Quantitative Structure-Pharmacokinetic Relationships for the Prediction of Renal Clearance in Humans^[S]

Rutwij A. Dave and Marilyn E. Morris

Department of Pharmaceutical Sciences, School of Pharmacy and Pharmaceutical Sciences, University at Buffalo, State University of New York, Buffalo, New York

Received August 18, 2014; accepted October 28, 2014

ABSTRACT

Renal clearance (CL_R), a major route of elimination for many drugs and drug metabolites, represents the net result of glomerular filtration, active secretion and reabsorption, and passive reabsorption. The aim of this study was to develop quantitative structure-pharmacokinetic relationships (QSPKR) to predict CL_R of drugs or drug-like compounds in humans. Human CL_R data for 382 compounds were obtained from the literature. Step-wise multiple linear regression was used to construct QSPKR models for training sets and their predictive performance was evaluated using internal validation (leave-one-out method). All qualified models were validated externally using test sets. QSPKR models were also constructed for compounds in accordance with their 1) net elimination pathways (net secretion, extensive net secretion, net reabsorption, and extensive net reabsorption), 2) net elimination clearances (net

secretion clearance, CL_{SEC} ; or net reabsorption clearance, CL_{REAB}), 3) ion status, and 4) substrate/inhibitor specificity for renal transporters. We were able to predict 1) CL_{REAB} ($Q^2 = 0.77$) of all compounds undergoing net reabsorption; 2) CL_{REAB} ($Q^2 = 0.81$) of all compounds undergoing extensive net reabsorption; and 3) CL_R for substrates and/or inhibitors of OAT1/3 ($Q^2 = 0.81$), OCT2 ($Q^2 = 0.85$), MRP2/4 ($Q^2 = 0.78$), P-gp ($Q^2 = 0.71$), and MATE1/2K ($Q^2 = 0.81$). Moreover, compounds undergoing net reabsorption/extensive net reabsorption predominantly belonged to Biopharmaceutics Drug Disposition Classification System classes 1 and 2. In conclusion, constructed parsimonious QSPKR models can be used to predict CL_R of compounds that 1) undergo net reabsorption/extensive net reabsorption and 2) are substrates and/or inhibitors of human renal transporters.

Introduction

Renal clearance (CL_R) is a major route of elimination for drugs with low to negligible metabolism and biliary excretion and for drug metabolites (Masereeuw and Russel, 2001; Fagerholm, 2007). Renal clearance (CL_R) is a net result of glomerular filtration, active tubular secretion and reabsorption, and passive reabsorption (Fagerholm, 2007). Active secretion and active reabsorption are transporter-mediated processes (Muller and Fromm, 2011). The International Transporter Consortium has emphasized the role of renal transporters in clinically significant drug-drug interactions (Giacomini et al., 2010). Simple allometry (Mahmood, 1998) and direct correlations (Paine et al., 2011) of human renal clearance (CL_R) from preclinical species serve as less robust predictive tools. The former approach results in underprediction of human CL_R due to interspecies differences in transporter expression, membrane localization, and substrate specificity and the availability of CL_R data limit the latter. In vitro–in vivo extrapolation approaches overcome species differences (as mentioned above) and can be used for the prediction of in vivo active renal secretion or active renal reabsorption from respective in vitro data

(Kusuhara and Sugiyama, 2009). Scaling factors such as absolute mRNA copy number and protein abundance of several transporters and enzymes have been successfully determined for use in in vitro–in vivo extrapolation of some transporter- and enzyme-activities (Uchida et al., 2011; Ohtsuki et al., 2012; Schaefer et al., 2012). However, scaling factors for the renal transporters need to be determined.

The quantitative structure-pharmacokinetic relationship (QSPKR) approach uses physicochemical and molecular properties of compounds to quantify a pharmacokinetic process of interest. This approach has an application at early stages of drug discovery because it is noninvasive and has high-throughput capacity. A high attrition rate of drug candidates during clinical trials has necessitated the development of predictive tools for determining the absorption, distribution, metabolism, and elimination of drug candidates in the early phase of drug discovery and development. Currently there has been limited application of allometric scaling and in vitro–in vivo extrapolation for the prediction of renal clearance in humans. QSPKR represents an alternative approach and is investigated in this study for the rapid prediction of renal clearance in humans. Previously, our laboratory has developed QSPKR models to predict biliary clearance and percent of administered dose excreted unchanged into bile in rats and humans (Yang et al., 2009). QSPKR models for the prediction of passive oral absorption have been deemed superior to labor-intensive in vitro and ex vivo permeability assays (Linnankoski et al., 2008). Applications of empirical and mechanism-based QSPKR models for the prediction of various

This work was supported by the National Institutes of Health National Institute on Drug Abuse [Grant DA023223]. R.A.D. was funded in part by a fellowship from Kapoor Foundation, University at Buffalo.

dx.doi.org/10.1124/dmd.114.059857.

[S] This article has supplemental material available at dmd.aspetjournals.org.

ABBREVIATIONS: BDDCS, Biopharmaceutics Drug Disposition Classification System; GFR, glomerular filtration rate; GMFE, geometric mean fold error; MATE, multidrug and toxin extrusion; MLR, multiple linear regressions; MRP, multidrug resistance-associated protein; NME, new molecular entity; OAT, organic anion transporter; OCT, organic cation transporter; PCA, principal component analysis; P-gp, p-glycoprotein; PSA, polar surface area; QSPKR, quantitative structure-pharmacokinetic relationship; ROC curve, receiver operating characteristic curve.

pharmacokinetic processes have been reviewed extensively (Xu and Mager, 2011). In the past, QSPKR models for the prediction of percent of dose excreted unchanged in urine have been moderately successful (Na'ngono Manga et al., 2003; Doddareddy et al., 2006). Varma et al. (2009) analyzed 391 compounds to relate their physicochemical properties to renal clearance in humans. The findings suggested that CL_R correlates positively with polar surface area, number of rotatable bonds, and sum of H-bond donors and acceptors and negatively with $cLog P$ and $Log D_{7.4}$. Moreover, neutral compounds predominantly undergo net reabsorption, whereas weak acids and bases undergo net secretion (Varma et al., 2009). To our best knowledge, QSPKR models for the prediction of renal clearance (CL_R) in humans have not been reported in the literature.

We hypothesized that a single QSPKR model cannot always accurately predict renal clearance of compounds that differ in their mechanism of elimination, ion status, and transporter specificity. The objectives of the present study were to 1) apply QSPKR to predict renal clearance of all compounds in humans; 2) develop QSPKR models for compounds in accordance with their I) net elimination pathways (net secretion, extensive net secretion, net reabsorption, and extensive net reabsorption), II) net elimination clearances (net secretion clearance, CL_{SEC} ; or net reabsorption clearance, CL_{REAB}), III) ion status, and IV) substrate/inhibitor specificity for the human renal transporters; and 3) implement classification methods to differentiate compounds that undergo net secretion from those that undergo net reabsorption.

Materials and Methods

Data Acquisition, Allocation, and Binning. Human renal clearance data collected for 382 drug or drug-like compounds, ranging from 0 (diazepam) to 12.5 ml/min per kg (nomifensine) from Varma et al. (2009) were used in our investigation. The fraction unbound in plasma values of these compounds ranged from 0 to 1 (Varma et al., 2009). Nine compounds from the dataset from Varma et al. (2009) were excluded because it was not possible to calculate the physicochemical descriptors for these compounds; the list is provided in the Supplemental Data. The authors had determined CL_R as the ratio of amount of administered dose excreted unchanged into the urine to the area under the plasma concentration-time curve after intravenous or oral administration of the compounds. Data were obtained from the population of healthy young volunteers, except chemotherapeutic and human immunodeficiency virus drugs where the data were only available from patient populations. Patient data were included from "the patients whose health and physiological conditions were not severely compromised with respect to total and renal clearance" (Varma et al., 2009).

First, the compounds were divided into two categories according to their mechanism of elimination: 1) net secretion and 2) net reabsorption. Mean physiologic glomerular filtration rate (GFR) of 1.78 ml/min per kg was used when necessary. Compounds with $CL_R > 1.2 \times f_u \times GFR$ and $CL_R < 0.8 \times f_u \times GFR$ were classified as those net-secreted and net-reabsorbed, respectively. This 20% tolerance value (subset 1) to separate net-secreted compounds from net-reabsorbed compounds was introduced previously (Varma et al., 2009) and was validated to ensure separation in our studies [compounds that undergo only filtration ($0.8 \times f_u \times GFR < CL_R < 1.2 \times f_u \times GFR$) would have a $CL_R \sim f_u \times GFR$]. Similarly, compounds with $CL_R > 1.5 \times f_u \times GFR$ and $CL_R < 0.5 \times f_u \times GFR$ were classified as extensively net-secreted and extensively net-reabsorbed compounds, respectively. This 50% tolerance value (subset 2) was adopted from the guidelines of the International Transporter Consortium (Muller and Fromm, 2011) to separate net-secreted compounds from net-reabsorbed compounds. For the two subsets, net secretion clearance ($CL_{SEC} = CL_R - f_u \times GFR$) and net reabsorption clearance ($CL_{REAB} = f_u \times GFR - CL_R$) were calculated. As 20% of compounds undergoing net reabsorption had $CL_R = 0$ ml/min per kg, a QSPKR model was developed for the prediction of net reabsorption clearance (CL_{REAB}) instead of CL_R , where CL_R can be calculated as $f_u \times GFR + CL_{REAB}$.

Compounds were further divided into smaller subsets according to their ion status. pK_a values for all compounds were obtained from SciFinder Scholar Database (American Chemical Society, Washington, DC). Ion status (anion,

cation, natural, or zwitterion) for compounds that undergo net secretion and net reabsorption were determined at pH 7.4 and pH 6.3, respectively.

Compounds were also classified as substrates and/or inhibitors of the following human renal transporter groups involved in active renal secretion: P-glycoprotein (P-gp), multidrug resistance associated proteins (MRP1/2/4), organic anion transporters (OAT1/3), organic cation transporter (OCT2), and multidrug and toxin extrusion (MATE1/2K) using information available on the UCSF-FDA TransPortal database (Morrissey et al., 2012). Compounds that were included in the OAT1/3 subset were both substrates and inhibitors; inhibitors alone were not included in this subset. Because of the limited number of substrates/inhibitors for an individual transporter, i.e., OAT1 or OAT3, compounds were grouped into a single subset of OAT1/3; a similar approach was used for all five transporter subsets. Owing to the overlapping specificity for the substrates and the inhibitors of OCT2 with P-gp, OCTN1/2, and MATE1/2K and the wide specificity for the substrates and the inhibitors of P-gp, CL_R was log transformed and $Log CL_R$ was predicted to ensure the universal applicability of the QSPKR models for these two subsets.

It is important to recognize that 1) given the CL_R of compounds, the extent of secretion cannot be distinguished from the extent of reabsorption; hence, compounds undergoing net reabsorption would possibly have an inherent secretion component, but the net result would predominantly be reabsorption and vice versa; 2) not all the compounds that were binned as filtered compounds ($0.8 \times f_u \times GFR < CL_R < 1.2 \times f_u \times GFR$) exclusively undergo glomerular filtration and may have the potential to undergo passive reabsorption or active secretion to some extent; 3) binning was done to separate all compounds undergoing net reabsorption (passive reabsorption) from those undergoing net secretion to achieve the statistical confidence; and (4) compounds in transporter subsets were included solely on the basis of literature evidence; in other words, QSPKR models for transporter subsets were not substitutes for the net secretion subset.

This original dataset ($N = 382$) was randomly divided into a training set ($N = 332$) and a test set ($n = 50$). Subsequently, the training set and the test set were further divided into respective subsets according to their mechanism of elimination and ion status. Because of the limited number of compounds in all transporter subsets, only the training set was generated. For the random division of compounds into training sets and test sets, 1) random number integers equal to the total number of compounds in the each subset were generated in Excel (Microsoft, Redmond, WA); 2) compounds with integers 1 to n (where n equaled to the predecided number of compounds for the test set) were separated to generate the test set; and 3) the remaining compounds made up the training set. This was repeated three times to ensure that compounds are truly selected at random and the fourth run was used for the binning of data. The number of the compounds was chosen to maximally use data for the model development process for all training sets and their respective test sets (Table 1 and Supplemental Data). All test sets were within the chemical space of the respective training sets (not shown). On average, 75% of compounds in all training and test sets had Tanimoto similarity index < 0.7 , making them structurally dissimilar (Luo et al., 2010) (not shown).

Calculation of Simple Physicochemical and Complex Molecular descriptors. Simple molecular input line entry specification strings for all compounds were obtained [PubChem substance and PubChem compound database (Bolton et al., 2008)]. Twelve simple physicochemical descriptors—molecular weight, molecular volume, $cLog P$, hydrogen bond donors and acceptors, sum of hydrogen bond donors and acceptors, polar surface area (PSA), $Log D_{(pH\ 6.3)}$, $Log D_{(pH\ 7.4)}$, number of rotational bonds, solvent accessible polar surface area, and relative surface area (PSA/solvent accessible polar surface area)—were calculated using Instant JChem version 4.1 (ChemAxon, Budapest, Hungary). Over 2500 complex molecular descriptors were calculated using QSARis version 1.2 software (SciVision-Academic Press, San Diego, CA), PaDEL-descriptor (Yap, 2011), and ADRIANA.code (Molecular Networks, Erlangen, Germany), of which 250 descriptors were used for model development after careful removal of redundant descriptors (with $> 60\%$ zero values and narrow range of values; S.D. $< 10\%$ of the mean). These two- and three-dimensional structural descriptors included: molecular connectivity chi indices, kappa shape indices, electrotopological state indices, information indices, subgraph count indices, and molecular polarizability, molecular weight, and volume.

Classification Methods. Receiver operating characteristic (ROC) curve analysis was used to determine a threshold value for each simple physicochemical property, which would distinguish compounds undergoing net

TABLE 1
Qualified QSPKR models

Name	<i>N</i>	QSPKR model	R^2_{adj}	Q^2	<i>n</i>	R^2_{test}	GMFE
Compounds that undergo net renal reabsorption ($CL_R < 0.8 \times f_u \times GFR$)							
All	121	$CL_{REAB} = 1.23 f_u - 0.04 HBD - 0.04 SaasC - 0.03 SdssC_{acnt} + 0.02 \text{Log } D_{(pH\ 7.4)} + 0.17$	0.79	0.77	20	0.67	1.69
Compounds that undergo extensive net renal reabsorption ($CL_R < 0.5 \times f_u \times GFR$)							
All	36	$CL_{REAB} = 0.52 f_u + 0.01$	0.93	0.81	10	0.98	1.01
Compounds that are substrates and/or inhibitors of renal transporters mediating active secretion							
OAT1/3	21	$CL_R = 0.61 SssCH2_{acnt} - 10.3 \text{MaxQP} - 5.50 H_{min} - 0.10 D_Z + 23.7$	0.83	0.81	—	—	—
MRP1/2/4	16	$CL_R = 5.03 f_u + 0.33 G_{max} - 0.12 D_Z + 0.27 SaasC_{acnt} + 12.6 \text{MaxNeg} - 7.81$	0.88	0.78	—	—	—
MATE1/2k	17	$CL_R = -19.7 \chi_{ch6} + 63.7 \text{MaxNeg} - 42.2 \text{Dipole} - 0.86 \text{ABSQ} + 4.41$	0.90	0.81	—	—	—
OCT2	24	$\text{Log } CL_R = 0.95 f_u - 0.44 \chi_{p9} + 0.28 \chi_{vc3} + 0.23 H_{max} - 0.58$	0.92	0.85	—	—	—
P-gp	17	$\text{Log } CL_R = 0.063 nCirc + 0.86 f_u - 0.24 G_{max} + 3.20 \text{RSA} + 1.59$	0.89	0.71	—	—	—

HBD, hydrogen bond donors; RSA, relative surface area.

secretion from net reabsorption, using the ROC Curve toolbox in SigmaPlot 11.0 (Systat Software, Inc., San Jose, CA). The recursive partitioning approach was also used to derive a decision tree that would classify compounds undergoing net secretion and net reabsorption based on their physicochemical properties using the rpart package in R 3.0.2 (Therneau et al., 2014). ROC curve analysis and recursive partitioning methods were also implemented to identify substrates and/or inhibitors of renal transporters. Lastly, principal component analysis (PCA) was used to determine the chemical space of compounds in accordance with net elimination pathways (net secretion, glomerular filtration, and net reabsorption) and ion status using PCA multivariate method in Minitab 16 Statistical Software (Minitab Inc., State College, PA). In brief, PCA was performed using 250 descriptors, and the scores of the first two resulting components were calculated and plotted as an XY scatterplot; each XY pair would then represent the chemical space of a respective compound.

Biopharmaceutics Drug Disposition Classification System Classification.

Biopharmaceutics Drug Disposition Classification System (BDDCS) provides insight into the importance of transporters in the gut and/or liver in the uptake and/or efflux of compounds, while classifying compounds into four classes according to their solubility and permeability rate/metabolism (Benet et al., 2011). As most of the compounds in our dataset were classified using the BDDCS classification system by Benet et al. (2011), we evaluated whether a trend existed between compounds' BDDCS class type and their mechanism of elimination and substrate/inhibitor specificity for renal transporters.

QSPKR Model Development and Validation. Simple linear regression method (not shown) did not establish a statistically significant linear association between CL_R and simple physicochemical properties. Independent variables (descriptors) were therefore selected into the QSPKR models using step-wise multiple linear regressions (MLR) using the step-wise MLR algorithm in Minitab 16 Statistical Software (Minitab Inc.). Step-wise MLR combines the forward selection (variable entry in the model; $P < 0.05$) and backward elimination (variable retention into the model; $P < 0.15$) of the most significant descriptors into a single step. One descriptor was added for every five compounds in the training set to avoid over-parameterization of models (Yang et al., 2009). Only those descriptors with Pearson's correlation coefficient of <0.7 were included in the final model to avoid any significant collinearity. The final model was generated in the form of:

$$Y = \sum_{i=1}^n a_i x_i + b$$

Where Y is the predicted value of CL_R , x_i are independent variables (structural and molecular descriptors), a_i are the coefficients associated with the independent variables, and b is the constant. A list of model selected independent variables and their description is given in Table 2.

Internal and external validation methods were used to evaluate predictive performance of all models (model-qualification criteria assigned a priori). The "leave-one-out" method was used for the internal validation of QSPKR models.

In this crossvalidation method, the property value (CL_R) for a single compound is excluded from the dataset, which is in turn predicted from the data for all remaining compounds using the constructed regression model. This process is repeated until all compounds in the training set are excluded once. The metric of leave-one-out is Q^2 , which is calculated using the equation:

$$Q^2 = 1 - \frac{\sum_i (y_i - y_{(i)})^2}{\sum_i (y_{(i)} - y_{mean})^2}$$

Where y_i is the predicted value for the i th compound using the model constructed with the i th compound excluded, $y_{(i)}$ is the observed value for the i th compound, and y_{mean} is the mean value of the dataset. Models that had Q^2 value of at least 0.5 (which translates to a P value of less than 0.05) were considered qualified. Moreover, the difference between values of the coefficient of determination R^2 and Q^2 was required to be <0.3 to ensure model stability (Van der Graaf et al., 1999).

All qualified models were further validated externally using the respective test sets. Where the number of compounds in the test sets was less than 10, the external validation was not performed to avoid the false-positive validation of the models. Metrics of external validation were R^2_{test} and geometric mean fold error (GMFE), where the former indicates the extent of linearity between observed and predicted values of CL_R and the latter indicates the fold-over-prediction or fold-under-prediction of CL_R values. GMFE gives equal weight to

TABLE 2
Definition of molecular descriptors selected into QSPKR models

Descriptor	Definition
SaasC	Sum of (aasC-) electro-topological states
SssCH2_acnt	Count of (-CH-) electro-topological states
SaasC_acnt	Count of (aasC-) electro-topological states
SdssC_acnt	Count of (=C<) electro-topological states
G_{max}	Maximum E-state value of an atom in a molecule
H_{min}	Maximum hydrogen E-state value of an atom in a molecule
H_{max}	Maximum hydrogen E-state value of an atom in a molecule
D_Z	The component of the displacement between the center-of-mass and center-of-dipole along the inertial Z-axis
MaxNeg	The largest negative charge over the atoms in a molecule
MaxQP	The largest positive charge over the atoms in a molecule
ABSQ	The sum of absolute values of the charges on each atom of a molecule, in electrons
Dipole	Dipole moment of the molecule
nCirc	Number of graph circuits, the total number of all cycles in the molecular graph
χ_{ch6}	Simple 6 th order chain chi index
χ_{vc3}	Valance 3 rd order cluster chi index
χ_{p9}	Simple 9 th order path chi index

both the overestimated and underestimated values (Parrott et al., 2005). GMFE was calculated using the equation:

$$GMFE = 10^{\frac{\sum_{i=1}^n \left| \log\left(\frac{Predicted}{Observed}\right) \right|}{n}}$$

A model that predicts the observed data without any bias will have a geometric mean fold-error of 1. A mean fold-error ≤ 2 is considered successful, indicating that the most of the predicted values fall within the twofold range of the observed values (between 0.5 and 2.0) (Ng et al., 2004).

Results

QSPKR Models. A QSPKR model was constructed for the training set consisting of 332 compounds ($Q^2 = 0.14$) but did not meet the criteria for internal and external validations. QSPKR models were then constructed for each of the four ion-status subsets, but all Q^2 values were < 0.5 . Thus, a single QSPKR model was not able to predict the renal clearance of all compounds and binning these compounds according to their ion-status did not improve model prediction of CL_R in all four subsets. Table 2 provides the definitions of the model selected descriptors.

Prediction of Net Reabsorption Clearance (CL_{REAB}). For this part of the analysis, two training sets comprised of compounds that undergo net reabsorption were used. The first training set ($N = 121$) included all compounds that undergo net reabsorption ($CL_R < 0.8 \times f_u \times GFR$), and the second training set ($N = 36$) included compounds that undergo extensive net reabsorption ($CL_R < 0.5 \times f_u \times GFR$).

As 20% of compounds undergoing net reabsorption had $CL_R = 0$ mL/min per kg, a QSPKR model was developed for the prediction of net reabsorption clearance (CL_{REAB}), where $CL_{REAB} = f_u \times GFR - CL_R$, using the first training set (Table 1; Fig. 1, A and B). Selected model descriptors were related to the lipophilicity of compounds. The external validation of this model suggested that this model can be used for predictive purposes. Predictive power of QSPKR models for all ion-status subsets were comparable to the base model (not shown). In accordance with Varma et al. (2009), compounds undergoing net reabsorption were predominantly neutral, suggesting passive reabsorption and about 82% of these compounds belonged to BDDCS classes 1 and 2.

A QSPKR model for the second training set ($CL_R < 0.5 \times f_u \times GFR$) (Table 1; Fig. 2, A and B) was constructed using a single

statistically significant descriptor, fraction unbound in plasma. This finding was interesting as the fraction unbound in plasma was the only rate limiting determinant of CL_{REAB} of compounds undergoing extensive net reabsorption. About 70% of the compounds belonged to BDDCS classes 1 and 2.

Prediction of Renal Clearance (CL_R) and Net Secretion Clearance (CL_{SEC}). All developed QSPKR models did not meet the criteria for internal and external validation and had poor predictive power ($Q^2 < 0.5$) (not shown). Moreover, categorizing them according to their ion status did not improve the predictive power of QSPKR models. QSPKR models were then developed for compounds that are known substrates and/or inhibitors of one or more renal transporters that mediate active secretion.

QSPKR Models for Transporter Subsets. Training sets were created for five renal transporters subsets that mediate active secretion: 1) OAT1/3 (Fig. 3A), 2) MRP1/2/4 (Fig. 3B), 3) MATE1/2K (Fig. 3C), 4) OCT2 (Fig. 4A), and 5) P-gp (Fig. 4B). QSPKR models for the prediction of CL_R of compounds that belong to specific transporter subsets are given in Table 1. About 52% of the compounds in all transporter subsets belonged to BDDCS classes 3 and 4. For the OAT1/3 subset, model selected descriptors H_{min} , MaxQP, and D_z indicate that minimum electrotopological state, ionic charge, and charge distribution, respectively, are important physicochemical determinants of CL_R of these compounds. Moreover, a positive charge of compounds (MaxQP) negatively impacts to its renal clearance. For the MRP1/2/4 subset, G_{max} , D_z , and MaxNeg indicate that maximum electrotopological state, charge distribution, and negative charge, respectively, are important physicochemical determinants of CL_R of these compounds. For the MATE1/2K subset, MaxNeg, ABSQ, and Dipole indicate that ionic charge, absolute charge, and charge distribution along with connectivity of atoms (χ_{ch6}) are important physicochemical determinants of CL_R of these compounds. For the OCT2 subset, fraction unbound in plasma, maximum electrotopological state (H_{max}), and atom connectivity (χ_{p9}) are important physicochemical determinants of CL_R of these compounds. For the P-gp subset, fraction unbound in plasma, relative surface area (relative surface area = PSA/solvent-accessible surface area), and maximum electrotopological state (G_{max}) are important physicochemical determinants of CL_R of these compounds. Similar model selected descriptors for 1) OAT1/3, MRP1/2/4, and MATE1/2K; 2) OCT2, MATE1/2K, and P-gp; and 3) MRP1/2/4 and P-gp further

Fig. 1. Compounds undergoing net reabsorption: plot of observed CL_{REAB} versus predicted CL_{REAB} of the training set ($N = 121$) (A) and plot of observed CL_{REAB} versus predicted CL_{REAB} of the test set ($N = 20$) (B). Dashed line in (A) and (B) is the line of unity.

Fig. 2. Compounds undergoing extensive net reabsorption: plot of observed CL_{REAB} versus predicted CL_{REAB} of the training set ($N = 36$) (A) and plot of observed CL_{REAB} versus predicted CL_{REAB} of the test set ($N = 10$) (B). Dashed line in (A) and (B) is the line of unity.

validate the observed overlapping specificity for the substrates and inhibitors of these transporter groups. The observed versus predicted plots of CL_R values in Fig. 3, A–C show strong agreement and have

strong predictive power of the QSPKR models for OAT1/3, MRP1/2/4, and MATE1/2K subsets. The QSPKR models for the prediction of Log CL_R of OCT2 and P-gp subsets also have strong predictive power and

Fig. 3. Plot of observed CL_R versus predicted CL_R of OAT1/3 subset ($N = 21$) (A), MRP1/2/4 subset ($N = 16$) (B), and MATE1/2K subset ($N = 17$) (C). Dashed line in (A–C) is the line of unity.

Fig. 4. Plot of observed Log CL_R versus predicted Log CL_R of OCT2 subset ($N = 24$) (A) and P-gp subset ($N = 17$) (B). Dashed line in (A) and (B) is the line of unity.

show good agreement between the observed and the predicted values, which is illustrated in Fig. 4, A and B.

For all transporter subsets, where available, K_m or K_i values were included as an independent variable (data not shown); however, these values were not selected into any QSPKR models. This suggests that the binding constant (K_m) may not be as critical as substrate specificity and transporter capacity (T_{max}) in predicting renal clearance of these compounds. However, more evidence is needed before any conclusions may be drawn. ROC curve analysis yielded threshold values with sensitivity ($<80\%$) and specificity ($<80\%$) that was not statistically significant (not shown). The decision tree derived using the recursive partitioning method had a great degree of misclassification and was not able to classify compounds with statistical significance (not shown).

Analysis of the Chemical Space of Compounds. The score plots from PCA in Fig. 5, A and B illustrate that a significant degree of overlap exists in the chemical space of compounds among all three net elimination groups and among all four ion status groups, respectively. The failure of the ROC curve analysis and recursive partitioning methods can thus be attributed to overlapping chemical spaces of compounds.

Application of QSPKR Models. If a new molecular entity (NMEs) belongs to BDDCS class 1 or 2 and is not a substrate and/or an inhibitor for any renal transporters, then the QSPKR model (general model) for all compounds undergoing net renal reabsorption can be used for the prediction of their CL_{REAB}, because extensive net reabsorption is a nested process of net reabsorption. CL_R can be calculated from the equation, $CL_R = f_u \times GFR - CL_{REAB}$. The underlying assumption is that the model predicts passive renal reabsorption of compounds, because there is no evidence for active reabsorption for these compounds. If a NME is a substrate and/or an inhibitor for one or more renal transporters, then its renal clearance can be predicted using one or more transporter models. Application of QSPKR models for the prediction of renal clearance of NMEs in humans is illustrated in Fig. 6.

Discussion

Renal clearance involves four processes. Glomerular filtration occurs at the glomerulus and it represents an ultrafiltrate of about 10% of total renal blood flow. Active secretion is a transporter-mediated process, where uptake of organic anions into proximal tubular cells is

Fig. 5. Analysis of the chemical space of compounds in accordance with their elimination pathways (A) and ion status (B).

mediated mainly by OAT1 and OAT3, and subsequent efflux into proximal tubular lumen is mediated by MRP2, MRP4, and OAT4 (Russel et al., 2002). Similarly, uptake of organic cations into proximal tubular cells is mediated mainly by OCT2, and subsequent efflux into proximal tubular lumen is mediated by P-gp, OCTN1, OCTN2, MATE1, and MATE2k (Muller and Fromm, 2011). Active renal reabsorption of compounds involves vectorial transport across proximal tubular cells from proximal tubular lumen into blood capillaries. Active renal reabsorption is mediated by proton-dependent and sodium-dependent monocarboxylate transporters (Morris and Felmlee, 2008), the ascorbic acid cotransporter (SVCT1) (Masereeuw and Russel, 2001), peptide transporters (PEPT1 and PEPT2), concentrative nucleoside transporters (CNT1 and CNT2), glucose transporters (SGLT2), the urate anion exchanger 1 (URAT1), and OAT4 (Lee and Kim, 2004), among others. Passive reabsorption takes place at the loop of Henle and at the distal tubules (Feng et al., 2010) and is influenced by urine pH, urine flow, and degree of ionization and lipophilicity of compounds. The International Transporter Consortium has recognized the importance of clinically important drug-transporter and drug-drug interactions that significantly affect exposure and disposition of drugs. QSPKR models developed in the present study enable prediction of renal clearance of compounds that interact with renal transporters.

Our dataset included compounds that undergo passive reabsorption (as there is no reported evidence for active reabsorption for these compounds) and compounds that are substrates and/or inhibitors of human renal transporters that mediate active secretion. We found the fraction unbound in plasma (f_u) to be the most important determinant of renal clearance as it appeared across multiple QSPKR models that were constructed in the present study. As expected, a single QSPKR model was unable to predict renal clearance, when their mechanism of elimination was not taken into account. We were able to predict renal clearance of compounds undergoing net reabsorption and extensive net reabsorption with statistically significant predictive power (Fig. 1, A and B and Fig. 2, A and B). Model selected descriptors for the latter relate to the lipophilicity of compounds and these compounds predominantly belong to BDDCS classes 1 and 2. Benet et al.

(2011) reported that percent of unchanged drug excreted into urine was statistically lower for compounds belonging to classes 1 and 2 than for compounds belonging to classes 3 and 4. Although, a low CL_R is not necessarily due to net reabsorption, lipophilicity of compounds is a significant determinant of extensive metabolism and distribution of compounds (Fagerholm, 2007). As well, compounds that undergo extensive metabolism are less likely to undergo biliary and renal excretion (Fagerholm, 2007; Benet et al., 2011). Moreover, compounds undergoing net reabsorption were predominantly neutral. Although compounds undergoing extensive net reabsorption were largely ionized, glomerular filtration is the rate limiting step for these compounds. The QSPKR model for extensive net reabsorption in Table 1 illustrates this point. It is important to note that variability in urine pH would affect the ionization status of compounds.

In the present study, QSPKR models—for the prediction of CL_R of compounds that were substrates and/or inhibitors for renal transporters—were also constructed with statistically significant predictive power and model performance (Table 1). Model selected variables for all transporter subsets were polar descriptors. Overlapping substrate and/or inhibitor specificity existed between OCT1/3, P-gp, and MRP1/2/4 subsets and OCT2, P-gp, and MATE1/2K subsets. However, percent contribution of these transporters to the active secretion of these compounds is not known and is beyond the scope of the present study. For all five transporter subsets, predicted values of CL_R were well within the twofold range of the observed CL_R values, suggesting that the overlapping substrate specificity of compounds for the renal transporters do not affect their model predicted CL_R values, as demonstrated in Table 3. Because of the limited number of compounds in all transporter subsets, the external validation could not be performed for the transporter subsets.

It was not possible to develop robust models (CL_R and CL_{SEC}) for compounds undergoing net secretion because these compounds had unique pathways for their secretion; accounting for drug-transporter interactions enabled us to predict CL_R of ~30% of these compounds. For the compounds undergoing net secretion, we recommend first establishing drug-transporter relationships rather than using a general QSPKR model of any nature for the prediction of renal clearance. Categorizing compounds according to their ion status did not improve predictive power of QSPKR models for 1) all compounds, 2) compounds undergoing net secretion, and 3) compounds undergoing net reabsorption. Therefore, the mechanism of elimination of compounds is statistically more important than their ion status in prediction of renal clearance. An interesting finding was that despite a great degree of overlap among the chemical space of compounds undergoing net reabsorption, glomerular filtration, and net secretion, the *in vivo* fate of these compounds was unique, which further establishes the importance of drug-transporter interactions.

To our best knowledge, the current study is the first to develop QSPKR models for the prediction of renal clearance (CL_R) for a large dataset of compounds. Paine et al. (2011) predicted ($R^2 = 0.84$) renal clearance of 36 compounds that undergo net secretion or net reabsorption in humans using direct correlation with renal clearance in dogs after correcting for species differences in plasma protein binding and kidney blood flow. The authors indicated that renal blood flow was similar between the two species, which could be a potential reason for good predictions. Simple allometry and Mahmood's renal clearance scaling methods were not able to predict renal clearance in humans from dogs and rats (Paine et al., 2011). Na'ngono Manga et al. (2003) developed a hierarchical QSAR model for urinary excretion of 100 drugs in humans to predict whether they are extensively or nonextensively metabolized (Na'ngono Manga et al., 2003). However, this model was a qualitative assessment of the extent of metabolism.

Fig. 6. Application of QSPKR models for the prediction of renal clearance of NMEs in humans.

TABLE 3

Comparison of observed and predicted renal clearance values (ml/min per kg or *log ml/min per kg) of compounds with overlapping substrate/inhibitor specificity for the renal transporters

Compound	Observed	OAT1/3	MRP2/4	P-gp	OCT2	MATE1/2K
Acyclovir	3.57	5.14	—	—	—	4.21
Adefovir	3.33	2.81	3.60	—	—	—
Cephalexin	3.54, 0.55*	—	—	—	0.68*	4.13
Chloroquine	2.29, 0.36*	—	—	—	0.07*	1.99
Cimetidine	7.90, 0.89*	6.80	—	—	0.50*	6.46
Ciprofloxacin	6.90	6.53	5.23	—	—	—
Cyclosporine	0.08, -1.12*	—	0.28	-1.21*	—	—
Dofetilide	3.12, 0.49*	—	—	—	0.45*	1.79
Famotidine	4.42, 0.64*	4.25	—	—	0.67*	6.07
Furosemide	1.70	1.24	0.84	—	—	—
Ganciclovir	4.14	3.36	—	—	—	3.54
Imipramine	0.26, -0.58*	—	—	—	-0.67*	0.38
Irinotecan	2.24, 0.35*	—	2.16	0.37*	—	—
Metformin	7.40, 0.86*	—	—	—	0.94*	5.81
Methotrexate	1.70	1.39	2.78	—	—	1.46
Pindolol	3.90, 0.59*	—	—	—	0.47*	4.83
Procainamide	4.50, 0.65*	—	—	—	0.61*	4.63
Quinidine	-0.22*	—	—	-0.12*	-0.17*	—
Ranitidine	7.40, 0.86*	—	—	—	0.90*	7.15
Tenofovir	2.70	4.76	3.83	—	—	—
Topotecan	5.20, 0.71*	—	5.16	0.54*	—	4.99
Verapamil	0.54, -0.27*	—	—	-0.37*	-0.18*	0.58
Vinblastine	2.63, 0.42*	—	3.10	0.43*	—	—
Zidovudine	5.67, 0.75*	5.16	—	—	0.62	—

*Log CL_R values, as predicted for P-gp and OCT2 subsets.

Doddareddy et al. (2006) performed partial least squared analysis ($Q^2 = 0.76$) to predict the percent of administered dose excreted unchanged in urine of 130 central nervous system and noncentral nervous system compounds. A step-wise MLR model to predict the log unbound renal clearance of 47 acidic compounds ($R^2 = 0.62$, $Q^2 = 0.51$) was recently reported (Zhivkova and Doytchinova, 2013); however, in our analyses, a statistically significant QSPKR model could not be constructed for all compounds that were acids ($N = 98$, $Q^2 = 0.16$) at the physiologic pH (not shown). Varma et al. (2009) analyzed 391 compounds to relate their physicochemical properties to renal clearance in humans. They found that CL_R correlates positively with polar surface area, number of rotatable bonds, and sum of H-bond donors and acceptors negatively with cLog P and Log D_{pH} = 7.4 (Varma et al., 2009); however, no statistical tests were performed to establish relationships between CL_R and physicochemical properties of compounds. Using their and other literature findings as our baseline, we extended these efforts in quantitatively predicting renal clearance using simple physicochemical and molecular properties of compounds.

Inasmuch as the compounds undergoing transporter mediated renal clearance have chemical space that is overlapping with that of the compounds undergoing net reabsorption, it is not possible to determine the fate of the compound a priori from its structure. Despite the significant overlap of chemical space, compounds display unique elimination pathways in vivo (Fig. 5A). Although CL_R is lower for the neutral compounds and higher for the ions (Varma et al., 2009), the chemical space of all compounds is similar (Fig. 5B). This emphasizes the role of the drug-transporter interactions and the physiologic conditions for the prediction of renal clearance in humans. The QSPKR models constructed in the present study are parsimonious and readily implementable (Fig. 6) and may thus find their application in drug discovery and development process.

In summary, QSPKR models were developed for the prediction of renal clearance in humans using a wide array of structural and molecular descriptors. The fraction unbound of compounds in plasma,

electrotopological state, dipole and quadrupole moments, size and charge of molecules, and simple physicochemical chemical properties were deemed important in prediction of CL_R. The constructed parsimonious QSPKR models can be used to predict CL_R of 1) compounds that undergo net reabsorption/extensive net reabsorption and 2) substrates and/or inhibitors of renal transporters. Moreover, compounds undergoing net reabsorption/extensive net reabsorption predominantly belonged to BDDCS classes 1 and 2. QSPKR models developed in this study have application in early clinical drug development, providing predictions of renal clearance.

Acknowledgments

The authors thank Dr. Yash Gandhi for his intellectual insight during QSPKR model development.

Authorship Contributions

Participated in research design: Morris, Dave.

Conducted experiments: Dave.

Contributed new reagents or analytic tools: Dave.

Performed data analysis: Dave.

Wrote or contributed to the writing of the manuscript: Morris, Dave.

References

- Benet LZ, Broccatelli F, and Oprea TI (2011) BDDCS applied to over 900 drugs. *AAPS J* **13**: 519–547.
- Bolton EE, Wang Y, Thiessen PA, and Bryant SH (2008) PubChem: integrated platform of small molecules and biological activities. *Ann Reports Computational Chem* **4**:217–241.
- Doddareddy MR, Cho YS, Koh HY, Kim DH, and Pae AN (2006) In silico renal clearance model using classical Vol surf approach. *J Chem Inf Model* **46**:1312–1320.
- Fagerholm U (2007) Prediction of human pharmacokinetics - renal metabolic and excretion clearance. *J Pharm Pharmacol* **59**:1463–1471.
- Feng B, LaPerle JL, Chang G, and Varma MV (2010) Renal clearance in drug discovery and development: molecular descriptors, drug transporters and disease state. *Expert Opin Drug Metab Toxicol* **6**:939–952.
- Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KL, Chu X, Dahlin A, Evers R, Fischer V, and Hillgren KM, et al.; International Transporter Consortium (2010) Membrane transporters in drug development. *Nat Rev Drug Discov* **9**:215–236.
- Kusuha H and Sugiyama Y (2009) In vitro-in vivo extrapolation of transporter-mediated clearance in the liver and kidney. *Drug Metab Pharmacokinet* **24**:37–52.
- Lee W and Kim RB (2004) Transporters and renal drug elimination. *Annu Rev Pharmacol Toxicol* **44**:137–166.
- Linnankoski J, Ranta VP, Yliperttula M, and Urtti A (2008) Passive oral drug absorption can be predicted more reliably by experimental than computational models—fact or myth. *Eur J Pharm Sci* **34**:129–139.
- Luo G, Johnson S, Hsueh MM, Zheng J, Cai H, Xin B, Chong S, He K, and Harper TW (2010) In silico prediction of biliary excretion of drugs in rats based on physicochemical properties. *Drug Metab Dispos* **38**:422–430.
- Mahmood I (1998) Interspecies scaling of renally secreted drugs. *Life Sci* **63**:2365–2371.
- Masereeuw R and Russel FG (2001) Mechanisms and clinical implications of renal drug excretion. *Drug Metab Rev* **33**:299–351.
- Morris ME and Felmlee MA (2008) Overview of the proton-coupled MCT (SLC16A) family of transporters: characterization, function and role in the transport of the drug of abuse gamma-hydroxybutyric acid. *AAPS J* **10**:311–321.
- Morrissey KM, Wen CC, Johns SJ, Zhang L, Huang SM, and Giacomini KM (2012) The UCSF-FDA TransPortal: a public drug transporter database. *Clin Pharmacol Ther* **92**:545–546.
- Müller F and Fromm MF (2011) Transporter-mediated drug-drug interactions. *Pharmacoeconomics* **12**:1017–1037.
- Na'ngono Manga JCD, Philip H, Rowe T, and Cronin D (2003) A hierarchical QSAR model for urinary excretion of drugs in humans as a predictive tool for biotransformation. *QSAR Comb Sci* **22**:263–273.
- Ng C, Xiao Y, Putnam W, Lum B, and Tropsha A (2004) Quantitative structure-pharmacokinetic parameters relationships (QSPKR) analysis of antimicrobial agents in humans using simulated annealing k-nearest-neighbor and partial least-square analysis methods. *J Pharm Sci* **93**: 2535–2544.
- Ohtsuki S, Schaefer O, Kawakami H, Inoue T, Liehner S, Saito A, Ishiguro N, Kishimoto W, Ludwig-Schwellinger E, and Ebner T, et al. (2012) Simultaneous absolute protein quantification of transporters, cytochromes P450, and UDP-glucuronosyltransferases as a novel approach for the characterization of individual human liver: comparison with mRNA levels and activities. *Drug Metab Dispos* **40**:83–92.
- Paine SW, Ménochet K, Denton R, McGinnity DF, and Riley RJ (2011) Prediction of human renal clearance from preclinical species for a diverse set of drugs that exhibit both active secretion and net reabsorption. *Drug Metab Dispos* **39**:1008–1013.
- Parrott N, Paquereau N, Coassolo P, and Lavé T (2005) An evaluation of the utility of physiologically based models of pharmacokinetics in early drug discovery. *J Pharm Sci* **94**: 2327–2343.
- Russel FG, Masereeuw R, and van Aubel RA (2002) Molecular aspects of renal anionic drug transport. *Annu Rev Physiol* **64**:563–594.
- Schaefer O, Ohtsuki S, Kawakami H, Inoue T, Liehner S, Saito A, Sakamoto A, Ishiguro N, Matsumaru T, and Terasaki T, et al. (2012) Absolute quantification and differential expression

- of drug transporters, cytochrome P450 enzymes, and UDP-glucuronosyltransferases in cultured primary human hepatocytes. *Drug Metab Dispos* **40**:93–103.
- Therneau T, Atkinson B, and Ripley B (2014) rpart: Recursive Partitioning and Regression Trees., R package version 4.1-5.
- Uchida Y, Ohtsuki S, Kamiie J, and Terasaki T (2011) Blood-brain barrier (BBB) pharmacoproteomics: reconstruction of in vivo brain distribution of 11 P-glycoprotein substrates based on the BBB transporter protein concentration, in vitro intrinsic transport activity, and unbound fraction in plasma and brain in mice. *J Pharmacol Exp Ther* **339**:579–588.
- Van der Graaf PH, Nilsson J, Van Schaick EA, and Danhof M (1999) Multivariate quantitative structure-pharmacokinetic relationships (QSPKR) analysis of adenosine A1 receptor agonists in rat. *J Pharm Sci* **88**:306–312.
- Varma MV, Feng B, Obach RS, Troutman MD, Chupka J, Miller HR, and El-Kattan A (2009) Physicochemical determinants of human renal clearance. *J Med Chem* **52**:4844–4852.
- Xu C and Mager DE (2011) Quantitative structure-pharmacokinetic relationships. *Expert Opin Drug Metab Toxicol* **7**:63–77.
- Yang X, Gandhi YA, Duignan DB, and Morris ME (2009) Prediction of biliary excretion in rats and humans using molecular weight and quantitative structure-pharmacokinetic relationships. *AAPS J* **11**:511–525.
- Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* **32**:1466–1474.
- Zhivkova Z and Doytchinova I (2013) Quantitative structure—clearance relationships of acidic drugs. *Molecular pharmaceutics* **10**:3758–3768.

Address correspondence to: Dr. Marilyn E. Morris, Department of Pharmaceutical Sciences, School of Pharmacy and Pharmaceutical Sciences, University at Buffalo, The State University of New York, Buffalo, NY 14214. E-mail: memorris@buffalo.edu