In Silico Modeling of Nonspecific Binding to Human Liver Microsomes

Hua Gao, Lili Yao, Heather W. Mathieu, Ying Zhang, Tristan S. Maurer, Matthew D. Troutman, Dennis O. Scott, Roger B. Ruggeri, and Jing Lin

Department of Chemistry (H.G., R.B.R.) and Department of Pharmacokinetics, Dynamics and Metabolism (L.Y., H.W.M., Y.Z., T.S.M., M.D.T., D.O.S., J.L.), Pfizer Global Research and Development, Groton, Connecticut

Received December 13, 2007; accepted July 3, 2008

ABSTRACT:
Estimation of unbound fraction of substrate in microsomal incubation media is important in accurately predicting hepatic intrinsic clearance and drug-drug interactions. In this study, the unbound fraction of 1223 drug-like molecules in human liver microsomal incubation media has been determined using equilibrium dialysis. These compounds, which include 27 marketed drug molecules, cover a much broader range of physicochemical properties such as hydrophobicity, molecular weight, ionization state, and degree of binding than those examined in previous work. In developing the in silico model, we have used two-dimensional molecular descriptors including cLogP, Kier connectivity, shape, and E-state indices, a subset of MOE descriptors, and a set of absorption, disposition, metabolism, and excretion structural keys used for our in-house ADMET, absorption, disposition, metabolism, excretion, and toxicity model. Hydrophobicity is the most important molecular property contributing to the nonspecific binding of substrate to microsomes. The prediction accuracy of the model is validated using a subset of 100 compounds, and 92% of the variance is accounted for by the model with a root mean square error (RMSE) of 0.10. For the training set of compounds, 99% of variance is accounted for by the model with a RMSE of 0.02. The performance of the developed model has been further tested using the 27 marketed drug molecules with a RMSE of 0.10 between the observed and the predicted unbound fraction values.

In drug discovery, the ultimate goal of a medicinal chemist is to rationally design compounds that are safe and efficacious at a marketable dose regimen. Such chemical design efforts frequently rely upon intrinsic clearance (CL_{int}) and P450 inhibitory potency (IC_{50}) estimates derived from human liver microsomal incubations. With relatively few assumptions, these estimates can be expected to have proportional implications for oral dose requirements and therapeutic windows against pharmacokinetic drug-drug interactions. For example, the classic well stirred model of hepatic clearance can be used to illustrate the direct proportional relationship between CL_{int} and oral dose by eq. 1 (Wagner et al., 1965). Other oral dose determinants include the unbound steady-state average concentration necessary for the desired pharmacologic response (C_{ss,u}), dose interval (τ), and fraction of dose absorbed (fa):

\[
\text{Oral dose} = \frac{C_{ss,u} \cdot CL_{int} \cdot \tau}{fa} \tag{1}
\]

In addition, by assuming that inhibition of P450 occurs via a competitive mechanism, eq. 2 is a valid and commonly used approach to calculate the therapeutic window against pharmacokinetic drug-drug interactions:

\[
\text{Therapeutic window} = \frac{C_{ss,u}}{IC_{50}} \tag{2}
\]

It has now been well established in the scientific literature that nonspecific binding of compounds into microsomal phospholipids will confound the experimental estimation of CL_{int} and IC_{50} by decreasing the fraction of unbound drug (f_{unbc}) to metabolizing enzymes within the incubation (Kalvass et al., 2001; Margolis and Obach, 2003; Obach, 1997, 1999; Tran et al., 2002). In such cases, estimates of CL_{int} and IC_{50} derived from microsomes are best deemed “apparent” (i.e., CL_{int,app} and IC_{50,app}) because they represent a mixture of kinetic determinants with both relevant (CL_{int} and IC_{50}) and irrelevant (f_{unbc}) implications for drug design by eqs. 3 and 4 (Obach, 1996):

\[
CL_{int,app} = CL_{int} \cdot f_{unbc} \tag{3}
\]

\[
IC_{50,app} = \frac{IC_{50}}{f_{unbc}} \tag{4}
\]

From these relationships, it follows that design efforts based simply on CL_{int,app} and IC_{50,app} values will be misguided in the presence of significant nonspecific binding (i.e., f_{unbc} \ll 1). This can be particularly damaging, considering the general influence of lipophilicity over CL_{int} (increases), IC_{50} (decreases), and f_{unbc} (decreases). Thus, it is important to note that microsomal binding may mask worsening...

ABBREVIATIONS: P450, cytochrome P450; LC, liquid chromatography; MS/MS, tandem mass spectroscopy; ADME, absorption, disposition, metabolism, and excretion; ADMET, absorption, disposition, metabolism, excretion, and toxicity; NN, nearest neighbor.
values of $CL_{int}$ and $IC_{50}$ and entice design efforts into increasingly lipophilic space. It is also important to note that extensive nonspecific micromolar binding is common among both marketed drugs and chemical libraries of the pharmaceutical industry, particularly among lipophilic chemical entities. Thus, it is essential for these kinetic parameters to be corrected for $fu_{mic}$ to ensure effective design in drug discovery (Obach, 1996; Tucker et al., 2001; Björnsson et al., 2003; Grime and Riley, 2006). Unfortunately, experimental methods for the determination of $fu_{mic}$ (e.g., equilibrium dialysis and ultracentrifugation) are far less amenable to high throughput screening than are micromolar stability incubations, thereby limiting the flow of information available to support design efforts. The need for this information has led to the development of several in silico methods for predicting $fu_{mic}$ from physicochemical properties (most notably, LogD and LogP) (Austin et al., 2002, 2005; Hallifax and Houston, 2006; Sykes et al., 2006). Although the number of compounds used to develop these published models is limited (e.g., 56) and the data are gathered from a variety of sources (e.g., different laboratories, conditions, and species), the results indicate that $fu_{mic}$ is highly amenable to prediction via physicochemical properties. This finding is perhaps not surprising, given that $fu_{mic}$ probably represents a passive partitioning into microsomal phospholipids. In this work, we developed a novel in silico model using Cubist with human $fu_{mic}$ data generated in our laboratory for 1223 compounds including 27 commercial available chemicals via a uniform 96-well equilibrium dialysis methodology. In addition to a relatively larger number of compounds, this data set encompasses a much broader range of physiochemical properties such as lipophilicity (LogP and LogD), molecular weight, ionization state ($pK_a$), and degree of binding ($\log[fu_{mic}/(1 – fu_{mic})]$) than were examined in previous work.

### Materials and Methods
**Materials.** Human liver microsome mix (HL-Mix-101) was generated from a pool of 59 individual livers and was obtained from BD Gentest (Woburn, MA). The protein content of the microsome mix was 20.4 mg/ml, and the cytochrome P450 concentration was determined to be 0.33 nmol/mg. The same batch of microsomes was used for all of the studies and was stored at $-80^\circ$C. Albendazole, amitriptyline, carbamazepine, chlorimipramine, chlorpromazine, clozapine, colchicine, desipramine, diazepam, erythromycin, imipramine, indomethacin, ketoprofen, ketocozalone, midazolam, pimozide, piroxicam, prednisone, promethazine, propranolol, quinidine, sulindac, tenidap, tolbutamide, trimiprazine, verapamil, and warfarin were obtained from the Pfizer compound collection (Table 1). All of the Pfizer proprietary compounds were obtained from the Pfizer internal sample bank (Groton, CT). A 96-well dialysis apparatus and cellulose dialysis membrane strips with a molecular weight cutoff at 12,000 to 14,000 were purchased from HTDialysis (Gales Ferry, CT).

**Experimental Methods.** Microsomal binding incubations. Microsomal binding in human liver microsomes was performed in triplicate using a 96-well equilibrium dialysis as adapted from a previous described method (Banker et al., 2003). Each dialysis membrane strip was conditioned sequentially for 15 min in deionized water, 30% ethanol, and then 0.1 M potassium phosphate buffer, pH 7.4. A conditioned membrane strip was then loaded into a 96-well dialysis apparatus. Experimental compound was diluted in dimethyl sulfoxide to a concentration of 100 $\mu$M. Human liver microsomes were thawed and diluted with 0.1 M potassium phosphate buffer (pH 7.4) to a final protein concentration of 0.76 mg/ml. Experimental compound was further diluted 100-fold to yield 1 $\mu$M final concentration with the diluted microsomal solution. The compound and microsomes were thoroughly mixed, and 150 $\mu$l of the resulting solution was added to one side of the dialysis well as a donor side and 20 $\mu$l of this solution was retained for analysis. An equivalent 150 $\mu$l of 0.1 M potassium phosphate buffer (pH 7.4) was added to the other side of the dialysis well as a receiving side. The 96-well HTDialysis apparatus was

### Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>$M_r$</th>
<th>cLogP</th>
<th>cLogD</th>
<th>Acidic $pK_a$</th>
<th>Basic $pK_a$</th>
<th>Measured $fu_{mic}$</th>
<th>Calculated $fu_{mic}$</th>
<th>$\Delta^a$</th>
<th>NN$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albendazole</td>
<td>265</td>
<td>3.46</td>
<td>3.06</td>
<td>11.2</td>
<td></td>
<td>5.62</td>
<td>0.69</td>
<td>0.02</td>
<td>0.65</td>
</tr>
<tr>
<td>Amitriptyline</td>
<td>277</td>
<td>4.85</td>
<td>3.15</td>
<td></td>
<td></td>
<td>9.18</td>
<td>0.27</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>Carbamazepine</td>
<td>236</td>
<td>2.38</td>
<td>2.67</td>
<td></td>
<td></td>
<td>0.97</td>
<td>0.04</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td>Chlorimipramine</td>
<td>315</td>
<td>5.92</td>
<td>3.54</td>
<td></td>
<td></td>
<td>9.46</td>
<td>0.09</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Chlorpromazine</td>
<td>319</td>
<td>4.05</td>
<td>2.87</td>
<td></td>
<td></td>
<td>9.41</td>
<td>0.15</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Clozapine</td>
<td>327</td>
<td>3.71</td>
<td>3.35</td>
<td></td>
<td></td>
<td>7.14</td>
<td>0.45</td>
<td>0.03</td>
<td>0.15</td>
</tr>
<tr>
<td>Colchicine</td>
<td>399</td>
<td>1.2</td>
<td>0.92</td>
<td></td>
<td></td>
<td>1.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Desipramine</td>
<td>266</td>
<td>4.47</td>
<td>1.43</td>
<td></td>
<td></td>
<td>10.4</td>
<td>0.35</td>
<td>0.01</td>
<td>0.26</td>
</tr>
<tr>
<td>Diazepam</td>
<td>285</td>
<td>2.96</td>
<td>2.96</td>
<td></td>
<td></td>
<td>3.4</td>
<td>0.77</td>
<td>0.03</td>
<td>0.55</td>
</tr>
<tr>
<td>Erythromycin</td>
<td>734</td>
<td>1.61</td>
<td>2.08</td>
<td></td>
<td></td>
<td>8.16</td>
<td>0.84</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Imipramine</td>
<td>280</td>
<td>5.04</td>
<td>2.79</td>
<td></td>
<td></td>
<td>9.49</td>
<td>0.35</td>
<td>0.02</td>
<td>0.57</td>
</tr>
<tr>
<td>Indomethacin</td>
<td>358</td>
<td>4.18</td>
<td>-0.16</td>
<td>3.96</td>
<td></td>
<td>9.01</td>
<td>0.93</td>
<td>0.03</td>
<td>0.72</td>
</tr>
<tr>
<td>Ketoconazole</td>
<td>531</td>
<td>3.38</td>
<td>2.71</td>
<td></td>
<td></td>
<td>6.33</td>
<td>0.27</td>
<td>0.08</td>
<td>0.27</td>
</tr>
<tr>
<td>Ketoprofen</td>
<td>254</td>
<td>2.76</td>
<td>-0.25</td>
<td>4.23</td>
<td></td>
<td>0.92</td>
<td>0.06</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Midazolam</td>
<td>326</td>
<td>3.42</td>
<td>3.92</td>
<td></td>
<td></td>
<td>5.59</td>
<td>0.65</td>
<td>0.01</td>
<td>0.64</td>
</tr>
<tr>
<td>Pimozide</td>
<td>462</td>
<td>6.4</td>
<td>4.77</td>
<td>12</td>
<td></td>
<td>9.42</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Piroxicam</td>
<td>331</td>
<td>1.89</td>
<td>-1.15</td>
<td>4.5</td>
<td></td>
<td>3.8</td>
<td>1.02</td>
<td>0.02</td>
<td>0.98</td>
</tr>
<tr>
<td>Prednisone</td>
<td>358</td>
<td>1.66</td>
<td>1.57</td>
<td>12.4</td>
<td></td>
<td>9.28</td>
<td>0.72</td>
<td>0.06</td>
<td>0.65</td>
</tr>
<tr>
<td>Promethazine</td>
<td>284</td>
<td>4.4</td>
<td>2.29</td>
<td></td>
<td></td>
<td>9.14</td>
<td>0.62</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td>Propranolol</td>
<td>259</td>
<td>2.75</td>
<td>1.27</td>
<td></td>
<td></td>
<td>9.14</td>
<td>0.62</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td>Quinidine</td>
<td>324</td>
<td>2.79</td>
<td>1.6</td>
<td>12.8</td>
<td></td>
<td>9.38</td>
<td>1.08</td>
<td>0.01</td>
<td>0.89</td>
</tr>
<tr>
<td>Sulindac</td>
<td>356</td>
<td>3.16</td>
<td>0.55</td>
<td>4.26</td>
<td></td>
<td>9.28</td>
<td>0.63</td>
<td>0.03</td>
<td>0.95</td>
</tr>
<tr>
<td>Tenidap</td>
<td>321</td>
<td>-2.3</td>
<td>0.16</td>
<td>4.5</td>
<td></td>
<td>9.34</td>
<td>0.96</td>
<td>0.06</td>
<td>0.98</td>
</tr>
<tr>
<td>Tolbutamide</td>
<td>270</td>
<td>2.5</td>
<td>0.39</td>
<td>5.12</td>
<td></td>
<td>9.34</td>
<td>0.23</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>Trimiprazine</td>
<td>298</td>
<td>4.8</td>
<td>2.53</td>
<td></td>
<td></td>
<td>9.34</td>
<td>0.59</td>
<td>0.03</td>
<td>0.43</td>
</tr>
<tr>
<td>Verapamil</td>
<td>455</td>
<td>4.47</td>
<td>2.33</td>
<td></td>
<td></td>
<td>8.97</td>
<td>0.59</td>
<td>0.03</td>
<td>0.43</td>
</tr>
<tr>
<td>Warfarin</td>
<td>308</td>
<td>2.9</td>
<td>0.61</td>
<td>4.5</td>
<td></td>
<td>8.97</td>
<td>0.80</td>
<td>0.08</td>
<td>0.85</td>
</tr>
</tbody>
</table>

$a$ Difference between measured and calculated $fu_{mic}$.

$b$ Number of nearest neighbors.
phase B. The system was operated with a flow rate of 0.75 ml/min. In the column was allowed to re-equilibrate for 0.2 min at 0% of mobile phase A, held at 95% for 0.2 min and returned to 0% over 0.1 min. After each injection the column was re-equilibrated for 0.2 min at 0% of mobile phase B and then linearly increased to 95% over next 0.75 min. Then mobile phase B consisted of 95% 10 mM ammonium formate (adjusted to pH 3.2 with formic acid). The proportion of mobile phase B was held at 0% for 0.45 min and then linearly increased to 95% over next 0.75 min. Then mobile phase B was held at 95% for 0.2 min and returned to 0% over 0.1 min. After each injection the column was allowed to re-equilibrate for 0.2 min at 0% of mobile phase B. The system was operated with a flow rate of 0.75 ml/min.

The frozen samples were thawed at room temperature and quenched by 200 μl of acetonitrile containing an appropriate internal standard compound. The plate of quenched samples was then centrifuged at 3000 rpm at 4°C for 20 min to precipitate proteins, and 10 μl of supernatant was injected into the liquid chromatograph-mass spectrometer. Approximately 10% of the eluent was introduced into the mass spectrometer source. The source temperature of the mass spectrometer was maintained at 450°C. The most abundant product ion of each molecular ion (M+H+) was chosen in multiple reaction monitoring mode on the basis of collision-induced dissociations of precursor occurring at a collision energy of 25, 35, or 45 eV, respectively, for accommodating the needs of a large quantity of compounds and the detection sensitivity of the mass spectrometer. Quantitation of each compound was achieved by comparison of the analyte/internal standard peak area ratios of the donor side (microsomal solution) and the receiving side (buffer). The unbound fraction of each compound in human liver microsomes (fu_mic) was calculated according to eq. 5:

\[ \text{fu}_{\text{mic}} = \frac{C_b}{C_m} \]  (5)

where \( C_b \) and \( C_m \) denote peak area ratio of a compound in the dialysis apparatus from buffer and microsomal solution chambers, respectively.

**Model Development.** For compounds with measurements done at different microsomal protein concentrations (internal historical data), the unbound fraction was normalized to a value at a microsomal protein concentration of 0.76 mg/ml using a modified version of an equation described previously (Kalvass and Maurer, 2002) (eq. 6):

\[ \text{fu}_{\text{mic2}} = \frac{0.76}{C_1} \times \left( \frac{1 - \text{fu}_{\text{mic1}}}{\text{fu}_{\text{mic1}}} \right) + 1 \]  (6)

where \( \text{fu}_{\text{mic2}} \) is the unbound fraction at a microsomal concentration of 0.76 mg/ml and \( \text{fu}_{\text{mic1}} \) is the unbound fraction at a microsomal concentration of \( C_1 \). A logit transformation of all the data was performed according to eq. 7, which will enable the predicted \( \text{fu}_{\text{mic}} \) within a boundary of 0 to 1:

\[ A = \log \left( \frac{\text{fu}_{\text{mic}}}{1 - \text{fu}_{\text{mic}}} \right) \]  (7)

The whole set of 1223 compounds was randomly divided into two subsets. A subset of 1123 compounds was selected as a training set to develop a model, and a subset of 100 compounds was used as a validation set to validate the developed model. In addition, a set of 27 drug molecules (Table 1) was used as an external validation set to test the model performance.

In this study, we have used two-dimensional molecular descriptors including cLogP (BioByte Corp., Claremont, CA), Kier connectivity, shape, and E-state indices (Kier, 1987, 1989, 1999), a subset of MOE descriptors (Chemical Computing Group Inc., 2004, MOE 2004.03, http://www.chemcomp.com), and a set of ADME keys that are structural features used for our in-house ADMET modeling (Tu and Li, 2004; Lee et al., 2007). Some of the descriptors such as Kier shape indices contain implicit three-dimensional information. Explicit three-dimensional molecular descriptors were not used to avoid bias of the analysis due to predicted conformational effects and speed of calculation for fast prediction as stated in a previous publication (Gao et al., 1999). cLogD and pK_a calculations in our analysis were performed using ACD software (version 9.03; ACD/Labs, Toronto, ON, Canada).

The statistical model was derived using Cubist committee plus composite descriptors.
models (Rulequest Research, 2006, http://www.rulequest.com). Cubist is a tool for generating rule-based predictive models and resembles a piecewise linear regression model, except that the rules can overlap. It does this by building a model containing one or more rules, where each rule is a conjunction of conditions associated with a linear regression. If a new case satisfies all the conditions, then the associated linear regression model is used to predict the target value. If two or more rules apply to a case, then the values are averaged to arrive at a final prediction. The predictive accuracy of a rule-based model can be improved by combining it with an instance-based or nearest-neighbor model. The latter predicts the target value of a new case by finding the \( n \) most similar cases in the training data and averaging their target values. Then Cubist combines the rule-based prediction with instance-based prediction to give a final predicted value. Prediction reliability was estimated using similarity to the training set of compounds. The similarity matrix used in the study is atom pair similarity (Carhart et al., 1985; Sheridan et al., 2004).

### Results

A set of 1223 Pfizer proprietary compounds was selected to cover a range of lipophilicity (cLogP from \(-1.45\) to \(9.20\) and cLogD from \(-4.6\) to \(9.0\)), molecular weight (from 200 to 650), ionization class (from an acidic \( pK_a \) of 2.0 to a basic \( pK_a \) of 11.2), and degree of binding (\( \log[fu_{mic}/(1 - fu_{mic})] \) from 6 to \(-8\)). The 27 commercial compounds selected cover a range of lipophilicity (cLogP from 1.2 to 6.4 and cLogD from \(-1.2\) to \(4.8\)) and molecular weight (from 236 to 734), ionization class (from an acidic \( pK_a \) of 3.96 to a basic \( pK_a \) of 10.4), and degree of binding (\( \log[fu_{mic}/(1 - fu_{mic})] \) from \(-4.6\) to \(5.0\)) (Table 1). An equilibrium dialysis in a 96-well format was used to determine the extent of binding in triplicate for all of the compounds in human liver microsomes. The 27 commercial molecules were used to validate the accuracy and precision.
of the experimental design and were used as quality control during the 4-month study period (S.E. <10% from quadruplicates). Data from studies with recovery >80%, and coefficient of variation <20% from triplicates were included in the data set.

A predictive model was developed using Cubist. The observed and predicted microsomal unbound fractions using the developed model are plotted in Fig. 1. Because the model was built using committee plus composite models, a molecular descriptor can be used in multiple regression equations. It is not possible to describe the importance of a particular descriptor using correlation coefficients. The 9 most important molecular descriptors are cLogP, ADME Key 179 (smarts: [O,o]:*.*), ADME Key 171 (smarts: [NH]C=([O,S,N]), kS_dO (Kier E-state for =O), vsa_don (VDW donor surface area), ADME Key 10 (smarts: *:=([c;H0,H1][O;C;H1,H2,H3])), ADME Key 302 (smarts: N~![H]~[N,O]), number of carboxylic groups, and d\(^4\) (Kier connectivity index). Hydrophobicity (cLogP) is the most important molecular property in this model. For the test set of 100 compounds, 84% of them with 1 or 2 NNs are predicted within an error of ±0.1, whereas only 55% of the compounds with 0 NNs are predicted within an error of ±0.1. Similarly, 52% of the 27 commercial molecules are predicted within an error of ±0.10. The relationship of prediction error against number of NNs is plotted in Fig. 3, and it is apparent that the precision of the prediction increases with the increase of NNs.

The nonspecific binding of compounds to in vitro microsomal incubation matrix is believed to be mainly due to the partitioning of compounds to the hydrophobic phospholipids compartment. It is expected that the most important molecular property underlying the nonspecific binding is molecular hydrophobicity. The correlations of unbound fraction with calculated hydrophobicity parameters cLogD and cLogP are shown in Fig. 4. A similar observation has also been found for rat microsomal binding (Austin et al., 2002). Within a similar cLogD range, basic compounds tend to have lower unbound fractions than neutral and acidic compounds, and this is mainly due to the fact that positively charged molecules have higher affinity to phospholipids. It has also been noted that there is no significant difference among the three classes of compounds when plotted against

![Fig. 4. Relationship of microsomal unbound fraction with hydrophobicity (a, ACD LogD; b, clogP) (blue +, basic; red △, acidic; C, neutral).](image-url)

**Discussion**

The distribution of measured \(f_{umic}\) is shown in Fig. 2, and \(f_{umic}\) from this study is evenly distributed across the whole range for neutral, basic, and acidic compounds. The applicability of quantitative structure-activity relationship models is limited by the chemistry space of data from which the models are derived. If a model is trained on a narrow chemistry space, we intuitively expect the model only to apply to a narrow range of chemical compounds. On another hand, if the training set is very diverse, covering broad chemistry spaces, we would expect the model to be applicable to a wider chemical scope and therefore more powerful in molecular design. Generally it is believed that if a new chemical structure is similar to the training set of compounds, it is more likely to be well predicted; otherwise there is a significant extrapolation and the estimated value contains high uncertainty. In this study, we have used atom-pair similarity as the similarity matrix to indicate whether a new molecule is similar to the training set of compounds. The numbers of nearest neighbors (NNs) were calculated using a similarity threshold of 0.7. This similarity index has been shown to perform well in many in-house ADMET models as well as in literature reports (Sheridan et al., 2004). For the test set of 100 compounds, 84% of them with 1 or 2 NNs are predicted within an error of ±0.1, whereas only 55% of the compounds with 0 NNs are predicted within an error of ±0.1. Similarly, 52% of the 27 commercial molecules are predicted within an error of ±0.10. The relationship of prediction error against number of NNs is plotted in Fig. 3, and it is apparent that the precision of the prediction increases with the increase of NNs.

![Fig. 5. Plot of unbound fraction in microsome (\(f_{umic}\)) and in plasma (\(f_{uplasma}\)).](image-url)
cLogP, which is different from the phenomena of plasma protein binding ($f_{\text{up}}$). In plasma protein, acidic compounds in general are more tightly bound than basic compounds. As Fig. 5 shows, $f_{\text{ub}}$ does not track well with $f_{\text{up}}$. This demonstrates that $f_{\text{ub}}$ and $f_{\text{up}}$ are unlikely to cancel out each other in the prediction of human clearance using well stirred or parallel-tube models (Obach, 1996). Both the unbound fraction in microsomal incubation media and the unbound fraction in human plasma are important in estimating in vivo clearance from in vitro metabolism data.

The free concentration of compound in microsomal incubation media significantly affects the observed in vitro $CL_{\text{int, app}}$ (eq. 3) and IC$_{50, \text{app}}$ (eq. 4). The observed $CL_{\text{int, app}}$ versus unbound $CL_{\text{int}}$ (corrected by $f_{\text{ub}}$, using eq. 3) is plotted in Fig. 6, and it has been noted that for compounds with low $f_{\text{ub}} (<1)$, the difference can be more than 1000-fold, where the observed $CL_{\text{int, app}}$ was adapted from a method described previously (Obach, 1999). Nonspecific binding to incubation matrix renders fewer molecules available being metabolized by P450 enzymes, thus masking the true metabolic liability of a compound. A retrospective Pfizer internal analysis has shown that incorporation of human microsomal and plasma binding parameters could significantly increase the prediction accuracy of human pharmacokinetic parameters, including hepatic metabolic clearance and P450 inhibitory potency (data not shown). This finding provides a strong driving force to develop a predictive in silico model to estimate nonspecific binding to microsomal incubation media that will further assist molecule design with appropriate pharmacokinetic profiles to achieve a desirable efficacious concentration and safety window.

References


Address correspondence to: Dr. Jing Lin, Department of Pharmacokinetics, Dynamics and Metabolism, MS8220–4018, Groton, CT 06340. E-mail: jing.lin@pfizer.com