

Insights into Molecular Basis of Cytochrome P450 Inhibitory Promiscuity of Compounds

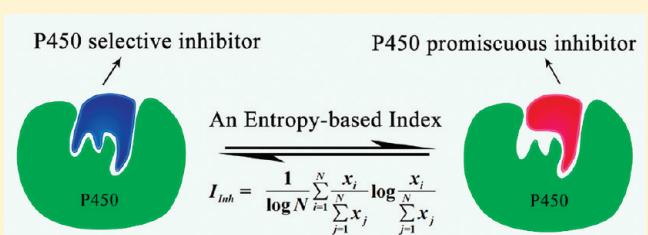
Feixiong Cheng,[†] Yue Yu,[†] Yadi Zhou,[†] Zhonghua Shen,[†] Wen Xiao,[†] Guixia Liu,[†] Weihua Li,^{*,†} Philip W. Lee,^{‡,§} and Yun Tang^{*,†}

[†]Department of Pharmaceutical Sciences, School of Pharmacy, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China

[‡]Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

 Supporting Information

ABSTRACT: Cytochrome P450 inhibitory promiscuity of a drug has potential effects on the occurrence of clinical drug–drug interactions. Understanding how a molecular property is related to the P450 inhibitory promiscuity could help to avoid such adverse effects. In this study, an entropy-based index was defined to quantify the P450 inhibitory promiscuity of a compound based on a comprehensive data set, containing more than 11,500 drug-like compounds with inhibition against five major P450 isoforms, 1A2, 2C9, 2C19, 2D6, and 3A4. The results indicated that the P450 inhibitory promiscuity of a compound would have a moderate correlation with molecular aromaticity, a minor correlation with molecular lipophilicity, and no relations with molecular complexity, hydrogen bonding ability, and TopoPSA. We also applied an index to quantify the susceptibilities of different P450 isoforms to inhibition based on the same data set. The results showed that there was a surprising level of P450 inhibitory promiscuity even for substrate specific P450, susceptibility to inhibition follows the rank-order: 1A2 > 2C19 > 3A4 > 2C9 > 2D6. There was essentially no correlation between P450 inhibitory potency and specificity and minor negative trade-offs between P450 inhibitory promiscuity and catalytic promiscuity. In addition, classification models were built to predict the P450 inhibitory promiscuity of new chemicals using support vector machine algorithm with different fingerprints. The area under the receiver operating characteristic curve of the best model was about 0.9, evaluated by 5-fold cross-validation. These findings would be helpful for understanding the mechanism of P450 inhibitory promiscuity and improving the P450 inhibitory selectivity of new chemicals in drug discovery.



INTRODUCTION

Over the past decades, developing drugs with high selectivity on a therapeutic target but less side effects or adverse drug reactions (ADR) is the main goal of drug discovery. However, drug promiscuity, such as polypharmacology that a drug acts on multiple proteins rather than a single target, is challenging the ‘one gene, one drug, one disease’ paradigm.¹ For example, a promiscuous drug NADH targets 95 proteins, and a promiscuous target adrenergic α_{1A} receptor could bind with 52 drugs in DrugBank database.² Compounds with promiscuous activities, i.e. “off-target” activities, usually link with ADR liabilities. In recent years, several drugs were withdrawn from the market due to inhibitory promiscuity on human cytochrome P450 (CYP), such as Seldane, Posicor, Hismanal, Propulsid, Lotronex, Baycol, and Seraone.^{3,4} Therefore, understanding the molecular basis of compound promiscuity could help to design novel drugs with less ADR profiles in the early drug discovery process.

Among ADR profiles, adverse drug–drug interactions (DDI) caused by CYP inhibition is an important issue in drug discovery. The CYP enzymes, particularly isoforms 1A2, 2C9, 2C19, 2D6, and 3A4, are responsible for about 90% oxidative metabolic reactions.⁵ As we know, the more CYP isoforms a given small

molecule inhibits, the more likely it will be involved in DDI with many other drugs. Thus, the CYP inhibitory promiscuity (i.e., a single compound inhibit multiple CYP isoforms) of a compound has potential effects on the occurrence of clinical DDI.

Degree of promiscuity refers to the level of specificity breach, namely, how diverse is the promiscuous activity of a given enzyme,^{6,7} and how different are the native and promiscuous functions. Recently, Nath et al. proposed an entropy-based index to quantitatively measure and compare the substrate catalytic promiscuity of three types of enzymes: eight serine and cysteine proteases, two glutathione S-transferase isoforms, and three CYP isoforms.⁸ In another study, Nath et al. quantified and modeled the inhibitory promiscuity and isoform specificity based on two high-throughput data sets (64×64 and 194×194 ligands matrix, respectively) of small molecular P450 inhibitors. They concluded that there was no essential correlation between a drug’s inhibitory potency and specificity.⁹ Recently, Foti et al. applied the same entropy-based index to investigate the catalytic versus inhibitory promiscuity in CYPs and found that inhibitory

Received: July 9, 2011

Published: August 29, 2011

promiscuity was not an obligate negative trade-off for catalytic promiscuity.¹⁰ However, a general investigation and modeling of CYP450 inhibitory promiscuity based on the comprehensive data set have not been done yet, although some progresses were made.^{8–10} Therefore, quantitatively steering chemical P450 inhibitory promiscuity and systemically investigating and understanding the relationship between P450 inhibitory promiscuity and molecular properties based on the comprehensive ligands matrix would be very helpful for understanding the mechanism of P450 inhibitory promiscuity and improving the P450 inhibitory selectivity of new chemicals in drug discovery.

Systematical high-throughput screening (HTS) of promiscuous enzymatic activities are not feasible at present, because no single method is available to detect the whole range of different substrates and reactions.⁷ In contrast, HTS of testing cross-reactivities are relatively straightforward.⁷ By using an *in vitro* bioluminescent assay of quantitative HTS, Auld's group recently determined the AC₅₀ values (the compound concentration leads to 50% of the activity of an inhibition control) of more than 17,000 compounds against five recombinant CYP isoforms (1A2, 2C9, 2C19, 2D6, and 3A4).¹¹ This comprehensive cross-screening data set provided us a possibility for large-scale quantifying the P450 inhibitory promiscuity, investigating, and modeling the relationship between molecular promiscuity and chemical properties.

In this study, an entropy-based index was defined to quantitatively measure P450 inhibitory promiscuity based on a large data set containing more than 11,500 drug-like compounds with inhibition against five major P450 isoforms, including CYP1A2, 2C9, 2C19, 2D6, and 3A4. The relationships between ten key physicochemical properties and chemical P450 inhibitory promiscuity were investigated. Then classification models were developed to predict the P450 inhibitory promiscuity of new chemicals using support vector machine (SVM) algorithm and different fingerprints. The results provided some novel perspectives of chemical P450 inhibitory promiscuity and could be used as useful tools for filtering DDI profiles in the earlier stage of drug discovery process.

MATERIALS AND METHODS

An Index of Inhibitory Promiscuity. In this study, the entropy-based index proposed by Nath et al.^{8,9} was used to quantify P450 inhibitory promiscuity of compounds. Considering a set of N enzymes, a compound displays an inhibitory potency, denoted as x_i , against isoform i . The ratio $(x_i)/(\sum_{i=1}^N x_i)$ can be thought as approximation of the probability that the compound will inhibit enzyme i out of all the enzymes in the set. Based on Shannon-Jaynes entropy¹²

$$H = - \sum_{i=1}^N p_i \cdot \log p_i \quad (1)$$

we can then define P450 inhibitory promiscuity as follows

$$I_{Inh} = - \frac{1}{\log N} \sum_{i=1}^N \frac{x_i}{\sum_{j=1}^N x_j} \log \frac{x_i}{\sum_{j=1}^N x_j} \quad (2)$$

As indicated in eq 2, if a compound inhibits all P450 isoforms with equal potency (i.e., It is completely promiscuous), $I_{Inh} = 1$. If a compound inhibits only one CYP isoform but none of the others

(i.e., It is completely specificity), $I_{Inh} = 0$. The detailed description of I_{Inh} definition can be found in Nath's work.^{8,9}

A Susceptibility Index for P450 Isoform Promiscuity. The I_{Inh} index can be inverted to describe the promiscuity with which a particular P450 isoform is inhibited by a panel of compounds or, in other words, the degree to which an isoform is susceptible to be inhibited by a wide range of compounds. We will call this quantitative susceptibility index^{8,9} of a given isoform and denote it as I_{sus}

$$I_{sus} = - \frac{1}{\log M} \sum_{i=1}^M \frac{x_i}{\sum_{j=1}^M x_j} \log \frac{x_i}{\sum_{j=1}^M x_j} \quad (3)$$

In eq 3, M is the number of compounds in the inhibitor panel, x_i is the inhibitory potency of compound i , and x_j is the inhibitory potency of compound j . Then I_{sus} was extended to account for chemical similarities among the compounds in the inhibitor panel by weighting the index based on Nath's reports.^{8,9} As shown in eq 4, a weighted enzyme susceptibility index, J_{Inh} , was defined as the following

$$J_{Inh} = - \frac{1}{\log M} \sum_{i=1}^M \langle d_i \rangle \frac{x_i}{\sum_{j=1}^M x_j} \log \frac{x_i}{\sum_{j=1}^M x_j} \quad (4)$$

Herein, $\langle d_i \rangle = \bar{d}_{ij}/d_{set}$ is the normalized mean chemical dissimilarity between compound i and all other members of the inhibitor panel, defined as follows. The Tanimoto distance, the complement of Tanimoto similarity coefficient, was applied.¹³ For a pair of compounds A and B, where a is the number of features present only in A, b is the number of features present only in B, and c is the number of features present in both A and B, the Tanimoto distance is $d_{ab} = (a + b)/(a + b + c)$. For inhibitors in a set, we can also define d_{ij} as the mean Tanimoto distance from a member i to all the other members in the set. The overall set dissimilarity d_{set} serves as an upper bound for d_{ij} : if k is the number of features presented in at least one but not all of the members of the set, and l is the number of features presented in all members of the set, then $d_{set} = k/(k + l)$. At last, we can obtain J_{Inh} from eq 4. The detailed descriptions about J_{Inh} definition can be found in original literature.^{8,9}

Data Collection and Preparation. The initial database (PubChem AID: 1851) in SMILES format with known P450 inhibitory data was provided by Dr. Auld.^{11,14} It contains 17,143 diverse substances which were measured by a standard protocol under the same experimental conditions. Entries containing inorganic compounds, noncovalent complexes, and mixtures were excluded. Salts were converted to the corresponding acids or bases; water molecules were removed from hydrates. If a compound shows no inhibitory effects for the five CYP isoforms at the same time, this compound is also excluded. At last, 11,578 compounds with AC₅₀ values were obtained.

When we defined the index of I_{Inh} or I_{sus} , we must define the inhibitory potency (x_i) at first. In this study, the inhibitory potency was defined as $x_i = -\log(AC_{50}/57 \mu M)$, with 57 μM as a reference value¹¹ to ensure that $x_i > 0$. In Auld's data set, some compounds did not display sufficient potency for accurate AC₅₀ determination on some isoforms at the highest test concentration (57 μM).¹¹ For such cases in this study, we used an approximate potency with AC₅₀ value of 57 μM for these compounds. The inhibitory potency (x_i) and inhibitory promiscuity index (I_{Inh})

were calculated based on in-house scripts (available upon request) written based on Nath's work.^{8,9} In order to model P450 inhibitory promiscuity, the classification models were built to predict high promiscuous P450 inhibitors versus low promiscuous P450 inhibitors. For example, if the inhibitory promiscuity index I_{Inh} of a compound was higher than 0.5, it was labeled as high promiscuous inhibitor. Otherwise, it was labeled as low promiscuous inhibitor. All the x_i, I_{Inh} values of 11,578 compounds against CYP1A2, 2C9, 2C19, 2D6, and 3A4 can be found in Table S1 of the Supporting Information.

Calculation of Molecular Descriptors. In this study, ten key molecular descriptors that were widely adopted in structure–activity relationship (SAR) studies of P450 inhibitor and substrate^{15,16} were evaluated. They are Ghose-Crippen LogK_{ow} (AlogP), XlogP, the number of aromatic atoms and bonds (naAromAtom and nAromBond, respectively), Topological polar surface area (TopoPSA), the number of hydrogen bond donors and acceptors (nHBDon and nHBAcc, respectively), Molecular Weight (MW), the number of heavy atoms (nHeavyAtom), and the number of rings (nRing). All descriptors were calculated using the open source software package of PaDEL-Descriptor.¹⁷

In addition, six kinds of fingerprints implemented in PaDEL-Descriptor¹⁷ were also evaluated, including CDK fingerprint (FP), CDK extended fingerprint (ExtFP), Estate fingerprint (EstateFP), MACCS keys (MACCS), PubChem fingerprint (PubChemFP), and Klekota-Roth fingerprint (KRFP). The detailed descriptions about these fingerprints can be found in original literature.^{17,18}

An advantage of fingerprints is that they can be easily translated into two-dimensional fragments. The representative substructure fragments were explored by information gain¹⁹ combined with substructure fragment analysis.²⁰ The frequency of a fragment in a high P450 inhibitory promiscuous class was defined as the following

$$\text{Frequency of a fragment} = \frac{(N_{\text{fragment_class}} \times N_{\text{total}})}{(N_{\text{fragment_total}} \times N_{\text{class}})} \quad (5)$$

where $N_{\text{fragment_class}}$ is the number of compounds containing the fragment in a high promiscuous chemical class, N_{total} is the total number of compounds, $N_{\text{fragment_total}}$ is the total number of compounds containing the fragment, and N_{class} is the number of compounds in the high P450 inhibitory promiscuous class.

Construction of Classification Models. SVM algorithm was used to build the classification models for predicting high or low P450 promiscuous inhibitors. SVM algorithm, originally developed by Vapnik²¹ for pattern recognition, aims at minimizing the structural risk under the frame of Vapnik-Chervonenkis (VC) theory. It has been successfully applied in modeling P450 inhibition and ADMET properties prediction by our group^{19,22,23} and some other groups.^{24–27} In this study, each molecule was expressed using an eigenvector \mathbf{t} , and the selected patterns t_1, t_2, \dots, t_n made up the components of \mathbf{t} . For SVM training, the category label y was added as following: $y_i = 1$ for the high P450 inhibitory promiscuous compound and $y_i = -1$ for the low P450 inhibitory promiscuous compound. So the i^{th} molecule in the data set is defined as $\mathbf{M}_i = (t_i, y_i)$. SVM gives a decision function (classifier)

$$f(\mathbf{t}) = \text{sgn}\left(\frac{1}{2} \sum_{i=1}^n \alpha_i K(t_i, \mathbf{t}) + b\right) \quad (6)$$

where α_i is the coefficient to be learned, and K is a kernel function. Parameter α_i is trained through maximizing the Lagrangian

expression given below

$$\begin{aligned} & \underset{\alpha_i}{\text{maximize}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{t}_i, \mathbf{t}_j) \\ & \text{subjectto : } \sum_i y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C \end{aligned} \quad (7)$$

A superiority of SVM is that it can deal with high dimensional space with the input of vectors from low dimensional space by introducing kernel function. In this study, the Gaussian radial basis function (RBF) kernel was used. To obtain a SVM model with optimal performance, the penalty parameter C and different kernel parameter γ were tuned based on the training set using grid search strategy based on 5-fold cross-validation. Herein, SVM algorithm was performed by the LIBSVM2.9 package.²⁸

Validation of Models. The k -fold cross-validation technique was used to evaluate all models. In a 5-fold cross-validation, the entire data set was equally divided into five cross-validation splits. Within each step of cross-validation, the model was trained on a set of four cross-validation splits together. The fifth subsample set was used as an internal validation set (test set).

All developed models were evaluated based on the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP represents the number of high P450 inhibitory promiscuous chemicals predicted as high P450 inhibitory promiscuous chemicals, TN is the number of low P450 inhibitory promiscuous chemicals predicted as low P450 inhibitory promiscuous chemicals, FP stands for the number of low P450 inhibitory promiscuous chemicals predicted as high P450 inhibitory promiscuous chemicals, and FN represents the number of high P450 inhibitory promiscuous chemicals predicted as low P450 inhibitory promiscuous chemicals. Furthermore, the sensitivity ($SE = TP / (TP + FN)$), which is the predictive accuracy of high P450 inhibitory promiscuous chemicals, and the specificity ($SP = TN / (TN + FP)$), which is the predictive accuracy of low P450 inhibitory promiscuous chemicals, were calculated. The overall predictive accuracy (Q), the Matthews correlation coefficient (C), and G-mean were calculated using eqs 8, 9, and 10, respectively

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$C = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad (9)$$

$$\text{G-mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \quad (10)$$

In addition, the receiver operating characteristic (ROC) curve was also plotted used in house python scripts (available upon request). The ROC curve was used to graphically present the model behavior in a visual way. It shows the separation ability of a binary classifier by iteratively setting the possible classifier threshold.²⁹

■ RESULTS AND DISCUSSION

Analysis of Relationships between Molecular Descriptors and P450 Inhibitory Promiscuity. In this study, an entropy-based index^{8,9} was used to quantitatively describe P450 inhibitory promiscuity of 11,578 compounds. The chemical space distribution

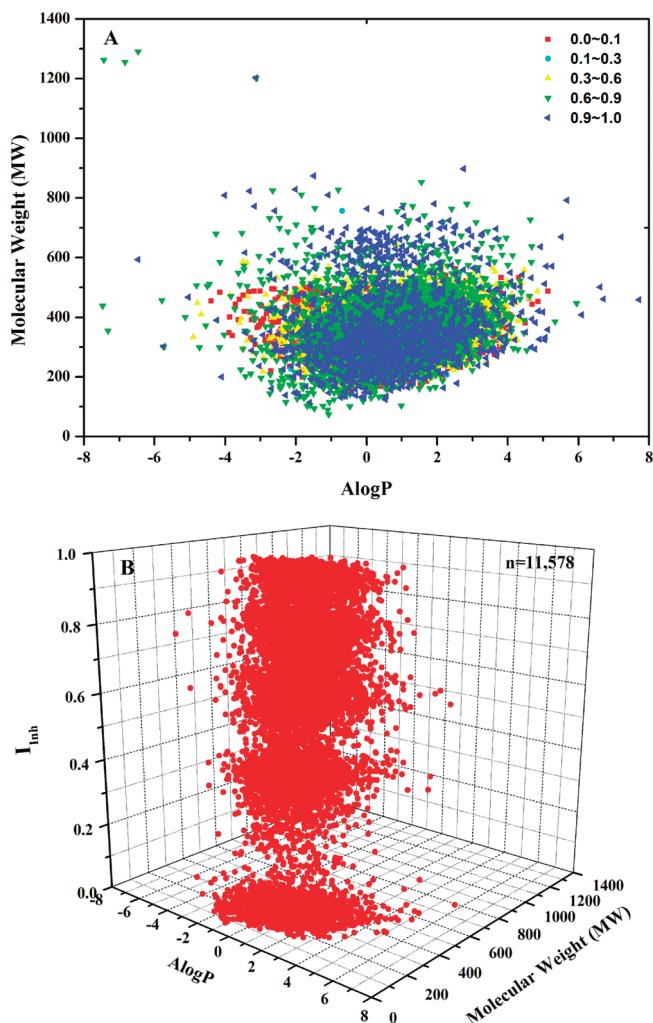


Figure 1. Diversity analysis of the entire data set ($n = 11,578$ compounds) with P450 inhibitory promiscuity index. (A) Chemical space defined by molecular weight (MW) and AlogP. The data are colored according to the chemical P450 inhibitory promiscuity index value. (B) Distribution of P450 inhibitory promiscuity index (I_{inh}) in a chemical space defined by MW and AlogP. One compound of AlogP = -12.65 was not displayed. Both figures are used the same color scheme.

of the entire data set ($n = 11,578$ compounds) was plotted in Figure 1, which is defined by MW and AlogP. As shown in Figure 1, there is a diverse chemical space and P450 inhibitory promiscuity index (I_{inh}) distribution for the 11,578 compounds.

The SAR of P450 inhibition and some common used molecular descriptors has been studied by several groups.^{15,16} Herein, we systematically investigated the relationships between chemical P450 inhibitory promiscuity and ten commonly used molecular descriptors, namely AlogP, XlogP, MW, TopoPSA, nHeavyAtom, nRing, nHBAcc, nHBDOn, naAromAtom, and nAromBond. The distributions of ten descriptors for high inhibitory promiscuity ($I_{inh} > 0.5$) class and low inhibitory promiscuity ($I_{inh} \leq 0.5$) class were shown in Figure 2. Student's *t* test was employed to evaluate the significance of the difference between paired-two samples and the means.

Among the ten molecular properties, AlogP and XlogP represent the lipophilicity of a compound. As shown in Figure 2, AlogP is distributed between -12.65 and 7.72, with a mean of

0.69. XlogP is distributed between -14.24 and 12.23, with a mean of 2.28. The distributions of AlogP and XlogP of high and low P450 inhibitory promiscuous classes were presented in Figure 2. The mean values of AlogP were 0.91 and 0.34 for 7016 high P450 inhibitory promiscuous compounds and 4562 low P450 inhibitory promiscuous compounds, respectively. The mean values of XlogP are 1.84 and 2.57 for high and low inhibitory promiscuous compounds, respectively. The data indicated that P450 inhibitory promiscuity was associated with chemical lipophilicity. As shown in Figure 2, the *p*-value linked with the difference in the mean AlogP and XlogP value of high inhibitory promiscuous compounds and those of low inhibitory promiscuous compounds were $2.37e^{-107}$ and $2.58e^{-137}$, respectively, at the 95% confidence level. It indicated that the two distributions were significantly different. As a complementary test, the linear correlation analysis of AlogP, XlogP, and P450 inhibitory promiscuity index (I_{inh}) were given in Figure 3. Both AlogP and XlogP indicated a minor linear correlation with P450 inhibitory promiscuity. Recently, Leeson and Springthorpe demonstrated a particularly compelling correlation between lipophilicity and drug promiscuity.³⁰ Our finding is consistent with Leeson and Springthorpe's work.

MW was regarded as a simple estimation of molecular size and complexity,³¹ which is generally thought to have an impact on polypharmacology.^{32,33} As shown in Figure 2, MW was distributed from 74.06 to 1308.50, with a mean of 359.92. The mean values of MW were 368.16 and 347.24 for high and low P450 inhibitory promiscuous compounds, respectively, which indicated that MW was not the primary determinant of P450 inhibitory promiscuity of chemicals. This conclusion was further supported by the low linear correlation ($R = 0.11$) between MW and I_{inh} in Figure 3. And the *p*-value linked with the difference in the mean MW of high promiscuous inhibitors and those of low promiscuous inhibitors was $4.37e^{-32}$ at the 95% confidence level. Recently, Yang et al. demonstrated that drug promiscuity correlates with the structural complexity of a compound based on the size of the molecular framework for large molecular framework.³⁴ In order further to investigate the relationship between molecular complexity and chemical P450 inhibitory promiscuity, we also investigated the relationship of nHeavyAtom and nRing versus chemical P450 inhibitory promiscuity. The mean values of nHeavyAtom were 26 and 25 for high and low P450 inhibitory promiscuous compounds, respectively. The mean value of nRing was about 3 for both high and low P450 inhibitory promiscuous compounds. And the *p*-value linked with the difference in the mean nHeavyAtom and nRing values of high and low P450 inhibitory promiscuous compounds was $1.88e^{-25}$ and $1.11e^{-46}$, respectively. Moreover, the linear correlation (R) was only 0.10 and -0.14 for I_{inh} with nHeavyAtom and nRing, respectively, in Figure 3. These data further indicated that there is no correlation between molecular complexity and chemical P450 inhibitory promiscuity.

Hydrogen bonding ability was commonly represented by nHBAcc and nHBDOn. As shown in Figure 2, the mean values of nHBAcc were 4 and 5 for high and low P450 inhibitory promiscuous compounds, respectively. The mean values of nHBDOn were about 1 for both high and low P450 inhibitory promiscuous compounds. The *p*-value linked with the difference in the mean nHBAcc and nHBDOn values was $4.84e^{-33}$ and $1.88e^{-25}$, respectively. And the low linear correlation of -0.14 and -0.11 were also observed between I_{inh} and nHBAcc and nHBDOn, respectively. The data indicated that there was no obvious relationship

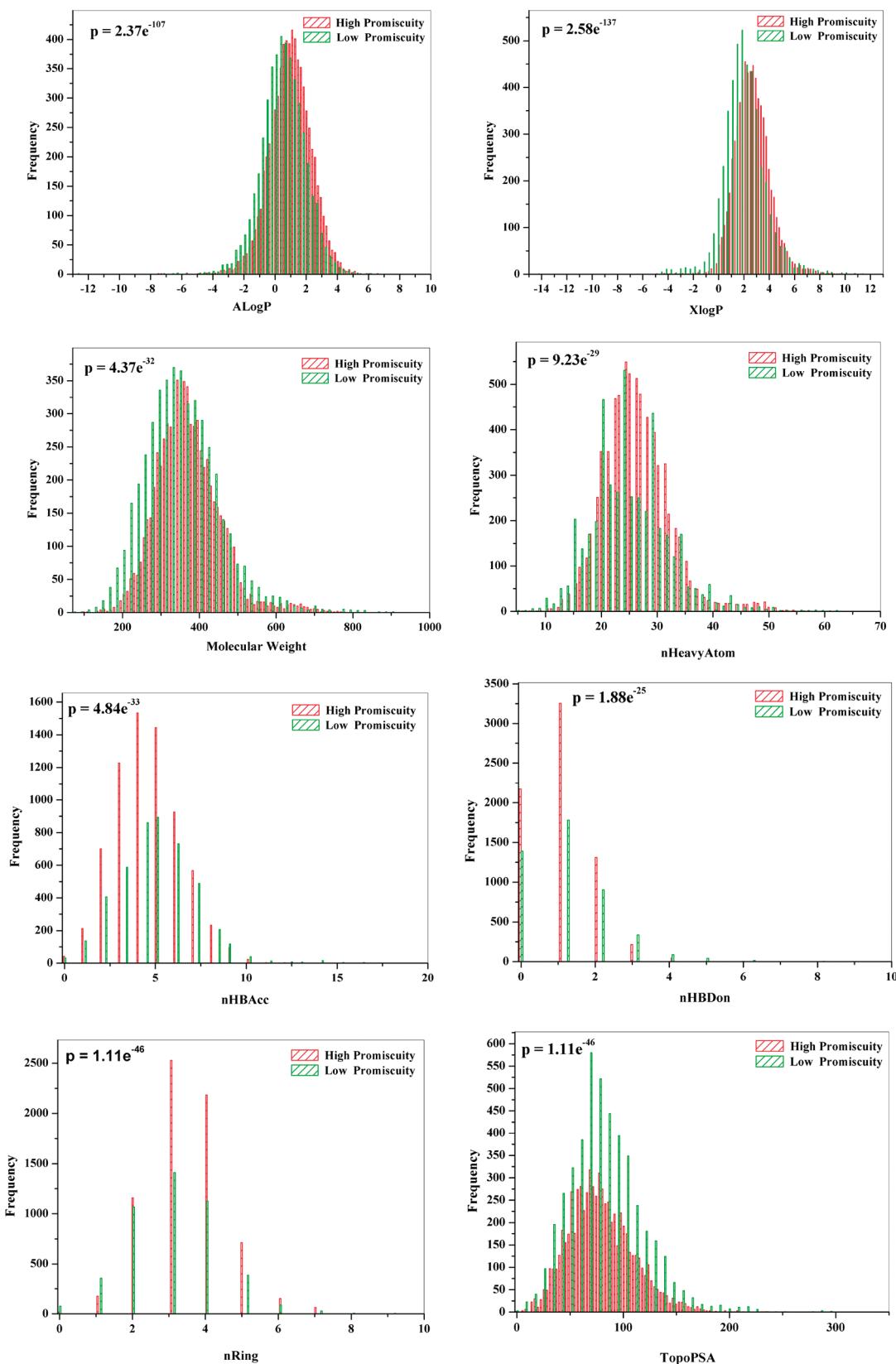


Figure 2. Continued

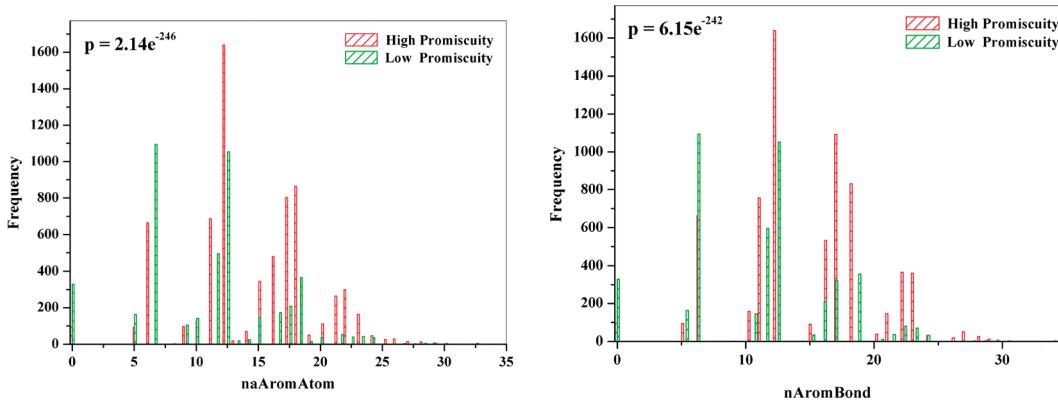


Figure 2. Distributions of ten molecular properties, including Ghose-Crippen LogKow (AlogP), XlogP, the number of Aromatic atoms and bonds (nAromAtom and nAromBond, respectively), TopoPSA, the number of hydrogen bond donors and acceptors (nHBDon and nHBAcc, respectively), Molecular Weight (MW), the number of heavy atoms (nHeavyAtom), and the number of rings (nRing) for the high P450 inhibitory promiscuous compounds ($n = 7016$) and low P450 inhibitory promiscuous compounds ($n = 4562$) classes, respectively.

between hydrogen bonding ability and chemical P450 inhibitory promiscuity. TopoPSA represents the electrostatic properties of a molecule. The mean values of TopoPSA were 80.09 and 78.08 for high and low P450 inhibitory promiscuous compounds, respectively. And there was no correlation ($R = -0.04$) between J_{inh} and TopoPSA (Figure 3). It indicated that TopoPSA did not show any capability to discriminate high P450 inhibitory promiscuous compounds and those of low P450 inhibitory promiscuous compounds.

Lewis et al. found that the specific substrates of P450s favorably included planar, polyaromatic structural features.³⁵ Herein, we also investigated the relationship between chemical P450 inhibitory promiscuity and the molecular aromaticity represented by nAromAtom and nAromBond. The mean values of nAromAtom were 14 and 11 for high and low P450 inhibitory promiscuous compounds, respectively. The mean values of nAromBond were 15 and 11 for high and low P450 inhibitory promiscuous compounds, respectively. The data demonstrated a particularly compelling correlation between molecular aromaticity and P450 inhibitory promiscuity. If a molecule has the higher aromaticity, this molecule has the higher P450 inhibitory promiscuity. Comparing the p -value of nAromAtom and nAromBond with those of others molecular properties, the extremely low p -value of $2.14e^{-246}$ and $6.15e^{-242}$ were observed for the difference in the mean nAromAtom and nAromBond value of high and low P450 inhibitory promiscuous compounds, respectively. And the reasonable moderate linear correlation of 0.34 was obtained between molecular P450 inhibitory promiscuity and nAromAtom or nAromBond.

P450 Susceptibility for Inhibition. Quantitatively steering the chemical P450 isoforms inhibitory promiscuity based on the comprehensive diverse P450 inhibitory compounds is important for exploring the DDI issue. Table 1 listed the susceptibility index (J_{inh}) weighted by 11,578 diverse compounds for five major CYP isoforms. All five CYP isoforms have J_{inh} within a high value and narrow range from 0.88 to 0.95 based on MACCS keys weighted, indicating that drug-metabolizing P450s are generally inhibited by a broad range of compounds. This result was in agreement with Nath's report.⁹ To validate the approach and the basis set of different subsets of compounds, we used the statistical "jack-knife" method. For each data set, each 100 compounds were randomly removed from the basis set and the J_{inh} value was

recalculated. If the basis set is adequately unbiased, then removal of any 100 compounds has no effect of the recovered J_{inh} value, if the enzyme is promiscuous. As shown in Figure 4, the plot validated that the basis set of compounds does not bias the calculation of J_{inh} weighted by MACCS keys. In addition, we also evaluated the influence of the J_{inh} value weighted by six different fingerprints. As given in Table 1, different fingerprints have minor influence for J_{inh} value, the maximum standard deviation (SD) was only 0.01 for CYP1A2. The data showed that the J_{inh} value was a robust index for quantitatively exploring the susceptibility of different CYP isoforms, because the J_{inh} did not influence by different fingerprints and the different subsets of compounds.

As given in Table 1, the average J_{inh} value was 0.9474 for 1A2, 0.9163 for 2C9, 0.9385 for 2C19, 0.8773 for 2D6, and 0.9272 for 3A4, respectively. Comparing the average J_{inh} value for different CYP isoforms, susceptibility to inhibition follows the rank-order: 2D6 < 2C9 < 3A4 < 2C19 < 1A2. CYP1A2 has the highest J_{inh} value, indicating that it is the most susceptible isoform to inhibition by diverse compounds. A possible explanation is that there is a high degree of inhibitory activity for CYP1A2 based on the comprehensive data set of 11,578 diverse compounds against CYP1A2, 2C9, 2C19, 2D6, and 3A4.¹¹ CYP2D6 has the lowest J_{inh} value, showing that it is more specific isoform which is slightly more resistant to inhibition by diverse compounds. Recently, Nath et al. found that CYP3A4 and 1A2 had the lowest J_{inh} value based on only 64 CYP inhibitors. Comparing our results with Nath's work, we confirmed that our results should be more reasonable than Nath's work. Because our results were calculated based on the more comprehensive data set of 11,578 compounds, and many experiments found that CYP3A4 is a promiscuous isoform for diverse substrate.^{10,36} Recently, Karaman et al. also found that small assay panels do not provide a robust measure of kinases selectivity.³⁷ It indicated that a comprehensive ligand matrix is very important for robustly exploring different P450 isoforms inhibitory susceptibility.

Classification Model of P450 Inhibitory Promiscuity. Are there any molecular descriptors that can discriminate P450 inhibitory specificity or promiscuity of small molecules? To answer this question, the classification models for predicting high versus low P450 inhibitory promiscuity of new chemicals were built using SVM algorithm. The performance of classification models were given in Table 2.

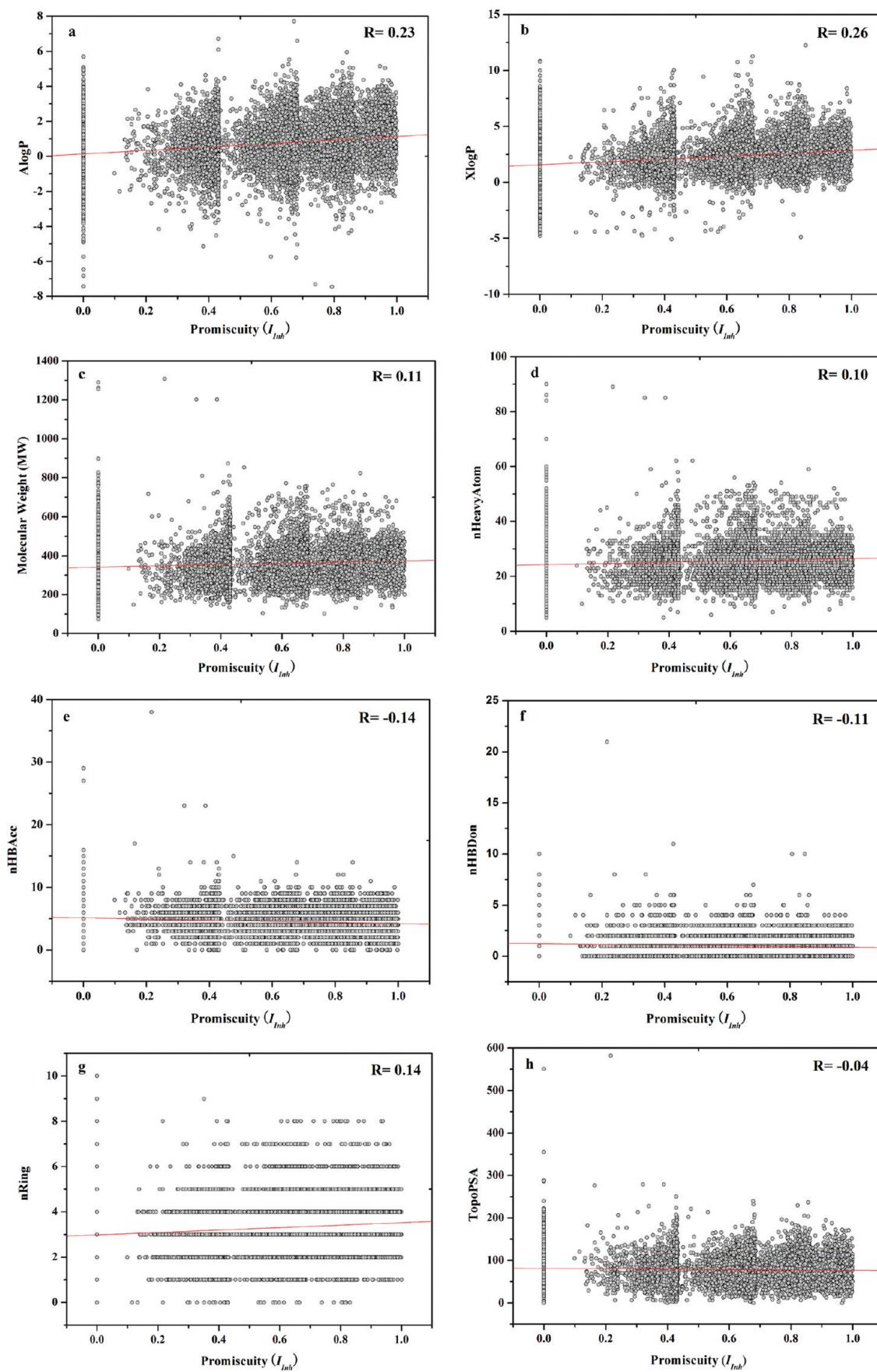


Figure 3. Continued

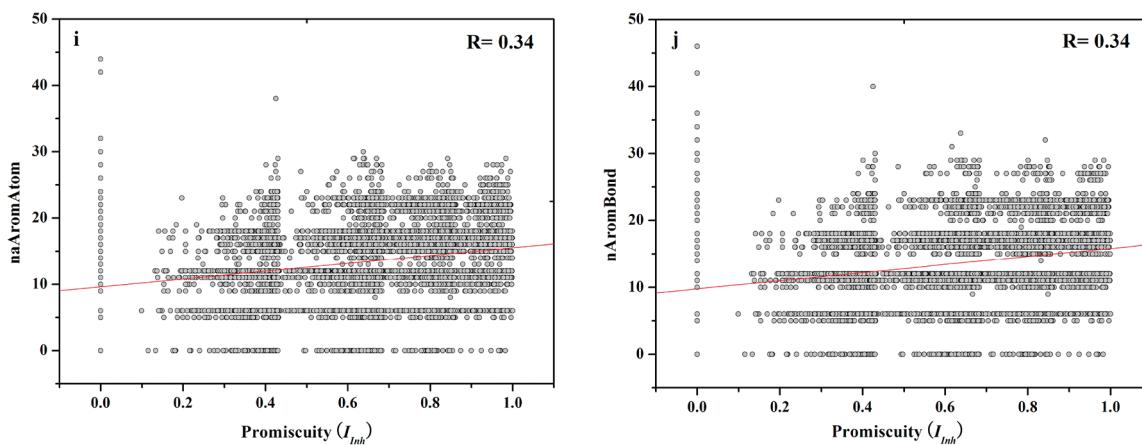


Figure 3. Correlation between ten molecular properties of Ghose-Crippen LogKow (AlogP), XlogP, the number of Aromatic atoms and bonds (naAromAtom and nAromBond, respectively), TopoPSA, the number of hydrogen bond donors and acceptors (nHBDon and nHBAcc, respectively), Molecular Weight (MW), the number of heavy atoms (nHeavyAtom), and the number of rings (nRing), and P450 inhibitory promiscuous index (J_{Inh}).

Table 1. Inhibitory Susceptibility Index (J_{Inh}) for Five P450 Isoforms Calculated from the Inhibitory Potencies Displayed by 11,578 Compounds^a

fingerprints	J_{Inh}				
	CYP1A2	CYP2C9	CYP2C19	CYP2D6	CYP3A4
MACCS	0.9502	0.9167	0.9417	0.8777	0.9247
PubChem	0.9295	0.9113	0.9300	0.8672	0.9181
EstateFP	0.9655	0.9196	0.9454	0.8855	0.9303
ExFP	0.9431	0.9122	0.9353	0.8790	0.9243
FP	0.9421	0.9127	0.9350	0.8785	0.9261
ERFP	0.9539	0.9250	0.9437	0.8756	0.9399
Average \pm SD	0.9474 \pm 0.0122	0.9163 \pm 0.0053	0.9385 \pm 0.0060	0.8773 \pm 0.0059	0.9272 \pm 0.0073

^a MACCS: MACCS keys, FP: CDK fingerprint, ExFP: CDK extended fingerprint, PubChem: PubChem fingerprints, EstateFP: Estate fingerprint, KRFP: Klekota-Roth fingerprint, SD: Standard Deviation.

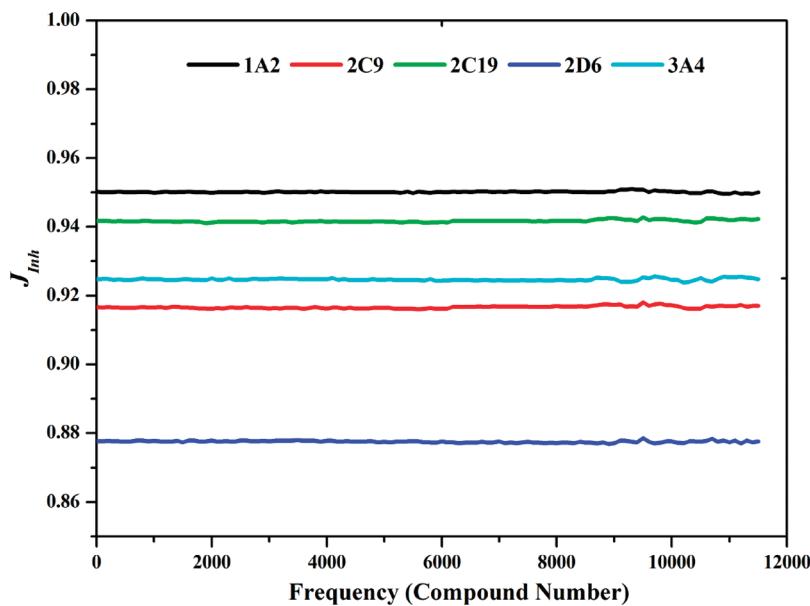


Figure 4. Jackknife analysis of J_{Inh} value for five CYP isoforms listed in Table 1. Each 100 compounds were randomly removed from the basis set and the J_{Inh} recalculated using the remaining compounds weighted by MACCS keys. The analysis validates that the basis set of inhibitors does not bias the calculation of J_{Inh} and further differentiates promiscuous enzymes from specific enzymes.

Table 2. Performance of High versus Low P450 Inhibitory Promiscuity Classification Models Using Different Threshold Value, Support Vector Machine, and Data Description Methods^a

threshold value	data description	performance					
		SE (%)	SP (%)	Q (%)	C	G-mean	AUC
$I_{Inh} = 0.5$	MD	30.6	76.2	48.6	0.084	0.483	0.576
	MACCS	84.8	56.3	73.6	0.433	0.691	0.784
	ExFP	85.9	55.6	73.9	0.440	0.691	0.790
	ERFP	85.3	57.1	74.2	0.446	0.698	0.798
	EStateFP	85.4	41.0	67.9	0.299	0.592	0.703
	PubChem	85.1	56.0	73.6	0.434	0.690	0.789
	FP	85.3	54.7	73.2	0.425	0.683	0.783
$I_{Inh} = 0.2 \text{ VS } 0.8$	MD	58.0	61.4	59.4	0.191	0.597	0.632
	MACCS	88.5	72.5	82.1	0.623	0.801	0.879
	ExFP	89.3	69.2	81.2	0.605	0.786	0.878
	ERFP	88.8	73.9	82.8	0.639	0.810	0.894
	EStateFP	84.6	57.7	73.8	0.443	0.699	0.790
	PubChem	88.5	72.4	82.1	0.623	0.801	0.880
	FP	88.9	68.6	80.7	0.594	0.781	0.878

^a MD: Molecular Descriptors, including AlogP, XlogP, Molecular Weight, the number of heavy atom, TopoPSA, the number of rings, the number of hydrogen bond acceptors and donors, and the number of aromatic atoms and bonds. MACCS: MACCS keys, FP: CDK fingerprint, ExFP: CDK extended fingerprint, PubChem: PubChem fingerprints, EStateFP: Estate fingerprint, KRFP: Klekota-Roth fingerprints.

As shown in Table 2, we evaluated the influence of different classification threshold (I_{Inh}) for classification models. For $I_{Inh} = 0.5$, if the I_{Inh} value of a chemical was higher than 0.5, this chemical was labeled as high P450 inhibitory promiscuous class (designated as +1). If the I_{Inh} value of a chemical was less than or equal 0.5, this chemical was labeled as low P450 inhibitory promiscuous class (designated as -1). At last, the total data set were divided into 7016 high P450 inhibitory promiscuous compounds and 4562 low P450 inhibitory promiscuous compounds. As shown in Table 2, the reasonable high performance was obtained. Comparing different data description methods, the performance of ten molecular descriptors (including AlogP, XlogP, MW, TopoPSA, nHeavyAtom, nRing, nHBAcc, nHBDOn, naRromAtom, and nAromBond) was extremely low. The reasonable high performance using MACCS, ExFP, ERFP, PubChem, and FP was obtained, and the performance of ERFP outperformed other fingerprints. ERFP had been proved an excellent substructure fingerprints enriching molecular biological activity in original literature.¹⁸

For the threshold value of $I_{Inh} = 0.2$ versus 0.8, if the I_{Inh} value of a chemical was higher than or equal 0.8, this chemical was labeled as high P450 inhibitory promiscuity class (designated as +1). If the I_{Inh} value of a chemical was less than or equal 0.2, this chemical was labeled as low P450 inhibitory promiscuity class (designated as -1). Compounds with intermediate P450 inhibitory promiscuity value were classified as inconclusive compounds and excluded in here to avoid uncertainty during models development. At last, the total data set were divided into 3269 high P450 inhibitory promiscuous compounds and 2192 low P450 inhibitory promiscuous compounds. As shown in Table 2, the higher performance was obtained for the threshold

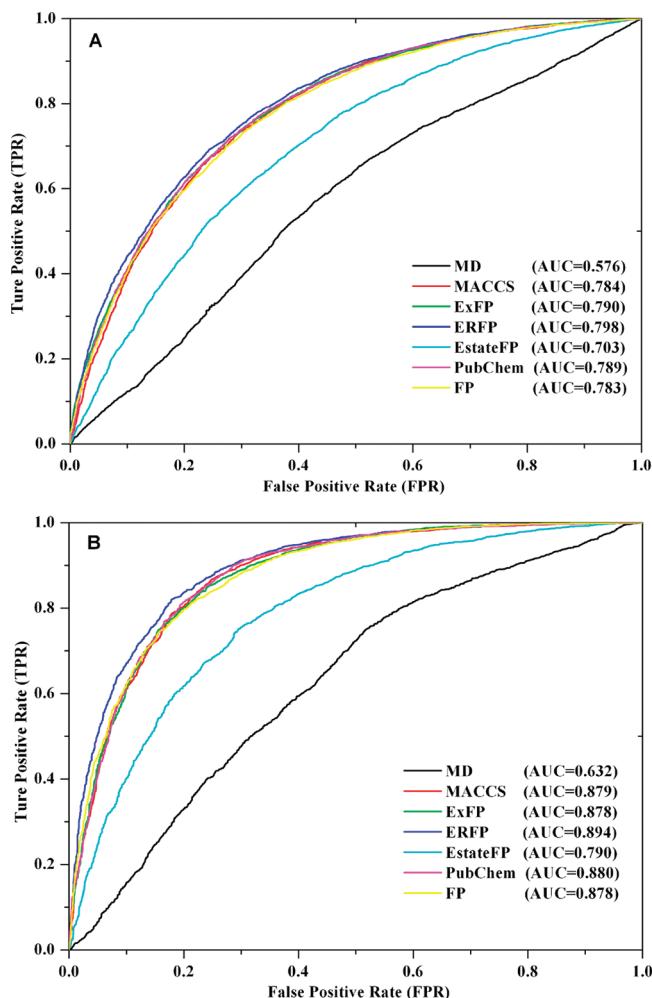


Figure 5. Representation of receiver operating characteristics (ROC) curve of test set with 5-fold cross-validation using support vector machine algorithm with different data description methods and different threshold value: $I_{Inh} = 0.5$ (A), $I_{Inh} = 0.2$ versus 0.8 (B). AUC: the area under ROC curve, MD: ten molecular descriptors (including AlogP, XlogP, MW, TopoPSA, nHeavyAtom, nRing, nHBAcc, nHBDOn, naRromAtom, and nAromBond), MACCS: MACCS keys, FP: CDK fingerprint, ExFP: CDK extended fingerprint, PubChem: PubChem fingerprints, EStateFP: Estate fingerprint, KRFP: Klekota-Roth fingerprint.

value of $I_{Inh} = 0.2$ versus 0.8. For the classification problem, the measurement of the area under the receiver operating characteristic (ROC) curve (AUC) was highly recommended.³⁸ The ROC curve with AUC values of different data description methods and threshold for test set of 5-fold cross-validation was given in Figure 5. Comparing different data description methods, the performance of molecular descriptors was also extremely low, and the ERFP substructure fingerprints (blue line) outperformed other methods with the highest AUC value of 0.894.

Interpretation of QSAR/QSPR models is the most important issue. In this study, some privileged substructure fragments characterizing high and low P450 inhibitory promiscuity with the classification threshold value of 0.5 were identified by combining information gain and ERFP substructure fragment analysis. As shown in Scheme 1, the fragments of No.1 to No.12 were frequently associated with high P450 inhibitory promiscuity. It indicated that the existence of these fragments is unfavorable for P450 inhibitory specificity. The fragments of No.13 to No.15 were

Scheme 1. Occurrence and Frequency of 15 Privileged Substructure Fragments for High or Low P450 Inhibitory Promiscuous Class Determined Using Substructure Fragment Analysis Method Based on Klekota-Roth Fingerprint^a

NO	SMARTS	General Structure	N _{HP}	N _{LP}	F _{HP}	F _{LP}	IG
1	SCc1ccccc1		203	60	1.27	0.57	0.002
2	SCC=O		520	171	1.24	0.63	0.004
3	Clc1ccc(NC(=O)cc1		70	20	1.28	0.56	0.001
4	Oc1ccccc1NC=O		103	33	1.25	0.62	0.001
5	OC(=O)C1CCCCN1		156	29	1.39	0.40	0.003
6	O=CNC=S		137	46	1.24	0.64	0.001
7	O=C1NC=CC(N1)c2ccccc2		13	1	1.53	0.18	0.001
8	O=C(NN=Cc1ccccc1)c2ccccc2		37	10	1.30	0.54	0.001
9	Nc1ncs1		163	33	1.37	0.43	0.003
10	NC(=S)NC=O		119	40	1.23	0.64	0.001
11	c1ccc(cc1)c2nccen2		262	51	1.38	0.41	0.005
12	c1nc(cs1)c2ccccc2		57	10	1.40	0.38	0.001
13	NC(=O)C1CCCCCC1		41	123	0.41	1.90	0.005
14	CC(O)CO		128	361	0.43	1.87	0.016
15	OCCOC=O		41	128	0.40	1.92	0.006

^a N_{HP} is the number of compounds in high P450 inhibitory promiscuous class with specified fragment *t*. N_{LP} is the number of compounds in low P450 inhibitory promiscuous class with specified fragment *t*. F_{HP} is the frequency of a specified fragment (*t*) in high P450 inhibitory promiscuous class. F_{LP} is the frequency of a specified fragment (*t*) in low P450 inhibitory promiscuous class.

frequently associated with P450 inhibitory specificity. The data indicated that if a chemical has these fragments of No.13 to No.15, this chemical was favorable for P450 inhibitory specificity. The detailed statistical results were given in Table S2 of the Supporting Information. P450 inhibitory specificity is an important functional parameter, and several companies are making efforts to develop isoform specific P450 inhibitors that would be administered as adjuvants to enhance the pharmacokinetic profiles of drugs that are highly susceptible to P450 metabolism.⁹ Extracting visual privileged substructure fragments characterizing high P450 inhibitory promiscuity versus high P450 inhibitory specificity could potentially provide scaffold modification for exploring P450 inhibitory specificity to medicine chemists.

P450 Inhibition versus P450 Inhibitory Promiscuity of Small Molecules. Herein, we used the entropy-based index (*I_{Inh}*) to quantitatively define P450 inhibitory promiscuous

matrix. There are two important characteristics for *I_{Inh}*: (i) *I_{Inh}* is dependent on the panel of enzymes, such as an AC₅₀ or IC₅₀, and (ii) *I_{Inh}* is completely independent of an inhibitory absolute level of potency. The second characteristics of *I_{Inh}* can be used to test a correlation between inhibitory potency and inhibitory promiscuity. As shown in Figure 6, there is no correlation (*R* = 0.08) between chemical inhibitory promiscuity and the principal component 1 (PC1) determined by principal component analysis for inhibitory potencies of CYP1A2, 2C9, 2C19, 2D6, and 3A4. And there is a slightly correlation (*R* = 0.28) between maximum inhibitory potency against any CYP isoform of CYP1A2, 2C9, 2C19, 2D6, and 3A4 and inhibitory promiscuity for 11,578 compounds (Figure 6B). The data confirmed that the inhibitory potency and inhibitory promiscuity should be considered independent quantities in drug-metabolizing P450s. This result was consistent with Nath's work.⁹

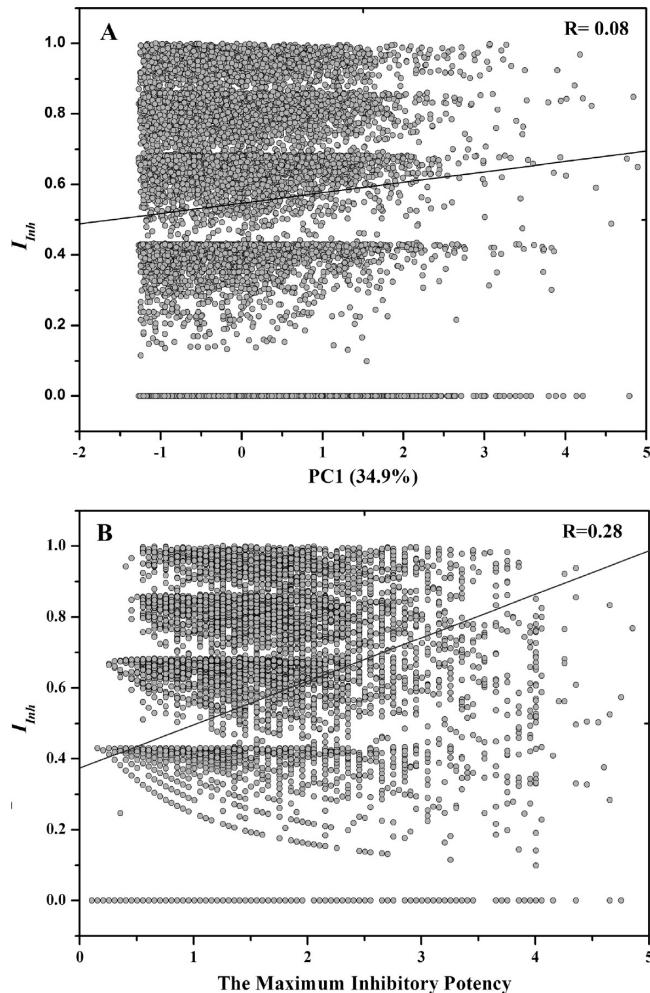


Figure 6. Correlation between inhibitory promiscuity and the principal component (PC1) determined by principal component analysis for inhibitory potency against five CYP isoforms (A) and correlation between the chemical inhibitory promiscuity and the maximum inhibitory potency against any CYP isoform of 1A2, 2C9, 2C19, 2D6, and 3A4 (B).

In the past decade, several indexes such as selectivity score,³⁷ Gini score,³⁹ and partition index⁴⁰ had been developed for quantifying the kinase selectivity. The selectivity score³⁷ simply divides the number of kinase hits at an arbitrary K_d or IC_{50} value. The disadvantage of selectivity score is difficult to select a suitable arbitrary parameter. Gini score was calculated based on the %-inhibition data.³⁹ The disadvantage of Gini score is no conceptual or thermodynamic meaning and it mainly depends on experimental conditions. Partition index is a K_d -based score with a thermodynamic underpinning.⁴⁰ But it does not characterize the complete inhibitor distribution in the imaginary kinase mixture, it just the fraction bound to the reference enzyme. Comparing with these traditional inhibitor selectivity quantitatively steering methods, the index-based on the Shannon-Jaynes entropy¹² used in this study had three advantages: (i) I_{Inh} is dependent on the panel of enzymes included in the calculation: just as an AC_{50} or a K_i is a functional parameter defined for a particular pair of enzymes and inhibitors. (ii) I_{Inh} is completely independent of an inhibitor's absolute level of potency, so it did not influence by different experimental condition. (iii) I_{Inh} was calculated based on the Shannon-Jaynes entropy theory: Lower

promiscuity (higher selectivity) results in lower entropy, higher promiscuity (lower selectivity) results in higher entropy. Recently, Uitdehaag et al. also developed a similar theoretical entropy score to quantitatively steer kinase selectivity.⁴¹

P450 Inhibitory Promiscuity versus Catalytic Promiscuity.

The calculation of catalytic promiscuity using 65 substrates were collected from Foti's work.¹⁰ As given in Table 3, all five isoforms have the enzyme susceptibility index of catalytic promiscuity (J_{Cat}) within a high value and narrow range from 0.65 to 0.83 weighted by MACCS keys, indicating that drug-metabolizing P450s generally catalyzed a broad range of compounds. Comparing the J_{Cat} value of different CYP isoforms, susceptibility to catalysis follows the rank-order: 2C9 < 2C19 < 1A2 ≈ 2D6 < 3A4. CYP3A4 has the highest J_{Cat} value, and CYP2C9 has the lowest J_{Cat} value for substrates catalysis. Comparing the P450 inhibitory promiscuity (J_{Inh}) and catalytic promiscuity (J_{Cat}) of CYP1A2, 2C9, 2C19, 2D6, and 3A4, it is strange that the high catalytic promiscuity isoform of CYP3A4 has the moderate inhibitory promiscuity. If a new promiscuous enzyme would have increased susceptibility to inhibition by ligands that do not inhibit the "original" enzyme, this would be a negative trade-off.⁷

A plot of J_{Inh} (y-axis) versus J_{Cat} (x-axis) was given in Figure 7. The diagonal line (dashed line) in Figure 7 with a slope of 1 is the expected line along which the data would lie if J_{Inh} and J_{Cat} were exactly correlated. As shown in Figure 7, the slope could in principle be 1, <1, or >1. If a slope is much greater than 1, it represents a significant negative trade-off. If a slope is much less than 1, it represents a minimal negative trade-off.¹⁰ There is a minimal negative trade-off of J_{Inh} with a slope of -0.38 for total data set of 11,578 compounds. In this study, the total data set of 11,578 compounds was collected from three different data sources: (i) 6107 compounds from the Molecular Libraries Small Molecule Repository (MLSMR; <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcsubstance&term=mlsmr>), including compounds chosen for diversity and rule-of-five compliance, synthetic tractability, and availability, (ii) 3015 compounds from biofocused libraries (Biofocused), and (iii) 2456 compounds from combinatorial chemistry libraries (POPC), containing privileged structures targeted at G protein-coupled receptors and kinase and containing purified natural products or related structures.¹¹ The detailed description can be found in Table S1. The Figure 7 gives the plots of J_{Inh} versus J_{Cat} for different databases. For POPC and Biofocused database, there is a significant negative trade-off of inhibitory promiscuity. But there is a minimal negative trade-off of inhibitory promiscuity for MLSMR database. Comparing the correlation between J_{Inh} and J_{Cat} for different databases, the POPC gave the highest correlation ($R = 0.91$, slope = 1.3), which indicated that the negative trade-off is nearly insignificant because the CYPs become more catalytic promiscuous. As one can see in Figure 7, the J_{Inh} values of given P450 isoforms were clustered at the right of the diagonal. It indicated that all J_{Inh} values were higher than J_{Cat} values for five P450 isoforms. As the data demonstrate, our results are in agreement with the general conclusion that more compounds in chemicals space are likely to be inhibitors of an enzyme than substrates for this enzyme.¹⁰

Functional promiscuity of proteins has become increasingly recognized in the past few years, a large scale investigating of the relationship between P450 inhibitory promiscuity, catalytic promiscuity, and physicochemical traits would be used to explore the enzyme evolutionary mechanisms.¹⁰ As shown in Figure 7, a different data source had different negative trade-offs, which indicated that a different ligand matrix had great impact for exploring

Table 3. Inhibitory Susceptibility Index (J_{Inh}) of Five Major P450 Isoforms for Different Data Source^b

isoform	I_{susc}				J_{Inh}				${}^a J_{Cat}$
	all	POPC	MLSMR	biofocused	all	POPC	MLSMR	biofocused	
1A2	0.9436	0.9567	0.9380	0.9157	0.9501	0.9543	0.9511	0.9381	0.81
2C9	0.9214	0.8323	0.9473	0.8653	0.9168	0.8242	0.9411	0.8611	0.65
2C19	0.9357	0.8745	0.9611	0.8749	0.9417	0.8944	0.9638	0.8785	0.70
2D6	0.8745	0.8774	0.8486	0.8666	0.8778	0.9091	0.8492	0.8593	0.81
3A4	0.9392	0.9477	0.9311	0.9200	0.9247	0.9612	0.9160	0.8962	0.83

^a The J_{Cat} data was collected from Foti's work.¹⁰ ^b P450 inhibitory susceptibility index (J_{Inh}) and P450 catalytic susceptibility index (J_{Cat}) values are weighted by the relative chemical similarities (MACCS keys) of the inhibitors, and I_{susc} value is unweighted. MLSMR: Molecular Libraries Small Molecule Repository, POPC: combinatorial chemistry libraries, Biofocused: biofocused libraries, All: MLSMR+ POPC+Biofocused.

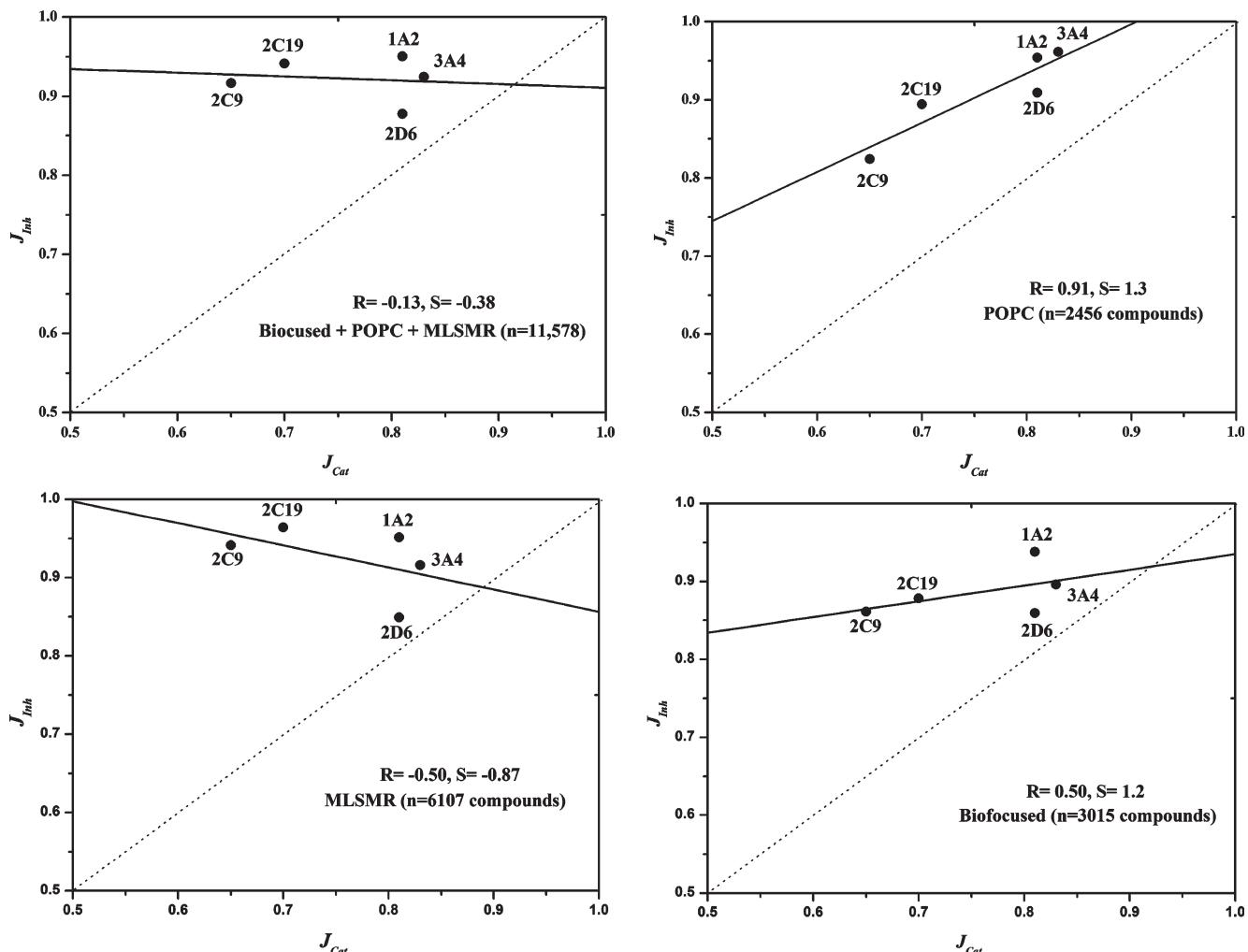


Figure 7. Plot of J_{Inh} (inhibitory promiscuity) versus J_{Cat} (catalytic promiscuity) weighted by MACCS keys for CYP1A2, 2C9, 2C19, 2D6, and 3A4. The gray dashed line represents a perfect correlation between J_{Inh} and J_{Cat} . The black solid line represents the real correlation between J_{Inh} and J_{Cat} . R: correlation coefficient. S: slope. MLSMR: Molecular Libraries Small Molecule Repository, POPC: combinatorial chemistry libraries; Biofocused: biofocused libraries.

the relationship between P450 inhibitory promiscuity and catalytic promiscuity. Overall, the slightly negative correlation ($R = -0.13$) between J_{Inh} and J_{Cat} was observed based on the comprehensive data set of 11578×11578 ligands matrix in Figure 7. These results were not consistent with Foti's conclusion that there was positive correlation $R^2 = 0.68$ using the 65×65

ligands matrix. It indicated that there might be other parameters, such as thermal stability, aggregation resulting in P450 evolutionary trajectory. In any case, the results presented here would be useful for people to understand the P450 isoform evolutionary mechanisms. Building a specific P450 isoform evolutionary mechanism was beyond the range of this article.

The classification model with reasonable predictive accuracy presented here could be used as a simple filter to explore the potential metabolism-related toxicological profiles, i.e., DDI, caused by CYP1A2, 2C9, 2C19, 2D6, and 3A4 promiscuous inhibition in drug discovery. Moreover, some privileged substructures were identified for high selective P450 inhibitors and high promiscuous inhibitors in Scheme 1, which was in good agreement with the qualitatively “known” aspects of P450 inhibition: for instance, the presence of certain functional groups on small molecules (such as triazoles and nitroaromatics) often increases inhibition for certain drug-metabolizing P450s. In addition, the relationships between ten physicochemical properties (AlogP, XlogP, MW, TopoPSA, nHeavyAtom, nRing, nHBAcc, nHBDon, naAromAtom, and nAromBond) and chemical P450 inhibitory promiscuity were systematically investigated. We found that molecular aromaticity was closely interrelated to chemical P450 inhibitory promiscuity. If a molecule had the higher aromaticity, this molecule would have the higher P450 inhibitory promiscuity. This conclusion was consistent with the known preference for planar, polyaromatic substrates of CYP isoforms, particularly for CYP1A2.³⁵

CONCLUSIONS

An entropy-based index was defined to quantitatively measure P450 inhibitory promiscuity of chemicals based on the comprehensive data set containing more than 11,500 drug-like compounds with inhibition against five major P450 isoforms, CYP1A2, 2C9, 2C19, 2D6, and 3A4. Five vital perspectives were obtained here: (i) among the ten physicochemical properties, molecular aromaticity correlates with chemical P450 inhibitory promiscuity, but there was little correlation or no correlation between chemical P450 inhibitory promiscuity and molecular lipophilicity, molecular complexity, hydrogen bonding ability, or TopoPSA. (ii) An index was used to quantify the susceptibilities of five different P450 isoforms to inhibition, and the data indicated that there was a surprising level of chemical P450 inhibitory promiscuity even for substrate specific P450 isoform, susceptibility to inhibition follows the rank-order: 1A2 > 2C19 > 3A4 > 2C9 > 2D6. (iii) We further confirmed that molecular P450 inhibitory potency would have no correlation with P450 inhibitory promiscuity. (iv) There is a minor negative trade-off between the P450 inhibitory promiscuity and catalytic promiscuity. For example, high catalytic promiscuous CYP3A4 only has moderate inhibitory promiscuity. (v) Robust classification models were built for predicting high or low P450 inhibitory promiscuity of new chemicals, which could be used as a filter to explore the potential DDI issues in early stage of drug discovery.

ASSOCIATED CONTENT

Supporting Information. Tables S1 and S2. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Phone: +86-21-64251052. Fax: +86-21-64253651. E-mail: whli@ecust.edu.cn (W.L.), ytang234@ecust.edu.cn (Y.T.).

ACKNOWLEDGMENT

We thank Dr. Douglas Auld (Genomic Assay Technologies, NIH Chemical Genomics Center, USA) for sharing their data sets with us. This work was supported by the National Natural Science Foundation of China (Grant No. 21072059), the Program for New Century Excellent Talents in University (Grant No. NCET-08-0774), Shanghai Natural Science Foundation (Grant No. 10ZR1407000), Fundamental Research Funds for the Central Universities (WY1014010), and the 111 Project (Grant No. B07023).

REFERENCES

- (1) Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4*, 682–690.
- (2) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906.
- (3) Wienkers, L. C.; Heath, T. G. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat. Rev. Drug Discovery* **2005**, *4*, 825–833.
- (4) Lasser, K. E.; Allen, P. D.; Woolhandler, S. J.; Himmelstein, D. U.; Wolfe, S. M.; Bor, D. H. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* **2002**, *287*, 2215–2220.
- (5) Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.* **2004**, *32*, 1201–1208.
- (6) van Loo, B.; Jonas, S.; Babtie, A. C.; Benjdia, A.; Berteau, O.; Hyvonen, M.; Hollfelder, F. An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proc. Natl. Acad. Sci.* **2010**, *107*, 2740–2745.
- (7) Khersonsky, O.; Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505.
- (8) Nath, A.; Atkins, W. M. A quantitative index of substrate promiscuity. *Biochemistry* **2008**, *47*, 157–166.
- (9) Nath, A.; Zientek, M. A.; Burke, B. J.; Jiang, Y.; Atkins, W. M. Quantifying and predicting the promiscuity and isoform specificity of small-molecule cytochrome P450 inhibitors. *Drug Metab. Dispos.* **2010**, *38*, 2195–2203.
- (10) Foti, R. S.; Honaker, M.; Nath, A.; Pearson, J. T.; Buttrick, B.; Isoherranen, N.; Atkins, W. M. Catalytic versus Inhibitory Promiscuity in Cytochrome P450s: Implications for Evolution of New Function. *Biochemistry* **2011**, *50*, 2387–2393.
- (11) Veith, H.; Southall, N.; Huang, R.; James, T.; Fayne, D.; Artemenko, N.; Shen, M.; Ingles, J.; Austin, C. P.; Lloyd, D. G.; Auld, D. S. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat. Biotechnol.* **2009**, *27*, 1050–1055.
- (12) Shannon, C. E. A mathematical theory of communication. *Bell. Syst. Tech. J.* **1948**, *27* (379–423), 623–656.
- (13) John, M. B.; Geoffrey, M. D.; Willett, P. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.
- (14) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Bryant, S. H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *W623–W633*.
- (15) Lewis, D. F.; Jacobs, M. N.; Dickins, M. Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discovery Today* **2004**, *9*, 530–537.
- (16) Lewis, D. F.; Dickins, M. Substrate SARs in human P450s. *Drug Discovery Today* **2002**, *7*, 918–925.
- (17) Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474.

- (18) Klekota, J.; Roth, F. P. Chemical substructures that enrich for biological activity. *Bioinformatics* **2008**, *24*, 2518–2525.
- (19) Shen, J.; Cheng, F.; Xu, Y.; Li, W.; Tang, Y. Estimation of ADME properties with substructure pattern recognition. *J. Chem. Inf. Model.* **2010**, *50*, 1034–1041.
- (20) Jensen, B. F.; Vind, C.; Padkjaer, S. B.; Brockhoff, P. B.; Refsgaard, H. H. In silico prediction of cytochrome P450 2D6 and 3A4 inhibition using Gaussian kernel weighted k-nearest neighbor and extended connectivity fingerprints, including structural fragment analysis of inhibitors versus noninhibitors. *J. Med. Chem.* **2007**, *50*, 501–511.
- (21) Corinna, C.; Vladimir, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (22) Cheng, F.; Yu, Y.; Shen, J.; Yang, L.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. Classification of Cytochrome P450 Inhibitors and non-Inhibitors using Combined Classifiers. *J. Chem. Inf. Model.* **2011**, *51*, 996–1011.
- (23) Cheng, F.; Shen, J.; Yu, Y.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. In silico prediction of Tetrahymena pyriformis toxicity for diverse industrial chemicals with substructure pattern recognition and machine learning methods. *Chemosphere* **2011**, *82*, 1636–1643.
- (24) Vasanthanathan, P.; Tabouret, O.; Oostenbrink, C.; Vermeulen, N. P.; Olsen, L.; Jorgensen, F. S. Classification of cytochrome P450 1A2 inhibitors and noninhibitors by machine learning techniques. *Drug Metab. Dispos.* **2009**, *37*, 658–664.
- (25) Eitrich, T.; Kless, A.; Druska, C.; Meyer, W.; Grotendorst, J. Classification of highly unbalanced CYP450 data of drugs using cost sensitive machine learning techniques. *J. Chem. Inf. Model.* **2007**, *47*, 92–103.
- (26) Michielan, L.; Terfloth, L.; Gasteiger, J.; Moro, S. Comparison of multilabel and single-label classification applied to the prediction of the isoform specificity of cytochrome p450 substrates. *J. Chem. Inf. Model.* **2009**, *49*, 2588–2605.
- (27) Yap, C. W.; Chen, Y. Z. Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines. *J. Chem. Inf. Model.* **2005**, *45*, 982–992.
- (28) Chang, C. C.; Lin, C.-J. LIBSVM: a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (accessed Jan. 18, 2010).
- (29) Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **2000**, *16*, 412–424.
- (30) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.
- (31) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **2006**, *16*, 127–136.
- (32) Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.
- (33) Peters, J. U.; Schnider, P.; Mattei, P.; Kansy, M. Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds. *ChemMedChem* **2009**, *4*, 680–686.
- (34) Yang, Y.; Chen, H.; Nilsson, I.; Muresan, S.; Engkvist, O. Investigation of the relationship between topology and selectivity for druglike molecules. *J. Med. Chem.* **2010**, *53*, 7709–7714.
- (35) Lewis, D. F.; Eddershaw, P. J.; Dickins, M.; Tarbit, M. H.; Goldfarb, P. S. Structural determinants of cytochrome P450 substrate specificity, binding affinity and catalytic rate. *Chem. Biol. Interact.* **1998**, *115*, 175–199.
- (36) Ekoos, M.; Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci.* **2006**, *103*, 13682–13687.
- (37) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.
- (38) Hanley, J. A.; McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36.
- (39) Graczyk, P. P. Gini coefficient: a new way to express selectivity of kinase inhibitors against a family of kinases. *J. Med. Chem.* **2007**, *50*, 5773–5779.
- (40) Cheng, A. C.; Eksterowicz, J.; Geuns-Meyer, S.; Sun, Y. Analysis of kinase inhibitor selectivity using a thermodynamics-based partition index. *J. Med. Chem.* **2010**, *53*, 4502–4510.
- (41) Uitdehaag, J. C.; Zaman, G. J. A theoretical entropy score as a single value to express inhibitor selectivity. *BMC Bioinf.* **2011**, *12*, 94.