Supervised Scoring Models with Docked Ligand Conformations for Structure-Based Virtual Screening

Reiji Teramoto* and Hiroaki Fukunishi

Fundamental and Environmental Research Laboratories, NEC Corporation, 34, Miyukigaoka, Tsukuba, Ibaraki 305-8501, Japan

Received April 3, 2007

Protein−ligand docking programs have been used to efficiently discover novel ligands for target proteins from large-scale compound databases. However, better scoring methods are needed. Generally, scoring functions are optimized by means of various techniques that affect their fitness for reproducing X-ray structures and protein−ligand binding affinities. However, these scoring functions do not always work well for all target proteins. A scoring function should be optimized for a target protein to enhance enrichment for structure-based virtual screening. To address this problem, we propose the supervised scoring model (SSM), which takes into account the protein−ligand binding process using docked ligand conformations with supervised learning for optimizing scoring functions against a target protein. SSM employs a rough linear correlation between binding free energy and the root mean square deviation of a native ligand for predicting binding energy. We applied SSM to the FlexX scoring function, that is, F-Score, with five different target proteins: thymidine kinase (TK), estrogen receptor (ER), acetylcholine esterase (AChE), phosphodiesterase 5 (PDE5), and peroxisome proliferator-activated receptor γ (PPARγ). For these five proteins, SSM always enhanced enrichment better than F-Score, exhibiting superior performance that was particularly remarkable for TK, AChE, and PPARγ. We also demonstrated that SSM is especially good at enhancing enrichments of the top ranks of screened compounds, which is useful in practical drug screening.

1. INTRODUCTION

Protein−ligand docking is widely used to efficiently discover novel ligands in structure-based drug design. Over the past 15 years, various docking programs have been developed and their performance has been evaluated in detail.[1−7] These docking programs attempt to predict the binding conformation of a ligand and the protein−ligand binding affinity using two computational steps: docking and scoring. Many ligand conformations are generated in the docking step. There are several conformation sampling methods, including genetic algorithms, Monte Carlo simulation, and simulated annealing. All sampling methods are guided by a function that evaluates the fitness between the protein and ligand. In the scoring step, a scoring function is used to evaluate the protein−ligand affinity. The scoring functions are important because the final predicted conformations are selected on the basis of the scores. There are three groups of scoring functions: force-field-based methods, empirical scoring functions, and knowledge-based potentials.

Force-field-based scoring functions apply classical molecular mechanics energy functions, approximating the binding free energy of protein−ligand complexes by summing the van der Waals and electrostatic interactions. Solvation is usually taken into account using a distance-dependent dielectric function, although solvent models based on continuum electrostatics have been developed.[8,9] Hydrophobic contributions are usually assumed to be proportional to the solvent-accessible surface area. A drawback is that the energy landscapes associated with force-field potentials are generally rugged, and therefore, minimization is required prior to any energy evaluation.

Empirical scoring functions estimate the binding free energy by summing interaction terms derived from weighted structural parameters. The weights are determined by fitting the scoring function to experimental binding constants of a training set of protein−ligand complexes. The main drawback is that it is unclear whether they are able to predict the binding affinity of a ligand that is structurally different from those used in the training sets.

Knowledge-based scoring functions represent binding affinity as a sum of protein−ligand atom-pair interactions. Those potentials are derived from the protein−ligand complexes with known structures, where probability distributions of interatomic distances are converted into distance-dependent interaction free energies of protein−ligand atom pairs. However, 3D structures of protein−ligand complexes do not provide a thermodynamic ensemble at equilibrium, and therefore, a knowledge-based potential should be considered as a statistical preference rather than a potential of mean force. A key ingredient of a knowledge-based potential is the reference state, which determines the weights between the various probability distributions. Several approaches to deriving these potentials have been proposed.[10−13] They differ in their definition of the reference state, the protein and ligand atom types, and the list of protein−ligand complexes from which they were extracted.

Many reports assessing the performance of docking programs have been published. Many of these reports concluded that docking algorithms reproduce binding modes

* Corresponding author phone: +81-29-850-1410; fax: +81-298-856-6136; e-mail: r-teramoto@bq.jp.nec.com.

10.1021/ci700116z CCC: $37.00
Published on Web 08/09/2007

Table 1. Protein−Ligand Complexes, Ligands, and Decoys Used in This Study

protein	PDB code	resolution (Å)	number of ligands	number of decoys
TK	1kim	2.1	21	888
ER	3ert	1.9	38	1429
AChE	1eve	2.5	105	3851
PDE5	1xp0	1.8	87	1970
PPARγ	1fm9	2.1	84	3071

very successfully and that scoring functions are less successful at identifying binding modes.[14−27] Therefore, scoring functions must be improved.

Generally, scoring functions are optimized by means of techniques that enhance their ability to reproduce X-ray structures and protein−ligand binding affinities. However, these scoring functions do not work well for all target proteins. Moreover, it is impossible to know in advance whether a scoring function performs well for a target protein. Accordingly, a scoring function should be optimized for any target protein it is applied to.

To address this problem, we propose the supervised scoring model (SSM), which takes into account the protein−ligand binding process using docked ligand conformations with supervised learning to optimize scoring functions for a target protein. SSM employs a rough linear correlation between the binding free energy and the root mean square deviation (RMSD) of a native ligand for predicting binding energy.[28]

A similar approach was proposed in the framework of consensus scoring, that is, supervised consensus scoring (SCS), and reported successful results.[29] However, SCS requires multiple scoring functions. In contrast, SSM requires only a single scoring function, and the objective of SSM is to enhance the performance of an original scoring function and provide a target-specified scoring model using supervised learning. Thus, there are differences of objectives between SCS and SSM.

To test SSM's effectiveness, we applied it to thymidine kinase (TK), estrogen receptor (ER), achetylcholine esterase (AChE), phosphodiesterase 5 (PDE5), and peroxisome proliferator-activated receptor γ (PPARγ). We evaluated SSM's performance on the basis of the enrichments with benchmarking sets for molecular docking and compared it to the FlexX scoring function, that is, F-Score.

2. METHODS

2.1. Preparation of Data Sets. Since a directory of useful decoys (DUD) provides a more stringent test by which to evaluate the performance of structure-based virtual screening, DUD is appropriate for a fair evaluation of ligand enrichment to avoid a bias of decoys. Moreover, since DUD is freely available, it is easy to compare the performance of other scoring methods by the same data sets. We collected 3D structures of five target proteins, that is, TK, ER, AChE, PDE5, and PPARγ; their ligands; and decoys from DUD to do a fair evaluation of ligand enrichment.[30] DUD stores many decoys that physically resemble ligands, so that enrichment is not simply a separation of gross features but is chemically distinct from them, so that they are unlikely to be binders. The test data sets are summarized in Table 1. Of the five target proteins, TK,[6,14,18,21,31−33] ER,[6,21,25,31,32,34−36] and AChE[25,30] have been used to benchmark and evaluate docking methods and scoring functions. PDE5 and PPARγ have not been thoroughly investigated yet. Detailed descriptions of DUD and all of the data sets are available online at http://blaster.docking.org/dud/.

Table 2. Number of Docked Conformations of Native Ligands

protein	ligand	number of ligand conformations
TK	deoxythymidine	236
ER	4-hydroxytamoxifen	108
AChE	E2020(aricept)	299
PDE5	vardenafil	293
PPARγ	GI262570	329

2.2. Docking Procedure. FlexSIS implemented in Sybyl7.1J is employed to generate an ensemble of docked conformations for each ligand.[37] Since we would like to dock diverse ligands with large variations in size and possible interactions, all water molecules in the active sites are removed to avoid biasing the docking to one particular binding mode. We generated up to 999 docked conformations for native ligands. The number of docked conformations of native ligands is summarized in Table 2. Figure 1 shows the RMSD distributions of each native ligand conformation generated by FlexSIS. We generated up to 100 docked conformations for other compounds, that is, known ligands and decoys.

2.3. Scoring Function. In this study, we used F-Score, which is given as[22]

$$\Delta G_{\text{bind}} = \Delta G_{\text{match}} F_{\text{match}} + \Delta G_{\text{lipo}} F_{\text{lipo}} + \Delta G_{\text{ambig}} F_{\text{ambig}} + \Delta G_{\text{clash}} F_{\text{clash}} + \Delta G_{\text{rot}} n_{\text{rot}} + \Delta G_0$$

Here, F_i are functions of the protein and ligand coordinates. The term F_{match} is the sum of the energy contributions from each hydrogen bond, metal contact, and specific aromatic interaction, where each contribution has been multiplied by two linear penalty functions for angle and distance deviations from predefined ideal values. Terms F_{lipo} and F_{ambig} provide a measure of hydrophobic contact as functions of protein−ligand atom pairs, F_{lipo} involving only pairs of nonpolar atoms and F_{ambig} involving pairs of one polar and one nonpolar atom. The term F_{clash} is a penalty function for protein−ligand overlap, and n_{rot} is equal to the number of rotatable bonds in the ligand. F-Score is an empirical scoring function.

2.4. Supervised Scoring Model (SSM). The overall SSM procedure is illustrated in Figure 2, and the binding free energy landscape when RMSD is used as a reaction coordinate is illustrated in Figure 3. As shown in Figure 3, we assume that protein−ligand binding has a funnel-shaped landscape, as discussed by Camacho and Vajda.[28] SSM employs this rough linear correlation between the binding free energy and RMSD of a native ligand. Thus, SSM does not estimate the binding energy directly, trying instead to estimate it indirectly via the predicted RMSD of the native ligand and provides a target-specified scoring model. In SSM, the binding energy prediction is formulated as supervised learning in which explanatory attributes and an objective variable are F-Score and the terms of F-Score, that is, F_{match}, F_{lipo}, F_{ambig}, and F_{clash}, and the RMSD between the docked conformations and the X-ray structures of ligands, respec-

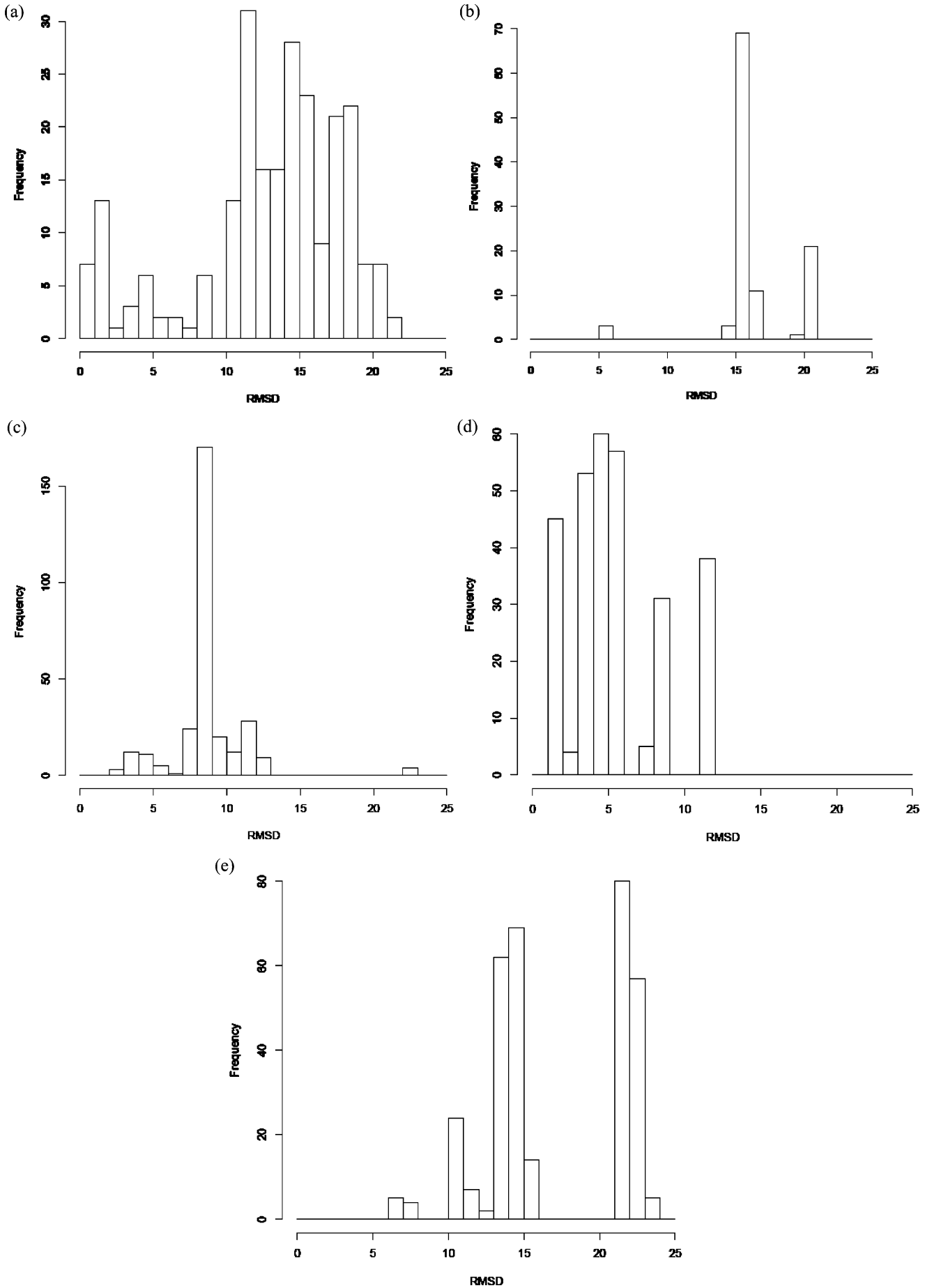

Figure 1. RMSD distributions of native ligand conformations. (a) TK (PDB code 1kim), (b) ER (PDE code 3ert), (c) AChE (PDB code 1eve), (d) PDE5 (PDB code 1xp0), and (e) PPARγ (PDB code 1fm9).

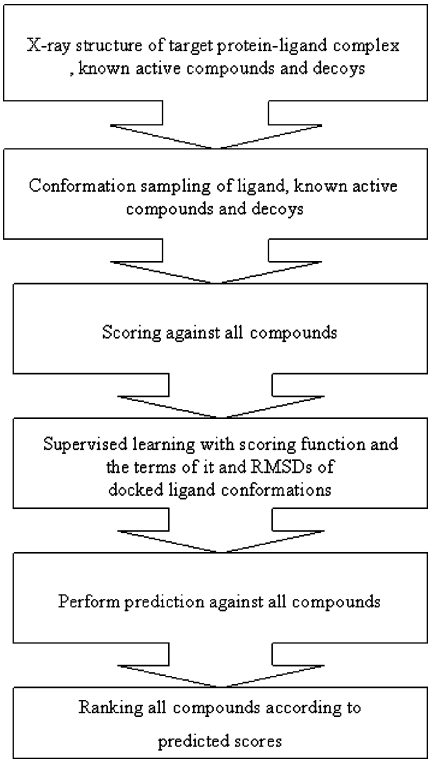

Figure 2. Overview of the SSM procedure.

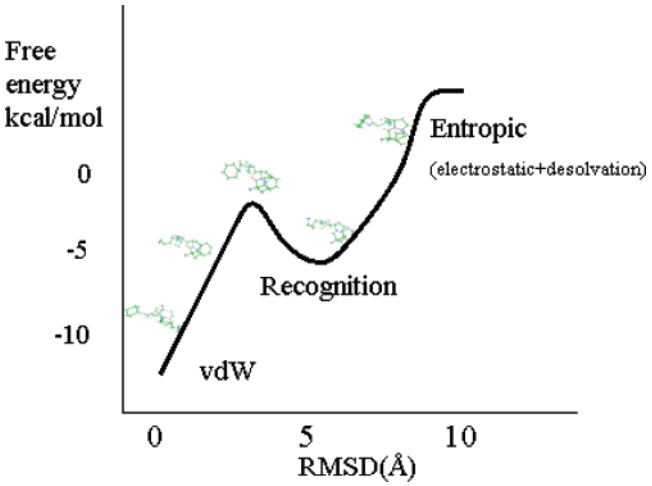

Figure 3. Binding free energy landscape when RMSD is used as a reaction coordinate. SSM employs a rough linear correlation between the binding free energy and RMSD of a native ligand for predicting binding energy. SSM takes into account the protein−ligand binding process using docked ligand conformations with supervised learning.

tively. The binding energy prediction is also formulated as a classification problem by defining ligand conformations in the range of the RMSD below the threshold as the bound state and other ligand conformations as the unbound state. For regression and classification problems, we use Random Forests as a supervised learning algorithm because it can handle both types of problems easily.[38,39] Other learning machines, such as support vector machines, decision trees, and artificial neural networks, are also available. However, they generally require time-consuming trial-and-error parameter tuning. Random Forests combines two machine-learning techniques: bagging and random feature subset selection using a decision tree and regression tree as the base learner. Bagging, which stands for *bootstrap aggregating*, uses resampling to produce pseudo-replicates to improve predictive accuracy. Random Forests can significantly improve the predictive accuracy through random feature subset selection. The Random Forests algorithm is illustrated in Figure 4. Note that F-Score itself is also included as an explanatory variable, because Random Forests is a robust learning machine and not affected by multicollinearity.[38,39]

The screening procedures are as follows. We employed F-Score, terms of F-Score, and the RMSD of docked native ligand conformations as a training set for each target protein: TK, ER, AChE, PDE5, and PPARγ. Then, SSM was applied to docked conformations of known ligands and decoys as a test set. The score of the top-ranked ligand conformation in each compound was selected as the score of each compound. Compounds were ranked in ascending order in regression and were ranked in ascending order of active class in classification.

We set the threshold on the basis of the RMSD distributions for each protein−ligand complex for classification models. Table 3 shows the RMSD thresholds for each protein−ligand complex. Since the RMSD distribution of ER is extremely imbalanced, as shown in Figure 1b, we set one threshold for ER and two thresholds for other target proteins, that is, TK, AChE, PDE5, and PPARγ. We investigated the threshold dependence of performance and the properties of each model in detail. We used the default parameters of Random Forests, except the number of trees, that is, 5000.

The performance of SSM was evaluated by ligand enrichments and compared to F-Score. The R package used in this study, randomForest, is available at http://cran-r-project.org.

3. RESULTS AND DISCUSSION

3.1. The RMSD Distributions of Native Ligand Conformations. As shown in Figure 1b and e, conformations of less than 5 Å are not sampled for ER and PPARγ. As can be seen in Figure 1a and c, the numbers of conformations less than 5 Å are relatively small, but conformations are sampled over a broad range for TK and AChE. In contrast, there are a lot of conformations less than 5 Å for PDE5 in Figure 1d. These imbalances in the RMSD distribution may be improved by preprocessing target protein structures. However, to avoid biasing performance with the intervention, we do not preprocess them. Note that SSM works even if docking calculations seem to have failed for ER and PPAR when a generally used criterion, which is RMSD < 2 Å, was employed, because SSM only requires docked native ligand conformations for the training set.

3.2. Correlation between Score and RMSD of Native Ligand Conformations. Since SSM does not estimate the binding energy directly, trying instead to indirectly estimate it via predicted RMSD, it is necessary for SSM to confirm the high correlation between the RMSD and predicted RMSD of native ligands. To achieve this, we investigate the correlation between the score of SSM's regression model (SSM_{reg}) and the RMSD of native ligand conformations. We also investigate the correlation between F-Score and RMSD to compare SSM_{reg}. We evaluate the correlation between the

(a)

Step 1. Sample with replacement to form N bootstrap samples $\{B_1, \ldots, B_N\}$.

Step 2. Use each sample B_k to construct a tree classifier T_k to predict those samples that are not in B_k (called *out-of-bag* samples). These predictions are called *out-of-bag* estimators.

Step 3. When constructing T_k, at each node splitting we first randomly select m variables, then we choose one best split from these m variables.

Step 4. Final prediction is the average or majority votes of *out-of-bag* estimators over all bootstrap samples.

(b)

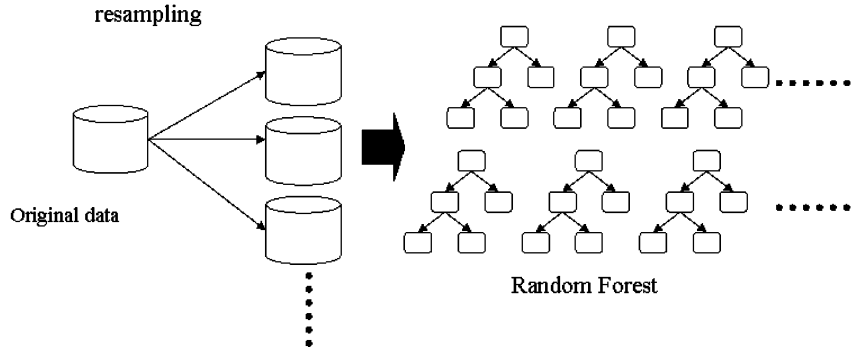

Figure 4. Random Forests algorithm. (a) Random Forests procedure, (b) illustration of Random Forests.

Table 3. RMSD Thresholds for Each Protein−Ligand Complex in Classification Models

protein	ligand	threshold (Å)
TK	deoxythymidine	1.2
ER	4-hydroxytamoxifen	6
AChE	E2020(aricept)	4.5
PDE5	vardenafil	2.3
PPARγ	GI262570	7.8

RMSD and score using the Spearman correlation coefficient (R_s), which calculates the correlation between two sets of rankings. R_s is formally defined as

$$R_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n-1)}$$

where n is the number of sets and d_i is the ranking difference of the ith ligand conformation under two criteria: the RMSD and score of SSM_{reg}. By definition, R_s ranges from -1 to $+1$. The larger R_s becomes, the more strongly two sets correlate. The score of SSM_{reg} was obtained using *out-of-bag* samples to compare F-Score fairly. *Out-of-bag* samples are samples left out of a bootstrap sample. This procedure is equivalent to n-fold asymptotic cross validation.[38,39]

Figures 5 and 6 show the correlation between the score of SSM_{reg} and RMSD of native ligand conformations and the correlation between the F-Score and RMSD, respectively. As shown in Figure 5, the lowest Spearman correlation coefficient between the RMSD and the score of SSM_{reg} is 0.83. This indicates that RMSD and SSM_{reg} are highly correlated. In contrast, as shown in Figure 6, the Spearman correlation coefficients between the RMSD and F-Score are generally low, the highest being 0.46. There are even negative correlations for ER and AChE. This indicates that RMSD and F-Score do not correlate. These results suggest that SSM may work well in virtual screening and that SSM is quite different from F-Score in terms of the concept of the scoring method. Since F-Score is designed to reproduce 3D structures of known protein−ligand complexes and not to reproduce RMSD, it is natural that F-Score does not correlate RMSD. Note that SSM does not predict and reproduce the RMSD of non-native ligands and provides scores for binding energy only.

3.3. Performance Evaluation of Virtual Screening. The main objective of our study is to compare the performance of SSM and the original F-Score as virtual screening tools when they are applied to the same target proteins. At a coarse level, virtual screening is a test of the ability of scoring methods to differentiate between active and inactive compounds. We tested five target proteins: TK, ER, AChE, PDE5, and PPARγ. Figure 7 shows the overall profile of the percentage of ligands found (y axis) plotted as a function of the percentage of the ranked docked database (x axis) for TK, ER, AChE, PDE5, and PPARγ by SSM and F-Score. As the other indicator of performance, we use the enrichment factor (EF). EF is defined as

$$\mathrm{EF} = \frac{\mathrm{Hits}_{\mathrm{sampled}}^{x\%}}{N_{\mathrm{sampled}}^{x\%}} \times \frac{N_{\mathrm{total}}}{\mathrm{Hits}_{\mathrm{total}}}$$

where $\mathrm{Hits}_{\mathrm{sampled}}^{x\%}$ is the number of hits found at x% of the database screened, $N_{\mathrm{sampled}}^{x\%}$ is the number of compounds screened at x% of the database, $\mathrm{Hits}_{\mathrm{total}}$ is the number of active compounds in the entire database, and N_{total} is the number of compounds in the entire database. EF is the relative enrichment of active compounds in the set of compounds predicted to be active in relation to the fraction of active compounds in the entire database. From the

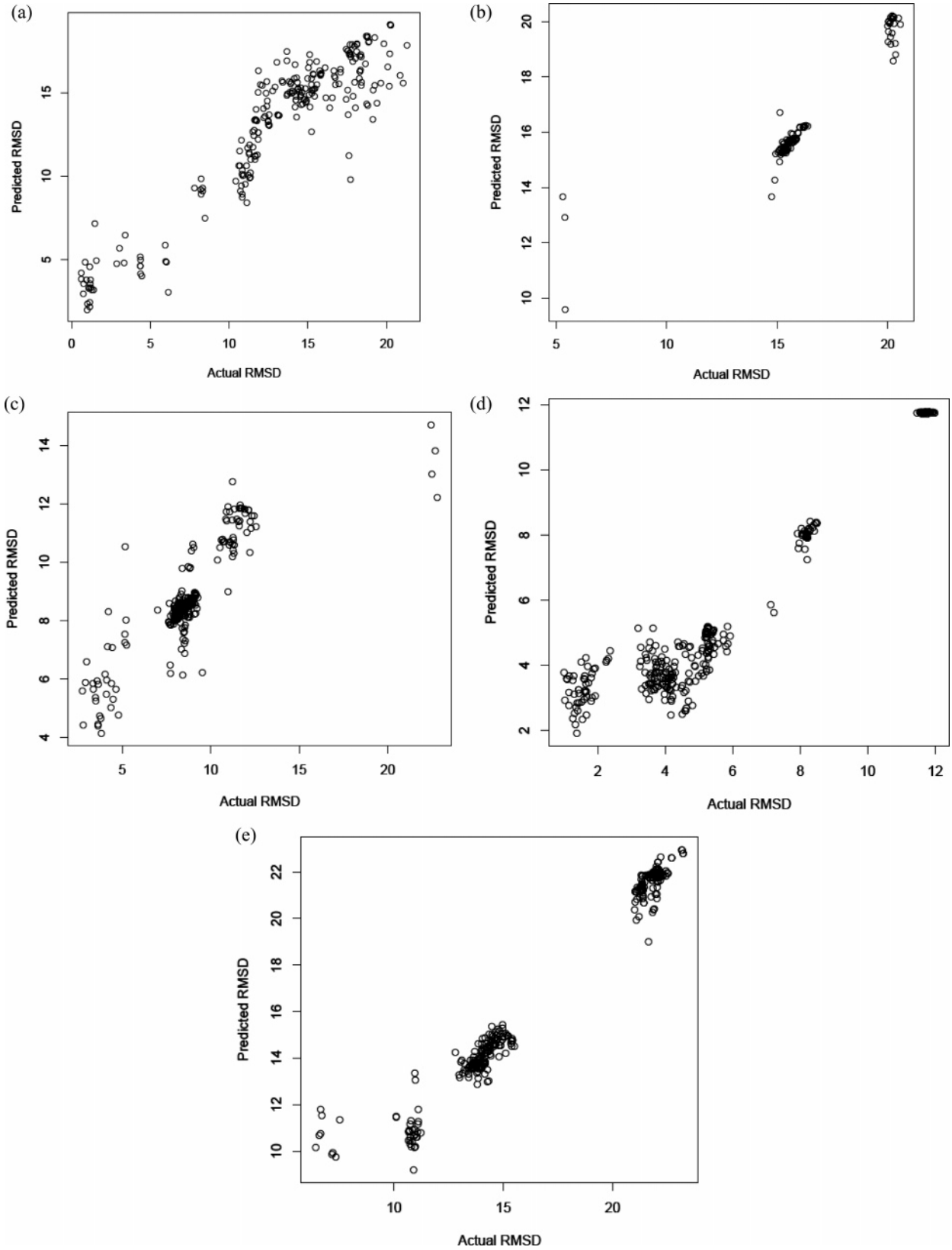

Figure 5. Correlation between RMSD and SSMreg score: (a) TK ($R_s = 0.85$), (b) ER ($R_s = 0.92$), (c) AChE ($R_s = 0.84$), (d) PDE5 ($R_s = 0.83$), and (e) PPARγ ($R_s = 0.95$).

definition of EF, the EF of random screening is 1. We calculated EF_1 (enrichment factor at 1% of the ranked database), EF_2 (enrichment factor at 2% of the ranked database), EF_5 (enrichment factor at 5% of the ranked database), EF_{10} (enrichment factor at 10% of the ranked database), and EF_{20} (enrichment factor at 20% of the ranked database) of SSM and F-Score. Enrichment factors are summarized in Table 4. As shown in Figure 7 and Table 4, SSM outperforms F-score overall for every target protein and EF. These results suggest that SSM's strategy works quite well and is effective for scoring to incorporate docked ligand conformations.

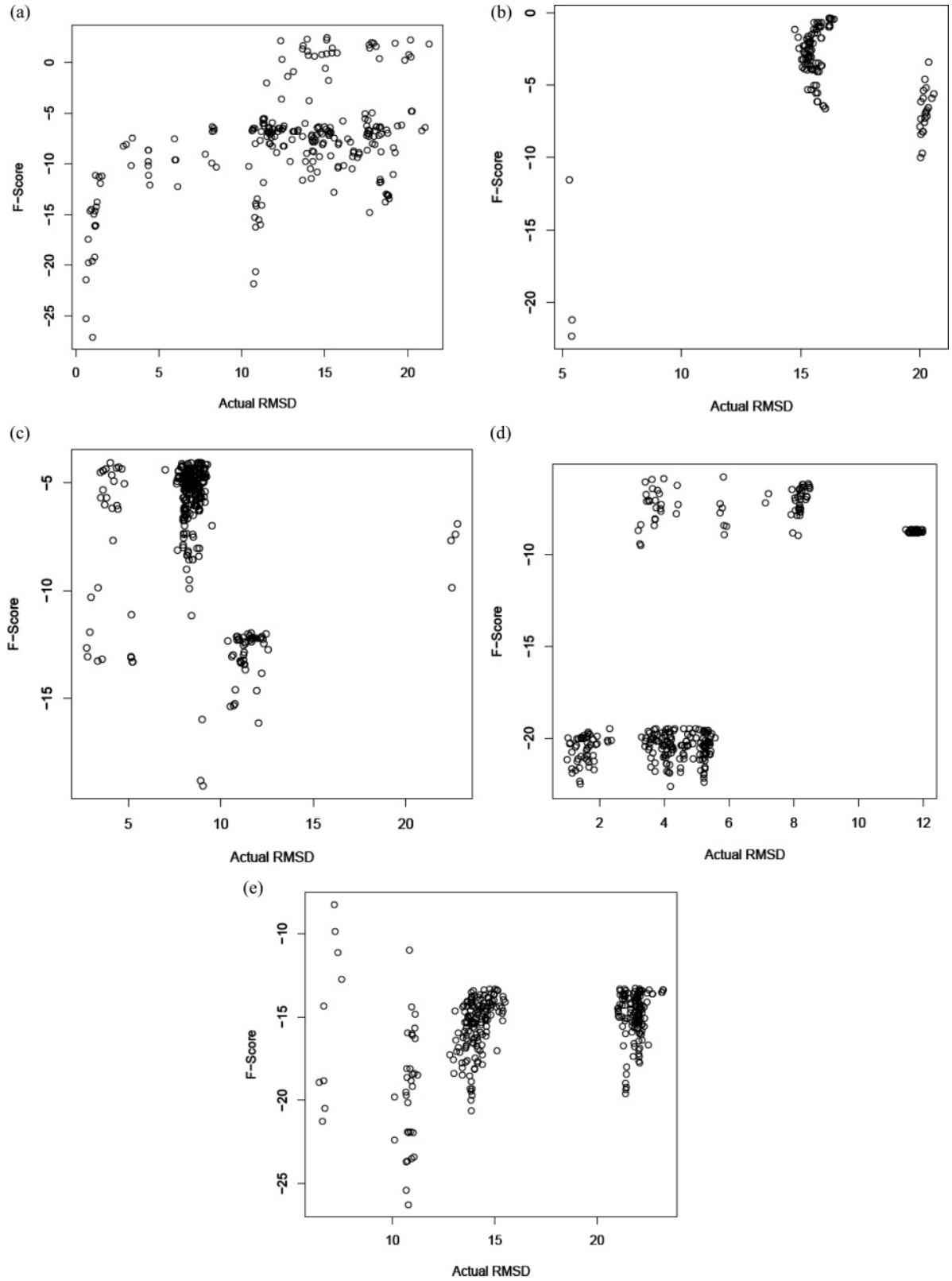

Figure 6. Correlation between RMSD and F-Score: (a) TK ($R_s = 0.38$), (b) ER ($R_s = -0.24$), (c) AChE ($R_s = -0.30$), (d) PDE5 ($R_s = 0.46$), and (e) PPARγ ($R_s = 0.43$).

Thus, although SSM's performance depends on target proteins, learning models, and thresholds of the classification model, SSM always outperforms F-Score from EF_1 to EF_5. Since the top ranks of screened compounds are useful in practical drug screening, this property is quite favorable and noteworthy.

3.4. Relationship between Score Distribution and Enrichment. As can be seen in Figures 5 and 7 and Table 4, the Spearman correlation coefficient and enrichment factors do not definitely correlate in SSM. Since the Spearman correlation coefficient for ER and PPARγ is higher than 0.9 and their enrichment factors are relatively high, it

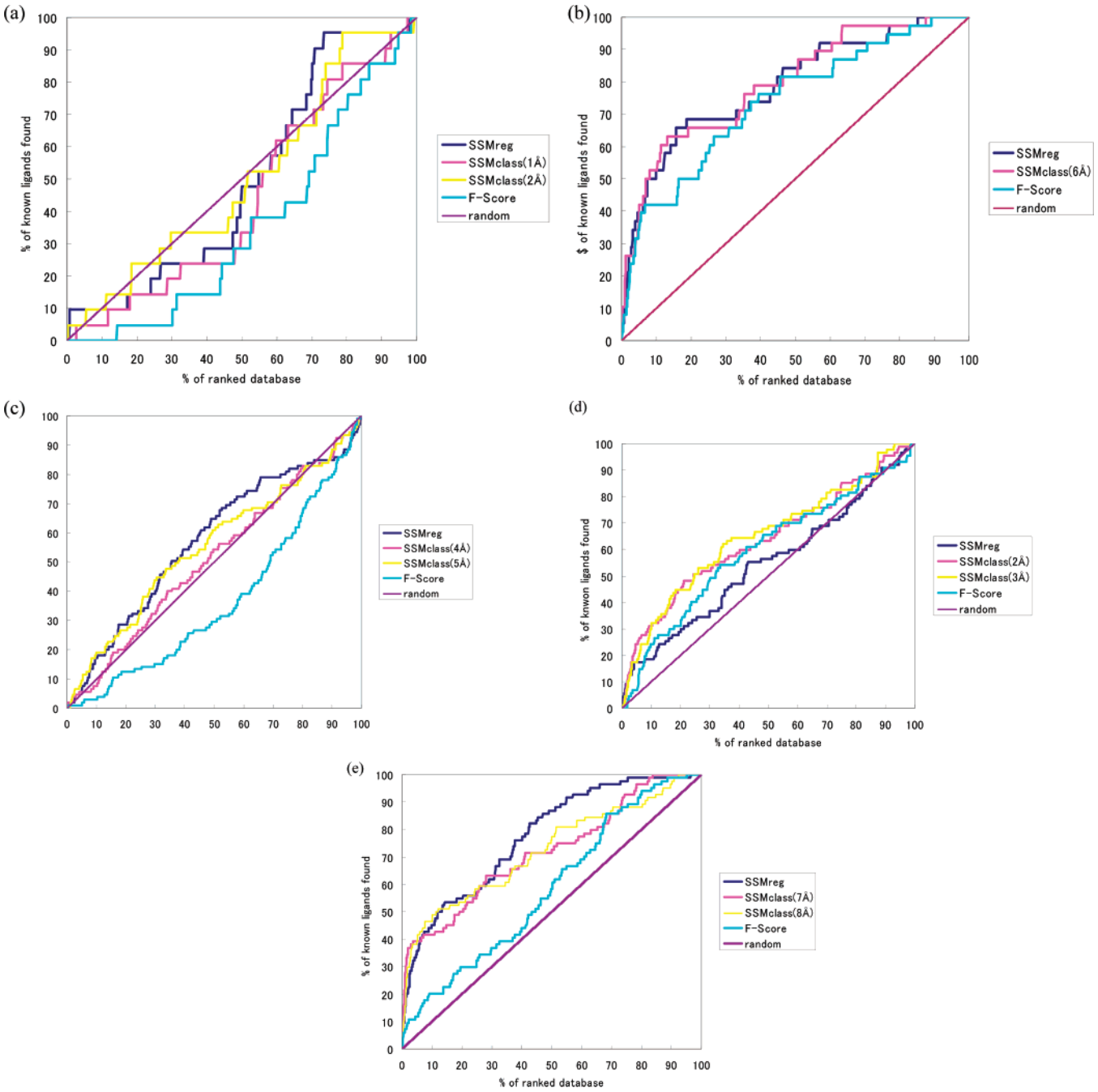

Figure 7. Docking enrichment plots for (a) TK, (b) ER, (c) AChE, (d) PDE5, and (e) PPARγ using SSM and F-Score. The docking ranked database (x axis) is plotted against the percentage of known ligands found by SSM or F-Score (y axis) at any given percentage of the ranked database. SSM_{reg} indicates a regression model of SSM, and SSM_{class} indicates a classification model of SSM with a threshold. The thresholds of SSM_{class} used in each target protein are shown in Table 3.

is possible that there is a weak correlation between them. If a broad range of native ligand conformations is sampled thoroughly, it will be possible to clarify this issue somewhat. As can be seen in Figures 6 and 7 and Table 4, there is no correlation between the Spearman correlation coefficient and enrichment factors in F-Score. This suggests that it is difficult to know the performance of virtual screening of a target protein in advance.

As can be seen in Figures 5 and 6, there is a significant difference between SSM and F-Score in the reproduction of binding modes of native ligands. For TK, both SSM and F-Score discriminate between docked conformations of small and large RMSDs. However, F-Score cannot discriminate between ligand conformations of less than 5 Å and about 10 Å. This means that SSM works better than F-Score among the top ranks of screened compounds. For ER, the distribution of docked conformations is extremely imbalanced. F-Score cannot discriminate between docked conformations in the range above 15 Å. On the other hand, SSM relatively discriminates between docked conformations in the range above 15 Å. This may explain why SSM works better than F-Score among the top ranks of screened compounds. For AChE, there is a relatively strong correlation between SSM and RMSD but no correlation between F-Score and RMSD. In addition, from Figure 5c, there are two clusters in the ranges 5 Å < RMSD < 10 Å and 10 Å < RMSD < 15 Å.

Table 4. Enrichment Factors of SSM and F-Score

scoring method	EF_1	EF_2	EF_5	EF_{10}	EF_{20}
		(a) TK			
F-Score	0	0	0	0	0.24
SSM_{reg}	**9.52**	**4.76**	**1.9**	**0.95**	0.71
SSM_{class} (1 Å)	0	0	0.95	0.48	0.71
SSM_{class} (2 Å)	4.76	2.38	0.95	**0.95**	**1.19**
		(b) ER			
F-Score	7.89	7.89	6.84	4.21	2.24
SSM_{reg}	10.53	10.53	**7.89**	**5.26**	**3.42**
SSM_{class} (6 Å)	**15.79**	**13.16**	7.37	**5.26**	3.29
		(c) AChE			
F-Score	0.95	0.48	0.19	0.29	0.62
SSM_{reg}	**1.9**	0.95	1.33	1.71	**1.43**
SSM_{class} (4 Å)	**1.9**	1.43	1.14	0.76	1.05
SSM_{class} (5 Å)	0	**2.38**	**1.9**	**1.9**	1.33
		(d) PDE5			
F-Score	0	1.15	1.38	2.41	1.55
SSM_{reg}	**5.75**	**5.75**	3.45	1.84	1.44
SSM_{class} (2 Å)	3.45	**5.75**	**4.83**	**3.1**	**2.24**
SSM_{class} (3 Å)	2.3	4.02	3.45	2.99	**2.24**
		(e) PPARγ			
F-Score	7.14	4.76	2.38	2.02	1.49
SSM_{reg}	14.29	10.71	7.14	4.52	**2.74**
SSM_{class} (7 Å)	**25**	**17.86**	**7.86**	4.17	2.5
SSM_{class} (8 Å)	14.29	14.88	7.62	**4.64**	2.68

The F-Score and RMSD of these clusters do not correlate. On the other hand, from Figure 6c, the SSM and RMSD of these clusters correlate well. These results may suggest that SSM works better than F-Score. For PDE5, both SSM and F-Score can discriminate well between docked conformations in the range lower than 6 Å. However, SSM is superior to F-Score in terms of the correlation between score and RMSD. This may explain why SSM works better than F-Score among top-ranked compounds. For PPARγ, SSM achieves much better correlation between the score and RMSD than F-Score. In addition, from Figure 5e, there are two clusters in the ranges 10 Å < RMSD < 15 Å and 20 Å < RMSD. The F-Score and RMSD of these clusters do not correlate. On the other hand, from Figure 6e, the SSM and RMSD of these clusters correlate well. These results may suggest that SSM works much better than F-Score.

In this study, SSM always outperformed F-Score, the original scoring function. However, since SSM constructs a scoring model from protein−ligand complexes, it is possible that the performance of virtual screening was not as good as that for an original scoring function. Moreover, if there is no correlation between the RMSD and SSM score, the relationship between binding free energy and RMSD, as shown in Figure 3, is not available for virtual screening because SSM does not estimate the binding energy directly, trying instead to estimate it indirectly via predicted RMSD. SSM may potentially be more suitable for finding the compounds with similar binding modes or binding specificity than simply finding lead compounds on the basis of the formulation of SSM. This will be both an advantage and a drawback, and it is necessary to choose the scoring method on the basis of the purpose of virtual screening.

4. CONCLUSION

To enhance the performance of the existing scoring function, we have proposed the SSM, which takes into account the protein−ligand binding process using docked ligand conformations with supervised learning to optimize the scoring function for a target protein. We constructed scoring models using F-Score and the terms of F-Score and RMSDs with Random Forests, which is a kind of supervised learning. To test SSM's effectiveness, we applied it to TK, ER, AChE, PDE5, and PPARγ. SSM's performance was evaluated on the basis of the enrichments with benchmarking sets for molecular docking.

We compared the performances of SSM and F-Score and demonstrated that SSM always outperforms F-Score. SSM particularly drastically outperforms F-Score for TK, AChE, and PPARg. We also demonstrated that SSM especially enhances enrichments of the top ranks of screened compounds, which is useful in practical drug screening. This property is quite valuable and noteworthy.

In this study, SSM always outperformed F-Score, the original scoring function. However, since SSM constructs a scoring model from protein−ligand complexes, it is possible that the virtual screening performs less well than the original scoring function. On the basis of its formulation, SSM may potentially be more suitable for finding compounds with similar binding modes or binding specificities than simply for finding lead compounds. This will be both an advantage and a drawback, and it is necessary to choose the scoring method on the basis of the purpose of the virtual screening.

Because of the limitations of the available docking program, we applied SSM only to F-Score. However, SSM can easily be applied to other scoring functions, and it seems that SSM also works well with them. Our proposed method will accelerate the drug discovery process.

ACKNOWLEDGMENT

We thank our colleagues at NEC Corp. for their fruitful discussions.

REFERENCES AND NOTES

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Sheridan, R. P.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(2) Abagyan, R. A.; Totrov, M. M.; Kuznetsov, D. A. ICM: a new method for structure modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488−506.

(3) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−89.

(4) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449−462.

(5) Jones, G.; Wilett, 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.

(6) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499−511.

(7) 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.

(8) Majeux, N.; Scarsi, M.; Apostolakis, J.; Caisch, A. Exhaustive docking of molecular fragments on protein binding sites with electrostatic salvation. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88−105.

(9) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.

(10) DeWitte, R.; Shakhnovich, E. SMoG: de novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733−11744.

(11) Mitchell, J. B. O.; Laskowski, R. A.; Alexander, A.; Thornton, J. M. BLEEP-Poteintial of mean force describing protein-ligand interactions: I. Generating potential. *J. Comput. Chem.* **1999**, *20*, 1165−1176.
(12) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(13) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.
(14) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57*, 225−242.
(15) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *57*, 225−242.
(16) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: Comparative data on docking algorithms. *J. Med. Chem.* **2004**, *56*, 235−249.
(17) Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J. Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlen, M.; Stouten, P. F. W. Assessment of docking poses: Interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871−881.
(18) Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* **2005**, *26*, 11−22.
(19) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.
(20) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(21) 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.
(22) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(23) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281−295.
(24) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(25) Jacobsson, M.; Liden, P.; Stjernschantz, E.; Bostrom, H.; Norinder, U. Improvement structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781−5789.
(26) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, T. D.; Watson, P. Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(27) Yang, J. M.; Chen, Y. F.; Shen, T. W.; Kristal, B. S.; Hsu, D. F. Consensus scoring criteria for improving enrichment in virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 1134−1146.
(28) Camacho, C. J.; Vajda, S. Protein docking along smooth association pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10636−10641.
(29) Teramoto, R.; Fukunishi, H. Supervised consensus scoring for docking and virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 526−534.
(30) Huang, N.; Schoichet, K. B.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(31) Pham, T. A.; Jain, A. N. Parameter estimation for scoring protein-ligand interactions using negative training data. *J. Med. Chem.* **2005**, *49*, 5856−5868.
(32) Li, H.; Li, C.; Gui, C.; Luo, X.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. GAsDock: A new approcah for rapid flexible docking based on a improved multi-population genetic algorithm. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4671−4676.
(33) Yang, J. M.; Shen, T. W. A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators. *Proteins* **2005**, *59*, 205−220.
(34) Schapira, M.; Abagyan, R.; Totrov, M. Nuclear hormone receptor targeted virtual screening. *J. Med. Chem.* **2003**, *46*, 3045−3059.
(35) Schulz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening: Evaluation of current docking tools. *J. Mol. Model.* **2003**, *9*, 47−57.
(36) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895−2907.
(37) *FlexSIS*; *Sybyl7.1J*; BioSolveIT GmbH: Sankt Augustin, Germany, 2005.
(38) Breiman, L. Random forests. *Machine Learning* **2001**, *45*, 5−32.
(39) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, C.; Sheridan, R.; Feuston, B. Random Forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Model.* **2003**, *43*, 1947−1958.

CI700116Z