Physicochemical Property Space of Hepatobiliary Transport and Computational Models for Predicting Rat Biliary Excretion

Manthena V. S. Varma, George Chang, Yurong Lai, Bo Feng, Ayman F. El-Kattan, John Litchfield, and Theunis C. Goosen

Pharmacokinetics Dynamics and Metabolism, Pfizer Global Research and Development, Pfizer, Inc., Groton, Connecticut

Received January 16, 2012; accepted May 11, 2012

ABSTRACT:

Biliary excretion (BE) is a major elimination pathway, and its prediction is particularly important for optimization of systemic and/or target-site exposure of new molecular entities. The objective is to characterize the physicochemical space associated with hepatobiliary transport and rat BE and to develop in silico models. BE of 123 in-house compounds was obtained using the bile-duct cannulated (BDC) rat model. Human and rat hepatic uptake transporters (hOATPs and rOatp1b2) and transporters involved in biliary transport and rat BE (%BE) were identified using transfected cells. Furthermore, the datasets were extended by adding BE of 163 compounds and 97 organic anion transporting polypeptide (OATP) substrates from the literature. Approximately 60% of compounds showing percentage of BE (%BE) ≥ 10 are anions, with mean BE of anions (36%) more than 3-fold higher than that of nonacids (11%). Compounds with %BE ≥ 10 are found to have high molecular mass, large polar surface area, more rotatable bonds, and high H-bond count, whereas the lipophilicity and passive membrane permeability are lower compared with compounds with %BE < 10. According to statistical analysis and principal component analysis, hOATPs and rOatp1b2 substrates showed physicochemical characteristics that were similar to those of the %BE ≥ 10 dataset. We further build categorical in silico models to predict rat BE, and the models (gradient boosting machine and scoring function) developed showed 80% predictability in identifying the rat BE bins (%BE ≥ 10 or < 10). In conclusion, the significant overlap of the property space of OATP substrates and rat BE suggests a predominant role of sinusoidal uptake transporters in biliary elimination. Categorical in silico models to predict rat BE were developed, and successful predictions were achieved.

Introduction

Biliary excretion (BE) is a major elimination pathway for many drugs and/or their metabolites and has a significant impact on exposure and pharmacological and toxicological effects of drugs (Gregus and Klaassen, 1987; Fagerholm, 2008). Estimating its contribution to the total systemic clearance is thus extremely valuable for predicting clinical pharmacokinetics and understanding the possible mechanisms of hepatobiliary toxicity and potential drug-drug interactions (DDIs) (Lavé et al., 2009). Many in vitro and in situ tools have been established to understand the hepatobiliary disposition of drugs. However, limited information exists on the ability of these models to predict human biliary disposition, primarily because of the absence of systematic clinical studies measuring the rate and extent of biliary elimination of drugs (Roberts et al., 2002; Ghibellini et al., 2006; Lavé et al., 2009). Therefore, despite the species differences and potential difficulties in extrapolating BE, the bile-duct-cannulated (BDC) rat model is often used for assessing biliary elimination (Yang et al., 2009; Luo et al., 2010). Unfortunately, this model is less amenable for generating data for a large number of compounds to support drug design efforts in the discovery setting, because of time-consuming and complicated protocols. Consequently, a computational model for predicting rat BE could significantly reduce laboratory efforts and expedite the decision process (Yang et al., 2009; Chen et al., 2010; Luo et al., 2010). Furthermore, such a model could enable scientists to determine the potential for BE of virtual compounds, thereby helping to prioritize synthetic efforts in drug discovery programs.

Multidrug resistance-associated protein (MRP) 2, multidrug resistance 1, and breast cancer resistance protein (BCRP) are expressed on the hepatic canalicular membrane and are thought to be responsible for the BE of drugs and metabolites (Lai, 2009; Kusuhara and Sugiyama, 2010). Organic anion-transporting polypeptides (hOATP1B1, hOATP1B3, hOATP2B1) and organic cation transporter 1 (OCT1) on the sinusoidal...
membrane are involved in the hepatic uptake of several drugs (Shitara et al., 2006; Fahrmaier et al., 2010; Fenner et al., 2012). Among these, OATP1B1 and OATP1B3 are selectively expressed in the human liver and exhibit reasonably broad substrate specificities, suggesting the importance of these transporters in the hepatic uptake of many clinically important anionic drugs.

In efforts to determine drivers for BE, molecular mass has been commonly identified as a physicochemical determinant in several preclinical species and human, where higher molecular mass compounds show a greater propensity for BE. For anions, a molecular mass threshold of 325 ± 50 for rats and 500 ± 50 for humans was suggested (Millburn et al., 1967a,b). Wright and Line (1980) studied 18 cephalosporin derivatives and noted that compounds with molecular mass over 450 undergo substantial BE in rats (Wright and Line, 1980). Despite varied molecular mass thresholds between different structural classes, a general trend is apparent within several chemical series. Yang et al. (2009) analyzed a comprehensive dataset compiled from published reports and suggested a molecular mass threshold of 400 and 475 for anions in rat and human, respectively. However, the reasons behind the molecular mass threshold are not well understood.

Kato et al. (2008) studied the substrate affinity of the cephalosporins for MRP2 and BCRP and suggested involvement of efflux pumps in molecular mass-dependent BE of β-lactam antibiotics in rats. Because of the broad substrate specificity of the canalicular efflux transporters [P-glycoprotein (P-gp), BCRP, and MRP2], we hypothesized that substrate specificity of sinusoidal uptake transporters, particularly OATPs, defines the molecular mass threshold and the general physicochemical space of BE. Furthermore, there are no systematic analyses of the physicochemical property space of compounds undergoing hepatobiliary transport.

The objectives of this study are as follows: 1) to investigate the physicochemical property space of rat BE and examine the relationship between easily accessible physicochemical properties and rat BE [percentage of BE (%BE)] and rat biliary clearance (CLb); 2) to investigate the physicochemical space of hepatic sinusoidal hOATP transporters substrates and rat Oatp1b2 (rOatp1b2) substrates and assess the physicochemical similarities relative to BE; and 3) to develop in silico models to predict rat BE. A set of 123 in-house (Pfizer, Inc., Groton, CT) compounds with diverse chemical space and therapeutic areas was evaluated in BDC rats after intravenous administration. Along with the literature compiled data, physicochemical analysis and in silico predictive models were built using a rat BE dataset of 286 compounds. The hOATP transporters substrates (123 in-house and 96 literature compounds) and the rOatp1b2 substrates (61 in-house compounds) were identified on the basis of uptake studies using transporter- and vector-transfected cell lines.

Materials and Methods

Materials. All proprietary compounds were synthesized in-house or at facilities affiliated with Pfizer Global Research and Development (Groton, CT). For certain compounds, the corresponding [3H]- or [14C]-radiolabeled materials were synthesized by the Radiochemical Synthesis Group at Pfizer Global Research and Development. Jugular vein-cannulated/BDC male and female Sprague-Dawley rats were procured from Charles River Laboratories, Wilmington, MA). Chemicals and solvents were high-performance liquid chromatography grade and were supplied by Aldrich Chemical Co. (Milwaukee, WI), Thermo Fisher Scientific (Waltham, MA), and Mallinckrodt Baker, Inc. (Phillipsburg, NJ).

In Vivo BE. The studies were conducted in accordance with the Institute of Laboratory Animal Resources (1996) and were approved by the Institutional Animal Care and Use Committee. Typically, 8-week-old jugular vein-cannulated/BDC male and female Sprague-Dawley rats were obtained and maintained for 5 days before the study. Rats (n = 2–5/gender) were administered the compounds via an intravenous bolus dose at 0.2 to 1 mg/kg. The rats were fasted overnight before the study but were allowed unlimited access to water throughout the study. Food was typically returned to the animals at 7 h after the dose. The formulation of the intravenous dose was a solution in sterile water, freshly prepared the morning of the study, and was delivered at 1 ml/kg via the jugular vein catheter. Blood samples (~0.3 ml) were collected via the jugular vein catheter at predetermined time points after the dose, and the catheter was flushed with sodium heparin (10 units/ml) after each sample. The collected blood samples were processed to obtain plasma samples, transferred to 1.2 ml polypropylene tubes, and stored at −20°C until analysis. Bile samples were collected via the bile-duct catheter immediately after dosing for several hours (typically 6–12 h). The bile volumes were measured at the end of each interval, and the aliquots were stored at −20°C until analysis.

The samples were quantified using quadruple mass spectrometer. Liquid chromatography/tandem mass spectrometry (LC/MS/MS) analysis was conducted on a Sciex Triple Quad 400 mass spectrometer (turbopry ionization source; Applied Biosystems/MDS Sciex, Foster City, CA) with a Shimadzu LC-10 high-performance liquid chromatography system (Shimadzu, Kyoto, Japan) and Gilson 215 autosampler (Gilson, Inc., Middleton, WI). The mass spectrometer was controlled by Analyst 1.4.2 software (Applied Biosystems/MDS Sciex, Carlsbad, CA). Triplicate aliquots of each sample of bile and plasma were transferred to a vial and were precipitated usually with 20 volumes of methanol containing an appropriate internal standard. The samples were vortex-mixed and centrifuged. The supernatant was analyzed using LC/MS/MS methodology with conditions optimized and validated for the individual compound. When radiolabeled material was available for quantification, triplicate gravimetric aliquots of each sample of bile and plasma were mixed with liquid scintillation cocktail (10 ml for bile and 5 ml for plasma) and were counted for 2 min in a model LS 6000 or LS 6500 liquid scintillation counter (Beckman Coulter, Fullerton, CA). The percentage of BE (%BE) was calculated on the basis of the intravenous dose, drug concentration in the bile (C bile), and the volume of the bile (V bile) collected over the collection time:

\[
\% BE = \frac{C_{\text{bile}} \times V_{\text{bile}} \times 100}{\text{Dose} \times \text{b.wt.}}
\]

Biliary clearance (CLb) was calculated from the total amount of unchanged compound excreted in the bile of the intravenously dosed animals and the plasma area under the curve (AUC) from zero to infinity, determined by noncompartmental analysis for each individual animal:

\[
\text{CLb (ml/min kg)} = \frac{\text{Total amount excreted (ng)}}{\text{AUC}_{0-\infty} (\text{ng} \cdot \text{h} / \text{ml}) \times 60 \text{ (min/h)} \times \text{b.wt. (kg)}}
\]

Hepatic OATP Transporter Uptake Assays. Human OATPs (hOATP1B1, hOATP1B3, and hOATP2B1) expressed in human embryonic kidney 293 cells were obtained from Prof. Dietrich Keppler (Deutsches Krebsforschungszentrum, Heidelberg, Germany). rOatp1b2 expressed in the Chinese hamster ovary cells were provided by Prof. Peter Meier (University of Zurich, Zurich, Switzerland). Cell culture and assay conditions used for these three major hepatic hOATPs and the rOatp1b2 substrate assays were described previously (Kaligutkar et al., 2007). Transporter-expressing or vector-transfected human embryonic kidney 293 cells were seeded onto 24-well poly-d-lysine-coated plates at a density of ~2.5 × 10^6 cells/well and were grown for 3 days. After the cells were confluent, they were washed three times with 1 ml of uptake buffer. The transporter substrate assay was initiated by incubation of the cells with 250 μM of compound (1 and 10 μM) for 2 min at 37°C. At completion of incubation, uptake was stopped by washing the cells three times with ice-cold buffer. The cells were then lysed in 1:1 methanol/water mixture containing internal standard and were processed for LC/MS/MS analysis. The hOATP1B1, hOATP2B1, hOATP1B3, and rOatp1b2 substrates’ estradiol 17β-d-glucuronide, estrone 3-sulfate, bromosulphthalin, and rosuvastatin, respectively, were used as quality controls for uptake studies. Compounds were considered to be a substrate of a transporter when the uptake ratio (uptake by transporter-transfected cells/uptake by vector-transfected cells) was ≥ 2.0.

Literature Rat BE/Clearance and OATP Transporters Substrate Datasets. A total of 163 rat BE and 48 rat biliary clearance data points were taken directly from the compilation by Yang et al. (2009). Data for all drugs or compounds in this dataset were obtained after intravenous administration in BDC rats. The list of 97 drugs or compounds, which were reported in the
the gradient boosting machine (GBM) regression and Molecular Operating Environment (MOE) two-dimensional (2D) descriptors (Chemical Computing Group, Montreal, QC, Canada). The GBM regression is an in-house implementation of the method described by Friedman (2001). The MOE descriptors were generated from MOE version 2007.09. The descriptors include Abraham descriptors, E-state keys, Kier and Hall connectivity and kappa shape indices, partial charge, pharmacophore features, and subdivided surface area descriptors. The breakpoint for the classification was set at 10% BE. The predictability of this model was assessed via a 5-fold cross-validation approach. In this approach, the data were sorted by activity (%BE), and every fifth compound was selected for a test set, whereas the remaining 80% was used to build the model. This process was repeated five times, with a different 20% selected at each time. Ultimately, five models were built, and every compound was part of a test set.

A Scoring Function Model for BE. Using cutoffs values derived from ROC analysis (Table 3), we build a scoring function based on three determinants: molecular mass, PSA, acidic pKa (Martínez-Bauer et al., 2006). For molecular mass and PSA, an increasing linear monotonic function was used with the limits set at 300 to 400 and 80 to 120, respectively. For the acidic pKa, a decreasing linear monotonic function was used with the limit at 6.4 and 8.4, respectively. Each score ranges from 0.1 to 1. The individual scores were summed. Thus, for example, a compound with a molecular mass < 300, PSA < 80, and acid pKa > 8.4 would have a summed score of 0.3. Conversely, a compound with molecular mass > 400, PSA > 120, and acid pKa < 6.4 would have a summed score of 3. Compounds with determinant values between these ranges would have a score between 0.3 and 3. A ROC analysis was then performed on the summed scores and %BE. An AUC of 0.847 was obtained, and the cutoff value was determined to be a score of 2.1.

### TABLE 2

<table>
<thead>
<tr>
<th>Properties</th>
<th>All (n = 286)</th>
<th>Acids (n = 116)</th>
<th>Nonacids (n = 170)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear Regression Slope (Intercept)</td>
<td>Correlation Coefficient</td>
<td>Linear Regression Slope (Intercept)</td>
</tr>
<tr>
<td>Molecular mass, Da</td>
<td>0.05 (−0.03)</td>
<td>0.29</td>
<td>0.15 (−0.50)</td>
</tr>
<tr>
<td>cLogP</td>
<td>−1.32 (+24.85)</td>
<td>−0.13</td>
<td>−1.76 (40.98)</td>
</tr>
<tr>
<td>cLogD&lt;sub&gt;7.4&lt;/sub&gt;</td>
<td>−3.02 (+24.87)</td>
<td>−0.34</td>
<td>−2.57 (36.69)</td>
</tr>
<tr>
<td>Permeability, 10&lt;sup&gt;−6&lt;/sup&gt; cm/s</td>
<td>−1.24 (+27.84)</td>
<td>−0.35</td>
<td>−2.56 (46.37)</td>
</tr>
<tr>
<td>Hydrogen bond acceptors</td>
<td>2.31 (+3.22)</td>
<td>0.34</td>
<td>4.88 (−1.71)</td>
</tr>
<tr>
<td>Hydrogen bond donors</td>
<td>4.00 (+10.60)</td>
<td>0.31</td>
<td>8.6 (13.21)</td>
</tr>
<tr>
<td>polar surface area, Å&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.20 (−0.42)</td>
<td>0.42</td>
<td>0.36 (−5.70)</td>
</tr>
<tr>
<td>%Relative PSA</td>
<td>0.55 (+8.14)</td>
<td>0.20</td>
<td>0.23 (29.89)</td>
</tr>
<tr>
<td>Andrews’ binding energy, kcal/mol</td>
<td>0.48 (+13.11)</td>
<td>0.22</td>
<td>2.01 (9.18)</td>
</tr>
<tr>
<td>Rotatable bonds</td>
<td>2.16 (+6.89)</td>
<td>0.30</td>
<td>4.78 (3.70)</td>
</tr>
<tr>
<td>%Relative RB</td>
<td>8.65 (+9.32)</td>
<td>0.20</td>
<td>13.16 (16.49)</td>
</tr>
</tbody>
</table>
Statistics. Standard statistical tests were used to analyze the differences in the pharmacokinetic parameters and physicochemical properties of various data subsets. The parametric t-test (unpaired, two-tailed, unequal variance) was used to determine the significance. However, because the distributions of some properties are skewed away from normality, data were also analyzed by the nonparametric Mann-Whitney test (two-tailed) (Varma et al., 2009, 2010b).

Results

Physicochemical Space of Rat BE. The dataset (Supplemental Table 1) was mined for trends between physicochemical properties and rat %BE values. The ionization state distribution of the compounds that showed BE in rat suggested that approximately 60% of the compounds showing %BE $\leq 10$ of intravenous dose are acids (Fig. 1A). The mean %BE of acids (36%) is more than 2-fold higher ($p < 0.01$) than that of other ionization states (Fig. 1B). Furthermore, above the mean value of rat %BE of all acids (%BE = 36), only approximately 27% of compounds are nonacids (bases, neutrals, and zwitterions). No distinct differences were noted in the distribution and mean BE for bases, neutrals, and zwitterions.

Because acids are predominantly excreted in bile, the whole dataset was further divided to acid ($n = 116$) and nonacid ($n = 170$) subsets for further trend analysis. In general, no single physicochemical property could be considered uniquely predictive of %BE. Nevertheless, trends were noted in which mean %BE values showed significant relationship with molecular mass, cLogD7.4, PSA, number of RBs, and Andrews’ binding energy (Fig. 2A; Tables 1 and 2). On average, molecular mass, polar descriptors (PSA, RPSA, and hydrogen bond count), number of RBs, and Andrews’ energy increased with an increase in %BE for both acidic and nonacidic compounds. In contrast, mean cLogP, cLogD7.4, and membrane permeability showed an inverse relationship with BE.

ROC curve analysis was performed to determine the relationships and define the cutoff values for some common physicochemical properties. There is a notable molecular mass cutoff at approximately 400, where 80% of all compounds with %BE $< 10$ are below this threshold value (true negatives). However, only 62% of all compounds with molecular mass $\geq 400$ showed %BE $< 10$ (true positives). A considerable improvement in true positives was observed when the same molecular mass cutoff was examined for only acids. Approximately 84% of acids excreted in bile showed molecular mass above 400; however, the true negatives for acids were reduced to

---

Supplementary Table 1: Results of ROC analysis for defining cutoff values of physicochemical properties, to predict compounds with negligible ($< 10\%$) or significant ($\geq 10\%$) rat BE.

<table>
<thead>
<tr>
<th>Properties</th>
<th>All Compounds ($n = 286$)</th>
<th>All Compounds with Molecular Mass $\geq 400$ ($n = 174$)</th>
<th>All Acids ($n = 116$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROC Curve AUC</td>
<td>Cutoff*</td>
<td>ROC Curve AUC</td>
<td>Cutoff*</td>
</tr>
<tr>
<td>Molecular mass, Da</td>
<td>0.72</td>
<td>400</td>
<td>0.81</td>
</tr>
<tr>
<td>cLogD7.4</td>
<td>0.67</td>
<td>2.9</td>
<td>0.79</td>
</tr>
<tr>
<td>Permeability, $10^{-6}$ cm/s</td>
<td>0.78</td>
<td>2.8</td>
<td>0.73</td>
</tr>
<tr>
<td>PSA, Å$^2$</td>
<td>0.77</td>
<td>85</td>
<td>0.73</td>
</tr>
<tr>
<td>%Relative PSA</td>
<td>0.64</td>
<td>19</td>
<td>0.76</td>
</tr>
</tbody>
</table>

* When ROC curve did not yield a point with both sensitivity and specificity above 0.8, the cutoff was assigned to the value that showed sensitivity of 0.8.

---

FIG. 1. A, ionization state distribution of compounds showing BE of $<10\%$ and $\geq 10\%$ in BDC rat model. B, ionization state versus rat BE for 286 compounds. Mean $\pm$ 95% CI is depicted for each bin. C, ionization state distribution of rat Oatp1b2 and human OATP transporters (hOATP1B1, hOATP1B3, hOATP2B1) substrates. Red, anions; blue, cations; yellow, neutrals; green, zwitterions.
Trends for biliary clearance suggested that only molecular mass and Andrews’ binding energy are important determinants for both acidic and nonacidic compounds (Fig. 2B). Although no significant relationship was observed for nonacids, the number of RBs showed a positive trend with biliary clearance for acids. In general, 76% of compounds with molecular mass ≥ 400 and an arbitrary cLogD7.4 cutoff at 2.0 were not significantly excreted in bile. In contrast, approximately 89% (27 of 30) compounds with molecular mass < 400 and cLogD7.4 ≥ 2.0 were not significantly excreted in bile. Likewise, data were also analyzed as a function of molecular mass and polar descriptors (Fig. 3B). Because an expected correlation exists between PSA or number of H-bonding groups and the molecular mass, RPSA (PSA/molecular mass%) was calculated to normalize the molecular mass influence. In general, 76% of compounds with molecular mass ≥ 400 and RPSA ≥ 20% showed BE in rat, whereas 90% (40 of 45) of compounds with molecular mass < 400 and RPSA < 20% were not excreted significantly (%BE < 10) in bile. Three of the remaining five compounds are quaternary amines with molecular mass < 300 and zero PSA. A scattered plot of molecular mass versus membrane permeability suggested that approximately 74% of the high molecular mass compounds with permeability < 2.5 × 10⁻⁶ cm/s were excreted in bile, whereas only 13% of low molecular mass and high permeability compounds showed BE (Fig. 3C).

Physicochemical Space of hOATPs and rOatp1b2 Transporters Substrates. From an analysis of 219 diverse compounds with hOATP substrate data (Supplemental Table 2), it is evident that ionization state plays a key role in substrate interaction with OATP transporters expressed on the hepatic sinusoidal membrane. Acidic compounds form the majority of the hOATP transporters substrates, with 130 of 219 (~60%) compounds being acids (Fig. 1C). Neutrals compose 26% of the substrates, whereas basic compounds interact the least with hOATPs. The mean molecular mass of hOATP substrates is 476,
which is higher and lower than the subset of $\%BE < 10$ and $\%BE \geq 10$, respectively (Table 1). However, the mean and median values for cLogP, cLogD7.4, transcellular permeability, and polar descriptors of hOATP substrates are similar to those of the compounds showing $\%BE \geq 10$ ($p > 0.05$). In contrast, the mean number of RBs for OATP substrates is similar to that of compounds with low rat $\%BE$. Analysis of rOatp1b2 substrates suggested physicochemical trends similar to those of hOATP substrates and the compounds with rat $\%BE \geq 10$ (Fig. 4; Table 1).

Scattered plots of molecular mass versus other physicochemical properties show distinct trends (Fig. 4). Whereas approximately 75% of hOATP transporters substrates have molecular mass $\geq 400$, the majority of the substrates have cLogD7.4 $< 2.0$ and RPSA $\geq 20\%$, the arbitrary cutoff values defined on the basis of the ROC analysis of the rat $\%BE$ dataset. Likewise, the majority of the rOatp1b2 substrates have molecular mass $\geq 400$, cLogD7.4 $< 2.0$, and RPSA $\geq 20\%$. The majority of the hOATP substrates and the rOatp1b2 substrates fall in the space of compounds that showed significant BE.

**Comparison between Rat BE and OATP Transporters Substrate Datasets.** To understand the similarities between rat $\%BE$ and uptake transporters substrate datasets, we performed a principal component analysis (PCA) on the datasets on the basis of six molecular descriptors, which included molecular mass, cLogD7.4, HBA, HBD, RPSA, and RB. These properties were chosen because of their significant correlation with rat BE (Fig. 2). Principal components (PCs) are linear combinations of the original variables, where the first component, PC1, attempts to maximally explain the residual variance data and PC2 tries to maximally explain the residual variance not explained by PC1. PCA showed that 99.6% of the variance is explained in the space of the first two PCs, and there is a notable overlap between the regions of chemical space occupied by the compounds in the $\%BE \geq 10$ subset and the hOATPs and rOatp1b2 substrate datasets (Fig. 5). On the contrary, a number of the compounds with $\%BE < 10$ are outside the space occupied by the compounds in the other three datasets. Although PCA is rather qualitative, it provides confirmation on the overlapping property space between hepatic sinusoidal uptake and BE.

**In Silico Models for Rat BE.** In evaluating options for building in silico models, numerous approaches were pursued—multiple linear regressions and continuous and categorical models. Multiple
linear regression [quantitative structure-activity relationship (QSAR)] models for BE have been reported previously (Yang et al., 2009; Chen et al., 2010). However, the descriptors used vary from model to model, and thus, it is unlikely that the equations are transferable from one dataset to another. It is appreciated that BE is a complex phenomenon with involvement of multiple processes. Hence, it seems unlikely that a widely applicable continuous and quantitative model could be developed. Furthermore, we envision that the purpose of an in silico BE model is to assist research teams in assessing the pursuit of a costly in vivo rat biliary study. Accordingly, we focused on categorical models with the cutoff at 10% of dose excreted in the bile.

**A GBM Model for Rat BE.** A binary classification model using a GBM algorithm and 183 MOE descriptors was built and validated via a 5-fold cross-validation approach. The breakpoint is set at 10% BE. The results presented in Fig. 6A are a composite of the 5-fold cross-validation. Overall, the model performed well, with a positive predictive value of 0.79 (100 of 127), a negative predictive value of 0.81 (129 of 159), and a kappa of 0.60. This indicates that given a compound from this dataset, the model can predict whether the compound would have rat BE of greater or less than 10% with an accuracy of approximately 80%.

**BE Scoring Function.** A scoring function based on three determinants, molecular mass, PSA, and acidic pKa, was built. The breakpoint for the scoring function was determined to be 2.1, meaning that compounds with a score $\geq 2.1$ were predicted to have $\%$BE $\geq 10$ and vice versa for compounds with a score $< 2.1$. Shown in Fig. 6B are the results from this scoring function. The positive predictive value was 0.72 (103 of 143), the negative predictive value was 0.81 (116 of 143), and the kappa was 0.53. Although the predictive and kappa values for this scoring function model are lower than the values for the GBM model, it is nevertheless a good predictive model given that the scoring function is built on three descriptors. Although acidic compounds show a higher propensity to be excreted in the bile, there are compounds in other ionic classes that also have $\%$BE $\leq 10%$. The three descriptors used in the scoring function do not adequately capture these compounds. Shown in Fig. 7 are the percentage correct and incorrect prediction separated by ionic class. For the GBM model, the percentage correct is uniform across all ionic classes (Fig. 7A), whereas for the scoring function model, the percentage correct for bases is only 64% (Fig. 7B).

**Discussion**

In vivo studies with the BDC rat model is often used to estimate the $\%$BE during the discovery phase. We analyzed the interrelation of widely used physicochemical properties and rat $\%$BE and further developed quantitative in silico models to predict BE. Physicochemical trends in relation to the rat biliary clearance were also evaluated. In addition, we studied the property space of the substrates of hOATPs and rOatp1b2 sinusoidal uptake transporters. Significant overlap was
noted between the substrates of sinusoidal uptake transporters and the compounds showing considerable BE (%BE ≥ 10) in rat.

The dominant molecular features for BE in rat, according to the current dataset, are ionization state, molecular mass, lipophilicity, and polar descriptors (PSA and RPSA). Higher values in molecular mass, PSA, and hydrogen bonding ability are clearly overrepresented among the compounds with significant BE (%BE ≥ 10), whereas acidic compounds are predominantly excreted in bile. Furthermore, mean %BE of acids is 2- to 6-fold higher than that of nonacids (Fig. 1B). The dataset was thus divided into acids and nonacids and analyzed for relationship between physicochemical properties and %BE. For acids, a significant relationship with correlation coefficient above 0.5 was observed for molecular mass, HBA, PSA, and Andrews’ binding energy; however, such relationships with nonacids yielded relatively low correlation coefficient values (Table 2). The physicochemical properties associated with extensive rat BE are apparently linked to the low passive membrane permeability (Varma et al., 2010b). For example, acids are generally ionized at physiological pH (7.4) resulting in low permeability compared with bases, which exist in a relatively un-ionized form. It is also intuitively obvious that the lipophilicity and polar descriptors are positively and inversely related to passive permeability, respectively. Taken together, trends noted here complement the hypothesis that compounds with high passive transmembrane transport have an increased potential of redistribution from hepatocytes back to blood and therefore are less likely to be eliminated in bile.

Molecular mass has been commonly identified as a physicochemical determinant in several preclinical species and humans, where higher molecular mass compounds show a greater propensity for BE (Millburn et al., 1967a,b; Wright and Line, 1980; Yang et al., 2009). With this dataset, we noted that approximately 84% of acids excreted in bile show molecular mass above 400 Da. However, the true negatives for acids were considerably lower than 80%, suggesting that this single descriptor may not be good enough to predict rat BE bins. Nevertheless, molecular mass, along with PSA and membrane permeability, emerged as a distinct feature for acids, with significant ROC curve AUC (Table 3).

Drug in bile is usually more concentrated than in hepatocytes and the plasma; thus, passive transport from hepatocytes to bile is expected to be negligible. Consequently, BE represents an active process involving efflux transporters embedded in the canalicular membrane of hepatocytes. The findings that ionization state, size, and polarity are important in the BE might suggest that these properties are closely associated with molecular interaction with at least the transporters mediating canalicular efflux (multidrug resistance 1, MRP2, and BCRP). However, the general physicochemical characteristics of the ATP-binding cassette transporters substrates were found to be associated with lipophilicity (Matsson et al., 2009). The most accepted model of substrate interaction with P-gp suggests ligand partitioning into the plasma membrane before accessing the transporter; thus, lipophilicity is proposed to be involved in the initial step of partitioning and membrane accumulation (Varma et al., 2003; Aller et al., 2009). Likewise, lipophilicity was found to be important for interaction with BCRP, as well as for MRP2 (Gandhi and Morris, 2009; Matsson et al., 2009; Varma et al., 2010a). These distinct physicochemical features between the compounds showing BE in rat and ATP-binding-cassette transporter substrates indicate the importance of other processes involved in the hepatobiliary transport. It should be noted that drugs with high lipophilicity are also prone to cytochrome P450-mediated metabolism, and thus, the extent of BE could be smaller (Obach et al., 2008; Varma et al., 2010b).

We hypothesized that sinusoidal uptake transporters play a determining role in BE and therefore analyzed the physicochemical space of hOATPs and rOatp1b2 substrates and compared the general characteristics with rat BE (%BE < 10 and %BE ≥ 10) datasets. OATPs substrates are predominantly anionic compounds with typically high molecular mass, and the majority of primary physicochemical properties are not significantly different (p > 0.05) from the dataset of compounds with %BE ≥ 10 (Fig. 4; Table 1). PCA further confirmed the physicochemical space overlap between the uptake transporter substrates and the %BE ≥ 10 dataset while unveiling a distinct difference from the %BE < 10 dataset (Fig. 5). rOatp1b2 substrates in the current dataset are confined to a small physicochemical space and significantly overlap with the hOATPs substrates and %BE ≥ 10 datasets. This significant overlap of the property space of rat and human uptake transporters substrates and rat %BE ≥ 10 dataset suggests a predominant role of sinusoidal uptake transporters in biliary elimination. Overall, the physicochemical features of BE are

**Fig. 5.** PCA of compounds with rat %BE < 10 (green squares), rat %BE ≥ 10 (red circles), rOatp1b2 substrates (yellow triangles), and the hOATP transporter substrates (blue circles). The analysis was performed with six physicochemical properties (molecular mass, cLogD7.4, HBA, HBD, percentage of relative PSA, and RB). Dotted lines represent arbitrary boundary between the compounds with rat %BE < 10 and rat %BE ≥ 10. The PCs 1 and 2 represent up to 99% variance in the data.

**Fig. 6.** Truth table of model performance for categorical prediction of rat BE by in silico model using GBM algorithm and 183 MOE descriptors (A) and by the scoring function model using three descriptors (molecular mass, acidic pKa, PSA) (B).
associated with a balance of limited passive permeability and hepatic metabolism and enhanced interaction with hepatic uptake as well as efflux transporters.

Based on apparent correlation between intestinal permeability and extent of drug metabolism, the biopharmaceutics drug disposition classification system has been proposed (Wu and Benet, 2005; Benet et al., 2008, 2011; Benet, 2009). Accordingly, the major route of elimination of high-permeability (class I and II) drugs is metabolism, predominantly cytochrome P450-mediated, whereas the major route of elimination for the poorly permeable (class III and IV) drugs is renal and/or BE of unchanged drug (Benet et al., 2011; Varma et al., 2012). This is generally consistent with our observations that permeability is inversely related to rat BE, and the compounds with high molecular mass (>400 Da) and low permeability (<2.5 × 10⁻⁶ cm/s) are considerably eliminated in bile. As noted in our previous (Varma et al., 2009) and the current study, low-permeable compounds with low molecular mass are primarily eliminated in urine, whereas larger size compounds are excreted in bile. The apparent molecular mass cutoff noted for BE may indicate that the potential substrate specificity of the transporters involved in hepatobiliary transport is associated with molecular mass. The fact that canalicular efflux transporters have a much wider substrate specificity implies that the secretion across canalicular membrane is not selective to a certain molecular size (Varma et al., 2005, 2010a; Aller et al., 2009; Gandhi and Morris, 2009; Matsson et al., 2009). However, our analysis using a sizable dataset showed that hOATP and rOatp1b2 substrates tend toward large molecular mass (Fig. 4; Table 1). Previously, using a dataset of human renal clearance for 391 compounds, we observed that the bases and acids (n=116) are extensively excreted in bile, and their overlapping physicochemical space between the datasets of BE (%BE ≥ 10) and the OATP transporters substrates conforms several premises. First, physicochemical properties responsible for interaction with canalicular efflux transporters are at least partially similar to those of sinusoidal uptake OATP transporters. Second, active hepatic uptake via OATPs or other sinusoidal uptake transporters is a prerequisite for a compound to be eliminated significantly in bile. Finally, sinusoidal active uptake clearance may be a rate-limiting step for biliary clearance of compounds with significant BE.

Although scatter plots of physicochemical properties and BE show strong trends for those compounds that reside in particular quadrants, they are less predictive for compounds in the other two quadrants (Fig. 3). Consequently, more comprehensive in silico models were pursued. Although QSAR models for BE with smaller datasets have been published (Yang et al., 2009; Chen et al., 2010), using our large collection of rat BE dataset, we were unable to obtain a meaningful QSAR model when limiting the number of 2D and physicochemical descriptors to 10 or fewer. For example, the most convincing QSAR model we arrived at for acids [based on a set of five 2D MOE descriptors (Chemical Computing Group) and ACD physicochemical descriptors (Advanced Chemistry Development) using JMP 9.0.1 (SAS, Cary, NC) for step-wise analysis], resulted in a correlation coefficient of only 0.58 (data not shown). Although further approaches using three-dimensional descriptors may help establish improved models, because of the complexity and multiplicity of processes involved in determining the extent of BE, it is likely that the predictability of quantitative models based on few 2D descriptors is limited. However, two accurate qualitative computational models for the prediction of BE were developed—a classification model and a novel scoring function. The GBM model was able to correctly classify %BE ≤ 10 or %BE ≥ 10 categories for a test set with an accuracy of approximately 80% for both categories (Fig. 6A). In addition, an easily interpretable scoring function that uses only three descriptors (molecular mass, PSA, and acidic pKa) was generated (Fig. 6B). This model correctly classified %BE ≤ 10 or %BE ≥ 10 with an accuracy of 81 and 72%, respectively. This scoring function model has the advantage of being more transparent than the GBM model. The predictability of the latter rules-of-thumb model may be valuable for rapid interpretation in the early drug discovery settings.

BE often contributes more to the elimination of drugs in rat than in human, possibly because of the higher expression levels of the hepatobiliary transporters and high bile flow of rats (Lai, 2009). According to the emerging absolute protein quantification data, MRP2, BCRP, and bile salt export pump are relatively highly expressed in rat compared with human. Although we note the importance of sinusoidal

![Figure 7](https://example.com/fig7.png)

**Fig. 7.** Predictability of the in silico model using GBM algorithm (A) and the scoring function model (B), as a function of ionization state. Gray, percentage compounds correctly predicted; white, percentage of mispredicted compounds.
uptake transporters in biliary clearance, species differences in the expression of major sinusoidal uptake transporters may also contribute to the differences in BE and clearance. Although the species differences need further investigation, analyzing the rat and human BE data reported by Yang et al. (2009), we noted that the extent of BE (%BE) is reasonably similar between the two species, with few exceptions that included predominantly β-lactam antibiotics, which showed higher %BE in rat. Taken together, it is expected that the physicochemical trends and the in silico models reported here can provide general consensus on the extent of BE in humans, which otherwise is rarely determined in clinical development.

The current data analysis suggests the potential involvement of not only canaliculal efflux pumps but also sinusoidal uptake transporters in BE. Although no comprehensive clinical reports exist, preclinical studies and the physiology-based pharmacokinetic modeling and simulations indicate that the inhibition of canaliculal efflux transporters can lead to hepatotoxicity due to accumulation of endogenous substrates and/or drugs and metabolites in the hepatocytes. However, several clinical DDI studies demonstrated increased systemic drug exposure when sinusoidal uptake transporters are inhibited (Niemi et al., 2011). Therefore, for compounds extensively excreted in bile, further investigation into the potential DDIs and transporter pharmacogenomics at the level of both sinusoidal and canaliculal membrane transport is needed to ensure drug safety and efficacy. Early prediction of BE by the observed distinct physicochemical profiles and the in silico models may be helpful in assessing the potential for hepatobi- lary transporter-mediated DDIs. As is evident, high molecular mass (>400 Da), hydrophilic, and low permeable anionic drugs such as rosuvastatin and pravastatin are primarily excreted in bile using OATPs and canaliculal efflux pumps, and a high degree of clinical DDIs are noted for these drugs when coadministered with OATPs inhibitors (Kalliokoski and Niemi, 2009; Karlgren et al., 2012).

In summary, the size and chemical diversity of the datasets put together here facilitated an extensive examination of the physicochemical space associated with the rat BE and hOATP and rOatp1b2 substrates. Although substrate affinity toward canaliculal efflux transporters is a prerequisite for secretion into bile, a significant overlap in the physicochemical features of compounds showing BE and the OATP substrates suggests hepatic sinusoidal uptake transport as a major determining factor. We further investigated to build a GBM-based classification model and a simple scoring function model for qualitative prediction of BE using our collection of rat BE database of 286 compounds. The scoring function model has the advantage of being more transparent than the GMB model. The predictability of these models may be valuable for rapid interpretation of clearance mechanisms in the early drug discovery settings.

Acknowledgments
We thank David B. Duignan, Iain B. Gardner, Robert Dow, Steven Hansel, and Dennis Scott for valuable suggestions regarding this work.

Authorship Contributions

Wrote or contributed to the writing of the manuscript: Varma, Chang, Lai, Feng, El-Kattan, Litchfield, and Goosen.

References


Address correspondence to: Dr. Manthena V. S. Varma, Department of Pharmacokinetics, Dynamics, and Metabolism, MS 8220-2451, Pfizer World Wide Research and Development, Pfizer, Inc., Groton, CT 06340. E-mail: manthena.v.varma@pfizer.com