

| scoring method | conformation | correlation ($r$) |
|---|---|---|
| GOLD | ER1 | 0.068 |
| | ER2 | 0.036 |
| | ER3 | 0.30 |
| | ER4 | 0.27 |
| | ER5 | 0.20 |
| | ER6 | 0.28 |
| | max(ER1 ,..,ER6) | 0.50 |
| AMBER | max(ER1 ,..,ER6) | −0.75 |

[a] For a given ligand, max(ER1,..,ER6) provides the highest GOLD score among those of ER1−ER6. The $r$ is based on Pearson's correlation coefficient. The negative sign of the $r$ value under AMBER is a consequence of the fact that GOLD scores are inherently positive while AMBER-calculated energies decrease (are more negative) for lower energy conformations.

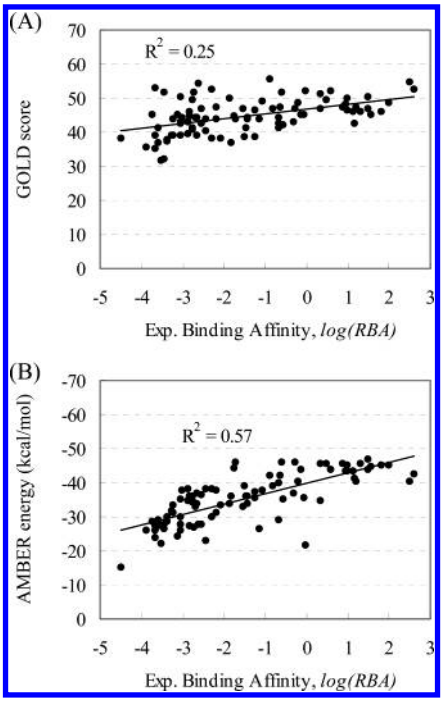


**Figure 5.** Prediction of binding affinity of 89 ERα ligands: (A) correlation between the GOLD score and the experimental binding affinity; (B) correlation between the AMBER-calculated molecular mechanics energy and the experimental binding affinity.

**Table 4.** Contribution of Individual Energy Terms to the Binding Affinity Prediction

| energy term | correlation coefficient[a] ($r$) | energy term | correlation coefficient[a] ($r$) |
|---|---|---|---|
| $E_{vdw}$ | −0.66 | $E_{vdw} + E_{GB}$ | −0.018 |
| $E_{elec}$ | −0.18 | $E_{vdw} + E_{elec} + E_{GB}$ | −0.74 |
| $E_{GB}$ | 0.20 | $E_{vdw} + E_{elec} + E_{surf.}$ | −0.37 |
| $E_{surf.}$ | −0.50 | $E_{vdw} + E_{elec} + E_{GB} + E_{surf.}$ | −0.75 |
| $E_{vdw} + E_{elec}$ | −0.36 | | |

[a] The correlation coefficient is between AMBER energy and experimental binding affinity.

When GOLD-max and AMBER re-scoring results were plotted against the experimental binding affinity, the multiple regression correlation coefficient for the least-squares fitting line ($R^2$) was 0.25 and 0.57, respectively (Figure 5A,B). The relatively broad distribution of weak binders along the *Y*-axis in Figure 5A indicates that many weak binders were overestimated in the GOLD-max. One main reason for this error should be that the binding pocket was more generously represented in the multiple-conformation approach than in a single-receptor method. This is a common drawback in flexible docking methods. We observed that AMBER re-scoring improved the correlation ($R^2 = 0.57$) by reducing the overestimated weak binders (Figure 5B). This result indicates that energy minimization and re-scoring using a sophisticated energy function can significantly improve the binding affinity prediction from flexible docking results.

Table 4 shows the contribution of individual energy terms to the correlation between the binding energy and experimental value in AMBER re-scoring. Inclusion of the $E_{vdw}$, $E_{elec}$, and $E_{GB}$ terms are chiefly responsible for the improved correlation, and omitting any of them resulted in significantly lower correlations. The most distinctive contribution available in the energy calculation but not in GOLD was the solvation term ($E_{GB}$). The GOLD scoring function consists of three terms: a hydrogen bonding term, a van der Waals term, and an internal energy term.[25] It has been optimized to correctly reproduce the experimental binding geometry of a protein−ligand complex system. Thus, it is generally not appropriate to use GOLD scores to predict and rank binding affinities of different ligands to a given receptor. However, GOLD can rapidly and extensively search conformational space with its simplified scoring function, thus making it a suitable choice for conducting multiple-conformation docking to optimize the binding geometry of a given ligand to the receptor.

Although a sophisticated energy function has been shown to improve the binding geometry optimization,[28] the calculation is almost impossible when applied to a large library of compounds. Instead, we used well-designated subset of receptor conformations to improve docking performance (binding mode optimization), and then, an energy function was used just for ranking. Considering the importance of receptor conformations in both docking and scoring steps, this approach looks very promising for virtual drug screening in terms of computational efficiency.

**Virtual Screening of 160 Compounds.** Our knowledge-based multiple conformation approach with two-tiered scoring was applied to virtual screening of 160 test compounds including both binders and nonbinders. A total of 89 compounds having log(RBA) > −5 were defined as binders according to Blair et al.[10] After conducting the multiple dockings of 160 compounds to the six selected ERα snapshots, we analyzed the results using receiver operating characteristic (ROC) curves.[29] A ROC curve describes the tradeoff between sensitivity and specificity. Sensitivity is defined as the ability of the model to avoid "false negatives", while specificity relates to its ability to avoid "false positives". The "area under the ROC curve" (AUC) is a measure of the test accuracy. For example, an AUC value of 0.5 represents a random prediction, while 1.0 represents a perfect prediction. In the analysis of virtual screening data from the mixture including 89 binders, AUC values ranged narrowly from 0.67 to 0.73, depending on the docking and scoring method (Figure 6A). The standard single-conformation receptor docking (GOLD-ER3) showed 0.67 AUC. Knowledge-based multiple-conformation docking (GOLD-max) and re-scoring by an energy function (AMBER) could only slightly enrich the virtual screening data better than the single-conformation receptor docking.

However, a compound exhibiting activity at concentrations in the millimolar or high-micromolar range is not considered as an initial lead in general drug discovery. Considering that the activity of "log(RBA) = −5.0" in ERα compounds actually represents millimolar $IC_{50}$ concentrations,[10] we felt it more practical for the present purpose of virtual screening to apply a more stringent definition of "active" versus "inactive" compounds. Since 10−20 μM activity is generally the minimum required for initial lead compounds in drug discovery programs, we applied "log(RBA) > −2" (corresponding to 10−20 μM activity for ERα compounds) as the cutoff for active compounds. On the basis of this criterion, 46 of the 160 test-set compounds were designated active. Results from docking and scoring by standard single-conformation receptor docking (GOLD-ER3), by application of the knowledge-based approach (GOLD-max) and by

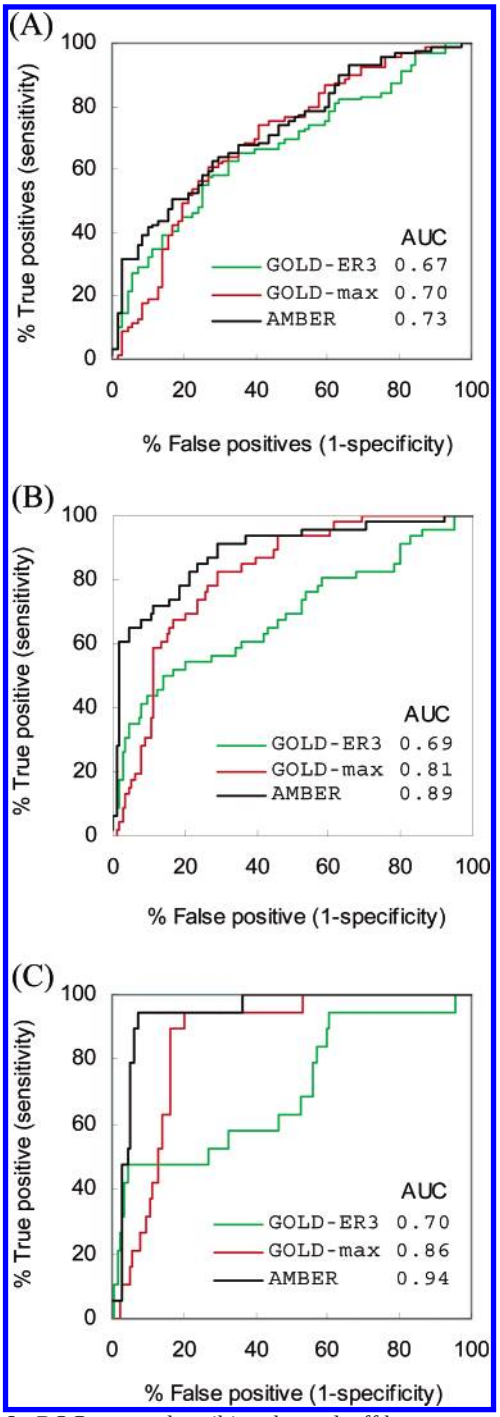


**Figure 6.** ROC curves describing the tradeoff between sensitivity and specificity for three different scoring procedures; typical single receptor docking using GOLD (GOLD-ER3), multiple receptor docking using GOLD (GOLD-max), and GOLD-max with AMBER re-scoring (AMBER). True positives include (A) 89 compounds with log(RBA) > −5.0, (B) 46 compounds with log(RBA) > −2.0, and (C) 19 compounds with log(RBA) > 0.0 out of a total of 160 compounds. The AUC value refers to the "area under the curve".

application of molecular mechanics re-scoring (AMBER), are compared in Figure 6B. GOLD-max and AMBER re-scoring markedly improved the performance of virtual screening. As measured by AUC values, GOLD-max (AUC = 0.81) and AMBER re-scoring (AUC = 0.89) performed substantially better than GOLD-ER3 (AUC = 0.69).

In Figure 6C, the definition for active compounds was set to the more stringent cutoff log(RBA) > 0 corresponding to binding affinity in the nanomolar range. This level of biological activity would correspond to a highly active and therefore promising lead compound in general drug discovery. On the basis of this criterion, only 19 of the 160 test-set compounds were designated active. The improvement in GOLD-max and AMBER re-scoring over GOLD-ER3 is more apparent. As measured by AUC values, the performance improvement of GOLD-max (ACU = 0.86) and AMBER re-scoring (AUC = 0.94) over GOLD-ER3 (AUC = 0.70) was even more pronounced.

As seen in Figure 6, knowledge-based multiple conformation docking and two-tiered scoring were able to enrich the virtual screening data more than single-conformation receptor docking. Moreover, these approaches performed even better when the criterion for active versus inactive compounds is set to criteria typically employed in drug discovery scenarios in the pharmaceutical industry. In case of single-conformation docking (GOLD-ER3), AUC values varied only from 0.67 to 0.70 by changing the definition for active compounds. However, GOLD-max improved the AUC from 0.70 to 0.86, and AMBER re-scoring improved the AUC from 0.73 to 0.94 by excluding weak binders from active compounds. These results demonstrate that our method should possess substantial advantages in distinguishing strong binders from weak binders or nonbinders in virtual screening using docking and scoring techniques.

An alternative way of comparing docking and scoring approaches for virtual drug screening is to plot the percentage of true hits versus the rank order as predicted by the particular docking and scoring method (Figure 7). Consider, for instance, the case corresponding to Figure 6B above in which 46 of 160 ERα ligands were designated as actives. In this case, the absolute best outcome (positive control) for the scoring method would be to rank the 46 actives in rank order 1−46 and to rank the remaining 114 inactives in rank order 47−160. In other words, the percentage of true hits would reach 100% when the rank order reaches 46. This type of analysis quantifies the advantages of the present method invention over traditional docking and scoring methods in a typical drug discovery scenario.

Using the cutoff (log(RBA) > −2.0) corresponding to 10−20 $\mu$M activity, a reasonable assessment of the ability of a particular scoring method would be to determine the rank order needed to encompass at least 60% of the true hits (i.e., 28 of the 46 actives from the test set of 160 compounds). Optimal performance (positive control) would require the minimal rank (rank = 28) to encompass these 28 true hits. By comparison, our methods required only rank = 43 without re-scoring (GOLD-max) and impressively rank = 30 with re-scoring (AMBER), whereas standard GOLD (GOLD-ER3) required rank = 70 (Figure 7A). Using the more stringent cutoff (log(RBA) > 0.0) corresponding to nanomolar activity (Figure 7B), a more demanding assessment of the ability of a particular scoring method would be to determine the rank order needed to encompass 95% of the true hits (i.e., 18 of the 19 actives from the test set of 160 compounds). By comparison, our methods required only rank = 46 without re-scoring (GOLD-max) and nearly perfect rank = 28 with re-scoring (AMBER), whereas standard GOLD (GOLD-ER3) required rank = 103. Considering the time and cost for chemical synthesis or biological evaluation of selected hits in the drug discovery process, this method represents an innovative and improved approach for virtual drug screening.

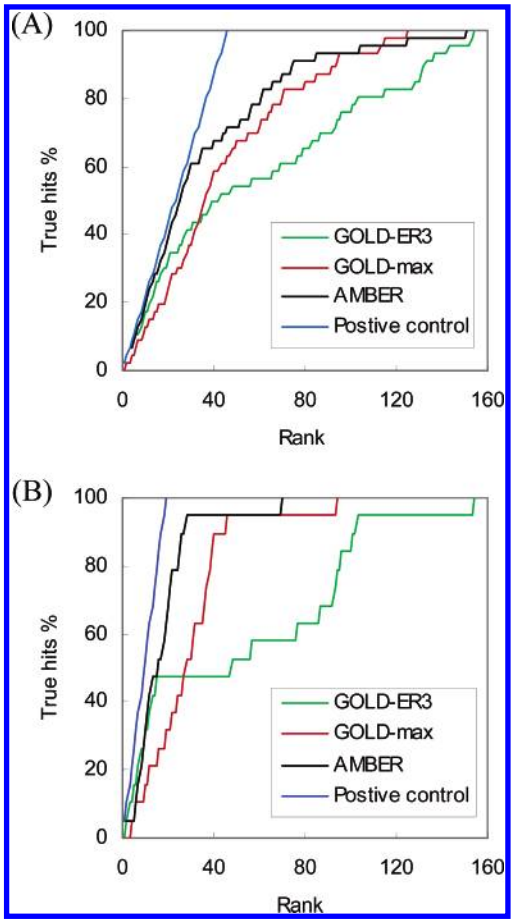


**Figure 7.** Cumulative ranking of ERα binding compounds using three different scoring methods: GOLD-ER3, GOLD-max, and AMBER. Ranking of (A) 46 compounds with log(RBA) > −2.0 and (B) 19 compounds with log(RBA) > 0.0 out of a total of 160 compounds.

**Table 5.** AUC Values of ROC Plots in Multiple-Conformation Dockings Using Six ERα Conformations[a]

| cutoff for active compds | random 1 | random 2 | random 3 | GOLD-max |
|---|---|---|---|---|
| log(RBA) > 0.0 | 0.83 | 0.78 | 0.84 | 0.86 |
| log(RBA) >-2.0 | 0.77 | 0.76 | 0.79 | 0.81 |

[a] In each case of random 1, random 2, and random 3, the six conformations were randomly chosen from 300 ERα MD snapshots. In the case of GOLD-max, the six conformations were selected from the knowledge of preliminary docking (see Table 2).

Finally, to understand the extent to which our knowledge-based approach can improve the performance of multiple-conformation docking, we also conducted multiple dockings of test compounds to randomly selected six snapshots of ERα conformations. We repeated virtual screening of 160 compounds using three different sets of randomly selected six snapshots (Table 5). In all three cases, the virtual screening results were better than the single receptor docking but always worse than that of the knowledge-based approach (GOLD-max). Considering the lack of simple theories/methods to predict the protein dynamics upon ligand binding, the well-designated subset of receptor conformations by a knowledge-based approach could complement current docking and scoring algorithms overcoming the receptor flexibility problem.

## CONCLUDING REMARKS

Although docking and scoring functions have been dramatically improved for the past decade, they still cannot handle the receptor flexibility problem properly when applied to a large library of compounds. In the flexible system, no clear relationship between docking and ranking has been found.[27] In this study, we found that multiple dockings to the well-designated subset of receptor conformations improved the scoring and ranking. These results suggest that the receptor conformation is of critical importance in both docking and scoring. We could apply this knowledge-based approach to the virtual screening of 160 ERα test compounds and successfully enrich the hit list. In addition, we employed a two-tiered scoring scheme in conjunction with the knowledge-based multiple-conformation docking procedure in order to achieve further improvements in the prediction of rank-order activity in a compound library. The two-step scoring scheme provides a workable compromise between time and accuracy in the binding affinity prediction. Consensus scoring methods connecting many simple scoring functions have outperformed single scoring whatever the target and the docking tool used, suggesting that finding a single scoring function to rank virtual hits was not a major concern.[27] Probably the binding energy calculation could be slower than the consensus scoring. However, we showed that it was a promising ranking method especially when the receptor−ligand structural configuration was further optimized by the knowledge-based multiple dockings.

## ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. Weida Tong (National Center for Toxicological Research, U.S. Food & Drug Administration, Jefferson, AR) for providing us the experimental data set and 2D structures used in this study. We also appreciate the helpful and stimulating discussions with Dr. Kiyean Nam (Merck). W.J.W. acknowledges the financial support for this research from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program.

## REFERENCES AND NOTES

(1) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.
(2) Drews, J. Drug discovery: A historical perspective. *Science* **2000**, *287*, 1960−1964.
(3) Onuchic, J. N. Contacting the protein folding funnel with NMR. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7129−7131.
(4) Shoemaker, B. A.; Wang, J.; Wolynes, P. G. Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 777−782.
(5) Ma, B.; Kumacr, S.; Tsai, C.-J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **1999**, *12*, 713−720.
(6) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213−218.
(7) Pang, Y. P.; Kozikowski, A. Prediction of the binding sites of huperizine A in acetylcholinesterase by docking studies. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 669−681.
(8) Carlson, H. A.; Masukawa, J.; McCammon, J. A. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem.* **1999**, *103*, 10213−10219.
(9) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE; Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377−395.
(10) Blair, R. M.; Fang, H.; Branham, W. S.; Hass, B. S.; Dial, S. L.; et al. The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands. *Toxicol. Sci.* **2000**, *54*, 138−153.

(11) Hong, H.; Tong, W.; Fang, H.; Shi, L.; Xie, Q.; et al. Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts. *Environ. Health Perspect.* **2002**, *110*, 29−36.
(12) Fang, H.; Tong, W.; Shi, L. M.; Blair, R.; Perkins, R.; et al. Structure−activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. *Chem. Res. Toxicol.* **2001**, *14*, 280−294.
(13) Gustafsson, J. A.; Kuiper, G. G. J. M.; Enmark, E.; Treuter, E.; Rafter, J. Receptor-mediated toxicity. *Arch. Toxicol.* **1998**, *20*, 21−28.
(14) Waller, C. L.; Oprea, T. I.; Chau, K.; Park, H. K.; Korach, K. S.; et al. Ligand-based identification of environmental estrogens. *Chem. Res. Toxicol.* **1996**, *9*, 1240−1248.
(15) Soto, A. M.; Sonnenschein, C.; Chung, K. L.; Fernandez, M. F.; Oka, N.; et al. The E-screen assay as a tool to identify estrogens−An update on estrogenic environmental-pollutants. *Environ. Health Perspect.* **1995**, *103*, 113−122.
(16) Breinholt, V.; Larsen, J. C. Detection of weak estrogenic flavonoids using a recombinant yeast strain and a modified MCF7 cell proliferation assay. *Chem. Res. Toxicol.* **1998**, *11*, 622−629.
(17) Garner, C. E.; Jefferson, W. N.; Burka, L. T.; Mattews, H. B.; Newbold, R. R. In vitro estrogenicity of the catechol metabolites of selected polychlorinated biphenyls. *Toxicol. Appl. Pharmacol.* **1999**, *154*, 188−197.
(18) Gantchev, T. G.; Ali, H.; Van Lier, J. E. Quantitative structure−activity relationships/comparative molecular field analysis (QSAR/CoMFA) for receptor-binding properties of halogenated estradiol derivatives. *J. Med. Chem.* **1994**, *1994*, 4164−4176.
(19) Yu, S. J.; Keenan, S. M.; Tong, W.; Welsh, W. J. Influence of the structural diversity of data sets on the statistical quality of three-dimensional quantitative structure−activity relationship (3D-QSAR) models: Predicting the estrogenic activity of xenoestrogens. *Chem. Res. Toxicol.* **2002**, *15*, 1229−1234.
(20) Tong, W.; Perkins, R.; Xing, L.; Welsh, W. J. QSAR models for binding of estrogenic compounds to estrogen receptor α and *β* subtypes. *Endocrinology* **1997**, *138*, 4022−4025.
(21) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; et al. *AMBER7*; University of California: San Francisco, 2002.
(22) Bostrom, J. Reproducing the conformations of protein-bound ligands: A critical evaluation of several popular conformational searching tools. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1137−1152.
(23) Sadowski, J.; Gasteiger, J. From atoms and bonds to three-dimensional atomic coordinates: Automatic model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(24) Oostenbrink, B. C.; Pitera, J.; van Lipzig, M.; Meerman, J.; van Gunsteren, W. Simulations of the estrogen receptor ligand-binding domain: Affinity of natural ligands and xenoestrogens. *J. Med. Chem.* **2000**, *43*, 4594−4605.
(25) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(26) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275−291.
(27) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical database. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(28) 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.
(29) Hand, D.; Mannila, H.; Smyth, P. *Principles of data mining*; The MIT Press: Cambridge, MA, 2001.

CI0341619