An Efficient in Silico Screening Method Based on the Protein−Compound Affinity Matrix and Its Application to the Design of a Focused Library for Cytochrome P450 (CYP) Ligands

Yoshifumi Fukunishi,*[,†,‡] Shinichi Hojo,[‡] and Haruki Nakamura[†,§]

Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan, Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1, Minamiosawa, Hachiouji, Tokyo 192-0397, Japan, and Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Received August 1, 2006

A new method has been developed to design a focused library based on available active compounds using protein−compound docking simulations. This method was applied to the design of a focused library for cytochrome P450 (CYP) ligands, not only to distinguish CYP ligands from other compounds but also to identify the putative ligands for a particular CYP. Principal component analysis (PCA) was applied to the protein−compound affinity matrix, which was obtained by thorough docking calculations between a large set of protein pockets and chemical compounds. Each compound was depicted as a point in the PCA space. Compounds that were close to the known active compounds were selected as candidate hit compounds. A machine-learning technique optimized the docking scores of the protein−compound affinity matrix to maximize the database enrichment of the known active compounds, providing an optimized focused library.

1. INTRODUCTION

Recently, virtual (in silico) screening has played an important role in drug screening. There are two main approaches: structure-based screening based on the 3D structure of the target protein and ligand-based screening, which is a type of similarity search based on 1D and/or 2D descriptors of the compound using available active compounds.

For the structure-based in silico drug screening, protein−ligand docking is a key technology, and many protein−ligand docking programs have been reported.[1−12] This approach has succeeded when the target protein structure is properly prepared and the size of the ligand is small.[13−20] The hit ratio of conventional single-target *i*n silico screening is low, and several multiple-target screening methods have been proposed to improve the database enrichment.[12,21,22] Multiple-target screening methods utilize a protein-compound affinity matrix, which is a matrix of docking scores or docking affinities between a set of proteins and a set of compounds.

For the ligand-based drug screening, usually a 1D or 2D descriptor of compounds is used to evaluate similarities between the compounds in the library and the known active compounds.[23,24] The affinity fingerprint approach developed by Kauvar et al.[25] is a new type of similarity search method based on a multiprotein/multicompound affinity matrix. In their study, the IC_{50} value of the target protein was estimated from the IC_{50} values of many other proteins. Later, the protein−compound docking score came to be used as the descriptor of the compound instead of the usual 1D or 2D descriptors, that is, mass, number of rotatable bonds, number of hydrogen donors/acceptors of the compound, and so forth.[26−30]

Recently, we reported a new ligand-based method for the classification of chemical compounds based on protein−compound docking scores instead of 1D/2D descriptors, and we applied it to a random screening experiment for a macrophage migration inhibitory factor (MIF).[31] Principal component analysis (PCA) was applied to the protein−compound interaction matrix to distinguish the active compounds of MIF from the negative compounds. A random screening experiment for MIF was performed, and our method revealed that the active compounds were localized in the PCA space of the compounds, which is a finite-dimension Hilbert space, while the negative compounds showed a wide distribution. In the PCA space, the compounds in a multidimensional sphere whose center was set to the average coordinates of known compounds were selected as a focused library, the database enrichment of which was equivalent to or better than that obtained by in silico screening. We call this method the docking score index (DSI) method.

The DSI method utilizes the subspace of the whole PCA space; thus, its database enrichment depends on the number of the principal components of the subspace that are used. Also, the active compounds are not always localized on the major principal component axes, but they form a cluster along the minor principal component axes. This is because the major principal component carries information mainly about the absolute values of the docking scores but does not

* Corresponding author tel.: +81-3-3599-8290; fax: +81-3-3599-8099; e-mail: y-fukunishi@jbirc.aist.go.jp.

[†] National Institute of Advanced Industrial Science and Technology.

[‡] Tokyo Metropolitan University.

[§] Osaka University.

10.1021/ci600334u CCC: $33.50
Published on Web 11/02/2006

always reflect information on the formation of clusters separating the active compounds from the inactive ones. Thus, the DSI method was first modified to select a set of best principle components in which the active compounds are well-separated from the inactive compounds. Second, the docking score of the protein−compound affinity matrix was modified to improve the database enrichment. In this study, we propose a machine-learning approach to modify the docking score with an automatic method to select a set of suitable principal component axes.

P450 (CYP) is an essential protein for drug metabolism, as most drugs are metabolized by CYP in the liver and small intestine. There are many subtypes of the CYP family, and each subtype metabolizes specific substrates. However, the selectivity of CYP family members is low. In fact, about half of the commonly prescribed drugs are metabolized by CYP3A4.[32]

Polymorphism of CYP3A4 is rare; by contrast, CYP2C9, CYP2D6, and CYP2C19 show frequent polymorphism, and their drug clearance depends on this polymorphism.[33] Thus, drugs that are metabolized by CYP2C9, CYP2D6, or CYP2C19 are not desirable. CYP metabolism has been predicted on the basis of 1D and/or 2D descriptors of compounds, such as quantitative structure−activity relationship, and statistical methods.[34,35] Also, comparative molecular field analysis, which is based on the 3D structure of the compound, has been used for such predictions.[35] These methods have successfully predicted absorption, distribution, metabolism, and excretion properties; however, the prediction performance is still not very satisfactory. Thus, we applied the machine-learning technique to our DSI method, to design a focused library for CYP metabolism. Our method requires neither the 3D structure of the target protein nor 1D/2D descriptors of the compound.

2. METHODS

2.1. Machine Learning Docking Score Index Method. Our in silico drug screening method is a type of similarity search based on known active compounds. The distance between two compounds is estimated on the basis of the protein−ligand interaction matrix, each element of which is the corresponding docking score. From the covariance matrix of compounds, PCA is performed to find similar clusters of compounds.[31] This DSI method was described in detail in our previous paper[31] and is briefly reviewed below.

We prepare a set of pockets P = {p$_1$, p$_2$, p$_3$, ... p$_{N_r}$}, where p$_i$ represents the *i*th pocket and N_r the total number of pockets, and a set of compounds X = {x^1, x^2, ... x^{N_c}}, where x^k represents the *k*th compound and N_c the total number of compounds. For each pocket p$_i$, all compounds of the set X are docked to the pocket p$_i$ with a score of s_i^k between the *i*th pocket and the *k*th compound. Here, s_i^k corresponds to the binding free energy.

The covariance matrix $\mathbf{M}^P$ of the proteins is defined as

$$\mathbf{M}^{P}_{ij} = \frac{1}{N_c}\sum_{k=1}^{N_c}(s_i^k - \bar{s}_i)(s_j^k - \bar{s}_j) \qquad (1)$$

and

$$\bar{s}_i = \frac{1}{N_c}\sum_{k}^{N_c} s_i^k \qquad (2)$$

where the upper bar represents an average. Let ϕ_j be the *j*th eigenvector of $\mathbf{M}^P$ with an eigenvalue ϵ_j, and let the order of ϵ_j be descendant. The vector of docking scores for the *k*th compound $\mathbf{X}_k$ = (s$_1^k$, s$_2^k$, ... s$_{N_r}^k$) is represented by the linear combination of ϕ_j

$$\mathbf{X}_k = \sum_{j=1}^{N_r} c_j^k \phi_j \qquad (3)$$

The coefficient $\{c_j^k\}$ represents the *j*th coordinate of the PCA space of the *k*th compound. In this study, we call this coefficient $\{c_j^k\}$ the docking score index.

Candidate hit compounds are selected using the following method. In the PCA space, compounds that are close to the known active compounds are selected as the candidate hit compounds. In the original version of the DSI method, the distance from the *k*th compound to the average position of the active compounds (D_k) is defined as

$$D_k = \sqrt{\sum_{j=1}^{N_{\mathrm{select}}} (c_j^k - \bar{c}_j)^2} \qquad (4)$$

and

$$\bar{c}_j = \sum c_j^{\mathrm{active}}/N_a \qquad (5)$$

where c_j^{active} and N_a are the DSI values of the active compounds and the total number of the active compounds, respectively.

The standard deviations (σ) of the DSI values were calculated for each axis, and DSI values more than 5σ distant from the origin were removed from the analysis. We adopt a standard Euclidian distance; namely, the DSI values were scaled to set the standard deviation of the distribution of compounds of each axis to 1.

Because the selection of principal components is effective at distinguishing particular data from others, in this study, the suffix *j* runs over the selected axes {α$_1$, α$_2$, ... α$_{N\mathrm{select}}$} in eq 4,[36,37] and the next modified distance D_k' is introduced.

$$D_k' = \sqrt{\sum_{j=\{\alpha_1,\alpha_2,...\alpha_{N_{\mathrm{select}}}\}} (c_j^k - \bar{c}_j)^2} \qquad (6)$$

The principal component axes are selected in the following manner. The contribution of each principal component is estimated using a database enrichment curve. The surface area under the database enrichment curve q_α is evaluated for the αth principal component axis; namely, the suffix *j* in eq 6 is set as α and N_{select} is set as 1, and the database enrichment curve f_α is calculated for the αth axis. The q_α values are calculated by

$$q_\alpha = \int_0^{100} f_\alpha(x)\,\mathrm{d}x \qquad (7)$$

where *x* and $f_\alpha(x)$ are the percentages of compounds that are selected from the total compound library and the database enrichment curve, respectively. A higher q_α value corresponds to better database enrichment, and the q_α value is always more than zero and less than 100. For the random screening, $q_\alpha = 50$.

The axes are sorted in descending order with respect to the q_α value. The surface area (q) under the total database enrichment curve (f) is a measure of the database enrichment as well as q_α in eq 7.

$$q = \int_0^{100} f(x)\,\mathrm{d}x \tag{8}$$

The q value is calculated by changing the number of axes (N_{select}) used in eq 6 to find the optimal N_{select} value, which gives the maximum q value.

The DSI method is applied only to cases in which the known active compounds are available; hence, the docking score can be modified to increase the database enrichment. If the new docking score is given by the linear combination of the docking scores of many proteins, we can optimize the coefficients of the linear combination to maximize the database enrichment. Let $s_a^{\mathrm{new}\,i}$, s_b^i, and $\mathbf{M}_a^b$ be the new docking score of the ith compound with the ath protein, the raw docking score of the ith compound with the bth protein, and the constant coefficient, respectively.

$$s_a^{\mathrm{new}\,i} = \sum_b s_b^i\,\mathbf{M}_a^b \tag{9}$$

The optimization procedure for $\mathbf{M}_a^b$ is as follows.

Step 1. The initial matrix **M** in eq 9 is set as a unit matrix ($\mathbf{M}_a^b = \delta_a^b$). The new docking scores are equal to the original docking scores. Then, the DSI method gives the q value by eq 8.

Step 2. Many new matrixes **M** are generated from the seed matrix **M** by using random numbers. In the first step, the seed matrix **M** is the initial matrix **M**, which is a unit matrix. The $a-b$ element of the new matrix **M** ($\mathbf{M}_a^{\mathrm{new}b}$) is given by $\mathbf{M}_a^{\mathrm{new}\,b} = \mathbf{M}_a^b + \eta_a^b$; here, η_a^b is a random number and $-1 < \eta_a^b < 1$. In this study, the number of newly generated matrixes is set at 20.

Step 3. When each newly generated matrix is used, the new docking score is calculated by eq 6. Then, the DSI method gives the q value by eq 8. The best matrix **M**, which gives the highest q value, is selected as the seed matrix for step 2.

Steps 2 and 3 are repeated until the q value shows convergence; in this study, the number of cycles is set at 40. This method is called the machine-learning docking score index (ML-DSI) method. When $\mathbf{M}_a^b$ in eq 9 is set as a unit matrix, the method is called the factor-selection docking score index (FS-DSI) method. Importantly, in the FS-DSI method, the important principal component axes are selected without machine learning. The ML-DSI method always gives a better result than the FS-DSI method does; however, we used both methods to demonstrate the performance of machine learning.

Protein−compound docking simulation was performed by the program Sievgene,[12] which is a protein−ligand flexible docking program for in silico drug screening. The protein−compound interactions accounted for by this program are van der Waals, Coulomb, hydrogen-bond, and hydrophobic interactions. In addition, the entropy change of the ligand is estimated according to the number of rotatable bonds of the ligand. This program generates many conformers (100 conformers by default) for each compound to select only one binding pose as the final result and keeps the target protein structure rigid, but with soft interaction forces adapting its slight structural change to some extent.[12] The docking program, Sievgene, reconstructed 18.9%, 50.8%, and 59.8% of a total of 132 complexes with root-mean-square deviation (RMSD) values of <1, 2, and 3 Å, respectively, where the RMSD was calculated between all atom positions, with the exception of the H atoms of each docked compound and the corresponding atom positions in the complex crystal structure.[12] When almost the same data set was used, DOCK, FlexX, and GOLD were reported to reconstruct 39%, 51%, and 56% of the complexes with RMSDs < 2 Å, respectively.[38] The results predicted by our program were almost the same as those predicted by other docking programs; as expected, the DSI/FS-DSI/ML-DSI results predicted by our program were almost the same as those predicted by other docking programs.[38] Our docking program, Sievgene, is part of the myPresto system, which is available from the Web site (http://www.jbic.or.jp/activity/st_pr_pj/mypresto/index_mypr.html) and is free for academic use.

3. PREPARATION OF MATERIALS

To evaluate our in silico screening method, we performed a protein−compound docking simulation based on the soluble protein structures registered in the Protein Data Bank (PDB). The protein−ligand complex structures in the PDB were suitable for our docking study, because the ligand pockets were clearly determined. A total of 180 proteins were selected from the PDB: 142 complexes out of 180 were selected from the database used in the evaluation of the GOLD and FlexX,[38] and 38 additional complexes were selected from the PDB. The former data set contains a rich variety of proteins and compounds whose structures were all determined by high-quality experiments with a resolution of less than 2.5 Å. Almost all of the atom coordinates are supplied, except for those of the hydrogen atoms, and the all-atomic structures around the ligand pockets are quite reliable. Thus, this data set was used in the clustering analysis of proteins and in silico screening. From the original data set, the complexes containing a covalent bond between the protein and ligand were removed, because our docking program cannot perform protein−ligand docking when a covalent bond exists between the protein and the ligand. The set of the other 38 structures included the human immunodeficiency virus protease-1 (HIV), cyclooxygenase-2 (COX-2), and glutathione S-transferase (GST). The PDB identifiers are summarized in Appendix A. All water molecules and cofactors were removed from the proteins, and all missing hydrogen atoms were added to form the all-atom models of proteins.

The binding pockets of these proteins were indicated by the positions of the ligands of the protein−compound complex structures. The test compound from the compound library can be moved around the binding pocket in the docking process while maintaining a minimum distance between the atoms of the ligand and the test compound of less than 6.5 Å.

Four subsets of proteins were selected from the entire set of 180 proteins by a clustering method.[12] The entire 180-

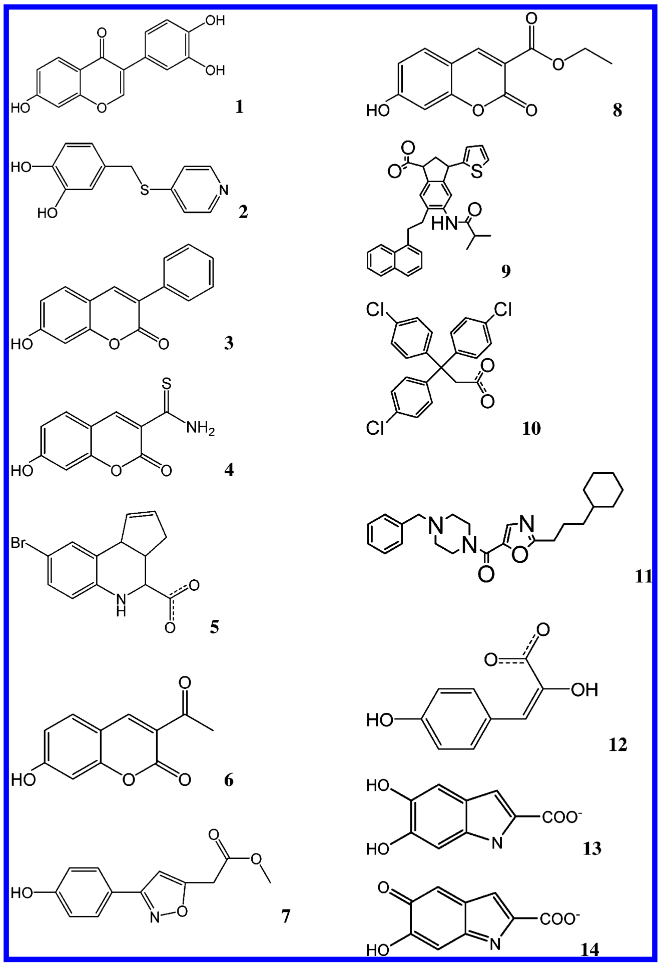

Figure 1. MIF-active compounds.

protein set was named protein set a. The four subsets were named protein sets b, c, d, and e, and these sets consisted of 123, 93, 63, and 24 proteins, respectively. The lists of the PDB codes of the four subsets are summarized in Appendix A.

Our target proteins were the macrophage MIF (PDB code: 1gcz), COX-2 (1cx2, 1pxx, 3pgh, 4cox, 5cox, and 6cox), HIV (1aid, 1hpx, and 1ivp), thermolysin (2tmn), GST (18gs, 2gss, and 3pgt), the histamine H1 receptor, the adrenaline β receptor, the serotonin receptor, and the dopamine D2 receptor. The compound set for validation tests of the ML-DSI, FS-DSI, and DSI methods consisted of 14 inhibitors of MIF, 28 inhibitors of thermolysin, 15 inhibitors of COX-2, 20 inhibitors of HIV, 12 inhibitors of GST, 10 antagonists of the histamine H1 receptor,[39] 12 agonists and 13 antagonists of the adrenaline β receptor,[40] eight agonists and nine antagonists of the serotonin receptor,[41] and six agonists and 15 antagonists of the dopamine D2 receptor[42] as the active compounds, along with 11 050 potential-negative compounds from the random compound library of the Coelacanth Chemical Corporation (East Windsor, NJ). Typically, only one hit compound could be found out of 10^4 randomly selected compounds; we therefore expected that there would be no more than a few, if any, hit compounds among these 11 212 compounds. The active compounds of MIF are depicted in Figure 1, and the other 148 active compounds are listed in Appendix B. In Figure 1, compounds **7** and **12** were newly selected from the PDB, while compounds **1**, **3**, **4**, **6**, and **8** had been reported in a previous study.[13] The others (compounds **2**, **5**, **9**, **10**, and **11**) were prepared in our previous study.[31] Compounds **13** and **14** are, respectively, D-dopachrome and 5,6-dihydroxyindole-2-carboxylic acid, which are the native ligands of MIF.[13]

To evaluate our method of constructing a CYP ligand library, we prepared a set of CYP inhibitors and a set of CYP substrates. The six target CYP proteins were CYP 1A2; CYP 2C19; CYP 2C8; CYP 2C9; CYP 2D6; and CYP 3A4, 5, 7. The inhibitors and substrates are listed in Appendix C. The compound set used for the study of CYP ligands consisted of the CYP ligands listed in Appendix C and 11 050 compounds of the random library described above.

The size distribution of compounds was as follows: the percentage of compounds with 0∼19 atoms, 0.1%; with 20∼29 atoms, 1.2%; with 30∼39 atoms, 1.6%; with 40∼49 atoms, 9.3%; with 50∼59 atoms, 22.5%; with 60∼69 atoms, 37.9%; with 70∼79 atoms, 20.5%; and with more than 80 atoms, 7.0%. The average compound size was 64.3 atoms.

The 3D coordinates of the above 11 050 random compounds were generated by the Concord program (Tripos, St. Louis, MO) from 2D Sybyl SD files provided by the Coelacanth Chemical Corporation. The 3D coordinates of the inhibitors, substrates, agonists, and antagonists were generated by ChemBats3D (Cambridge Software, Cambridge, MA). The atomic charges of each compound were determined by the Gasteiger method.[43,44] The atomic charges of proteins were the same as the atomic charges of AMBER parm99.[45]

4. RESULTS

4.1. In Silico Drug Screening Results by the DSI, FS-DSI, and ML-DSI Methods.

Figure 2a−d show the PCA results for 14 MIF-active compounds and another 11 198 potential negative compounds by the DSI and ML-DSI methods, respectively. In Figure 2a and b, the active compounds show slight localization in the PCA space. The q_α values for the first, second, third, and fourth principal components were 44.9, 76.3, 53.5, and 50.0, respectively. In contrast, in Figure 2c and d, the active compounds are localized in the PCA space. The q_α values for the first, second, third, and fourth principal components were 85.0, 77.7, 77.0, and 75.6, respectively. Thus, the ML-DSI method worked well to modify the docking score and select the appropriate PC axes on which the active compounds are well-localized.

The ML-DSI, FS-DSI, and DSI methods were applied to drug screening for the 12 targets. Figure 3a−e show the average database enrichment curves for the 12 targets. The q values of these 12 target proteins are summarized in Table 1.

To evaluate the efficiency of this method, the Jack-knife test was applied; namely, the active compounds of each target protein were divided into two sets, the known active compounds for machine learning and the unknown active compounds discovered by the software. For each target protein, the number of known active compounds was equal to that of the unknown active compounds. Ten pairs of these active-compound sets were prepared for each target protein. Thus, a total of 120 (= 12 targets × 10 trials) database enrichment curves were calculated for these 12 target proteins, and the results were averaged.

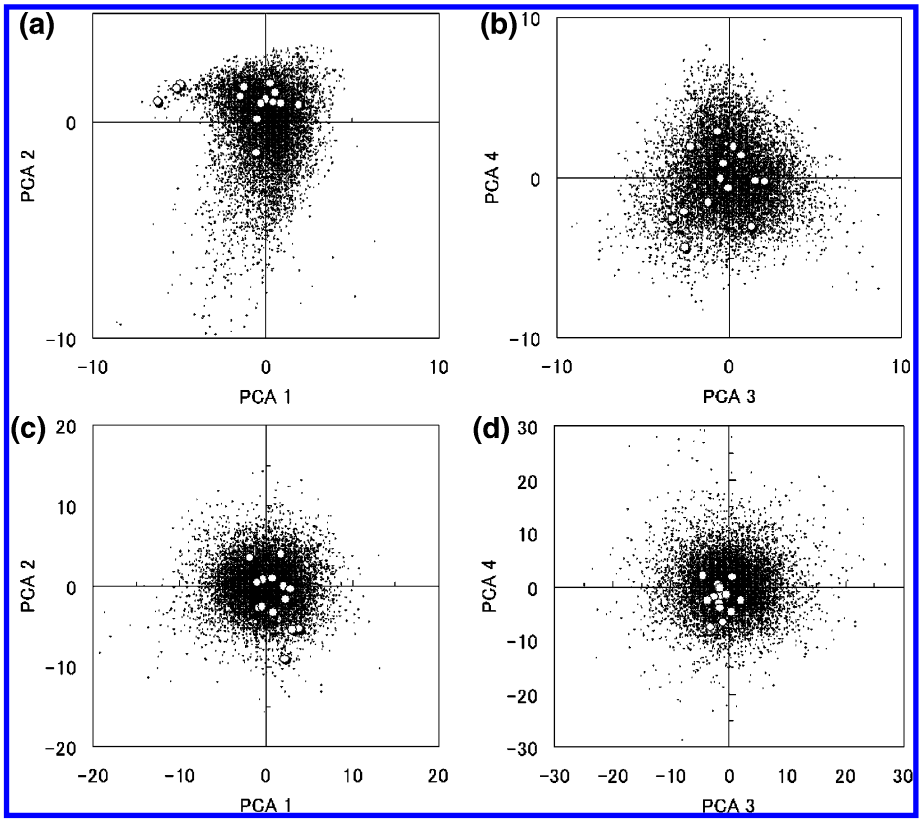

Figure 2. PCA results for the MIF-active compounds and the other compounds. The open circles and black dots represent the MIF-active compounds and the other compounds, respectively. (a) PCA results by the DSI method. “PCA 1” and “PCA 2” represent the first and the second principal component axes. (b) PCA results by the DSI method. “PCA 3” and “PCA 4” represent the third and the fourth principal component axes. (c) PCA results by the ML-DSI method. “PCA 1” and “PCA 2” represent the seventh and 26th principal component axes, which gave the best and the second-best q_α values. (d) PCA results by the ML-DSI method. “PCA 3” and “PCA 4” represent the 10th and second principal component axes, which gave the third and the fourth best q_α values.

When all 180 proteins (protein set a) were used, the database enrichment was drastically improved by the ML-DSI method compared to the results by the FS-DSI and DSI methods. The factor selection worked well to improve the database enrichment; the machine-learning technique also worked well. The important part of the database enrichment curve is the slope around the origin of the axis, because the purpose of the in silico screening is to select a small number of compounds from the large number of compounds in the library. The slopes by the ML-DSI method were much steeper than those by the DSI and FS-DSI methods. In fact, 12.4%, 43.4%, and 67.5% of the active compounds were found within the first 1% of the database by the DSI, FS-DSI, and ML-DSI methods, respectively. The average *q* value by the DSI method was 71.6, which was better than that of the random screening, while the average *q* values by the FS-DSI and ML-DSI methods reached 90.5 and 98.5, respectively.

The database enrichment was mostly independent of the number of proteins used, and the trends in *q* values by the DSI, FS-DSI, and ML-DSI methods were entirely independent of the number of proteins used. When 123 proteins (protein set b) were used, 10.8%, 41.6%, and 65.7% of the active compounds were found within the first 1% of the database by the DSI, FS-DSI, and ML-DSI methods, respectively. The average *q* values by the DSI, FS-DSI, and ML-DSI methods were 69.6, 90.3, and 98.0, respectively. When 93 proteins (protein set c) were used, 10.9%, 43.0%, and 63.3% of the active compounds were found within the first 1% of the database by the DSI, FS-DSI, and ML-DSI methods, respectively. The average *q* values by the DSI, FS-DSI, and ML-DSI methods were 69.8, 89.3, and 97.8, respectively. When 63 proteins (protein set d) were used, 10.3%, 43.0%, and 62.6% of the active compounds were found within the first 1% of the database by the DSI, FS-DSI, and ML-DSI methods, respectively. The average *q* values by the DSI, FS-DSI, and ML-DSI methods were 70.4, 89.3, and 97.8, respectively. These values for protein sets b, c, and d were slightly worse than those for protein set a, but the differences among these results were negligible.

When 24 proteins (protein set e) were used, 14.2%, 26.5%, and 41.5% of the active compounds were found within the first 1% of the database by the DSI, FS-DSI, and ML-DSI methods, respectively. The *q* values by the FS-DSI and ML-DSI methods were drastically reduced by about 15% and 20%, respectively. The average *q* values by the DSI, FS-DSI, and ML-DSI methods were 69.4, 85.2, and 95.8, respectively. When the ML-DSI method was sued, the average *q* value by protein set e was not much reduced, but the database enrichment factor within the first 1% of the database was reduced.

4.2. Designing a CYP Ligand Library by the ML-DSI Method. 4.2.1. Distinguishing CYP Inhibitors from CYP Substrates. In this section, we examine how to construct a CYP-focused compound library using the ML-DSI method. In addition, we are also interested in the performance of our method, that is, whether it can distinguish the inhibitors from the substrates of CYP. This is important because inhibition of CYP is a serious problem, and the CYP substrate is more

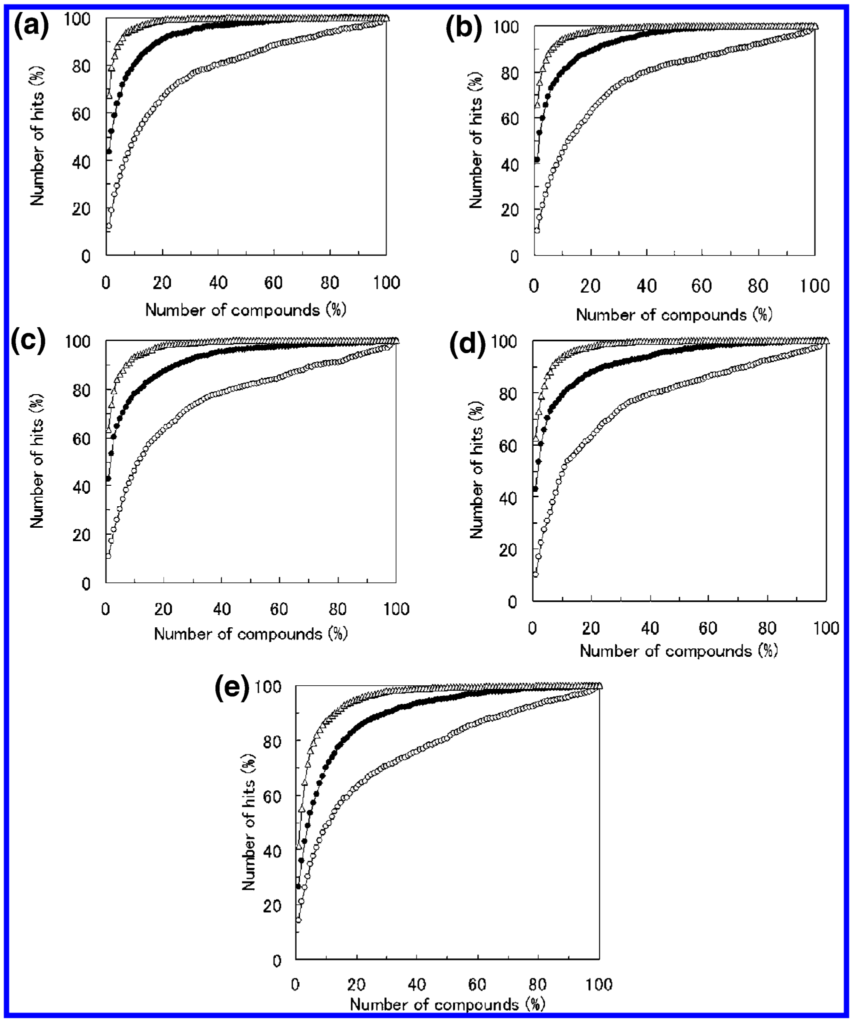

Figure 3. Averaged database enrichment curves of 12 targets using the affinity matrix of (a) 180 proteins (protein set a), (b) 123 proteins (protein set b), (c) 93 proteins (protein set c), (d) 63 proteins (protein set d), and (e) 24 proteins (protein set e). Open circles, filled circles, and open triangles represent the averaged database enrichments by the DSI, FS-DSI, and ML-DSI methods, respectively.

Table 1. Database Enrichments of 12 Target Proteins and Their Averages Using the DSI, FS-DSI, and ML-DSI Methods and the Dependence on the Number of Proteins[a]

number of proteins	180 (protein set a)			123 (protein set b)			93 (protein set c)			63 (protein set d)			24 (protein set e)		
PDB	DSI	FS-DSI	ML-DSI	DSI	FS-DSI	ML-DSI	DSI	FS-DSI	ML-DSI	DSI	FS-DSI	ML-DSI	DSI	FS-DSI	ML-DSI
MIF	63.5	88.4	97.7	64.8	80.3	93.7	60.0	80.3	92.6	58.3	81.0	92.7	53.5	78.9	91.2
Ther	71.6	84.5	93.3	68.6	89.4	92.8	66.8	86.8	93.3	67.1	84.7	92.8	65.5	80.8	93.0
GST	35.5	71.3	91.0	31.2	70.9	92.2	34.6	70.8	97.0	36.2	73.4	82.0	31.3	57.7	90.4
HIV	37.9	78.6	87.5	35.5	79.1	91.5	38.4	78.5	89.2	38.2	76.1	87.9	34.1	67.8	89.7
COX-2	77.6	92.6	97.7	73.5	88.5	96.6	80.9	91.4	97.0	82.3	89.8	97.5	83.0	90.4	95.9
Hant	83.3	97.6	100.0	81.1	98.3	99.9	82.8	98.8	99.9	82.7	94.8	99.9	85.7	92.0	93.2
Aago	88.8	97.9	99.9	87.1	96.9	99.9	85.9	97.1	99.7	87.3	96.8	99.8	90.1	95.8	94.7
Aant	85.1	94.0	98.8	82.2	95.5	98.9	81.5	93.0	98.7	83.6	93.7	98.3	80.0	87.6	91.9
Sago	82.2	98.5	100.0	79.1	99.3	99.9	76.2	98.5	100.0	78.9	98.8	100.0	80.3	94.1	93.0
Sant	87.2	96.7	99.9	85.0	96.6	99.9	83.7	95.7	99.8	83.1	98.5	99.9	87.1	96.5	92.8
Dago	66.7	93.4	99.8	67.3	97.9	99.9	66.5	90.4	99.7	67.4	97.8	99.6	69.3	92.1	95.8
Dant	80.2	92.2	97.6	80.2	90.8	97.9	79.7	89.8	97.6	80.2	86.6	97.6	73.1	88.5	95.2
average	71.6	90.5	98.5	69.6	90.3	98.0	69.8	89.3	97.8	70.4	89.3	97.8	69.4	85.2	95.8

[a] DSI: *q* value by the DSI method. FS-DSI: *q* value by the FS-DSI method. ML-DSI: *q* value by the ML-DSI method. Ther: thermolysin; HIV: HIV protease-1; Hant: histamine H1 receptor antagonists; Aago: adrenalin β receptor agonists; Aant: adrenalin β receptor antagonist; Sago: serotonin receptor agonist; Sant: serotonin receptor antagonist; Dago: dopamine D2 receptor agonist; Dant: dopamine D2 receptor antagonist.

desirable than the CYP inhibitor unless the reaction rate constant k_{cat} is relatively low.

We tried two approaches for selection of the CYP substrates and inhibitors. The first approach (approach A) was to select the CYP inhibitors from the compound library. The percentage of inhibitors among the selected compounds must be low. In contrast, the second approach (approach B) was to select CYP substrates. By removing the CYP inhibitors or substrates from the compound library, we could construct a desirable compound library, in which the percent-

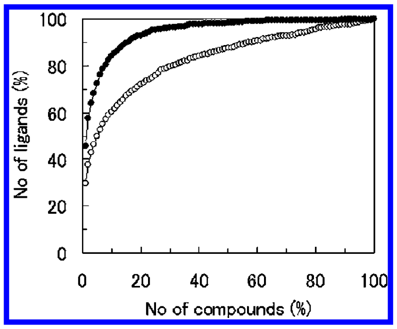

Figure 4. Averaged database enrichment curves of inhibitors and substrates of six CYPs using the affinity matrix of 180 proteins (protein set a) by approach A with the ML-DSI method. Filled circles represent the averaged database enrichments of inhibitors selected by the inhibitor screening. Open circles represent the averaged database enrichments of substrates as contaminants of the selected compound by the inhibitor screening.

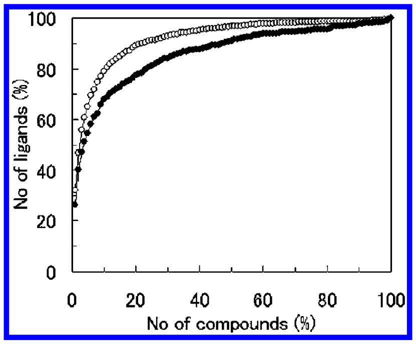

Figure 5. Averaged database enrichment curves of substrates and inhibitors of six CYPs using the affinity matrix of 180 proteins (protein set a) by approach B with the ML-DSI method. Open circles represent the averaged database enrichments of substrates selected by the substrate screening. Filled circles represent the averaged database enrichments of inhibitors as contaminants of the selected compound by the substrate screening.

age of CYP inhibitors or substrates was low. Thus, both the selection of CYP inhibitors and the selection of CYP substrates were found to be useful.

We evaluated the two approaches A and B by calculating the database enrichment. In approach A, the purpose was to find a new CYP inhibitor on the basis of the known CYP inhibitor; thus, the Jack-knife test was applied to the known CYP inhibitors. Specifically, the inhibitors of each CYP were divided into two sets, the known inhibitors for machine learning and the unknown inhibitors, as discovered by the software. The compounds in either set constituted half of all the known inhibitors for each CYP. Ten pairs of these inhibitor sets were prepared for each CYP. Thus, a total of 60 (= 6 CYPs × 10 trials) database enrichment curves were calculated for these six CYPs, and the results were averaged. Also, to check how many substrates were contaminating the compound set of the predicted CYP inhibitors, the degree of database enrichment of substrates was calculated. The same method was applied to evaluate approach B. The results in the above section suggested that protein set a was large enough to achieve high database enrichment; thus, in this study, only protein set a was used.

Figures 4 and 5 show the database enrichment results for approaches A and B, respectively. The purpose of this calculation was to construct a focused library for further screening; thus, we selected only 10−20% of the compounds of the whole library. By approach A, 84.1% of the inhibitors and 60.1% of the substrates were found within the first 10% of the database, and 92.7% of the inhibitors and 71.8% of the substrates were found within the first 20% of the database. The known CYP inhibitors were clearly selected from the compound library, and more inhibitors were found than substrates, although the inhibitors and substrates could not be separated clearly. When a smaller number of compounds was selected, the separation became clear; within the first 1% of the database, 45.7% of the inhibitors and 29.8% of the substrates were found.

By approach B, the known CYP substrates were clearly selected from the compound library, while the separation between the inhibitors and the substrates was worse than that by approach A. Within the first 10% of the database, 78.8% of the substrates and 67.8% of the inhibitors were found. The difference between the two values was 11.0% (= 78.8% of the substrates − 67.8% of the inhibitors) by approach B; the corresponding difference by approach A was 24.0% (= 84.1% of the inhibitors − 60.1% of the substrates). And 89.0% of the substrates and 77.2% of the inhibitors were found within the first 20% of the database. More substrates were found than inhibitors, but the substrates and inhibitors could not be separated clearly, as in approach A. When a smaller number of compounds was selected, the separation became slightly clear; within the first 1% of the database, 32.1% of the substrates and 26.1% of the inhibitors were found. The hit ratio for substrates was 1.2 times (= 32.1%/26.1%) that for inhibitors. Still, this value was lower than that by approach A, in which the hit ratio for inhibitors was 1.5 times (45.7%/29.8%) that for substrates.

4.2.2. Distinguishing Ligands of One CYP from Ligands of Other CYPs. The CYP 3A4 substrate is more desirable than the CYP 2C9, CYP 2D6, and CYP 2C19 substrates, because CYP 2C9, CYP 2D6, and 2C19 have frequent polymorphism.[32] In some cases, we wanted to remove the ligands of particular CYPs from the compound library. When two different drugs were processed by one CYP, the k_{cat} of these drugs became slow because of the drug−drug interaction, and the concentration of these drugs became high. Thus, we wanted to distinguish the ligands (substrates and inhibitors) of one CYP from the ligands of other CYPs.

Again, we tried two approaches. The first approach (approach C) was selection of the inhibitors of a particular CYP from the compound library. Among the selected compounds, the number of ligands of the other CYPs must be small. The second approach (approach D) was the selection of substrates of a particular CYP. If the number of ligands of the other CYPs among the selected compounds was small, we could remove the selected compound from the original compound library to generate a suitable focused library, in which the percentage of inhibitors or substrates of the particular CYP is low.

As with approaches A and B, we evaluated approaches C and D by calculating the database enrichment. In approach C, the purpose was to find a new inhibitor of a particular CYP on the basis of the known CYP inhibitors; thus, the Jack-knife test was applied to the known CYP inhibitors, in the same manner as described above. Ten pairs of these inhibitor sets were prepared for each CYP. Thus, a total of 60 (= 6 CYPs × 10 trials) database enrichment curves were calculated for these six CYPs, and the results were averaged. Also, to examine how many ligands of other CYPs were

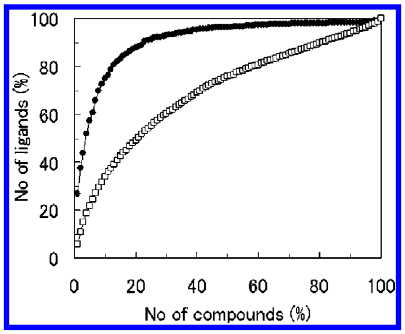

Figure 6. Averaged database enrichment curves of inhibitors of a CYP and ligands of the other five CYPs using the affinity matrix of 180 proteins (protein set a) by approach C with the ML-DSI method. Filled circles represent the averaged database enrichments of inhibitors of a CYP selected by the inhibitor screening for the CYP. Open squares represent the averaged database enrichments of ligands of the other five CYPs as contaminants of the selected compound by the inhibitor screening. The calculations were performed for each of the six CYPs, and these six results were averaged.

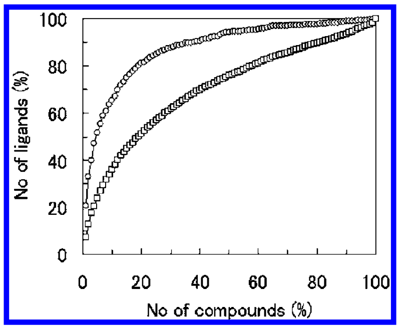

Figure 7. Averaged database enrichment curves of substrates of a CYP and ligands of the other five CYPs using the affinity matrix of 180 proteins (protein set a) by approach D with the ML-DSI method. Open circles represent the averaged database enrichments of substrates of a CYP selected by the substrate screening for the CYP. Open squares represent the averaged database enrichments of ligands of the other five CYPs as contaminants of the selected compound by the substrate screening. The calculations were performed for each of the six CYPs, and these six results were averaged.

contaminating the compound set of the predicted CYP substrates, the database enrichment curves of these ligands were calculated. The same method was applied to evaluate approach D.

Figures 6 and 7 show the database enrichment results by approaches C and D, respectively. The purpose of this calculation was to construct a focused library for further screening; thus, we selected 10−20% of the compounds of the whole library.

By approach C, 74.9% of the inhibitors of each CYP and 33.8% of the ligands of the other CYPs were found within the first 10% of the database, and 88.0% of the inhibitors of each CYP and 49.3% of the ligands of the other CYPs were found within the first 20% of the database. The hit ratio of the target CYP inhibitor was 1.8 (= 88.0%/49.3%) to 2.2 (74.9%/33.8%) times that of the other CYP ligands. The target CYP inhibitors were clearly distinguished from the ligands of the other CYPs. When a smaller number of compounds was selected, the separation became clear; within the first 1% of the database, 26.8% of the inhibitors of the target CYP and 5.7% of the ligands of the other CYPs were found. In this case, the hit ratio of the target CYP substrate was 4.7 (= 26.8%/5.7%) times that of the other CYPs ligands.

By approach D, 65.3% of the substrates of each CYP and 35.7% of the ligands of the other CYPs were found within the first 10% of the database, and 81.0% of the substrates of each CYP and 51.3% of the ligands of the other CYPs were found within the first 20% of the database. The hit ratio of the target CYP substrate was 1.6 (= 81.0%/51.3%) to 1.8 (65.3%/35.7%) times that of the other CYPs ligands. When a smaller number of compounds was selected, the separation became clear; within the first 1% of the database, 20.6% of the substrates of the target CYP and 7.3% of the ligands of the other CYPs were found. In this case, the hit ratio of the target CYP substrate was 2.8 (= 20.6%/7.3%) times that of the other CYPs ligands. The separation between the target CYP substrates and the ligands of the other CYPs by approach D was worse than the separation between the target CYP inhibitors and the ligands of the other CYPs by approach C. However, the target CYP substrates were clearly distinguished from the ligands of the other CYPs.

5. DISCUSSION

Figure 2 shows how the ML-DSI method works. The descriptor of the DSI/FS-DSI/ML-DSI methods is a set of docking scores. If the 3D shape and spatial distribution of charges of the binding site of a protein match those of the compound, the docking score will show a strong affinity. In contrast, a weak affinity is expected when the shapes and spatial distributions of charges of the binding site do not match those of the compound.

Figure 3a−e and Table 1 suggest that the ML-DSI method could be applied to general drug targets when the active compounds of the targets are known. This method provides high database enrichment for both the soluble proteins and the G-protein-coupled receptors. These results for different targets were obtained from a single protein-compound affinity matrix, suggesting that, once an appropriate protein-compound affinity matrix is prepared, it can be applied to drug screening against general targets. Thus, although some CPU resources are required for the initial construction of the protein-compound affinity matrix, the method is not very time-consuming and could offer a rational means of in silico drug screening.

Figure 3a−e show that the database enrichments by the ML-DSI, FS-DSI, and DSI methods were saturated by using more than 24 proteins; thus, in this study, the number of proteins was large enough to show the theoretical upper limit of the database enrichment. Table 1 shows that the *q* values by the ML-DSI method were 87.5−100, those by the FS-DSI method were 71.3−98.5, and those by the DSI method were 35.5−88.8, when protein set a was used. The range of the *q* values by the ML-DSI method was also the smallest among the three methods. Thus, the ML-DSI method was robust and always superior to the other two methods, showing the highest database enrichment for all target proteins.

As shown in our previous works,[31] the first and the second principal components by the PCA correspond to the molecular sizes and the solvation free energies of the compounds, respectively, while the physical meanings of the other principal components have mostly remained unclear. However, the DSI/FS-DSI/ML-DSI methods work well in scaffold hopping.[46] In Figure 1, compounds 3, 4, 6, and 8 are coumarin-like compounds. It would be difficult to find compounds other than coumarin-like compounds by the ordinary fingerprint approach, because the 2D structures of coumarin-like compounds would differ from those of non-coumarin-like compounds. Some other chemometric approach could be used to extract information about the binding features and pharmacophores of the active compounds selected by the DSI/FS-DSI/ML-DSI methods.

The CYP ligands predicted here are somewhat different from those predicted by conventional screening methods, which usually judge whether a given candidate compound can be a ligand of the target CYP. In contrast, the current methods perform a similarity search of chemical compounds, and the results strongly depend on the selection of known active compounds. As we prepare sets of active compounds with greater diversity, a wider range of new active compounds will be discovered. Thus, using our present methods, it may be possible to discover new candidate compounds that would never have been discovered by conventional methods.

The probabilities of finding CYP substrates and CYP inhibitors in a random library are unknown, because CYP metabolism is usually observed for popular drugs and toxins. Many compounds other than CYPs could be processed by the more than 100 enzymes present in blood.[47] Figure 4 shows that 84.1% of inhibitors were found within the first 10% of the compounds of the whole library, and Figure 5 shows that 78.8% of substrates were found within the first 10% of the compounds of the whole library. These enrichment results depend on the concentrations of the inhibitors and substrates. If all the compounds of the library were inhibitors or substrates, 10% of them would be found within 10% of the compounds of the whole library by any screening method. Figures 4 and 5 show that the known CYP inhibitors and substrates were clearly selected from the compound library, although the separation between the inhibitors and the substrates was unclear. However, the above discussion suggests that an inhibitor-free focused library could not be designed by the ML-DSI method.

There are several reasons for CYP inhibition. One is the formation of a covalent bond between the CYP and the ligand due to a chemical reaction (suicidal reaction).[32] Our docking study cannot consider the type of chemical reaction; thus, our method cannot predict this type of inhibition. The other reason for inhibition mechanisms is a strong binding affinity between the CYP and the ligand. In this case, the difference between the definitions of the inhibitor and substrate becomes unclear. Thus, it is difficult to distinguish the inhibitors from the substrates by a docking study.

Figures 4−7 show that the separation of ligands of different CYPs is easier than the separation of inhibitors from substrates of a single CYP. This is because the definition of ligands of different CYPs is clear. For instance, it is known that the size of the CYP 3A ligands is bigger than that of others. The first principal component of the DSI method corresponds to the size of compounds; thus, the ML-DSI method can distinguish the size of compounds, and the CYP 3A ligands can be separated from the other ligands.[31]

In this study, we applied our new method to the prediction of CYP metabolism, but it could be applied to a general-purpose compound search even if the target protein structure is unknown. Another important area of application is drug transportation, which is mediated by a family of proteins called “transporters”[33,48] and plays an important role in pharmacodynamics, that is, the adsorption of a drug at the small intestine, distribution at the blood−brain barrier, and excretion at the kidneys. The origin of drug transportation is the protein−compound interaction, to which our screening method could be applied to discover the putative interactions between transporters and drugs.

6. CONCLUSION

We developed an ML-DSI method that is a similarity search method based on a protein−compound affinity matrix supported by a machine-learning approach. Principal components of the protein−compound docking scores are utilized as descriptors of compounds, and then, a portion of the principal components are selected to maximize the separation between the active compounds and the inactive compounds. Finally, the machine-learning approach modifies the docking score to maximize the database enrichment of the known active compounds. These steps are repeated until the database enrichment result converges.

The ML-DSI method always gave better results than our previous DSI method and showed stable database enrichments for the 12 targets. Almost 70% of the active compounds could be found within the first 1% of the compound of the whole compound library, when protein set a, composed of 180 proteins, was used.

The ML-DSI method was applied to CYP ligands and clearly distinguished the CYP ligands from other compounds. Moreover, it could also distinguish the ligands of a particular CYP from the ligands of other CYPs. Thus, it could be used to design a focused library, i.e., one that contains or excludes the ligands of a particular CYP. However, the inhibitors of the CYPs could not be well-separated from their substrates by the current method, although slight but meaningful differences were found in the database enrichment curves.

The prediction of the CYP ligands was based on the same protein−compound affinity matrix that was used in the above in silico drug screening. Once the protein−compound affinity matrix was prepared, we could perform in silico drug screening for the target protein and simultaneously evaluate the CYP metabolism without additional docking calculations.

The ML-DSI method showed high database enrichment for both soluble proteins and membrane proteins (G-protein-coupled receptors). Also, these screening results and the focused library for CYP ligands could be obtained from one common protein−compound affinity matrix, which could be reused for many other purposes. Thus, the ML-DSI method could provide a useful tool for in silico drug screening.

ACKNOWLEDGMENT

This work was supported by the New Energy and Industrial Technology Development Organization of Japan (NEDO), the Ministry of Economy, Trade, and Industry (METI) of Japan, and the Japan Biological Informatics Consortium (JBiC).

APPENDIX A

The 180 selected proteins (protein set a) were as follows: 12as, 1a28, 1a42, 1a4g, 1a4q, 1abe, 1abf, 1aco, 1ady, 1aer, 1ai5, 1aoe, 1apt, 1apu, 1aqw, 1asz, 1atl, 1aux, 1b58, 1b76, 1b9v, 1bdg, 1bma, 1byb, 1byg, 1c1e, 1c5c, 1c83, 1cbs, 1cbx, 1cdg, 1ckp, 1com, 1coy, 1cps, 1cqe, 1csn, 1cbx, 1cdg, 1ckp, 1com, 1coy, 1cps, 1cqe, 1csn, 1cvu, 1cx2, 1d0l, 1d3h, 1dd7, 1dg5, 1dhf, 1dog, 1dr1, 1ebg, 1eed, 1efv, 1ejn, 1epb, 1epo, 1eqg, 1eqh, 1ets, 1f0r, 1f0s, 1f3d, 1fen, 1fkg, 1fki, 1fl3, 1glg, 1glp, 1gol, 1gtr, 1hck, 1hdc, 1hfc, 1hos, 1hpv, 1hsb, 1hsl, 1htf, 1hyt, 1hfc, 1hos, 1hpv, 1hsb, 1hsl, 1htf1, 1htf2, 1hyt, 1ida, 1ivb, 1jap, 1lah, 1lcp, 1ldm, 1lic, 1lna, 1lst, 1mbi, 1mdr, 1gc7, 1mld, 1mmq, 1mmu, 1mrg, 1mts, 1mup, 1nco, 1ngp, 1nis, 1nks, 1okl, 1pbd, 1pdz, 1phd, 1phg, 1poc, 1ppc, 1pph, 1pso, 1pxx, 1pyg, 1qbr, 1qbu, 1qh7, 1qpq, 1rds, 1rne, 1pxx, 1pyg, 1qbr, 1qbu, 1qh7, 1qpq, 1rds, 1rne, 1rnt, 1rob, 1s2a, 1s2c1, 1s2c2, 1ses, 1snc, 1so0, 1srj, 1tlp, 1tmn, 1tng, 1tnh, 1tni, 1tnl, 1tyl, 1xid, 1xie, 1yee, 2aac, 2aad, 2ack, 2ada, 2cht, 2cmd, 2cpp, 2ctc, 2fox, 2gbp, 2gbp, 2ifb, 2pk4, 2qwk, 2tmd, 2tmn, 3cla, 3cpa, 3erd, 3ert, 3pgh, 3r1r, 3tpi, 4cox, 4est, 4lbd, 4phv, 5abp, 5cpp, 5er1, 6cox, 6rnt, 7tim, 1l3f, 3hvp, 5cox, 1aid, 1hpx, 1ivp, 18gs, 2gss, 3pgt, and 16gs. For 1abe, 1abf, 5abp, and 1htf, two receptor pockets were prepared, because these proteins bind two ligands each.

The 123 selected proteins (protein set b) were as follows: 3pgh, 4cox, 5cox, 6cox, 1pxx, 1cx2, 18gs, 3pgt, 2gss, 1aid, 1hpx, 1ivp, 4phv, 1hpv, 1hos, 1qbr, 1tnl, 1tng, 1tlp, 1lna, 1a4q, 1a4g, 1abf, 1glp, 1srj, 1d0l, 1r55, 1c1e, 2cht, 1fki, 2qwk, 2pk4, 1ets, 1aqw, 1hfc, 1f0s, 1hdc, 4est, 1bma, 2pk4, 1ets, 1aqw, 1hfc, 1f0s, 1hdc, 4est, 1bma, 3cla, 1ppc, 2ifb, 1tnh, 1htf, 1byg, 1ckp, 1aoe, 1pph, 1mts, 1a42, 1cdg, 1lic, 1f0r, 1dhf, 1f3d, 1qpq, 1tni, 1lst, 1dg5, 1nco, 1hsb, 1ejn, 1cvu, 1atl, 1cle, 1rne, 1mmq, 1bqq, 1rob, 1ivb, 1coy, 1byb, 3ert, 1dd7, 1bkc, 1ai5, 2tmn, 2fox, 1coy, 1byb, 3ert, 1dd7, 1snc, 1jap, 1hyt, 1epb, 1cbs, 2gbp, 1c5c, 1ngp, 1poc, 1yee, 1cps, 1pbd, 1mbi, 1com, 1xid, 1okl, 3erd, 1a28, 1xie, 1b58, 1d3h, 1fl3, 1hsl, 4lbd, 1fen, 1mdr, 1c83, 1ldm, 3tpi, 1lcp, 2ada, 1dog, 1gc7, 1pdz, 1lah, 3cpa, 4aah, 2ack, 1ebg, 1mrg, 1cbx, 1nis, 1aco, 2ctc, and 1mup.

The 93 selected proteins (protein set c) were as follows: 3pgh, 4cox, 5cox, 6cox, 1pxx, 1cx2, 18gs, 3pgt, 2gss, 1aid, 1hpx, 1ivp, 1f3d, 1qpq, 1tnl, 1tni, 1glp, 1lst, 1aoe, 1dg5, 1srj, 1nco, 1a42, 1hsb, 1ets, 1ejn, 1r55, 1cvu, 1hfc, 1atl, 2cht, 1cle, 1hos, 1rne, 1mmq, 1bqq, 1ppc, 1rob, 1ivb, 1coy, 1byb, 3ert, 1dhf, 1dd7, 1bkc, 1ai5, 2tmn, 2fox, 1snc, 1jap, 1hyt, 1epb, 1cbs, 2gbp, 1c5c, 1ngp, 1poc, 1yee, 1cps, 1pbd, 1mbi, 1com, 1xid, 1okl, 3erd, 1a28, 1xie, 1b58, 1d3h, 1fl3, 1hsl, 4lbd, 1fen, 1mdr, 1c83, 1ldm, 3tpi, 1lcp, 2ada, 1dog, 1gc7, 1pdz, 1lah, 3cpa, 4aah, 2ack, 1ebg, 1mrg, 1cbx, 1nis, 1aco, 2ctc, and 1mup.

The 63 selected proteins (protein set d) were as follows: 3pgh, 4cox, 5cox, 6cox, 1pxx, 1cx2, 18gs, 3pgt, 2gss, 1aid, 1hpx, 1ivp, 2tmn, 1a28, 1ai5, 1b58, 1bqq, 1c83, 1cbx, 1cdg, 1com, 1coy, 1cvu, 1d3h, 1dog, 1epb, 1fen, 1fki, 1fl3, 1hfc, 1hos, 1jap, 1lcp, 1ldm, 1mbi, 1mdr, 1gc7, 1mld, 1mmq, 1mrg, 1mup, 1ngp, 1okl, 1pbd, 1pdz, 1pso, 1qbu, 1qpq, 1tng, 1xie, 1yee, 2ack, 2ada, 2cmd, 2ctc, 2fox, 2ifb, 2pk4, 3cpa, 3ert, 3tpi, 4aah, and 4lbd.

The 24 selected proteins (protein set e) were as follows: 3pgh, 4cox, 5cox, 6cox, 1pxx, 1cx2, 18gs, 3pgt, 2gss, 1aid, 1hpx, 1ivp, 1gc7, 2tmn, 2ada, 1ngp, 1hfc, 1mup, 1fl3, 2ctc, 4aah, 2cmd, 1pbd, and 1d3h.

APPENDIX B

As COX-2-active compounds, 12 inhibitors and two natural ligands were selected. The two natural ligands were arachidonic acid and prostaglandin H2. The 12 inhibitors were diclofenac, etodolac, suprofen, diflunisal, piroxicam, sulindac, indomethacin, ketoprofen, naproxen, nimesulide, rofecoxib, and 1-phenylsulfonamide-3-trifluoromethyl-5-parabromophenylpyrazole.

The names of the thermolysin inhibitors used in the present study are as follows, with the PDB code in parentheses representing the complex structure from which the compound originated: l-benzylsuccinate (1hyt), phenylalanine phosphinic acid—deamino-methyl-phenylalanine (1os0), (6-methyl-3,4-dihydro-2H-chromen-2-Yl)methylphosphonate (1pe5), 2-(4-methylphenoxy)ethylphosphonate—3-methylbutan-1-amine (1pe7), 2-ethoxyethylphosphonate—3-methylbutan-1-amine (1pe8), (2-sulfanyl-3-phenylpropanoyl)-Phe-Tyr (1qf0), [2(*R*,*S*)-2-sulfanylheptanoyl]-Phe-Ala (1qf1), [(2*S*)-2-sulfanyl-3-phenylpropanoyl]-Gly-(5-phenylproline) (1qf2), *n*-(1-(2(*R*,*S*)-carboxy-4-phenylbutyl)cyclopentylcarbonyl)-(*S*)-tryptophan (1thl), (*R*)-retrothiorphan (1z9g), (*S*)-thiorphan (1zdp), hydroxamic acid (4tln), phenylalanine phosphinic acid (4tmn), Honh-benzylmalonyl-L-alanylglycine-P-nitroanilide (5tln), Cbz-Gly^P-Leu-Leu (Zg^{P}Ll) (5tmn), Cbz-Gly^P-(O)-Leu-Leu (Zg^P(O)Ll) (6tmn), $CH_2CO(N\text{-}OH)Leu\text{-}OCH_3$ (7tln), benzyloxycarbonyl-D-Ala (1kto), benzyloxycarbonyl-L-Ala (1kl6), benzyloxycarbonyl-D-Thr (1kro), benzyloxycarbonyl-L-Thr (1kj0), benzyloxycarbonyl-D-Asp (1ks7), benzyloxycarbonyl-L-Asp (1kkk), benzyloxycarbonyl-D-Glu (1kr6), benzyloxycarbonyl-L-Glu (1kjp), aspartame, aspartic acid, and phenyl alanine.

The names of the GST inhibitors used in the present study are as follows, with the PDB code in parentheses representing the complex structure from which the compound originated: benzylcysteine—phenylglycine (10gs), glutathione—[2,3-dichloro-4-(2-methylene-1-oxobutyl) phenoxyacetic acid (11gs), *S*-nonyl-cysteine (12gs), 1-(*S*-glutathionyl)-2,4-dinitrobenzene (18gs), glutamyl group—*S*-(4-bromobenzyl)cystine (1aqv), glutamyl group—*S*-(2,3,6-trinitrophenyl)cysteine (1aqx), *S*-hexylglutathione (1pgt), cibacron blue (20gs), chlorambucil (21gs), ethacrynic acid (2gss), (9r,10r)-9-(*S*-glutathionyl)-10-hydroxy-9,10 dihydrophenanthrene (2pgt), 2-amino-4-[1-(carboxymethyl-carbamoyl)-2-(9-hydroxy-7,8-dioxo-7,8,9,10-tetrahydro-benzo[def]chrysen-10-ylsulfanyl)-ethylcarbamoyl]-butyric acid (3pgt).

The names or the SMILES of the HIV protease-1 inhibitors used in the present study are as follows, with the PDB code in parentheses representing the complex structure from which the compound originated: C1(c2ccc(F)cc2)(SCCS1)CCCN3CCC(c4ccc(Cl)cc4)(O)CC3 (1aid), c1(OCC2N(S(N(C(C(C2O)O)COc3ccccc3)Cc4ccccc4)(=O)=O)Cc5ccccc5)ccccc1 (1ajv), C1(N(C(C(C(C(N1Cc2ccccc2)COc3ccccc3)O)O)COc4ccccc4)Cc5ccccc5)=O (1ajx), [4r-(4α,5α,6*β*,7*β*)]-3,3′-[[tetrahydro-5 6-dihydroxy-2-oxo-4,7-bis(phenylmethyl)-1h-1,3-diazepine-1,3(2h)-diyl] bis(methylene)]bis[N-2-thiazolylbenzamide (1bv7), C(N(Cc1ncccc1)C)(=O)NC(C(=O)NC(C(C(C(NC(=O)C(C(C)C)NC(N(Cc2ncccc2)C)=O)Cc3ccccc3)(O)O)(F)F)Cc4ccccc4)C(C)C (1dif), C(N1C(C(=O)NC(C)(C)C)CSC1)(=O)C(C(NC(=O)C(NC(=O)COc2[c]3[c](cncc3)ccc2)CSC)Cc4ccccc4)O (1hpx), C(=O)(C(NC(=O)C(CC(C)C)N)CCC(=O)N)NC(C(=O)NC(C(=

O)O)CO)CCC(=O)O (1hte), C(=O)(C1C(SC(C(C(=O)NCc2ccccc2)NC(=O)Cc3ccccc3)N1)(C)C)NC(Cc4ccccc4)CO (1htf), c12c(cccc1)NC(=N2)CNC(=O)CC(C(NC(=O)C3C(SC(C(C(=O)NCc4ccccc4)NC(=O)Cc5ccccc5)N3)(C)C)Cc6ccccc6)O (1htg), 2-phosphoglycolic acid (1hvi), C1(N(C(C(C(C(N1Cc2c[c]3[c](cc2)cccc3)Cc4ccccc4)O)O)Cc5ccccc5)Cc6c[c]7[c](cc6)cccc7)=O (1hvr), 2-carbonylquinoline—phenylalaninol group—decahydro-1-methylisoquinoline-2-carbonyl—tertiary-butylamino group (1hxb), ritonavir (1hxw), naphthyloxyacetyl—cyclohexyl-Ala-*ψ*(Choh-Choh)-Val—2-aminomethyl-pyridine (1ivp), 2-carbonylquinoline—phenylalanylmethane—3-(carboxyamide(2-carboxyamide-2-tertbutylethyl))penta (1jld), C1(N(C(C(C(C(N1Cc2ccc(cc2)CO)Cc3ccccc3)O)O)Cc4ccccc4)Cc5ccc(cc5)CO)=O (1mes), tertiary-butoxyformic acid—phenylalaninol group—dimethylamine—phenylalaninol group—tertiary-butoxyformic acid (1odw), (5r,6r)-2,4-bis-(4-hydroxy-3-methoxybenzyl)-1,5-dibenzyl-3-oxo-6-hydroxy-1,2,4-triazacycloheptane (1pro), C1(C(=C(C=C(O1)C(Cc2ccccc2)CC)O)C(c3cc(ccc3)NC(=O)CCNC(=O)OC(C)(C)C)C4CC4)=O (2upj), and N,N-bis-(2(*R*)-hydroxy-1-(*S*)-indanyl-2,6-(*R*,*R*)-diphenylmethyl-4-hydroxy-1,7-heptandiamide (4hpv).

The following compounds are the antagonists of the histamine H1 receptor: astemizole, cetirizine, chlorpheniramine, clemastine, cyprohrptadine, diphenhydramine, homochlorcyclizine, mequitazine, olopatadine, and promethazine.

The following compounds are the agonists of the adrenaline *β* receptor: clenbuterol, dobutamine, epinephrine, fenoterol, isoprenaline, mabuterol, methylephedrine, norepinephrine, procatelol, salbutamol, terbutaline, and trimetoquinol.

The following compounds are the antagonists of the adrenaline *β* receptor: alprenolol, arotinolol, atenolol, betaxolol, bisoprolol, bopindolol, carteolol, metoprolol, nadlol, pindolol, propranolol, tilisolol, and timolol.

The following compounds are the agonists of the serotonin receptor: 3-(2-aminopropyl)indol-5-ol, 7-(dipropylamino)-5,6,7,8-tetrahydoronaphthol, (4-fluorophenyl)-N-[3-(4-methylpiperidyl)indol-5-yl]carboxamido, 3-(2-aminopropyl)indol-5-ol, (3-chlorophenyl){imino[(iminoethyl)amino]methl}amine, 2-piperidyl 4-amino-5-chloro-2-methoxybenzoate, 1-(4-amino-5-chloro-2-methoxyphenyl)-3-(1-{2-[(methylsulfonyl)amino]ethyl}(4-piperidyl))propan-1-one, and sumatriptan.

The following compounds are the antagonists of the serotonin receptor: azasetron, (1-{2-[(methelsulfonyl)amino]ethyl}-4-piperidyl)methyl 1-methylindoline-3-carboxylate, granisetron, ketanserin, mesulergine, ondansetron, ramosetron, tropisetron, and cyclohexy-N-{2-[4-(2-methoxyphenyl)piperazinyl]ethyl}-N-(2-pyridyl)carboxamide.

The following compounds are the agonists of the dopamine D2 receptor: apomorphine, bromocriptine, denopamine, dobutamine, quinpirole, and 1-phenyl-1H,2H,3H,4H,5H-benzo[*d*]azepine-7,8-diol.

The following compounds are the antagonists of the dopamine D2 receptor: benperidol, chlorpromazine, clozapine, fluphenazine, haloperidol, metiapine, molindone, primozide, prochlorperazine, promazine, spiperone, sulpiride, thioproperazine, thioridazine, and trazodone.

APPENDIX C

The ligands of CYPs were taken from the CYP database at Indiana University (http://medicine.iupui.edu/flockhart/).

The following compounds are the inhibitors of CYP 1A2 (seven compounds): amiodrone, cimetidine, fluoroquinolones, fluvoxamine, furafylline, methoxsalen, and mibefradil.

The following compounds are the substrates of CYP 1A2 (15 compounds): caffeine, clozapine, cyclobenzaprine, olanzapine, rilizole, ropivacaine, tacrine, theophylline, tizanidine, zileuton, zolmitriptan, fluoroquinoline, fluvoxamie, furafylline, and methoxsalen.

The following compounds are the inhibitors of CYP 2C8 (five compounds): gemfibrozil, glitazones, montelukast, quercetin, and trimethoprim.

The following compounds are the substrates of CYP 2C8 (eight compounds): paclitaxel, amodiquine, repaglinide, trimethoprim, quercetin, glitazone, gemfibrozil, and montelukast.

The following compounds are the inhibitors of CYP 2C9 (12 compounds): fenofibrate, fluconazole, fluvastatin, isoniazid, lovastatin, phenylbutazone, probenicid, sertraline, sulfamethoxazole, sulfaphenazole, teniposide, and zafirlukast.

The following compounds are the substrates of CYP 2C9 (21 compounds): diclofenac, ibuprofen, lornoxicam, meloxixam, piroxixam, suprfen, tolbutamide, glipizide, losartan, glyburide, glimepiride, fluvastatin, rosiglitazone, S_warfarin, fenofibrate, isoniazid, phenylbutazone, sulfamethoxazole, sulfaphenazole, trimethoprim, and zafirlukast.

The following compounds are the inhibitors of CYP 2C19 (11 compounds): R_fluoxetine, chloramphenicol, felbamate, indomethacin, ketoconazole, lansoprazole, modafinil, omeprazole, oxcarbazepine, and topiramate.

The following compounds are the substrates of CYP 2C19 (16 compounds): lansoprazole, omeprazole, pantoprazole, rabeprazole (E3810), S_mephenytoin, phenobarbital, cyclophosphamide, hexobarbital, R_mephobartital, nilutamide, primidone, proguanil, felbamate, modafinil, oxcarbazepin, and topiramate.

The following compounds are the inhibitors of CYP 2D6 (27 compounds): buproprion, celecoxib, chlorpheniramine, citalopram, clemastine, clomipramine, cocaine, chlorpromazine, diphenhydramine, doxepine, doxorubicin, duloxetine, escitalopram, halofantrine, hydroxyzine, levomeptomazine, methadone, metoclopramide, moclobemide, paroxetine, perphenazine, quinidine, ranitidine, ritonavir, terbinafine, tripelennamine, and metoclopramide.

The following compounds are the substrates of CYP 2D6 (32 compounds): carvedilol, paroxetin, perphenazine, alprenolol, amphetamine, atomoxetine, bufuralol, chlorpromazine, debrisoquine, dexfenfluoramine, duloxetine, flecainide, metoclopramide, minaorine, nebinvolol, nortriptyline, perhexiline, phenformin, sparteine, P_tramadol, venlafaxine, doxepine, doxorubicin, duloxetine, escitalopram, halofantrine, levomeptomaxine, ranitidine, terbinafine, diphenhydramine, clemastine, and hydroxyzine.

The following compounds are the inhibitors of CYP 3A4, 3A5, and 3A7 (14 compounds): aprepitant, ciprofloxacin, clarithromycin, delaviridine, diltiazem, erythromycin, gestodene, indinavir, itraconazole, mifepristone, nefazodone, nelfinavir, norfloxacin, and verapamil.

The following compounds are the substrates of CYP 3A4, 3A5, and 3A7 (51 compounds): clarithromycin, erythromycin, telithromycin, alprazolam, midazolam, triazolam, cyclosporine, tacrolimus, indinavir, saquinavir, M_cisapride, astemizole, terfenadine, amlodipine, felodipine, lercanidipine, nifedipine, nisoldipine, verapamil, atorvastatin, simvastatin, hydrocortisone, testosterone, buspirone, cafergot, dapsone, docetaxel, domperidone, eplerenone, fentanyl, finasteride, gleevec, irinotecan, levacetylmethadol (LAAM), ondansetron, pimozide, quinine, sildenafil, trazodone, vincristine, zaleplon, zolpidem, apreoutant, ciprofloxacin, gestodene, itraconazol, mifepristone, nefazodone, norfloxacin, mibefradil, and delavirid.

REFERENCES AND NOTES

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule−Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(3) Jones, G.; Willet, 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.

(4) Paul, N.; Rognan, D. ConsDock: A New Program for the Consensus Analysis of Protein−Ligand Interactions. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 521−533.

(5) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 367−382.

(6) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian Docking Functions. *Biopolymers* **2003**, *68*, 76−90.

(7) Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Struct., Funct., Genet.* **1990**, *8*, 195−202.

(8) Taylor, J. S.; Burnett, R. M. DARWIN: A Program for Docking Flexible Molecules. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 173−191.

(9) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM: A New Method for Structure Modeling and Design: Application to Docking and Structure Prediction from the Disordered Native Conformation. *J. Comput. Chem.* **1994**, *15*, 488−506.

(10) Colman, P. M. Structure-Based Drug Design. *Curr. Opin. Struct. Biol.* **1994**, *4*, 868−874.

(11) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: A Novel Scoring Function for Predicting Binding Affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395−407.

(12) Fukunishi, Y.; Mikami, Y.; Nakamura, H. Similarities among Receptor Pockets and among Compounds: Analysis and Application to in Silico Ligand Screening. *J. Mol. Graphics Modell.* **2005**, *24*, 34−45.

(13) Orita, M.; Yamamoto, S.; Katayama, N.; Aoki, M.; Takayama, K.; Yamagiwa, Y.; Seki, N.; Suzuki, H.; Kurihara, H.; Sakashita, H.; Takeuchi, M.; Fujita, S.; Yamada, T.; Tanaka, A. Coumarin and Chomen-4-one Analogues as Tautomerase Inhibitors of Macrophage Migration Inhibitory Factor: Discovery and X-ray Crystallography. *J. Med. Chem.* **2001**, *44*, 540−547.

(14) Cotesta, S.; Giordanetto, F.; Trosset, J. Y.; Crivori, P.; Kroemer, R. T.; Stouten, P. F. W.; Vulpetti, A. Virtual Screening to Enrich a Compound Collection with CDK2 Inhibitors Using Docking, Scoring, and Composite Scoring Models. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 629−643.

(15) Schellhammer, I.; Rarey, M. FlexX-Scan: Fast, Structure-Based Virtual Screening. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 504−517.

(16) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J. Med. Chem.* **2005**, *48*, 5448−5465.

(17) Howard, M. H.; Cenizal, T.; Gutteridge, S.; Hanna, W. S.; Tao, Y.; Totrov, M.; Wittenbach, V. A.; Zheng, Y.-J. A Novel Class of Inhibitors of Peptide Deformylase Discovered through High-Throughput Screening and Virtual Ligand Screening. *J. Med. Chem.* **2004**, *47*, 6669−6672.

(18) Godden, J. W.; Stahura, F. L.; Bajorath, J. POT-DMC: A Virtual Screening Method for the Identification of Potent Hits. *J. Med. Chem.* **2004**, *47*, 5608−5611.

(19) Zhao, L.; Brinton, R. D. Structure-Based Virtual Screening for Plant-Based ER$^{\beta}$-Selective Ligands as Potential Preventative Therapy against Age-Related Neurodegenerative Diseases. *J. Med. Chem.* **2005**, *48*, 3463−3466.

(20) Mestres, J.; Veeneman, G. H. Identification of "Latent Hits" in Compound Screening Collections. *J. Med. Chem.* **2003**, *46*, 3441−3444.

(21) Vigers, G. P. A.; Rizzi, J. P. Multiple Active Site Corrections for Docking and Virtual Screening. *J. Med. Chem.* **2004**, *47*, 80−89.

(22) Fukunishi, Y.; Mikami, Y.; Kubota, S.; Nakamura, H. Multiple Target Screening Method for Robust and Accurate in Silico Ligand Screening. *J. Mol. Graphics Modell.* **2005**, *25*, 61−70.

(23) Pickett, S. In *Protein−Ligand Interactions from Molecular Recognition to Drug Design − Methods and Principles in Medicinal Chemistry*; Boehm, H. J., Schneider, G., Mannhold, R., Kubinyi, H., Folkers, G., Eds.; Wiley−VCH: Weinheim, Germany, 2003; pp 88−91.

(24) Pearlman, R. S.; Smith, K. M. Metric Validation and the Receptor-Relevant Subspace Concept. *J. Chem. Inf. Compt. Sci.* **1999**, *39*, 28−35.

(25) Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. Predicting Ligand Binding to Proteins by Affinity Fingerprinting. *Chem. Biol.* **1995**, *2*, 107−118.

(26) Briem, H.; Kuntz, I. D. Molecular Similarity Based on DOCK-Generated Fingerprints. *J. Med. Chem.* **1996**, *39*, 3401−3408.

(27) Lessel, U. F.; Briem, H. Flexsim-X: A Method for the Detection of Molecules with Similar Biological Activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 246−253.

(28) Briem, H.; Lessel, U. F. In Vitro and in Silico Affinity Fingerprints: Finding Similarities beyond Structural Classes. *Perspect. Drug Discovery Des.* **2000**, *20*, 231−244.

(29) Weber, A.; Teckentrup, A.; Briem, H. Flexsim-R: A Virtual Affinity Fingerprint Descriptor to Calculate Similarities of Functional Groups. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 903−916.

(30) Hsu, N.; Cai, D.; Damodaran, K.; Gomez, R. F.; Keck, J. G.; Laborde, E.; Lum, R. T.; Macke, T. J.; Martin, G.; Schow, S. R.; Simon, R. J.; Villar, H. O.; Wick, M. M.; Beroza, P. Novel Cyclooxygenase-1 Inhibitors Discovered Using Affinity Fingerprints. *J. Med. Chem.* **2004**, *47*, 4875−4880.

(31) Fukunishi, Y.; Mikami, Y.; Takedomi, K.; Yamanouchi, M.; Shima, H.; Nakamura, H. Classification of Chemical Compounds by Protein-Compound Docking for Use in Designing a Focused Library. *J. Med. Chem.* **2006**, *49*, 523−533.

(32) Kamataki, T. In *P450 no Bunshiseibutugaku*; Oomura, T., Ishimura, T., Fujii, Y., Eds.; Koudansya: Tokyo, 2005; pp 148−149.

(33) Iwasawa, Y. In *Medicinal Chemistry*; Yamakawa, K., Kanaoka, Y., Iwasawa, Y., Eds.; Koudansya: Tokyo, 2004; pp 73−85.

(34) Butina, D.; Segall, M. D.; Frankcombe, K. Predicting ADME Properties in Silico: Methods and Models. *Drug Discovery Today* **2002**, *7*, S83−S88.

(35) Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A.; Wikel, J. H. Progress in Predicting Human ADME Parameters in Silico. *J. Pharmacol. Toxicol. Methods* **2000**, 251−272.

(36) Cattel, R. B. The Scree Test for the Number of Factors. *Multivar. Behav. Res.* **1966**, *1*, 245−276.

(37) Abdi, H. In *Encyclopedia for Research Methods for the Social Science*; Lewis-Beck, M., Futing, T., Eds.; Sage: Thousand Oaks, CA, 2003; pp 978−982.

(38) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A New Test Set for Validating Predictions of Protein−Ligand Interaction. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 457−471.

(39) Watanabe, T.; Fukui, Y. In *Saiboumaku no Jyuyoutai*; Takayanagi, I., Ed.; Nanzandou: Tokyo, 1998; pp 121−131.

(40) Koike, K.; Nagatomo, T. In *Saiboumaku no Jyuyoutai*; Takayanagi, I., Ed.; Nanzandou: Tokyo, 1998; pp 103−118.

(41) Sasa, M.; Ishihara, K. In *Saiboumaku no Jyuyoutai*; Takayanagi, I., Ed.; Nanzandou: Tokyo, 1998; pp 135−147.

(42) Nakata, Y.; Inoue, A. In *Saiboumaku no Jyuyoutai*; Takayanagi, I., Ed.; Nanzandou: Tokyo, 1998; pp 169−182.

(43) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(44) Gasteiger, J.; Marsili, M. A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **1978**, 3181−3184.

(45) Case, D. A.; Darden, T. A.; Cheatham, T. R., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(46) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.

(47) Beresford, A. P.; Selick, H. E.; Tarbit, M. H. The Emerging Importance of Predictive ADME Simulation in Drug Discovery. *Drug Discovery Today* **2002**, *7*, 109−116.

(48) Sugiyama, M. In *Mechanisms of Drug Interactions*; Sugiyama, M., Kamiya, H., Eds.; Iyakusyuppan: Tokyo, 2004; pp 46−88.

CI600334U