Prediction of Steady-State Volume of Distribution of Acidic Drugs by Quantitative Structure–Pharmacokinetics Relationships

ZVETANKA ZHIVKOVA, IRINI DOYTCHINOVA

Faculty of Pharmacy, Medical University of Sofia, 1000 Sofia, Bulgaria

Received 19 September 2011; revised 17 October 2011; accepted 28 October 2011

Published online 13 December 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.22819

ABSTRACT: The volume of distribution (VD) is one of the most important pharmacokinetic parameters of drugs. The present study employs quantitative structure–pharmacokinetics relationships (QSPkR) to derive models for VD prediction of acidic drugs. The steady-state volume of distribution (VD_{ss}) values of 132 acidic drugs were collected, the chemical structures were described by 178 molecular descriptors, and QSPkR models were derived after variable selection by genetic algorithm and stepwise regression. Models were validated by cross-validation procedures and external test set. According to the molecular descriptors selected as the most predictive for VD_{ss}, the presence of seven- and nine-member cycles, atom type P^{5+}, SH groups, and large nonionized substituents increase the VD_{ss}, whereas atom types S^{2+} and S^{4+} and polar ionized substituents decrease it. Cross-validation and external validation studies on the QSPkR models derived in the present study showed good predictive ability with mean fold error values ranging from 1.58 (cross-validation) to 2.25 (external validation). The model performance is comparable to more complicated methods requiring *in vitro* or *in vivo* experiments and superior to the existing QSPkR models concerning acidic drugs. Apart from the prediction of VD in human, present models are also useful as a curator of available pharmacokinetic databases. © 2011 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 101:1253–1266, 2012

Keywords: computational ADME; distribution; *in silico* modeling; multivariate analysis; pharmacokinetics; QSPR; steady-state volume of distribution (VD_{ss})

INTRODUCTION

The application of combinatorial chemistry methods in drug development has led to extensively growing number of chemical structures with drug-like activities. Unfortunately, majority of these compounds fail to become marketable products because of the lack of efficacy in humans, most often due to their inappropriate pharmacokinetic characteristics. Reliable prediction of the pharmacokinetic behavior is able to identify the improper drug candidates in terms of inadequate absorption, distribution, metabolism, and excretion (ADME) and eliminate them at early stages of drug design, before the expensive clinical trials. Thus, the optimization of drug pharmacokinetic properties in humans has become an obligatory step in modern drug discovery process.

Correspondence to: Zvetanka Zhivkova (Telephone: +359-2-9236514; Fax: +359-2-9879874; E-mail: zzhivkova@pharmfac.acad.bg)

Journal of Pharmaceutical Sciences, Vol. 101, 1253–1266 (2012)

One of the most important pharmacokinetic parameters is the volume of distribution (VD). VD is an apparent pharmacokinetic parameter, which relates the amount of a drug in the body to the measured concentration in a relevant biological fluid (typically plasma). It is defined as the volume in which a drug will appear to be distributed if it is presented throughout at the same concentration as that in plasma. VD can be calculated on the basis of the experimental data for the change in plasma concentration with time. Several types of VD are defined. The volume of the central compartment (VD_c) is equal to the ratio between the intravenous (i.v.) administered dose D and the initial plasma concentration C_0:

$$VD_c = D/C_0$$

The VD of drugs that follow multiexponential decay (multicompartmental behavior) is usually expressed as VD_{area} or steady-state volume of distribution (VD_{ss}). VD_{area} represents the ratio between the plasma clearance CL and the elimination rate

constant β determined from the final phase of the logarithmic concentration versus time (C/t) curve:

$$VD_{area} = \frac{CL}{\beta} = \frac{D}{\beta AUC_{0-\infty}}$$

where D is the i.v. administered dose and $AUC_{0-\infty}$ is the area under the C/t curve from $t = 0$ to $t = \infty$.

Steady-state volume of distribution corresponds to the equilibrium distribution of a drug between the central and tissue compartments (at steady state), and is given by the following equation:

$$VD_{ss} = VD_c(1 + {}^{k_{12}}/_{k_{21}})$$

where k_{12} and k_{21} are the constants of distribution between the central and tissue compartments. VD_{ss} is considered as the most reliable indicator for drug partitioning between plasma and tissues.[1]

Although VD does not correspond to any physiological space and, hence, has no physical meaning, it is of great practical importance. It determines the residence time of the drug in the body and serves as a key parameter for setting up a suitable dosing regimen.[2]

Various methods have been proposed for the prediction of VD in humans: extrapolation from animal data (allometric method), physiologically based pharmacokinetic modeling, biomimetic binding experiments (phospholipids or plasma protein binding at immobilized artificial membrane), and computational (*in silico*) approaches. Recently, Mager[3] and Fagerholm[4] summarized and critically analyzed the methods for the prediction of human pharmacokinetics.

Computational methods for quantitative structure–pharmacokinetic relationships (QSPkR) are increasingly gaining popularity and utility owing to their uncontested advantages: They can be based solely on easily calculated structural descriptors, ensure high efficiency with respect to time, labor, and cost, and eliminate any inevitable experimental errors. Several studies have been published proving the utility of QSPkR models for the prediction of VD. Most of them focused on congeneric series of drugs,[5–10] and their predictive power is restricted within the respective structural class.

Generation of adequate QSPkR models for heterogeneous series of structurally diverse drugs is a rather complicated task. A great variety of approaches have been used to develop predictive models for VD, starting from the Øie—Tozer equation[11] combined with multiple linear regression (MLR),[12–14] through the genetic algorithm (GA) and stepwise regression,[15–16] to partial least squares (PLS)[9,17–18], artificial neural networks[18–19] Bayesian neural networks,[20] classification and regression trees,[20] random forest,[21–22] and mechanism-based models.[23] The problems stem not only from the complex nature of the distribution processes determined by membrane permeability, specific and nonspecific binding to plasma and tissue proteins, metabolism, and so on, but also from highly varying numeral values of VDs published in literature. Any inaccuracy of the data can confound the performance of the models.

Sometimes, the statistics of QSPkR models is improved by replacing VD with related parameters such as unbound VD (corrected for protein binding)[15] or fractal volume v_t (related to physiological volume),[17] but these volumes are of less practical importance.

Most of the models developed for the prediction of VD are based on datasets including acidic, basic, neutral, and amphoteric drugs. There are only few models developed on subsets of acids only and on subsets of bases only.[15–16] It is well established that acids and bases follow different distribution patterns. In general, acids have relatively small VDs due to extensive binding to plasma proteins, whereas bases penetrate preferably in tissues and possess higher VD values. Surprisingly, the separation of drugs into acids and bases did not improve the predictive ability of the models.[16] Even more, the QSPkR models for acids were statistically poor. This could be due to the limited number of drugs involved in the study, the incorrect allocation of drugs to acidic or basic group, or the narrow range of VD values of acids.

The aim of the present study was to develop a QSPkR model for the prediction of apparent VD_{ss} of structurally and pharmacologically diverse acidic drugs in humans based on calculated structural descriptors. A dataset of 132 acidic drugs were collected, the chemical structures were described by 178 molecular descriptors, and QSPkR models were derived after variable selection by GA and stepwise regression. Models were validated by cross-validation procedures and external test set.

MATERIALS AND METHODS

Datasets

Two datasets were used in the present study—one training and one external test set. The training dataset was compiled from Obach's database,[1] which is considered as the largest and the best curated database for pharmacokinetic parameters in humans.

The mol files of the drugs from Obach's database were retrieved from DrugBank[24] and ChemicalBook[25] and their p*K*a values were calculated by the ACD/LogD software (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada). The fraction of a drug ionized at pH 7.4 as acid (f_A) or base (f_B) were

 DOI 10.1002/jps

calculated according to the following equations:

$$f_A = \frac{1}{1 + 10^{(pKa - 7.4)}};$$

$$f_B = \frac{1}{1 + 10^{(7.4 - pKa)}}$$

In case of more than one acidic or basic center, the p*K*a of the strongest one was taken. A drug was considered as an acid if it met at least one of the following two conditions: (1) f_A exceeded 10% and $f_B \approx 0$ and/or (2) f_A was considerably higher than f_B and close to 1. One hundred and thirty-two drugs met one or both of these conditions and entered the training set.

The VD values in Obach's database refer to VD_{ss} after i.v. administration of drugs. This volume term describes the overall distributional behavior most generally.[1] For quantitative structure–activity relationship (QSAR) purpose, the VD_{ss} values were used as log VD_{ss} and log (VD_{ss}/MW), where MW is the molecular weight of drugs. The training set was used for the development of the QSPkR models and for cross-validation.

The external test set was compiled from Berrelini et al.[22] and consisted of 10 acidic drugs not included in Obach's database. Similar to the training set, the drugs in the test set were classified as acidic if it met at least one of the two conditions described above. The test set was used for external validation of the models derived in the present study.

Molecular Descriptors

The chemical structure of the drugs used in the present study was described by 178 molecular descriptors computed using the software packages ACD/LogD version 9.08 (Advanced Chemistry Development, Inc.) and MDL QSAR version 2.2 (MDL Information Systems, Inc., San Leandro, California). The descriptors were grouped into five types. The first type included the molecular connectivity χ indices,[26] which represent molecular structure by encoding significant topological features of the whole molecule. There are five categories of structural information described by χ indices: degree of branching (low-order χ indices), variable branching pattern (path χ indices), position and influence of heteroatoms (valence χ indices), patterns of adjacency (χ cluster and path/cluster indices), and degree of cyclicity (χ chain indices). The second type of descriptors—the κ shape indices—is a family of graph-based structure descriptors that represent shape.[27] The third type comprised the electrotopological state (E-state) indices, which are atom-level molecular descriptors computed for each atom in the molecule.[27] They represent the electron density at each atom and the ability of those electrons to participate in intermolecular interactions. A variety of molecular properties—weight, log *p*, log $D_{7.4}$, number of rings, number of hydrogen bond donors (HBD) and acceptors (HBA), and so on—was defined as the fourth type. The last type consisted of three-dimensional (3D) molecular properties such as polarizability, surface area, volume, and so on.

Variable Selection

A GA,[28] as implemented in the MDL QSAR package, was used as a variable selection procedure in the present study. GA allows one to select a subset of the most significant predictors using two evolutionary operations: random mutation and genetic recombination (crossover). The algorithm was used in the study with default values for the size of the initial population (32), choice of parents (tournament selection), types of crossover (uniform crossover) and mutation (one-point mutation), and fitness function (Friedman's lack-of-fit scoring function with two parameters).[29] The selected variables entered a stepwise linear regression.

QSPkR by Stepwise Linear Regression

The QSPkR models in the present study were derived by a stepwise linear regression, as implemented in the MDL QSAR package. It was used in a forward mode with default values for the Fisher ratio *F*-to-enter (4.00) and *F*-to-remove (3.99). Final models were assessed by explained variance (r^2) and standard error of estimate (SEE) according to the following equations:

$$r^2 = 1 - \frac{\sum_{i=1}^{n} (\log VD_{ss,obs,i} - \log VD_{ss,calc,i})}{\sum_{i=1}^{n} (\log VD_{ss,obs,i} - \log VD_{ss,obs,mean})},$$

$$SEE = \sqrt{\frac{\sum_{i=1}^{n} (\log VD_{ss,obs,i} - \log VD_{ss,calc,i})}{n - p - 1}},$$

where $VD_{ss,obs,i}$ is the observed VD_{ss} of the *i*-th drug, $VD_{ss,calc,i}$ is calculated by the model VD_{ss} of the *i*-th drug, *n* is the number of drugs in the dataset, and *p* is the number of molecular descriptors in the model. Fisher statistics (*F*) for models also was calculated.

Cross-Validation

Two cross-validation procedures were applied to the training set in order to access the predictive ability of the QSPkR models—leave-one-out cross-validation (LOO-CV) and leave-many-out cross-validation (LMO-CV). The models were assessed by cross-validated coefficients $q^2_{LOO\text{-}CV}$ and $q^2_{LMO\text{-}CV}$, respectively, according to the following

equation:

$$q^2 = 1 - \frac{\sum_{i=1}^{n} (\log \mathrm{VD}_{\mathrm{ss,obs},i} - \log \mathrm{VD}_{\mathrm{ss,pred},i})}{\sum_{i=1}^{n} (\log \mathrm{VD}_{\mathrm{ss,obs},i} - \log \mathrm{VD}_{\mathrm{ss,obs,mean}})}$$

where $\mathrm{VD}_{\mathrm{ss,pred},i}$ is predicted by the model $\mathrm{VD}_{\mathrm{ss}}$ of the i-th drug. In the LOO-CV procedure, one drug was excluded from the initial dataset and the model was derived based on the remaining n–1 drugs and was used to predict the $\mathrm{VD}_{\mathrm{ss,pred},i}$ of the excluded i-th drug. In the LMO-CV procedure, the initial set was divided randomly into training (80%) and test (20%) sets, and the test set was used to validate the QSPkR model derived on the training set. The average value of 30 runs was given as $q^2_{\mathrm{LMO\text{-}CV}}$.

Fold error (FE) of prediction for the test sets was calculated according to the following equation:

$$\mathrm{FE} = 10^{|\log \mathrm{VD}_{\mathrm{ss,obs},i} - \log \mathrm{VD}_{\mathrm{ss,pred},i}|}$$

and the average value of 30 runs was given as mean fold error (MFE).

The accuracy of prediction is given as the percent of the total number of drugs predicted with less than twofold error.

External Validation

The external test set was used for external validation of the QSPkR models derived in the present study. The $\mathrm{VD}_{\mathrm{ss}}$ values of the tested drugs were predicted by the models and compared with the observed $\mathrm{VD}_{\mathrm{ss}}$ in terms of FE and accuracy at twofold and threefold error.

RESULTS

Dataset Analysis

The training and test datasets of acidic drugs used in the present study encompasses a wide variety of structures, therapeutic actions, and physicochemical properties. A drug was considered an acid if it met at least one of the following two conditions: (1) f_{A} exceeded 10% and $f_{\mathrm{B}} \approx 0$ and/or (2) f_{A} was considerably higher than f_{B} and close to 1. The training set consisted of 132 drugs and the test set included 10 drugs.

The MW of the analyzed drugs varies between 126 and 1297 g/mol, with an average value of 372 g/mol. MW exceeds 500 g/mol for 14 drugs and two of them (micafungin and suramin) have MW higher than 1200 g/mol.

The values of the lipophilicity parameters log p and log $D_{7.4}$ also vary significantly. Log p ranges between –7.48 (micafungin) and 8.39 (hypericin), and the same drugs have extreme values of log $D_{7.4}$ (–11 and 7.64, respectively).

The values of $\mathrm{VD}_{\mathrm{ss}}$ range from 0.04 (suprofen) to 15 L/kg (artesunate), with an average value of 0.541 L/kg and median value of 0.22 L/kg. Log $\mathrm{VD}_{\mathrm{ss}}$ ranges between –1.40 and 1.176, with an average value of –0.576 and median value of –0.658. The histogram of the data distribution of the experimental log $\mathrm{VD}_{\mathrm{ss}}$ values is shown in Figure 1.

QSPkR Model for Log $\mathrm{VD}_{\mathrm{ss}}$

One hundred and seventy-eight molecular descriptors were calculated for 132 acidic drugs. Variable selection procedure by GA was applied to select the most predictive descriptors followed by a stepwise linear regression. Several models were derived and validated by LOO-CV and LMO-CV. The best performing one in terms of r^2 and q^2 is given as model 1:

$$\begin{aligned}\log \mathrm{VD}_{\mathrm{ss}} = {} & 28.01(\pm 3.82)\mathit{xch9} + 0.53(\pm 0.06)\mathit{SdsssP\_acnt} \\ & -0.12(\pm 0.03)\mathit{SsssS\_acnt} - 0.81(\pm 0.16)\mathit{SdssS\_acnt} \\ & +0.35(\pm 0.10)\mathit{SHsSH} - 0.41(\pm 0.07)H_{\max} \\ & +0.05(\pm 0.01)G_{\min} - 0.17(\pm 0.03)\mathit{knotpv} \\ & +0.38(\pm 0.12) \qquad \text{(Model 1)}\end{aligned}$$

For model 1, $n = 125$, $r^2 = 0.661$, SEE = 0.194, and $F = 28.32$.

No intercorrelation between the descriptors in model 1 was observed ($r < 0.65$) (the intercorrelation matrix is given in Appendix 1). Seven drugs were considered as outliers having residuals greater than ±0.5 log unit. The plot calculated by model 1 log $\mathrm{VD}_{\mathrm{ss}}$ versus observed log $\mathrm{VD}_{\mathrm{ss}}$ is given in Figure 2.

The molecular descriptors of model 1 belong to three types: molecular connectivity χ indices, E-state indices, and molecular properties. The molecular connectivity index *xch9* accounts for the presence and characteristics (type and position of substituents) of a nine-member cycle. The E-state indices are presented in the model by three groups: E-state indices, atom-type count indices, and hydrogen E-state indices. All E-state indices have positive values. Descriptor $G_{\min}$ gives the minimum E-state index in the molecule. The second group accounts for the number of atoms of a given type present in molecule. Descriptors *SdsssP_acnt*, *SsssS_acnt*, and *SdssS_acnt* reflect the number of (=P≡), (—S—), and (=S<) atom types in the molecule, respectively. Descriptors *SHsSH* and $H_{\max}$ are hydrogen E-state indices. The first one equals the sum of all hydrogen E-state indices for all —SH groups in the molecule, and the second one shows the maximum hydrogen E-state index in the molecule. The molecular properties are presented by

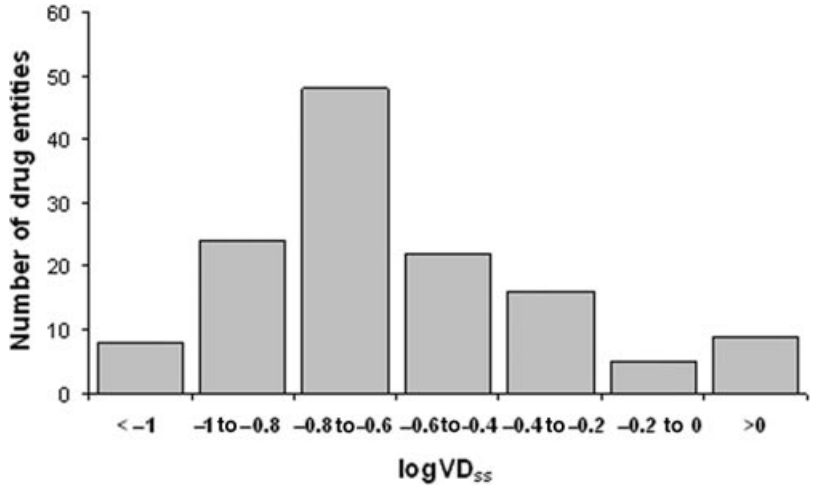

Figure 1. Histogram of the data distribution of experimental log VD_{ss}.

the descriptor *knotpv*, which accounts for the intermolecular accessibility.

QSPkR Model for Log (VD_{ss}/MW)

The ADME parameter log (VD_{ss}/MW) was defined as VD_{ss} per unit weight in order to eliminate the influence of MW. The best performing QSPkR model for log (VD_{ss}/MW) is given below as model 2:

$$\begin{aligned}\log(VD_{ss}/MW) = {} & 6.71(\pm 0.74)xch7 + 27.15(\pm 4.19)xch9 \\ & + 0.64(\pm 0.06)SdsssP\_acnt \\ & - 0.87(\pm 0.18)SdssS\_acnt \\ & + 0.22(\pm 0.04)SsSH - 0.64(\pm 0.07)H_{max} \\ & - 0.17(\pm 0.03)nelem - 0.78(\pm 0.15)\end{aligned} \quad \text{(Model 2)}$$

For model 2, $n = 126, r^2 = 0.687, SEE = 0.221$, and $F = 36.99$.

No intercorrelation between the descriptors in model 2 was observed ($r < 0.65$) (the intercorrelation matrix is given in Appendix 1). Six drugs were considered as outliers having residuals greater than ±0.5 log unit. The plot calculated by model 2 log (VD_{ss}/MW) versus observed log (VD_{ss}/MW) is given in Figure 3.

Model 1 and model 2 have four identical descriptors: *xch9*, *SdsssP_acnt*, *SdssS_acnt*, and H_{max}. Model 2 contains three additional descriptors: *xch7*, *SsSH*, and *nelem*. The molecular connectivity index *xch7* accounts for the presence of a seven-member ring. The E-state index *SsSH* represents the sum of all (—SH) E-state values in the molecule. The molecular property *nelem* corresponds to the number of chemical elements in a molecule.

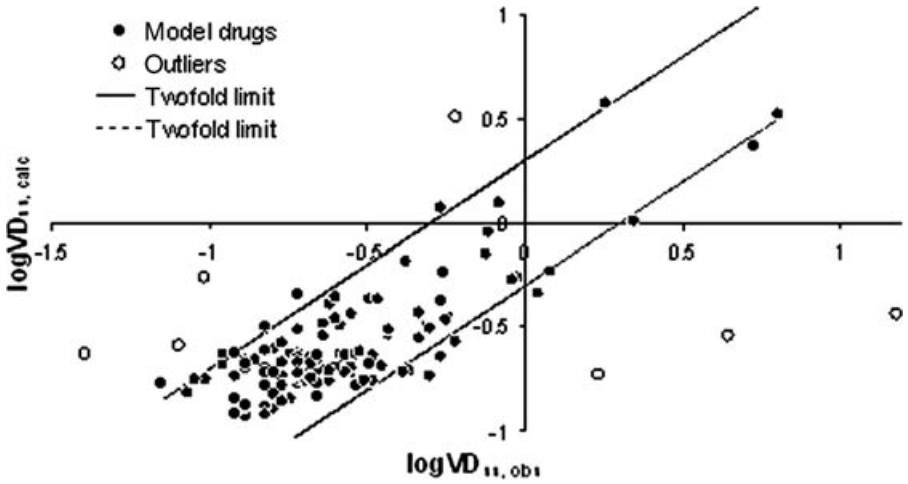

Figure 2. Log VD_{ss} values calculated by model 1 versus observed log VD_{ss} values for 132 acidic drugs. Seven outliers are given as blank points. The dotted lines represent the twofold error limits.

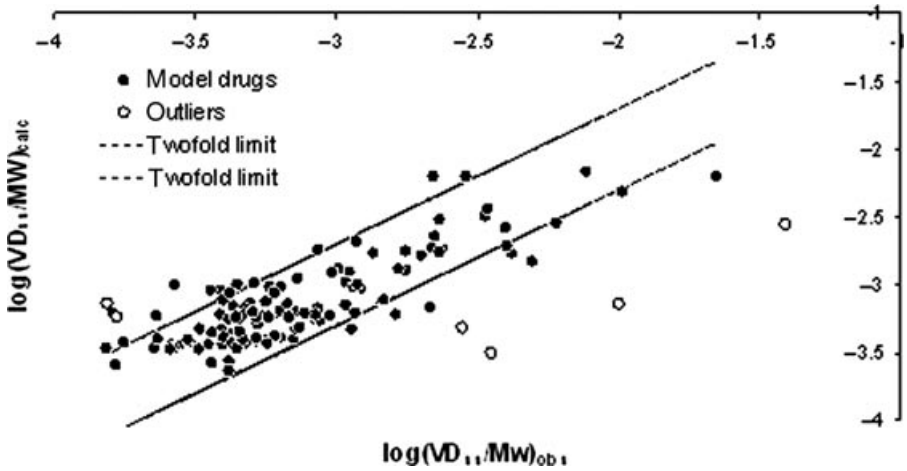

Figure 3. Log (VD_{ss}/MW) values calculated by model 2) versus observed log (VD_{ss}/MW) values for 132 acidic drugs. Six outliers are given as blank points. The dotted lines represent the twofold error limits.

Cross-Validation of the QSPkR Models

The QSPkR models derived in the present study were validated as described in the *Materials and Methods* section. Parameters $q^2_{\text{LOO-CV}}$, $q^2_{\text{LMO-CV}}$, MFE, and accuracy of prediction are given in Table 1.

Models show good predictive ability, having 82%–83% of the drugs predicted with less than twofold error and MFE between 1.581 and 1.726.

The observed, calculated, and predicted VD_{ss} values of the drugs used in the present study are given in Appendix 2.

Table 1. Cross-Validation of the QSPkR Models Derived in the Present Study

Parameter	Model 1	Model 2
$q^2_{\text{LOO-CV}}$	0.581	0.625
$q^2_{\text{LMO-CV}}$	0.540	0.600
MFE	1.581 ± 0.584	1.716 ± 0.981
Accuracy	83%	82%

External Validation of the QSPkR Models

The predicted VD_{ss} values of the drugs by model 1 and model 2 from the external test set are given in Table 2. The MFE for model 1 is 2.04 and the accuracies at twofold and threefold error are 50% and 80%, respectively. For model 2, the external validation showed MFE of 2.25, accuracy of 60% at twofold error, and accuracy of 80% at threefold error.

Comparison with Other QSPkR Models for VD_{ss} Prediction

Models derived in the present study were compared with other QSPkR models for the prediction of VD_{ss} in humans based on datasets of acidic drugs only (Table 3). Models were compared in terms of the predictive ability measured by r^2, q^2, MFE, and accuracy.

Karalis et al.[9] applied PLS and MLR and nonlinear regressions to generate QSPkR models on a dataset of 23 cephalosporines; 14 of them are presented in our dataset. The PLS model had $r^2 = 0.592$ and $q^2_{\text{LMO-CV}} = 0.554$. The linear regression analysis led to one parameter model with $r^2 = 0.523$ and SD = 4.752, and the nonlinear one-parameter model was slightly superior with $r^2 = 0.571$ and SD = 4.609.

Ghafourian et al.[16] used GA and stepwise regression to derive models for VD prediction of acidic drugs. Two models were developed with MFE 1.79 and 1.84 and *accuracy* of 71% at twofold error threshold.

As is evident from Table 2, the QSPkR models for VD_{ss} prediction of acidic drugs derived in the present study outperform the existing models. They show the least MFE and the highest accuracy.

DISCUSSION

The relationship between the pharmacokinetic parameter apparent VD_{ss} and the structure of 132 acidic drugs was examined by four sets of 2D-QSAR and one set of 3D-QSAR descriptors. VD_{ss} was presented as log VD_{ss} and log (VD_{ss}/MW) in the QSPkR models. The most predictive variables were selected by

 DOI 10.1002/jps

Table 2. Observed and Predicted VD_{ss} Values of the Acidic Drugs from the External Test Set

		Model 1		Model 2	
Drug	$VD_{ss,obs}$ (L/kg)	$VD_{ss,pred}$ (L/kg)	FE	$VD_{ss,pred}$ (L/kg)	FE
Acenocoumarol	0.24	0.22	1.18	0.19	1.26
Amphotericin	0.62	0.26	2.35	0.46	1.36
Candoxatrilat	0.25	0.26	1.06	0.25	1.01
Clazosentan	0.23	0.18	1.31	0.28	1.21
Deferasirox	0.23	0.21	1.11	0.22	1.02
Gavestinel	0.12	0.44	3.64	0.32	2.67
Levocabastine	1.17	0.44	2.69	0.18	6.53
Pelrinone	0.39	0.39	1.01	0.47	1.21
Ridogrel	0.43	0.14	3.18	0.18	2.44
Treprostinil	0.23	0.66	2.88	0.87	3.79
MFE			2.04		2.25
Accuracy at two-fold error			50%		60%
Accuracy at three-fold error			80%		80%

GA and stepwise linear regression. Two significant QSPkR models containing similar descriptors were derived.

The molecular connectivity indices *xch7* and *xch9* account for the presence and characteristics of seven-member and nine-member cycles, respectively. Both descriptors have positive contributions in the models; increasing the values of *xch7* and *xch9* increases the VD_{ss} of the molecules. Ghafourian et al.[16] also found a positive correlation between *xch7* and log VD. Obviously, *xch7* contributes to tissue distribution. *xch9* is common for both models with great influence on VD. Recently, Yang et al.[30] found that the same descriptor correlates positively with the biliary clearance. Apparently, biliary excretion may contribute to the extension of VD.

E-state indices are widely presented in the models derived in the present study. This is not surprising, as these descriptors represent the polarity of compounds and the polarity governs the drug distribution in the body.

Descriptor *SdsssP_acnt* corresponds to the number of phosphonate groups in molecule. It contributes positively to the VD_{ss}, although the presence of phosphonate groups increases the polarity of compounds significantly. At physiological pH, the phosphonate moieties are negatively charged ions unable to cross membranes and distribute around the body. Indeed, bisphosphonate drugs used to prevent osteoporosis have poor absorption and low oral bioavailability.[31] Nevertheless, relatively high values for VD_{ss} have been observed for pamidronate (1.8 L/kg), risedronate (6.3 L/kg), and fludarabine (2.2 L/kg). The former two drugs are bisphosphonates, which are proved to bind with high affinity to bones.[32] Fludarabine also reveals extensive tissue binding, preferably in kidney, liver, and spleen.[33] The positive contribution of the number of phosphonate groups to the VD_{ss} should be related to specific binding to various tissues rather than to transmembrane transport.

The presence of sulfur and sulfuric groups greatly affect the VD_{ss} of acidic drugs. Different sulfuric groups, however, have different influence on the drug distribution. For example, descriptor *SssS_acnt*, representing the number of sulfuric atoms connected to other atoms with two simple σ-bonds, namely, atom type (—S—), contributes negatively to VD_{ss}. Similarly, descriptor *SdssS_acnt*, which corresponds to the number of SO groups, also has a negative contribution to VD_{ss}. SO groups are highly polar structures often involved in the binding of drugs to human serum albumin (HSA).[34] The extensive binding to plasma proteins is considered as the major reason for low values of VD_{ss}. Oppositely, the number of SH groups

Table 3. QSPkR Models for Prediction of VD of Acidic Drugs

Dataset	Method	Statistics	MFE and Accuracy	Reference
Cephalosporins, $n = 23$	PLS, MLR, nonlinear	$r^2 = 0.592$, $q^2_{LMO\text{-}CV} = 0.554$; $r^2 = 0.523$, SD = 4.752; $r^2 = 0.571$, SD = 4.609	ND	Karalis et al.[9]
Structurally diverse acids ($f_A > f_B$)	MLR	$n = 69$, $r^2 = 0.554$, SEE = 0.268, $q^2_{LMO\text{-}CV} = 0.448$; $n = 68$, $r^2 = 0.575$, SEE = 0.266, $q^2_{LMO\text{-}CV} = 0.470$	1.79, 71%; 1.84, 71%	Ghafourian et al.[16]
Structiurally diverse acids ($f_A > 0.1$ and $f_B \approx 0$ and/or $f_A \gg f_B$ $f_A \approx 1$)	Variable selection by GA and stepwise MLR	$n = 125$, $r^2 = 0.661$, SEE = 0.194, $q^2_{LOO\text{-}CV} = 0.581$, $q^2_{LMO\text{-}CV} = 0.540$; $n = 126$, $r^2 = 0.687$, SEE = 0.221, $q^2_{LOO\text{-}CV} = 0.625$, $q^2_{LMO\text{-}CV} = 0.600$	1.58, 83%; 1.72, 82%	Model 1, this study; Model 2, this study

ND, not determined; PLS, partial least squares; MLR, multiple linear regression; SEE, standard error of estimate.

(described by *SHsSH* and *SsSH*) affects VD_{ss} positively. The SH groups are capable of forming hydrogen and disulfide bonds with tissue proteins.

The polarity of the drugs is also encoded by descriptors G_{min} and H_{max}. Descriptor G_{min} represents the minimum E-state value in the molecule and is a quantitative measure for the most electrophilic atom. The drugs with the most negative G_{min} values (suramin, risedronate, zoledronate, micafungin, and pamidronate) are extremely polar and their log $D_{7.4}$ values range from –6.3 to –11. As G_{min} is negative, it has a positive coefficient in the models, that is, polar drugs have low VD_{ss}. Descriptor H_{max} equals the maximum hydrogen atom E-state value in the molecule and reflects its ionizability. It has positive values, with the lowest H_{max} corresponding to weak nonionized acids (with p*K*a > 7.4) and the highest values corresponding to stronger ionized acids (with p*K*a < 4.2). The negative coefficient of this descriptor in the models is in good agreement with the general consideration that the ionized acids have low VD_{ss} values.

The molecular descriptor *knotpv* represents the difference between two connectivity indices: χ valence cluster 3 and χ valence path/cluster 4. Both relate positively to the size of the molecule and the number of HBD and HBA. The molecular connectivity indices could be interpreted in terms of intermolecular accessibility.[35] The intermolecular accessibility is as higher as lower is the value of *knotpv*. For small compact members of the dataset, the value of *knotpv* is close to zero (–0.20 to +0.22), whereas for large molecules with a high degree of branching, distal substituents, or adjacent (conjugated) rings, *knotpv* is highly negative (the lowest value is –3.46). The latter molecules face difficulties in crossing membranes, but they are rich in atoms and HBD and/or HBA groups, potential participants in drug–tissue protein interactions. The negative coefficient of *knotpv* in model 1 reflects the positive relationship between intermolecular accessibility and VD_{ss}.

Descriptor *nelem* appears in model 2 and its value influences log (VD_{ss}/MW) negatively. Most probably, *nelem* affects the dependent variable through MW rather than through VD_{ss} because the presence of heteroatoms such as F, Cl, and P results in a higher MW.

Although lipophilicity (expressed as log *p* or log $D_{7.4}$) is considered as a key property governing pharmacokinetic behavior, a direct relationship between VD_{ss} and lipophilicity parameters was not found in the present study. It seems intuitive to expect a relationship between VD_{ss} and log *p* or log $D_{7.4}$ with the VD_{ss} increasing with increasing lipophilicity (approaching a maximum or assuming a parabolic profile); however, such correlations are rarely observed, even for a homogenous series of compounds.[3] A brief look at the database shows that drugs with opposite lipophilicity have similar VD_{ss} values and vice versa; drugs with similar lipophilicity have quite different VD_{ss}s. For example, hypericin is a highly lipophilic drug (log p = 8.39, log $D_{7.4}$ = 7.64) with a VD_{ss} of 0.25 L/kg, close to VD_{ss} of micafungin (VD_{ss} = 0.21 L/kg), which is a very hydrophilic drug (log p = −7.48, log $D_{7.4}$ =–11.02). Another example, artesunate (log p = 2.94, log $D_{7.4}$ =–0.09, VD_{ss} = 15 L/kg) and bromfenac (log p = 3.17, log $D_{7.4}$ = 0.1, VD_{ss} = 0.11 L/kg) have similar lipophilicity, but very different VD_{ss}. Instead of lipophilicity, the polarity of drugs is widely expressed in our models. Actually, polarity relates closely to lipophilicity.

Special attention should be paid to the outliers in the models (Table 4). Four compounds (artesinate, ifetroban, rosuvastatin, and suprofen) are the common outliers in the two models; three (glyburide, indomethacine, and zoledronate) are additional outliers in model 1 and two (acivicin and mezlocillin)–additional outliers in model 2. The outliers differ in their structure (Fig. 4) and physicochemical properties. Their incompatibility with the models can be due to several reasons: presence of structural features not captured by the descriptors used, high interindividual variability in metabolism, specific multicompartment distribution, and inaccurate experimental values.

The experimental value of VD_{ss} for acivicin is 0.5 L/kg; the calculated value according to model 2 is significantly lower −0.086 L/kg. Acivicin is a negatively charged amino acid with p*K*a = 1.89 (calculated by ACD/LogD; Advanced Chemistry Development, Inc.), but it follows biexponential kinetics.[36] It could be hypothesized that as an amino acid, acivicin is involved in active transport-mediated distribution.

The VD_{ss} for artesunate is 15 L/kg, reported as a median value for a group of 11 patients with malaria, varying from 2.2 to 39 L/kg.[37] Artesunate has a great interindividual variability in clearance, VD, and half-life due to variability in its metabolism.[37] The calculated values for VD_{ss} are 2.49 (model 1) and 1.07 (model 2), both closer to the lower limit of the observed value. In addition, it was suggested that the primary route of artesunate elimination is biliary excretion with enterohepatic circulation,[38] which might be responsible for the higher observed value for VD_{ss}.

The VD_{ss} for glyburide in Obach's database is given as 0.08 L/kg; the calculated value according to our model 1 is 0.34 L/kg. Jaber et al.[39] have studied the pharmacokinetics of glyburide in 12 obese patients with mean body weight of 100 kg and eight non-obese patients with mean body weight of 73 kg and have found that the observed VD for the obese group is 47.0 L (0.47 L/kg) and for the non-obese is 56.8 L (0.78 L/kg). Our calculated value is close to Jaber's

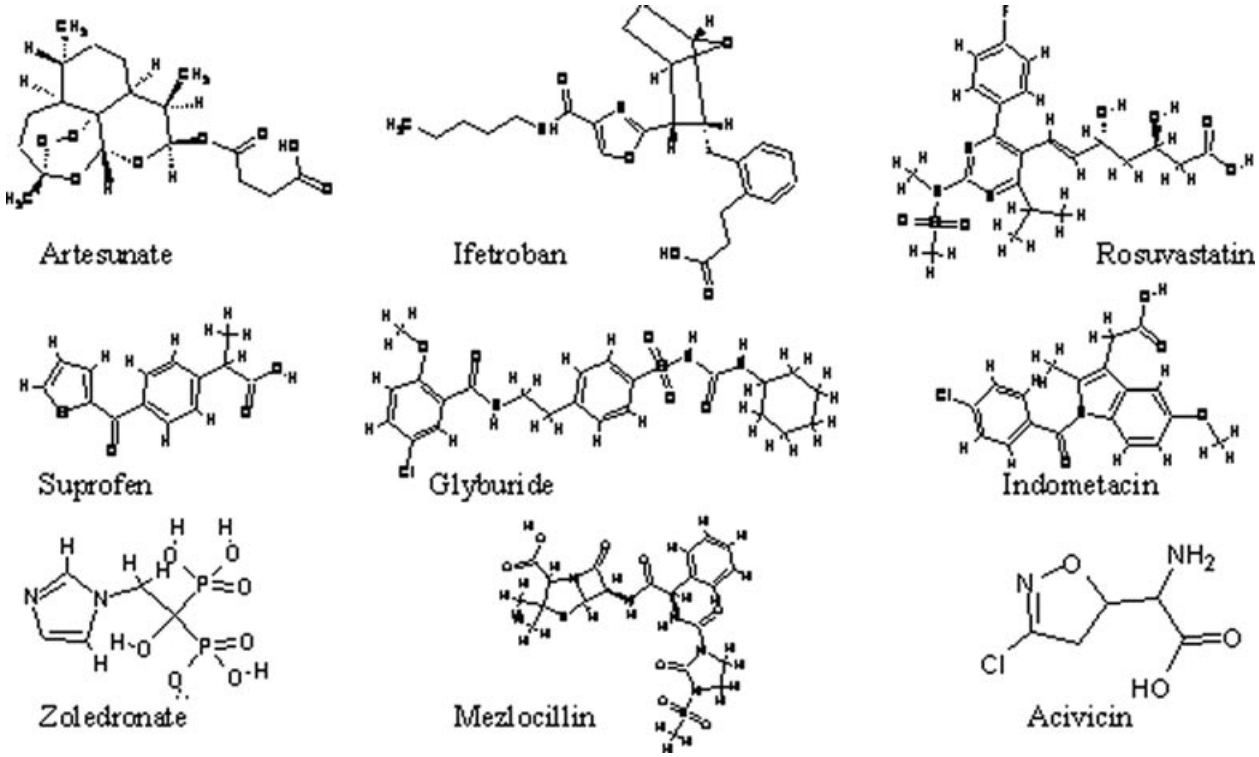

Figure 4. Chemical structures of the outliers.

observed VD values falling within the twofold error interval.

Ifetroban is known to be extensively distributed throughout the body, with concentration in tissues (except for heart and brain) exceeding that in plasma and $VD_{ss} = 4.4$ L/kg.[40] The calculated values for VD_{ss} are 0.42 (model 1) and 0.32 L/kg (model 2), that is, around 10 times lower than the observed one. Concentrations of ifetroban in human plasma showed secondary peaks at about 6 h after administration, suggesting that the drug undergoes enterohepatic recirculation.[40] This explains the observed extended VD.

The VD_{ss} for indomethacine in Obach's database is 0.096 L/kg, other authors report 0.927,[41] 0.95,[16] and 1 L/kg.[42] The calculated value according to model 1 is 0.62 L/kg, which is closer to the last three observed VD_{ss}.

Similarly, the VD_{ss} value for mezlocilline in Obach's database is 0.09 L/kg, whereas other authors gave 0.19,[43] 0.34,[44] and 0.38–0.55 L/kg.[42] Our calculated value is 0.31 L/kg, which is in good agreement with the last several observed.

Rosuvastatin is extensively distributed in tissues with apparent preference to the liver. This is supported by the high proportion of nonrenal clearance (over 70% from the total clearance) and the high extraction coefficient of the liver.[45] The drug is excreted to a significant extent into the bile. About 11% of the administered dose was recovered in the bile, and biliary concentration exceeded the plasma levels considerably.[46] Its observed VD_{ss} in Obach's database

Table 4. Outliers in the QSPkR Models

		Model 1		Model 2	
Drug	$VD_{ss,obs}$ (L/kg)	$VD_{ss,calc}$ (L/kg)	FE	$VD_{ss,calc}$ (L/kg)	FE
Acivicin	0.5	–	–	0.086↓	5.81
Artesunate	15	2.49↓	6	1.07↓	14
Glyburide	0.08	0.34↑	4.25	–	–
Ifetroban	4.40	0.42↓	10.5	0.32↓	13.90
Indomethacin	0.096	0.62↑	6.5	–	–
Mezlocillin	0.09	–	–	0.31↑	3.44
Rosuvastatin	1.70	0.17↓	10	0.15↓	11.21
Suprofen	0.04	0.22↑	5.5	0.18↑	4.65
Zoledronate	0.60	2.18↑	3.6	–	–

is 1.70 L/kg. A similar value is cited also by other authors.[47–48] The calculated VD_{ss} is 0.17 (model 1) and 0.15 L/kg (model 2), around ten times less than the observed. The extensive biliary secretion and enterohepatic circulation could explain the great difference between the observed and calculated VD_{ss}, as already has been proposed for artesunate and ifetroban.

Suprofen's VD_{ss} value in Obach's database is extremely low—0.04 L/kg. The lowest possible VD_{ss} of acidic drugs is 0.1 L/kg, as this is the VD of HSA itself.[24] Other authors give values for suprofen's VD_{ss} from 0.1 to 0.3 L/kg[49] and 0.17 L/kg[24] which are closer to our calculated values of 0.22 L/kg (model 1) and 0.18 L/kg (model 2).

The value of VD_{ss} for zoledronate in Obach's database is 0.6 L/kg. Like other bisulfonates, zolendronate binds rapidly and extensively to bones from where it is gradually released for a relatively long period of time.[50] Zoledronate is structurally similar to pamidronate and risedronate, and the three drugs possess comparable MW and physicochemical properties. However, the observed VD_{ss} for zoledronate (0.6 L/kg) is considerably lower than those of pamidronate (1.8 L/kg) and risedronate (6.3 L/kg). The value of 0.6 L/kg seems unreliably small because it is lower than the total body water even and is in a contradiction to the established distribution patterns for this class of drugs.[51] The calculated value for VD_{ss} is 3.6 L/kg, being between the observed VD_{ss} of pamidronate and risedronate.

Cross-validation and external validation studies on the QSPkR models derived in the present study showed good predictive ability with MFE values ranging from 1.58 (cross-validation) to 2.25 (external validation).

Having the experience from the outliers in the training set, we cross-checked the VD_{ss} values for the outliers in the external test set (Table 2). Gavestinel and ridogrel are outliers according to model 1, whereas levocabastine and treprostinil are outliers in model 2. Hoke et al.[52] found VD_{ss} for gavestinel in the range 0.16–0.28 L/kg, which falls in the twofold error of our prediction. Wilson and Quest[53] gave VD_{ss} of 0.1–0.2 L/kg for ridogrel, which is also in the twofold error of our predicted value. Levocabastine follows two-compartment distribution.[54] The VD_{ss} of treprostinil is 0.5 L/kg,[55] being very close to the values predicted by our models.

The comparative analysis of the existing models for VD prediction of acidic drugs available in the literature showed that the QSPkR models derived in the present study performs better than others (Table 3).

CONCLUSION

The QSPkR models for VD_{ss} prediction derived in the present study have clear physical sense and allow unambiguous interpretation of the chemical features concerning the distribution of acidic drugs in the body. They are useful guide for lead optimization in the drug design process. Even more, apart from their usage as predictors of VD_{ss}, they could be used as a curator of existing pharmacokinetic databases.

APPENDIX A

Table A1. Intercorrelation Matrix for Descriptors of Model 1

	Log VD_{ss}	knotpv	*xch9*	*SdsssP_ac*	SssS_ac	*SdssS_ac*	*SHsSH*	G_{min}
Knotpv	–0.1452							
xch9	0.4181	0.00039						
SdsssP_ac	0.4888	–0.129	0.1217					
SssS_ac	–0.2975	–0.4449	–0.1458	–0.1447				
SdssS_ac	–0.0992	0.02862	0.2289	–0.03139	–0.07497			
SHsSH	0.1591	0.1166	–0.03161	–0.03139	–0.07496	–0.01626		
G_{min}	–0.1029	0.1579	–0.07057	–0.5152	0.172	–0.02073	0.07797	
H_{max}	–0.1427	–0.3387	0.02332	0.173	0.2933	–0.3613	0.0008456	–0.1189

Table A2. Intercorrelation Matrix for Descriptors of Model 2

	Log (VD_{ss}/MW)	*nelem*	*xch7*	*xch9*	*SdsssP_ac*	*SdssS_ac*	*SsSH*
nelem	–0.2305						
xch7	–0.104	0.293					
xch9	0.2889	0.0435	–0.099				
SdsssP_ac	0.5003	0.1106	–0.098	0.0896			
SdssS_ac	–0.1014	0.1411	–0.048	0.2211	–0.033		
SsSH	0.2219	0.0539	–0.048	–0.0333	–0.033	–0.016	
H_{max}	–0.2552	0.0049	–0.165	0.0295	0.1953	–0.357	0.0015

APPENDIX B

Table B1. The VD_{ss} Values for Acidic Drugs: Observed, Calculated and Predicted by Developed QSPkR Models

		Model 1				Model 2			
		Calculated		Predicted (Mean Value)[a]		Calculated		Predicted (Mean Value)[a]	
Drug	VD_{obs} (L/kg)	VD_{calc} (L/kg)	FE	VD_{pred} (L/kg)	MFE	VD_{calc} (L/kg)	FE	VD_{pred} (L/kg)	MFE
5-Aminosalicylic acid	0.33	0.18	1.83	0.18	1.83	0.10	3.20	0.10	3.25
5-Fluorouracil	0.23	0.33	1.43	0.34	1.48	0.17	1.39	0.16	1.47
Acetazolamide	0.37	0.29	1.28	0.28	1.32	0.29	1.28	0.29	1.28
Acetylcysteine	0.55	0.58	1.05	0.56	1.02	0.52	1.06	0.36	1.55
Acetylsalicylic acid	0.22	0.17	1.29	0.17	1.29	0.17	1.29	0.16	1.35
Acivicin	0.50	0.19	2.63	0.18	2.78			Outlier	
Amidotrizoate	0.26	0.33	1.27	0.33	1.27	0.25	1.04	0.26	1.01
Amoxicillin	0.25	0.21	1.19	0.21	1.19	0.22	1.12	0.20	1.23
Ampicillin	0.22	0.21	1.05	0.21	1.05	0.22	1.02	0.20	1.09
Atovaquone	0.60	0.27	2.22	0.25	2.40	0.22	2.70	0.21	2.80
Azlocillin	0.26	0.22	1.18	0.21	1.24	0.27	1.04	0.25	1.02
Aztreonam	0.18	0.14	1.29	0.14	1.29	0.17	1.04	0.15	1.23
Betamipron	0.28	0.19	1.47	0.19	1.47	0.15	1.92	0.14	1.94
Bosentan	0.29	0.17	1.74	0.16	1.81	0.28	1.02	0.31	1.06
Bromfenac	0.11	0.21	1.91	0.22	2.00	0.16	1.43	0.16	1.45
Bumetanide	0.16	0.13	1.23	0.12	1.33	0.14	1.11	0.14	1.10
Captopril	0.75	0.71	1.06	0.56	1.34	0.79	1.06	0.42	1.78
Carbenicillin	0.17	0.19	1.12	0.19	1.12	0.21	1.24	0.20	1.16
Cefadroxil	0.23	0.20	1.15	0.20	1.15	0.15	1.55	0.15	1.58
Cefamandole	0.16	0.15	1.07	0.15	1.07	0.17	1.05	0.17	1.07
Cefatrizine	0.22	0.15	1.47	0.14	1.57	0.17	1.29	0.17	1.27
Cefazolin	0.12	0.14	1.17	0.15	1.25	0.17	1.40	0.17	1.44
Cefepime	0.28	0.37	1.32	0.38	1.36	0.46	1.66	0.49	1.75
Cefetamet	0.28	0.19	1.47	0.19	1.47	0.16	1.77	0.16	1.76
Cefixime	0.24	0.17	1.41	0.18	1.33	0.17	1.41	0.18	1.36
Cefmetazole	0.13	0.12	1.08	0.12	1.08	0.17	1.29	0.17	1.33
Cefoperazone	0.17	0.18	1.06	0.18	1.06	0.22	1.29	0.22	1.31
Cefotaxime	0.19	0.17	1.12	0.17	1.12	0.16	1.16	0.16	1.18
Cefotetan	0.13	0.13	1.00	0.13	1.00	0.20	1.51	0.20	1.52
Cefoxitin	0.17	0.18	1.06	0.20	1.15	0.15	1.11	0.15	1.11
Cefprozil	0.21	0.19	1.11	0.19	1.11	0.15	1.37	0.15	1.37
Ceftazidime	0.31	0.18	1.72	0.18	1.72	0.21	1.46	0.22	1.41
Ceftizoxime	0.20	0.17	1.18	0.17	1.18	0.15	1.30	0.16	1.26
Ceftriaxone	0.09	0.15	1.76	0.16	1.88	0.19	2.24	0.11	1.34
Cefuroxime	0.15	0.17	1.13	0.17	1.13	0.15	1.01	0.15	1.03
Cephalexin	0.21	0.20	1.05	0.20	1.05	0.14	1.47	0.15	1.44
Cephaloridine	0.46	0.37	1.24	0.36	1.28	0.50	1.10	0.52	1.13
Cephalothin	0.07	0.17	2.43	0.18	2.57	0.15	2.14	0.16	2.31
Cephapiriin	0.13	0.13	1.00	0.14	1.08	0.16	1.22	0.16	1.25
Cephradine	0.21	0.20	1.05	0.20	1.05	0.14	1.45	0.15	1.42
Cerivastatin	0.33	0.23	1.43	0.23	1.43	0.22	1.53	0.22	1.51
Chlorambucil	0.26	0.22	1.18	0.23	1.13	0.16	1.59	0.17	1.57
Chlorazepate	0.20	0.18	1.11	0.17	1.18	0.20	1.01	0.20	1.01
Chlorpropamide	0.19	0.24	1.26	0.24	1.26	0.19	1.02	0.19	1.02
Cidofovir	0.49	0.32	1.53	0.35	1.40	0.50	1.01	0.47	1.05
Cilastatin	0.15	0.13	1.15	0.13	1.15	0.16	1.05	0.16	1.06
Cilomilast	0.23	0.29	1.26	0.29	1.26	0.25	1.09	0.25	1.09
Clavulanic acid	0.22	0.20	1.10	0.20	1.10	0.21	1.06	0.16	1.38
Dexloxiglumide	0.18	0.24	1.33	0.24	1.33	0.21	1.18	0.22	1.22
Dichloroacetic acid	0.19	0.18	1.06	0.18	1.06	0.10	1.91	0.10	1.96

(Continued)

Table B1. Continued

		Model 1				Model 2			
		Calculated		Predicted (Mean Value)[a]		Calculated		Predicted (Mean Value)[a]	
Drug	VD_{obs} (L/kg)	VD_{calc} (L/kg)	FE	VD_{pred} (L/kg)	MFE	VD_{calc} (L/kg)	FE	VD_{pred} (L/kg)	MFE
Diclofenac	0.22	0.22	1.00	0.22	1.00	0.14	1.56	0.14	1.55
Dicloxacillin	0.11	0.24	2.18	0.24	2.18	0.19	1.72	0.19	1.73
Diflunisal	0.10	0.18	1.86	0.19	1.96	0.15	1.56	0.15	1.55
Doxifluridine	0.28	0.20	1.40	0.20	1.40	0.12	2.43	0.11	2.47
Entacapone	0.27	0.19	1.42	0.19	1.42	0.17	1.55	0.17	1.59
Epristeride	0.54	1.18	2.19	1.24	2.30	0.68	1.25	0.65	1.21
Eprosartan	0.17	0.21	1.24	0.21	1.24	0.17	1.03	0.18	1.04
Ethacrinic acid	0.26	0.24	1.08	0.23	1.13	0.20	1.28	0.20	1.32
Flucloxacillin	0.19	0.23	1.21	0.23	1.21	0.12	1.52	0.12	1.64
Fludarabine	2.20	1.02	2.16	0.82	2.68	1.03	2.13	0.65	3.40
Fluvastatin	0.42	0.65	1.55	0.68	1.62	0.53	1.27	0.53	1.25
Folinic acid	0.25	0.20	1.25	0.20	1.25	0.28	1.12	0.27	1.10
Foscarnet	0.50	0.31	1.61	0.28	1.79	0.33	1.53	0.31	1.62
Fosfluconazole	0.15	0.32	2.13	0.35	2.33	0.34	2.28	0.39	2.62
Fosfomycin	0.32	0.43	1.34	0.47	1.47	0.41	1.29	0.43	1.33
Furosemide	0.12	0.14	1.17	0.15	1.25	0.09	1.37	0.09	1.32
Glimepiride	0.19	0.31	1.63	0.31	1.63	0.44	2.33	0.44	2.33
Glipizide	0.16	0.25	1.56	0.25	1.56	0.41	2.53	0.44	2.77
Glyburide	0.08			Outlier		0.30	3.81	0.34	4.27
Hypericin	0.25	0.35	1.40	0.33	1.32	0.36	1.45	0.35	1.41
Ibuprofen	0.15	0.20	1.33	0.21	1.40	0.23	1.53	0.22	1.47
Indomethacin	0.096			Outlier		0.36	3.72	0.40	4.11
Irbesartan	0.94	0.55	1.71	0.51	1.84	0.79	1.19	0.79	1.19
Isoxicam	0.19	0.19	1.00	0.17	1.12	0.12	1.55	0.12	1.53
Ketoprofen	0.13	0.22	1.69	0.22	1.69	0.26	2.01	0.25	1.96
Ketorolac	0.11	0.24	2.18	0.24	2.18	0.18	1.60	0.18	1.61
Levosimendan	0.24	0.41	1.71	0.42	1.75	0.50	2.10	0.53	2.19
Losartan	0.37	0.31	1.19	0.31	1.19	0.23	1.60	0.23	1.61
Meloxicam	0.15	0.19	1.27	0.19	1.27	0.13	1.15	0.13	1.14
Methicillin	0.32	0.21	1.52	0.21	1.52	0.23	1.40	0.20	1.57
Methohexital	1.10	0.46	2.39	0.43	2.56	0.45	2.45	0.42	2.64
Methotrexate	0.43	0.20	2.15	0.19	2.26	0.27	1.60	0.27	1.60
Mezlocillin	0.09	0.18	2.00	0.18	2.00			Outlier	
Micafungin	0.21	0.18	1.17	0.17	1.24	0.32	1.54	0.35	1.65
Milrinone	0.25	0.44	1.76	0.47	1.88	0.44	1.75	0.45	1.82
Montelucast	0.15	0.24	1.60	0.24	1.60	0.20	1.31	0.21	1.42
Moxalactam	0.17	0.17	1.00	0.17	1.00	0.17	1.01	0.18	1.03
Nafcillin	0.22	0.22	1.00	0.22	1.00	0.25	1.13	0.24	1.10
Nateglinide	0.15	0.25	1.67	0.25	1.67	0.21	1.40	0.21	1.41
Nitrofurantoin	0.57	0.35	1.63	0.34	1.68	0.43	1.31	0.41	1.38
Olsalazine	0.07	0.17	2.43	0.18	2.57	0.18	2.54	0.18	2.50
Oxacillin	0.19	0.22	1.16	0.18	1.06	0.24	1.26	0.24	1.27
Pamidronate	1.80	3.76	2.09	6.40	3.56	1.60	1.13	1.00	1.79
Pantoprazole	0.17	0.17	1.00	0.58	3.41	0.13	1.31	0.39	2.30
Penicillin G	0.24	0.20	1.20	0.20	1.20	0.21	1.15	0.21	1.16
Pentobarbital	0.91	0.54	1.69	0.53	1.72	0.43	2.11	0.39	2.32
Perindoprilat	0.76	0.91	1.20	1.12	1.47	0.77	1.01	0.69	1.10
Phenobarbital	0.54	0.43	1.26	0.41	1.32	0.40	1.33	0.40	1.35
Phenoxymethylpenicillin	0.41	0.19	2.16	0.19	2.16	0.22	1.90	0.18	2.23
Piperacillin	0.27	0.23	1.17	0.23	1.17	0.30	1.11	0.25	1.07
Piretanide	0.17	0.14	1.21	0.14	1.21	0.14	1.19	0.14	1.17
Pravastatin	0.46	0.28	1.64	0.28	1.64	0.44	1.04	0.40	1.16
Probenecid	0.13	0.20	1.54	0.20	1.54	0.13	1.00	0.13	1.01
Propylthiouracil	0.34	0.43	1.26	0.45	1.32	0.28	1.23	0.28	1.22
Pseudohypericin	0.56	0.34	1.65	0.30	1.87	0.37	1.52	0.36	1.57
Quercetin	0.12	0.18	1.50	0.18	1.50	0.23	1.92	0.23	1.91
Repaglinide	0.35	0.21	1.67	0.21	1.67	0.28	1.26	0.28	1.25
Risedronate	6.30	3.32	1.90	1.91	3.30	1.78	3.53	1.14	5.51

(Continued)

 DOI 10.1002/jps

Table B1. Continued

		Model 1				Model 2			
		Calculated		Predicted (Mean Value)[a]		Calculated		Predicted (Mean Value)[a]	
Drug	VD_{obs} (L/kg)	VD_{calc} (L/kg)	FE	VD_{pred} (L/kg)	MFE	VD_{calc} (L/kg)	FE	VD_{pred} (L/kg)	MFE
Roquinimex	0.21	0.23	1.10	0.23	1.10	0.18	1.19	0.18	1.19
Sulbenicillin	0.15	0.12	1.25	0.12	1.25	0.18	1.22	0.17	1.16
Silfadiazine	0.29	0.24	1.21	0.23	1.26	0.24	1.20	0.24	1.21
Sulfamethoxazole	0.30	0.24	1.25	0.24	1.25	0.25	1.19	0.25	1.19
Sulfinpyrazone	0.12	0.12	1.00	0.21	1.75	0.16	1.31	0.68	5.64
Sulfisoxazole	0.17	0.27	1.59	0.27	1.59	0.26	1.50	0.25	1.49
Suramin	0.54	0.23	2.35	0.17	3.18	0.30	1.80	0.31	1.75
Telmisartan	5.30	2.36	2.25	1.70	3.12	2.45	2.17	1.35	3.93
Tenofovir	0.83	1.24	1.49	1.49	1.80	1.79	2.16	1.89	2.28
Tenoxicam	0.19	0.46	2.42	0.52	2.74	0.25	1.34	0.27	1.41
Thiopental	1.20	0.59	2.03	0.54	2.22	0.35	3.39	0.33	3.67
Ticarcillin	0.16	0.19	1.19	0.19	1.19	0.21	1.34	0.23	1.45
Tolbutamide	0.12	0.24	2.00	0.25	2.08	0.27	2.26	0.27	2.29
Tolcapone[b]	0.12	0.18	1.54			0.16	1.31	0.15	1.27
Torsemide	0.21	0.21	1.01			0.30	1.45		
Valproic acid	0.14	0.22	1.58			0.18	1.26		
Valsartan	0.22	0.23	1.06			0.27	1.24		
Warfarin	0.13	0.21	1.62			0.27	2.05		
Zoledronate	0.60			Outlier		1.70	2.83		
Zonampanel	0.19	0.19	1.00			0.19	1.01		

[a]The values are averaged from a few predictions, as each compound randomly falls in several test sets.
[b]The last few compounds did not fall in any test set.

REFERENCES

1. Obach RS, Lombardo F, Waters NJ. 2008. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab Dispos 36:1385–1405.
2. Rowland M, Tozer TN. 2010. Multiple-dose regimens. In Clinical pharmacokinetics and pharmacodynamics: Concepts and applications; Rowland M, Tozer TN, Eds. 4th ed. Baltimore Maryland: Lippincott Williams & Wilkins, pp 293–329.
3. Mager DE. 2006. Quantitative structure–pharmacokinetic/pharmacodynamic relationships. Adv Drug Deliv Rev 58:1326–1356.
4. Fagerholm U. 2007. Prediction of human pharmacokinetics—Evaluation of methods for prediction of volume of distribution. J Pharm Pharmacol 59:1181–1190.
5. Gobburu J, Shelver W. 1995. Quantitative structure–pharmacokinetic relationships of beta blockers derived using neural networks. J Pharm Sci 84:862–865.
6. Van der Graaf PH, Nilsson J, Van Schaick EA, Danhof M. 1999. Multivariate quantitative structure–pharmacokinetic relationships (QSPKR) analysis of adenosine A1 receptor agonists in rat. J Pharm Sci 88:306–312.
7. Mager D, Jusko W. 2002. Quantitative structure–pharmacokinetic/pharmacodynamic relationships of corticosteroids in man. J Pharm Sci 91:2441–2451.
8. Turner JV, Maddalena DJ, Culter DJ, Agatonovic-Kustrin S. 2003. Multiple pharmacokinetic parameter prediction for a series of cephalosporins. J Pharm Sci 92:552–559.
9. Karalis V, Tsantili-Kakoulidou A, Macheras P. 2003. Quantitative structure–pharamacokinetic relationships for disposition parameters of cephalosporins. Eur J Pharm Sci 20:115–123.
10. Chee NN, Xiao Y, Putnam W, Lum B, Tropsha A. 2004. Quantitative structure–pharamacokinetic parameters relationships analysis of antimicrobial agents in human using simulated annealing K-nearest neighbor and PLS amalysis methods. J Pharm Chem 93:2535–2544.
11. Øie S, Tozer TN. 1979. Effect of altered plasma protein binding on apparent volume of distribution. J Pharm Sci 68:1203–1205.
12. Lombardo F, Obach RS, Shalaeva MY, Gao F. 2002. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. J Med Chem 45:2867–2876.
13. Lombardo F, Obach RS, Shalaeva MY, Gao F. 2004. Prediction of volume of distribution values in humans for neutral and basic drugs. 2. Extended data set and leave-class-out statistics. J Med Chem 47:1242–1250.
14. Sui X. Sun J, Li H, Wang Y, Liu J, Liu X, Zhang W, Chen L, He Z. 2009. Prediction of volume of distribution values in human using immobilized artificial membrane partitioning coefficients, the fraction of compound ionized and plasma protein binding data. Eur J Med Chem 44:4455–4460.
15. Ghafourian T, Barzeger-Jalali M, Hakimiha N, Cronin MTD. 2004. Quantitative structure–pharamacokinetic relationship modeling: Apparent volume of distribution. J Pharm Pharmacol 56:339–350.
16. Ghafourian T, Barzeger-Jalali M, Dasmalchi S, Khavari-Khosarani T, Hakimiha N, Nokhodchi A. 2006. QSPR models for the prediction of apparent volume of distribution. Int J Pharm 319:82–97.
17. Karalis V, Tsantili-Kakoulidou A, Macheras P. 2002. Multivariate statistics of disposition pharmacokinetic parameters for structurally unrelated drugs used in therapeutics. Pharm Res 19:1827–1934.
18. Wajima T, Fukumura K, Yano Y, Oguma T. 2003. Prediction of human pharmacokinetics from animal data and molecular structural parameters using multivariate regression analysis: Volume of distribution at steady state. J Pharm Pharmacol 55:939–949.

19. Turner JV, Maddalena DJ, Culter DJ. 2004. Pharmacokinetic parameter prediction from drug structure using artificial neural networks. Int J Pharm 270:209–219.
20. Gleeson MP, Waters NJ, Paine SW, Davis AM. 2006. *In silico* human and rat Vss quantitative structure–activity relationship models. J Med Chem 49:1953–1963.
21. Lombardo F, Obach RS, DiCapua FM, Bakken GA, Lu J, Potter DM, Gao F, Miller MD, Zhang Y. 2006. A hybrid mixture discriminant amalysis–random forest computational model for prediction of volume of distribution of drugs in human. J Med Chem 49:2262–2257.
22. Berrelini G, Springer C, Waters NJ, Lombardo F. 2009. *In silico* prediction of volume of distribution in human using linear and nonlinear models on a 669 compound data set. J Med Chem 52:4488–4495.
23. Poulin P, Theil FP. 2002. Prediction of pharmacokinetics prior to *in vivo* studies. 1. Mechanism-based prediction of volume of distribution. J Pharm Sci 91:129–156.
24. http://www.drugbank.ca Last accessed 11 November 2011.
25. http://www.chemicalbook.com Last accessed 11 November, 2011.
26. Hall LH, Kier LB. 1999. Molecular connectivity chi indices for database analysis and structure–property modeling. In Topological indices and related descriptors in QSAR and QSPR; Devillers J, Balaban A, Eds. London, UK: Gordon and Breach, pp 307–360.
27. Hall LH, Kier LB. 1999. Electrotopological state: Structure modeling for QSAR and database analysis. In Topological indices and related descriptors in QSAR and QSPR; Devillers J, Balaban A, Eds. London, UK: Gordon and Breach, pp 491–562.
28. Leardi R, Boggia R, Terrile M. 1992. Genetic algorithms as a strategy for feature selection. J Chemom 6:267–281.
29. Friedman J. 1988. Multivariate adaptive regression splines. Technical report No. 102. Stanford University, Stanford, California: Laboratory for Computational Statistics, Department of Statistics.
30. Yang X, Gandhi YA, Duignan DB, Morris ME. 2009. Prediction of biliary excretion in rats and humans using molecular weight and quantitative structure–pharmacokinetic relationships. AAPS J 11:511–525.
31. Wingen F, Schmahl D. 1987. Pharmacokinetics of the osteotropic diphosphonate 3-amino-1-hydroxypropane-1,1-diphosphonic acid in mammals. Arzneimittel Forschung-Drug Research 37:1037–1042.
32. Van Beek E, Löwik C, Van der Pluijm G, Papapoulos S. 1999. The role of geranylgeranylation in bone resorption and its suppression by bisphosphonates in fetal bone explants *in vitro*: A clue to the mechanism of action of nitrogen-containing bisphosphonates. J Bone Miner Res 14:722–729.
33. Keating MJ, Talpaz KM, Redman J, Koller C, Barlogie B, Velasquez W, Plunkett W, Freireich EJ, McCredie KB. 1989. Fludarabine: A new agent with major activity against chronic lymphocytic leukemia. Blood 74:19–25.
34. Kragh-Hansen U, Riisom T. 1976. Complexes of aliphatic sulfates and human-serum albumin studied by 13C nuclear-magnetic-resonance spectroscopy. Eur J Biochem 70:15–23.
35. Kier LB, Hall LH. 2000. Intermolecular accessability: The meaning of molecular connectivity. J Chem Inf Comput Sci 40:792–795.
36. McGovern JP, Pratt EA, Belt KJ, Taylor SA, Benjamin RS, Ardalan B, Ohnuma T. 1985. Pharmacokinetic and biochemical studies on acivicin in phase I clinical trials. Cancer Res 45:4460–4463.
37. Newton PN, Barnes KI, Smith PJ, Evans AC, Chierakul W, Ruangveerayuth R, White NJ. 2006. The pharmacokinetics if intravenous artesunate in adults with severe falciparum malaria. Eur J Clin Pharmacol 62:1003–1009.
38. Li Q, Xie LH, Haeberle A, Zhang J, Weina P. 2006. The evaluation of radiolabeled artesunate on tissue distribution in rats and protein binding in humans. Am J Trop Med Hyg 75:817–826.
39. Jaber LA, Antal EJ, Slaughter RL, Welshman IR. 1993. The pharmacokinetics and pharmacodynamics of 12 weeks of glyburide therapy in obese diabetics. Eur J Clin Pharmacol 45:459–463.
40. Dockens RC, Santone KS, Mitroka JG, Morrison RA, Jemal M, Green DS, Barbhaiya RH. 2000. Disposition of radiolabeled ifetroban in rats, dogs, monkeys, and humans. Drug Metab Dispos 28:973–980.
41. Baber N, Halliday L, Sibeon R, Littler T, Orme ML. 1978. The interaction between indomethacin and probenecid. A clinical and pharmacokinetic study. Clin Pharmacol Ther 24:298–307.
42. http://www.druginfosys.com Last accesed 11 November 2011.
43. Pancoast SJ, Neu HC. 1978. Kinetics of mezlocillin and carbenicillin. Clin Phramacol Ther 24:108–116.
44. Verbist L, Tjandramaga TB, Vebesselt R, de Schepper PJ. 1982. Pharmacokinetics of mezlocillin. Comparison with ampicillin and influence of probenecid. Nouv Presse Med 11:347–352.
45. Nezasa K, Higaki K, Yukawa T, Inazawa K, Hasegawa H, Nakano M. 2002. Liver-specific distribution of rosuvastatin in rats: Comparison with pravastatin and simvastatin. Drug Metab Dispos 30:1158–1163.
46. Bergman E, Forsel P, Tevell A, Persson EM, Hedeland M, Bondesson U, Knutson I, Lennernäs H. 2006. Biliary secretion of rosuvastatin and bile acids in humans during the absorption phase. Eur J Pharm Sci 29:205–214.
47. Martin PD, Warwick MJ, Dane AL, Brindley C, Short T. 2003. Absolute oral bioavailability of rosuvastatin in healthy white adult male volunteers. Clin Ther 25:2553–2563.
48. White CM. 2002. A review of the pharmacologic and pharmacokinetic aspects of rosuvastatin. J Clin Pharmacol 42:963–970.
49. Davies NM, Skjodt NM. 2000. Choosing the right nonsteroidal anti-inflammatory drug for the right patient: A pharmacokinetic approach. Clin Pharmacokinet 38:377–392.
50. Weiss HM, Pfaar U, Schweitzer A, Wiegand H, Skerjanec A, Schran H. 2008. Biodistribution and plasma protein binding of zoledronic acid. Drug Metab Dispos 36:2043–2049.
51. Li EC, Davis LE. 2003. Zoledronic acid: A new parenteral bisphosphonate. Clin Ther 25:2669–2708.
52. Hoke JF, Dyker AG, Barnaby RJ, Lees KR. 2000. Pharmacokinetics of a glycine site antagonist (gavestinel) following multiple dosing in patients with acute stroke. Eur J Clin Pharmacol 55:867–872.
53. Wilson TW, Quest DW. 2000. Ridogrel: An antiplatelet agent with antihypertensive properties. Cardiovasc Drug Rev 18:222–231.
54. Heykants J, Van Peer A, Woestenborghs R, Geuens I, Rombaut N, Vanden Bussche G. 1985. Pharmacokinetics and bioavailability of levocabastine (R 50547) in man. Arch Int Pharmacodyn Ther 274:329–330.
55. http://www.drugs.com Last accesed 11 November 2011.