Predicting Total Clearance in Humans from Chemical Structure

Melvin J. Yu

Eisai Incorporated, 4 Corporate Drive, Andover, Massachusetts 01810

Received January 19, 2010

A conceptually simple, fully in silico model to predict total clearance of new compounds in humans is described. Based on the premise that similar molecules will exhibit similar pharmacokinetic properties, we used a *k*-nearest-neighbors (*k*NN) technique to predict total clearance by comparison with known reference agents. Molecular similarity was defined using readily calculated one- and two-dimensional molecular descriptors, and the reference set was obtained by combining the Obach and Berellini sets of human pharmacokinetic data. Neutral molecules and drugs whose biological activity is associated with a metal center were removed from the combined set. The remaining 462 compounds were partitioned into a training and external test set of 370 and 92 compounds, respectively. For acids, bases, zwitterions, and quaternary ammonium/pyridinium ions, average prediction accuracy was within two-fold of observed for the external test set ($n = 92$). Using a collection of 20 drugs from the literature with ≥3 preclinical animal species allometric scaling data, accuracy of the in silico *k*NN model was not as good as the rule of exponents, but better than simple allometry (SA), and approached that of combination multiexponential allometry (ME) as defined by the number of predictions with ≤50% error. For a collection of 18 drugs with two species (rat−dog) data, the *k*NN model outperformed both SA and combination ME using the same performance standard. Since the model is fully in silico and, therefore, capable of generating total clearance predictions in the absence of any experimental data, it can be used to help guide early drug discovery research efforts, such as virtual compound library screening, and analogue prioritization prior to chemical synthesis and biological evaluation. Model validation was accomplished using the external test set, three- and five-fold cross-validation and two different *y*-randomization techniques (*y*-shuffling and random number pseudodescriptors).

INTRODUCTION

The ability to easily predict a compound's pharmacokinetic profile in humans from a two-dimensional chemical structure has substantial potential to broadly impact preclinical discovery efforts from guiding virtual compound library design during lead identification to facilitating analog prioritization during lead optimization. By helping medicinal chemists prioritize new compounds to those most likely to exhibit a desirable pharmacokinetic profile before chemical synthesis and biological evaluation, both speed and efficiency can potentially be increased. Thus, to maximize impact on early stage drug discovery, we established the following criteria for moving forward: (1) The models must be fully in silico and may not rely on any experimental data for input. (2) The molecular descriptors must not require three-dimensional chemical structures and cannot necessitate more computationally intensive semiempirical or ab initio calculations (e.g., thermochemical parameters and/or molecular orbital energies). Under these conditions the model would be capable of rapidly screening virtual compound libraries and supporting both lead identification and lead optimization efforts.

Total clearance (CL) is a critical pharmacokinetic parameter that influences both dose size and dosing regimen. As a result, predicting CL in humans for new chemical entities represents an important step in selecting a first-in-human dose.[1,2] Experimental methods to predict this pharmacokinetic property include interspecies scaling,[3] which is resource intensive, time-consuming, and therefore not appropriate for compound library screening, and in vitro-in vivo extrapolation,[4,5] which requires a variety of experimental values. Given the importance of this parameter and the expense associated with conducting multiple pharmacokinetic studies, particularly in non-rodent species, a number of efforts to predict total clearance in human using allometric methods with as few as one or two preclinical animal species have been reported.[6,7,8] However, these methods have not been widely accepted,[9] and even so would still not be appropriate for virtual compound screening or testing large numbers of analogues. Computational methods to predict intrinsic,[10,11] hepatic,[12] biliary,[13] renal[14] and total[15] clearance have been reported in the literature. Yet, despite these reports there remains a clear need for simple, fully in silico models that could be used to help guide early drug design efforts.

The process of drug clearance involves a number of mechanistic pathways including cytochrome P450-mediated metabolism, passive and active transport processes, such as biliary excretion and renal secretion, and phase II metabolism, such as glucuronidation. Membrane permeability plays an important role, particularly in cases where the drug or drug candidate exhibits low metabolic and biliary clearance. For example, of the various physicochemical properties examined by Varma et al.,[7] ionization state, lipophilicity, and polar descriptors were reported to be the most important determinants for renal clearance. A qualitative heuristic giving rise to the "golden triangle" employed molecular weight (MW) and lipophilicity as the most important factors for simultaneously optimizing CL and oral absorption of drug

10.1021/ci1000295
Published on Web 07/09/2010

candidates,[16] reinforcing the importance of MW (either directly or through a surrogate) and polarity in these type of models. Similarly, these two were reported to be the most important factors in determining permeability of drug candidates.[17] Molecular weight, in particular, is a commonly identified parameter used to predict biliary excretion of drugs.[6,18] Hosea et al.,[19] categorized compounds based on the predominant clearance mechanism (e.g., P450-mediated vs non-P450 processes and passive vs active transport) when comparing prediction methods. Thus, different molecular properties may play more or less important roles depending on the drug's predominant clearance mechanism.[20]

If similar compounds exhibit similar properties[21−23] and physical properties strongly influence pharmacokinetic behavior,[20] then it follows from transitive reasoning that similar compounds may exhibit similar pharmacokinetic behavior. Based on this premise, we explored the possibility that total clearance of new compounds could be predicted by comparative analysis with known reference agents. Application of this approach would require: (1) a comprehensive reference data set of human clearance values and (2) a collection of appropriate molecular descriptors to define compound similarity. To satisfy the first requirement, we utilized the set of human intravenous pharmacokinetic data published by Obach and co-workers for 670 drugs and drug candidates.[24] In addition, Berellini and co-workers published data for an additional 29 that were not included in the original 670.[25] To our knowledge, the combination of these two represents the largest public domain compilation of human pharmacokinetic data reported to date that encompasses a wide range of chemotypes including macrolides, peptides, heterocycles, steroids, and polysaccharides. To address the second requirement, we developed an algorithm to stochastically sample subsets of m descriptors from a pool of M descriptors and integrated it with a *k*-nearest neighbors (*k*NN) quantitative structure−property relationship (QSPR) technique[26−29] to build conceptually simple regression models based on compound similarity.

This report details a successful implementation of this procedure to provide a fully in silico *k*NN model for predicting total clearance in humans. Since the model does not require any experimental data for input, it is amenable to guiding early drug discovery efforts (e.g., lead identification and optimization) where little to nothing may be known experimentally about a compound beyond its two-dimensional chemical structure.

METHODS

The Obach and Berellini intravenous pharmacokinetic data sets were combined to afford a total of 699 nonduplicate compounds. Two drugs (lithium carbonate, Chemical Abstract Service (CAS) 554−13−2, and carboplatin, CAS 41575−94−4) whose biological activity is associated with a metal center were removed. The remaining 697 compounds were stripped of counterions, if any, prior to molecular descriptor calculation by Pipeline Pilot.[30] The compounds were also manually classified into one of five charge-type chemical groups (acid, base, neutral, zwitterion, and quaternary ammonium/pyridinium ion). For purposes of this study, acids and bases were defined as being greater than 50% ionized at pH = 7. In this regard, imidazoles (pK_a of the conjugate acid at 25 °C is 6.99)[31] were classified as neutrals. The hierarchical classification scheme used in this study can be found in the Supporting Information.

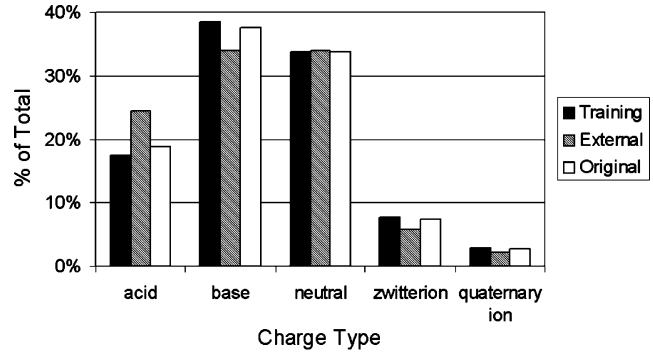

Figure 1. Compound representation in the original (n = 697), training (n = 558), and external test (n = 139) sets by charge type.

Table 1. Full In Silico Descriptor Pool Calculated by Pipeline Pilot

name	details
MW	calculated molecular weight
AlogP	lipophilicity
Sol	solubility
HBA	no. hydrogen-bond acceptors
HBD	no. hydrogen-bond donors
PSA	total polar surface area (N, O, P, S atoms considered polar)
kappa3	Hall and Kier shape index
Jx	Balaban topological index
chiv3c	Hall and Kier connectivity index (cluster)
RotB	no. rotatable bonds
delta	logD (pH 7.4) − logD (pH 1.2)
delta2	AlogP − logD (pH 7.4)

The compound set was then randomly divided into two groups in a 4:1 ratio. The larger group consisting of 558 compounds was used as the training set. The smaller group of 139 compounds was set aside and used as an external test set. The distribution of compound classes was essentially identical for the training and external test sets, and both were representative of the original 697 (Figure 1).

Table 1 lists the pool of 12 descriptors employed in this study. The balaban distance-based topological index (Jx) is the average distance sum connectivity value.[32] The Kier and Hall kappa3 descriptor is a shape index that reflects the degree of branching at the center of a molecule and assumes all atoms are equivalent.[33] The chiv3c connectivity descriptor refers to the chiv order 3 cluster type.[33] In addition to the 10 primary descriptors, two secondary ones were used. Delta and delta2 were calculated as the difference of logD(7.4) − logD(1.2) and AlogP − logD(7.4), respectively, and would be sensitive to changes in pH-dependent ionization states. Inspection of the scatterplot matrix for the 12 descriptors reveals little cross-correlation as indexed by the shape of the 95% bivariate normal density ellipse overlaid in each scatterplot (Figure 2).

Each model was built using a subset of descriptors selected from the pool of 12. Prediction accuracy was determined using a leave-one-out (LOO) procedure: One compound at a time was systematically removed from the training set, and a value assigned based on proximity to the nearest neighbors found among the remaining N-1 compounds (N is the number

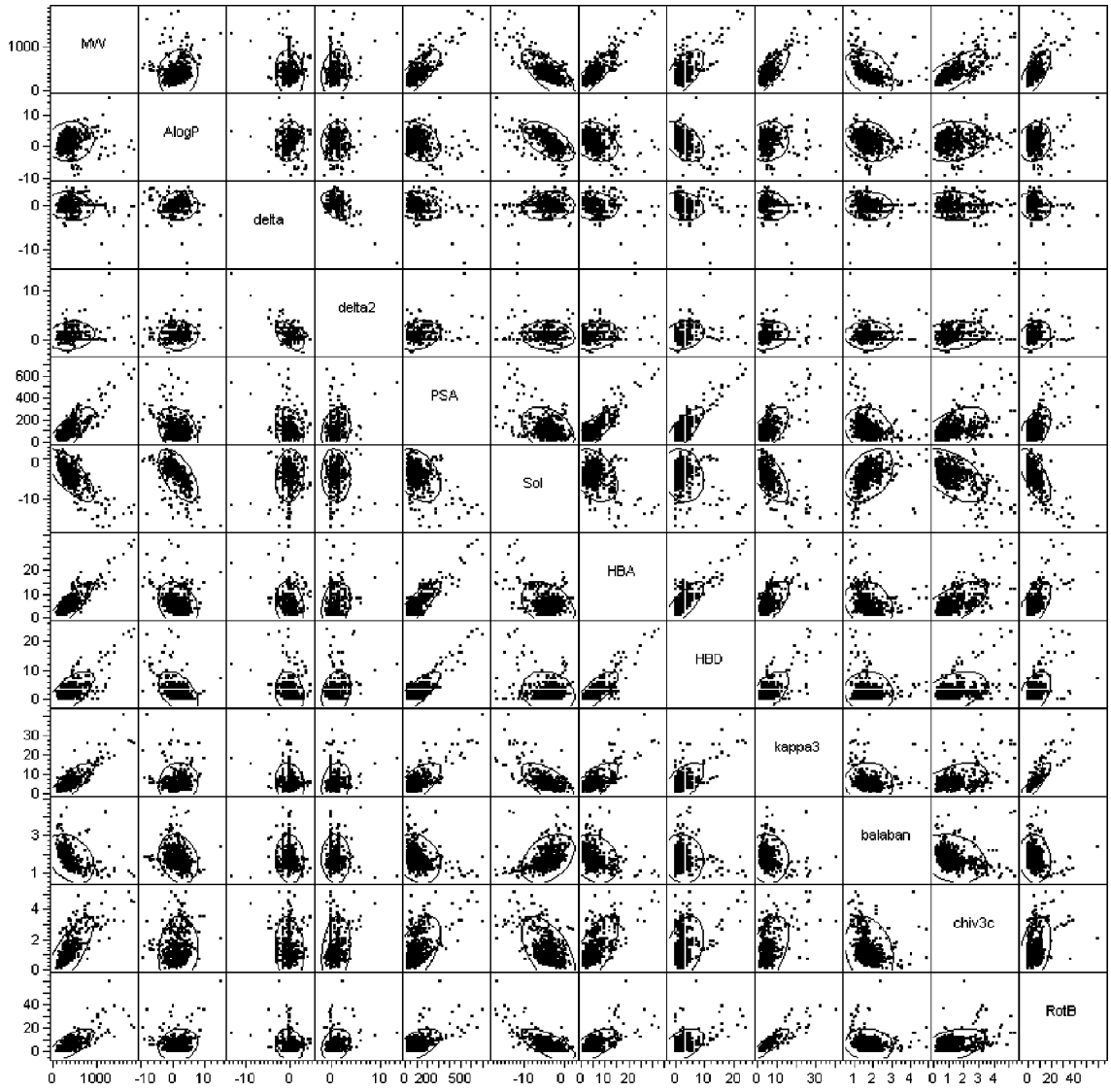

Figure 2. Scatter plot matrix of the 12 calculated molecular descriptors used in this study. Overlaid in each scatter plot is the 95% bivariate normal density ellipse.

of training set compounds). The process was iterated N times so that each compound was predicted once. A new subset of descriptors was then chosen, and the entire process repeated until prediction accuracy no longer improved. The descriptor subset that gave the best performance against the internal training set by the LOO procedure was then used to predict values for the external test compounds. The algorithm flow scheme can be found in the Supporting Information.

Model performance was assessed by calculating the prediction fold error and the absolute average fold error between predicted and observed values as defined by Mahmood[14] and Yang[6] (geometric mean fold error, GMFE, eqs 1−2). To simplify comparison, fold error was taken as the ratio $CL_{predicted}/CL_{observed}$, only if $CL_{predicted}$ was greater than $CL_{observed}$. If the opposite was true, then the reciprocal was used so that fold error was always ≥ 1.[5,13] The number of correct predictions within a two- and three-fold error window was also summed and used to index model accuracy. To compare model performance with in vivo interspecies scaling methods, the percent error and the percent error classification scheme as defined by Goteti, Garner, and Mahmood[34] was used with a threshold value of 50%, representing the maximum acceptable prediction error as proposed by Mahmood (eq 3).[14]

$$\text{fold error} = \frac{CL_{predicted}}{CL_{observed}} \tag{1}$$

$$GMFE = 10^{\sum|\log(\text{fold error})|/N} \tag{2}$$

$$\%\text{error} = \frac{CL_{observed} - CL_{predicted}}{CL_{observed}} \times 100 \tag{3}$$

where

$$\text{very good to fair prediction} = |\%\text{error}| \leq 50\%$$
$$\text{not good to poor prediction} = |\%\text{error}| > 50\%$$

Distance to a reference compound was defined by the difference sum of squares (DSS) as expressed by eq 4, where v_t = ith descriptor value for the test compound, and v_r = ith descriptor value for the reference compound in the training set. Compound pairs that exhibited the smallest DSS were matched, and the corresponding CL value returned. A

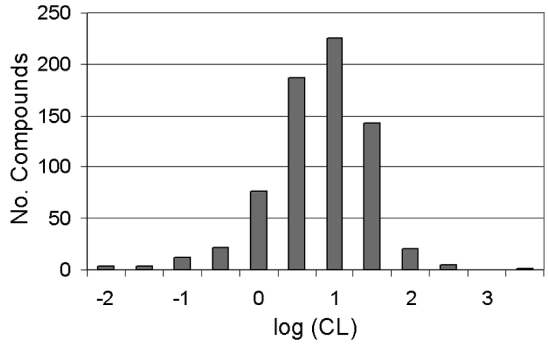

Figure 3. Histogram of observed clearance vs number of compounds for the original data set ($n = 697$).

diagrammatic representation of the algorithm used for model development can be found in the Supporting Information.

$$\mathrm{DSS} = \sum_{i=1}^{j} (v_t - v_r)_i^2 \tag{4}$$

Since descriptor values in the pool exhibit different orders of magnitude (e.g., maximum MW is 1875, whereas maximum solubility is 2.9), the numbers were normalized by descriptor to cover the range 0−10 inclusive. Without normalization, the DSS value would have been dominated by those descriptors (e.g., MW) exhibiting the largest absolute values.

An inverse weighting method was used in building weighted *k*NN models (eq 5). To avoid undefined weights in cases where DSS fell to 0, the parameter ε was introduced and set to a value of 1.

$$\text{Prediction} = \frac{\sum_{i=1}^{k} w_i \cdot \mathrm{CL}_i}{\sum_{i=1}^{k} w_i},$$

$$\text{where } w_i = \frac{1}{\mathrm{DSS}_i + \varepsilon} \text{ and } \mathrm{CL}_i = \text{clearance for the } x_i^{\text{th}} \text{ nearest neighbor in the training set} \tag{5}$$

A general limitation to the nearest-neighbors approach is the need for adequate coverage of response variable space by the training set both in terms of uniformity and density. The combined Obach and Berellini data sets do provide good representation for the most part, although examples at the extreme high and low end of the spectrum are less well represented (Figure 3). In particular, human total clearance values greater than 100 mL/min/kg and less than 0.1 mL/min/kg are poorly represented. This potential shortcoming, however, is offset by the ease with which new reference agents (either in-house proprietary data or newly published public domain data) could be added to the training set.

The algorithms described in this paper were written in C and compiled with a GNU C compiler. All calculations were performed on a standard windows computer.

RESULTS AND DISCUSSION

Model Development. A preliminary one nearest-neighbor test run for predicting total clearance in humans gave a LOO internal training set GMFE of 3.07 ($n = 558$) and an external test set GMFE of 3.04 ($n = 139$). While weak, the correlation between predicted and observed values for both the training and external test sets was statistically significant ($p < 0.0001$). Given the complexity of multiple serial and parallel mechanistic pathways involved in the total clearance of drugs from the body, it is perhaps surprising that the preliminary test run for CL gave a GMFE of 3.07. Nevertheless, encouraged by this initial result, we investigated stratifying the data set by descriptor value. An algorithm was, therefore, devised to band compounds according to either MW or molecular lipophilicity. However, in our hands these modifications only led to modest improvement in prediction accuracy (data not shown) and, as a result, were abandoned in favor of other refinement strategies.

Table 2. Results from LMO Three-Fold Cross Validation

validation set	GMFE
1	2.51 ($n = 186$)
2	2.39 ($n = 186$)
3	2.48 ($n = 186$)
Concatenated sets 1−3	2.46 ($n = 558$)

A consensus approach[35] was next considered where 10 models were built using the internal LOO method. The predictions from each of the models were then averaged, and the mean value used as output. In this manner, the consensus model gave an internal LOO training set GMFE of 2.61 ($n = 558$) with somewhat improved predictive capability as indexed by the external test set GMFE of 2.65 ($n = 139$). The correlation between predicted and observed values was again found to be statistically significant ($F_{\text{ratio}} = 38.2$, $p < 0.0001$). As an initial test of the approach, the observed CL values in the 558 compound training set were randomly shuffled, and the entire model building procedure repeated (including automatic descriptor selection). This *y*-shuffling and model building process were iterated a total of 25 independent times, and the results averaged to afford a LOO training set GMFE of 3.37 ± 0.08 and an external test set GMFE of 3.32 ± 0.14 (mean ± standard deviation). Thus, there is a clear performance difference between the original and the *y*-shuffled models. However, the original GMFE values were still relatively high suggesting that an important aspect or molecular feature had been overlooked in the initial analysis.

To help identify the missing component, a leave-many-out (LMO) three-fold cross-validation was explored by randomly dividing the original 558 compound pool into three groups of 186 compounds each. One group was set aside, and a consensus model built using the remaining two. The resulting model was then used to predict CL values for the excluded group. This was iterated a total of three times so that each of the 558 compounds in the original set was predicted once. In addition to examining the individual set results, the predictions were concatenated to provide an overall composite assessment. In all cases, the GMFE results were consistent with that found earlier using the LOO method (Table 2).

Examining the concatenated LMO three-fold cross-validation set predictions by charge-type chemical class revealed some striking differences. For example, neutral compounds were the least well predicted with a GMFE of 2.99, while the zwitterions and the quaternary salts were the best (Table

Table 3. Charge Type Analysis of Concatenated Sets 1−3

charge type	GMFE
acid	2.42 ($n = 97$)
base	2.22 ($n = 214$)
neutral	2.99 ($n = 188$)
zwitterion	2.02 ($n = 43$)
quaternary ion	1.89 ($n = 16$)

Table 4. LCO Model Performance

charge type	GMFE
acid	2.91 ($n = 97$)
base	2.79 ($n = 214$)
neutral	3.61 ($n = 188$)
zwitterion	2.09 ($n = 43$)
quaternary ion	1.53 ($n = 16$)

Table 5. Model Performance by Charge-Type Chemical Class

chemical class	training set LOO GMFE	external set GMFE
acid	2.46 ($n = 97$)	1.81 (34)
base	2.15 ($n = 214$)	2.02 ($n = 47$)
neutral	3.04 ($n = 188$)	3.47 ($n = 47$)
zwitterion	2.05 ($n = 43$)	1.30 ($n = 8$)
quaternary ion	1.49 ($n = 16$)	1.49 ($n = 3$)

3). The latter two, however, were also the smallest subgroups, and as a result, the data for these must be interpreted with caution. Nevertheless, these results raised the possibility for enhancing prediction accuracy by taking advantage of a charge-type effect.

To confirm this possibility, we employed a leave-class-out (LCO) procedure where all compounds belonging to a particular charge-type chemical class were removed from the original 558 compound training set and used as the validation set. For example, if the training set was comprised of bases, neutrals, zwitterions, and quaternary ions, then it would be used to predict CL for acids. Using this procedure, individual consensus models were built for each of the five charge types. Except for the zwitterion and quaternary ion groups, model performance was uniformly poor, and values for compounds across chemical classes could not be accurately predicted (Table 4). It is interesting to note that the LCO results mirrored those from the LMO charge-type analysis, where neutral molecules were the most difficult to predict and zwitterions and quaternary ions the easiest. Taken together, these results suggest that compounds must first be matched by chemical class before evaluating nearest-neighbor distances.

To test this hypothesis, the acids, bases, neutrals, zwitterions, and quaternary ions from the original 558 compound training set were separated and individually grouped. A consensus model for each of these five was built and used to predict CL for the external test set based on charge type, e.g., the model trained using acids was used to predict CL for acids in the external set. As shown in Table 5, with one exception these chemical class-based models delivered improved performance with the external test set GMFE values ranging from a low of 1.49 for quaternary ions ($n = 3$) to a high of 3.47 for neutral molecules ($n = 47$). The training set LOO cross-validation GMFE values with larger numbers of compounds in each class were found to be comparable to those observed for the external test set.

Table 6. Effect of Parameter k on Nearest-Neighbors Model Performance

	training set LOO GMFE ($n = 370$)		external set GMFE ($n = 92$)	
k	nonweighted	weighted	nonweighted	weighted
1	2.26	−	1.90	−
3	2.21	2.19	1.92	1.90
10	2.32	2.26	1.92	1.89
20	−	2.31	−	1.92
30	−	2.35	−	1.94
entire training set	−	2.45	−	2.06

Consistent with the LMO and LCO results, neutral molecules could not be accurately predicted suggesting a limitation to this particular nearest-neighbors approach.

Given this constraint, the acid, base, zwitterion, and quaternary ion predictions were concatenated for the training and external sets to provide the corresponding composite sets for overall assessment. By excluding neutral molecules, GMFE values of 2.18 ($n = 370$) and 1.85 ($n = 92$) were obtained for the LOO internal training and external test sets, respectively.

Final Model. To balance model complexity with prediction accuracy, we investigated the possibility of generating a *single* unified model that could be used to predict CL irrespective of compound charge type. Thus, based on the chemical class effect we modified the algorithm so that new compounds could only be matched with those in the training set that belong to the same chemical class (see Supporting Information for the modified algorithm). This would necessitate the use of only one set of molecular descriptors that would be applicable to all molecular charge types (with the exception of neutral molecules), rather than the paradigm of one descriptor set for each chemical class.

To test this approach, neutral molecules were removed from the original training array to afford a redacted set of 370 compounds comprised of the remaining charge-type chemical classes. Neutral molecules were removed from the original external set to provide a redacted test set of 92 compounds. By excluding neutral molecules from consideration, the one nearest-neighbor model exhibited good performance with GMFE values of 1.90 ($n = 92$) and 2.26 ($n = 370$) for the external test and internal training sets, respectively. Thus, the revised algorithm allows a single model to be cast, whose performance is equivalent to that of the four individual chemical class-based models described earlier.

We next investigated the effect of parameter k on the nearest-neighbors analysis. Up to a value of $k = 10$, there appeared to be little if any difference in model performance. Since there are only 16 quaternary/pyridinium ions in the training set, higher values of k were not explored using a nondistance weighted kNN approach. To explore the performance of a weighted kNN model, we used simple inverse DSS weight values (eq 5) and repeated the model building process with various values of k. As before, performance as measured by average fold error was found to be similar, although there may be a weak trend with larger values of k being associated with a larger average fold error (Table 6).

To help distinguish between the kNN models, we compared performance of the in silico models with that of various interspecies scaling methods. This represents an extreme

Table 7. Number of Compounds from Goteti Set Grouped by % Error and Fold Error[a]

method	number of compounds			
	very good to fair prediction[b]	not good to poor prediction[c]	<two-fold error	<three-fold error
SA[d]	9	11	12	19
ROE[e]	19	1	19	20
ME[f]	9	11	11	16
ME[g]	15	5	15	19
$k = 1$ NN[h]	12	8	13	17
$k = 3$ NN[h]	14	6	15	16
$k = 10$ NN[h]	12	8	14	16
w$k = 3$ NN[i]	12	8	14	16
w$k = 10$ NN[i]	12	8	14	17
w$k = 20$ NN[i]	14	6	15	17
w$k = 30$ NN[i]	12	8	14	17
w$k =$ n NN[j]	10	10	12	17

[a] For various allometric methods (≥3 preclinical animal species) and in silico models ($n = 20$). See Table 10 for list of drugs. [b] Percent error ≤50% (see eq 3). [c] Percent error >50% (see eq 3). [d] Simple allometry. [e] Rule of exponents. [f] Multiexponential allometry. [g] Combination multiexponential allometry (ME applied only when the exponents from SA were >0.7). [h] Unweighted *k*NN model. [i] Distance weighted *k*NN model. [j] Distance weighted *k*NN model and $n =$ entire training set.

comparison since multiple species allometry is typically conducted to support selection and nomination of a clinical candidate, at which time a considerable amount of experimental data is available and, therefore, much is known about the compound. The in silico model, on the other hand, uses only calculated molecular descriptors as input and may be applied in situations where little to nothing might be known experimentally beyond a two-dimensional chemical structure.

Goteti and co-workers recently published a comparative study of four allometric methods using two and more than two preclinical animal species to predict total clearance in humans.[34] Compounds in their data set were comprised of 45 drugs selected from the literature and administered either orally (6 compounds) or intravenously (39 compounds). The authors indicated that these particular drugs were specifically chosen since they lacked significant biliary excretion or renal secretion and covered a range of slopes calculated from simple allometry. To directly compare performance of our fully in silico model with the allometric methods described by Goteti and co-workers for ≥3 preclinical animal species, we narrowed their select list of 45 drugs to only those that were administered intravenously.[36] The neutral molecules were then removed to afford a final list of 20 compounds (5 acids, 11 bases, and 4 zwitterions). Fifteen of these were found to be present in the 370 compound training set and were, therefore, removed from that array prior to running the *k*NN models built earlier.

Using eq 3, the percent prediction error was calculated for each compound. These were then grouped according to the performance criteria proposed by Goteti for allometric method comparison, where ≤50% error was considered a very good to fair prediction and >50% error was regarded as a not good to poor prediction. For the 20 drugs described in this comparative analysis, the *k*NN in silico models outperformed simple allometric scaling (SA) that used ≥3 animal species in terms of the number of very good to fair predictions (Table 7). In addition, performance of the unweighted $k = 3$ and the distance weighted $k = 20$ NN models was similar to that of combination multiexponential (ME) allometry[37] (ME applied only when the exponents of SA were >0.7, SA predictions otherwise) and superior to ME when applied to all drugs irrespective of SA exponent value. Similarly, using fold error as the performance metric for the 20 drugs revealed that the fully in silico models compared favorably with the SA and combination ME allometric methods. However, the in silico model was less accurate than the rule of exponents (ROE) method,[38] which was designed for ≥3 species scaling.

Table 8. Number of Compounds from Goteti Set Grouped by % Error and Fold Error[a]

method	number of compounds			
	very good to fair prediction[b]	not good to poor prediction[c]	<two-fold error	<three-fold error
SA[d]	7	11	10	14
ME[e]	8	10	9	12
$k = 3$ NN[f]	13	5	14	15

[a] For various prediction methods using rat−dog preclinical data and the $k = 3$ NN in silico model ($n = 18$). See Table 11 for list of drugs. [b] Percent error ≤50% (see eq 3). [c] Percent error >50% (see eq 3). [d] Simple allometry. [e] Combination multiexponential allometry (ME applied only when the exponents from SA were >0.7). [f] Unweighted *k*NN model.

Limitations often exist in the pharmaceutical industry regarding the number and type of species to include in allometric scaling for predicting total clearance in humans. As a result, a number of studies have been reported in the literature that explore various scaling methods that require only one or two preclinical animal species.[11,12] Although the accuracy of these approaches has been the subject of recent debate,[39] the issue driving the need for alternative methods is not. We, therefore, compared the performance of our in silico model with two species scaling techniques. Using either percent error or fold error, the in silico model was found to be more accurate than either SA or combination ME for the set of 20 drugs taken from the Goteti study that used preclinical data from rat and dog (Table 8). Although the number of compounds is small, the results are highly promising since the comparison with allometric scaling is extreme, i.e., the two approaches (fully in silico models and interspecies scaling) are at the opposite ends of the spectrum in terms of what must be known experimentally about a compound in order to generate a prediction. Further study with additional drugs will be needed to confirm and extend these initial findings.

Plotting predicted vs observed clearance values for the various methods indicates that more scatter is associated with the in silico predictions relative to those derived from animal allometry for this particular set of drugs (Figure 4). It would be interesting to compare performance of the in silico model with various methods using drugs that undergo significant biliary excretion[40] or renal secretion,[41,42] since clearance predictions in humans for these types of drugs from interspecies scaling, for example, are more difficult or require physiological correction factors. At this time, however, the number of compounds available for such a comparison is very small. Nevertheless, the model is sufficiently accurate to support early drug discovery where pharmacokinetic data

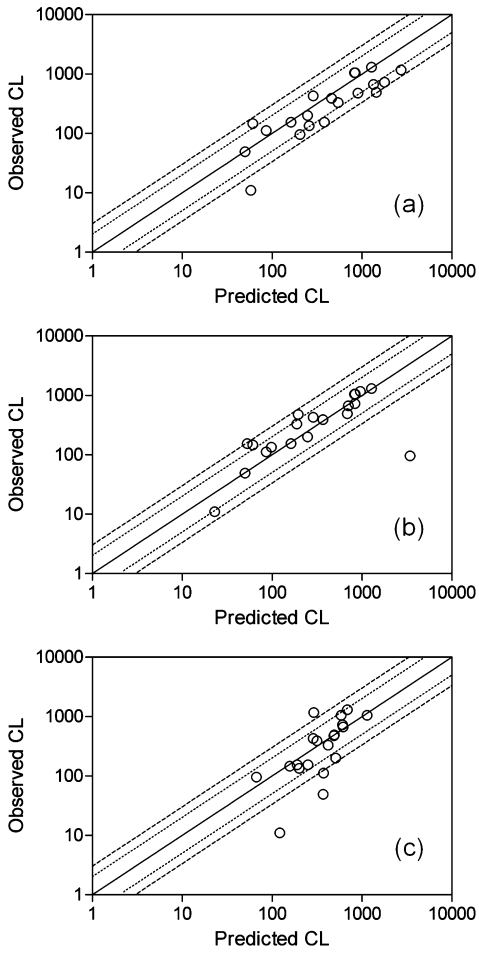

Figure 4. Plot of predicted vs observed clearance values for several allometric methods reported by Goteti[34] (≥3 species), and the unweighted $k = 3$ NN in silico model ($n = 20$). (a) Simple allometry. (b) Combination multiexponential allometry. (c) Unweighted $k = 3$ NN in silico model.

from ≥3 animal species may not be available or experimental data to support in vitro−in vivo extrapolation may be limited. In particular, the in silico model is capable of rapidly predicting total clearance in humans of virtual compounds based solely on two-dimensional chemical structure and may, therefore, find utility in helping guide the design of new analogues *before* chemical synthesis and biological evaluation. Thus, the model may help fill the gap for total clearance projections in humans that currently exists earlier in the drug discovery process. Optimization of pharmacokinetic parameters in animals is important for demonstrating activity in preclinical disease models, but ultimately compounds must be optimized for human use if they are to become clinically useful drugs. Information to help achieve this goal early in the drug discovery process could be used to supplement in vitro metabolic stability and animal pharmacokinetic data and to help scientists make better informed decisions in the design and prioritization of new analogues.

On the basis of parsimony and predictive performance against both the 92 compound external test set and the Goteti set, the unweighted $k = 3$ NN consensus model was selected as the final one. Molecular descriptors used by the final model appear in Table 9. Observed, allometry and in silico clearance values for the Goteti set appear in Tables 10 (≥3 species scaling) and 11 (rat−dog scaling).

Table 9. Molecular Descriptors Used by the Final 10 Member Unweighted $k = 3$ NN Consensus Model

model no.	descriptors[a]
1	HBD, chiv3c
2	AlogP, HBA, chiv3c, RotB, delta2
3	MW, Sol, kappa3, delta
4	HBD, chiv3c
5	JX, delta, delta2
6	MW, AlogP, HBD, chiv3c, RotB
7	PSA, Jx
8	AlogP, PSA, chiv3c, RotB
9	AlogP, HBD, delta2
10	MW, HBD, kappa3, RotB, delta

[a] See Table 1 for details regarding each descriptor.

Inspection of Table 10 reveals that the in silico *k*NN model tended to over predict clearance for this set of compounds. In addition, the error does not appear to be associated with a particular range of exponents from SA. Although the accuracy of the in silico model approaches that of interspecies scaling using simple and multiexponential allometric methods, there were four compounds (20% of total) in particular whose clearance values were poorly predicted with values greater than three-fold from observed (acivicin, gabapentin, iloprost, and caspefungin).

To help elucidate the source of this error, we analyzed the external test set predictions (Figure 5). Within a three-fold window, the model achieved an 86% rate of correct predictions ($n = 92$). However, visual inspection of the plot reveals a number of outliers, particularly CAS 113-45-1, 99-66-1, and 95399-71-6. Of the 92 compounds in the external set, 13 (14%) had a fold error rate greater than 3, which did not appear to be charge-type specific (Table 12). This rate of correct predictions is in line with that reported for drug exposure predictions made prior to the first in-human studies for 35 proprietary Bristol-Myers Squibb compounds.[43]

Plotting fold error against observed CL for the external test set revealed that prediction error appears to be largely associated with the two extremes, as illustrated by the U-shaped data scatter (Figure 6). Compound representation at these points is the lowest (cf. Figures 3 and 6), suggesting an inverse relationship between number of compounds available for comparison in the training set for a particular CL range and average prediction accuracy. Figure 7 illustrates this more clearly in a series of combined two-axes histogram and average fold error plots. In this series, charge types were matched, e.g., acids in the external test set were plotted with the number of acids available in the training set for comparison. Greater prediction accuracy was generally associated with larger numbers of comparator compounds in each CL value range. The one exception was isoglycirrhizinate (CAS 691410-17-0), which was predicted more accurately than might be expected based on the observed trend. Prediction accuracy of new compounds will, therefore, likely be highest for those that fall midrange of the training set CL values where the density of comparator compounds is highest. This trend appears to be general irrespective of charge-type chemical class.

Using this information to investigate the in silico prediction error associated with the Goteti data set, fold error for the four compounds identified earlier was plotted with the training set CL histogram according to charge type (Table

Table 10. Predicted and Observed Clearance Values by Several Allometric Methods[a]

name	observed[b]	exponent[b]	SA[b,c]	ROE[b,d]	ME[b,e]	ME[b,f]	*k*NN[g]
metoprolol	1050	0.428	826	826	3619	826	1140
ofloxacin	146	0.483	61	61	185	61	157
enoxacin	427	0.518	286	286	722	286	286
moxifloxacin	154	0.541	162	162	355	162	251
acivicin	49	0.595	50	50	79	50	369
propranolol	1050	0.662	840	840	246	840	585
minaprine	201	0.668	248	248	82	248	511
gabapentin	112	0.675	86	86	851	86	373
morphine	1300	0.684	1272	1272	1147	1272	687
irinotecan	392	0.704	458	458	367	367	315
carumonem	96	0.773	204	75	3424	3424	67
amlotriptan	667	0.784	1342	830	706	706	616
erythromycin	492	0.807	1445	418	688	688	492
fentanyl	730	0.812	1782	868	832	832	608
lamifiban	134	0.886	259	134	98	98	200
iloprost	1169	0.923	2715	801	957	957	290
quinidine	329	0.944	543	297	188	188	420
furosemide	154	1	381	163	53	53	190
actisomide	475	1.066	899	531	195	195	487
caspefungin	11	1.13	58	10.9	23	23	122

[a] Using ≥3 preclinical animal species reported by Goteti and predicted values from the final unweighted $k = 3$ NN in silico model ($n = 20$). Clearance values eported as mL/min and assumes a 70 kg person. [b] Observed and allometry values taken from Goteti et al. See ref 34. [c] Simple allometry. [d] Rule of exponents. [e] Multiexponential allometry. [f] Combination multiexponential allometry (ME applied only when the exponents from SA were >0.7). [g] $k = 3$ NN model.

Table 11. Predicted and Observed Clearance Values by Several Allometric Methods[a]

name	observed[b]	exponent[b]	SA[b,c]	ME[b,d]	*k*NN[e]
enoxacin	427	0.225	131	131	286
moxifloxacin	154	0.413	92	92	251
metoprolol	1050	0.459	1070	1070	1140
ofloxacin	146	0.49	65	65	157
quinidine	329	0.587	232	232	420
acivicin	49	0.627	45	45	369
minaprine	201	0.66	234	234	511
erythromycin	492	0.666	767	767	492
gabapentin	112	0.691	103	103	373
propranolol	1050	0.754	1600	974	585
amlotriptan	667	0.763	1162	676	616
carumonem	96	0.766	257	1225	67
fentanyl	730	0.821	1875	55	608
morphine	1300	0.829	2800	154	687
lamifiban	134	0.925	438	685	200
actisomide	475	1.022	682	186	487
furosemide	154	1.05	554	55	190
iloprost	1169	1.46	9784	5790	290

[a] Using rat−dog preclinical data reported by Goteti and predicted values from the final unweighted $k = 3$ NN in silico model ($n = 18$). Clearance values reported as mL/min and assumes a 70 kg person. [b] Observed and allometry values taken from Goteti et al. See ref 34. [c] Simple allometry. [d] Combination multiexponential allometry (ME applied only when the exponents from SA were >0.7). [e] $k = 3$ NN model.

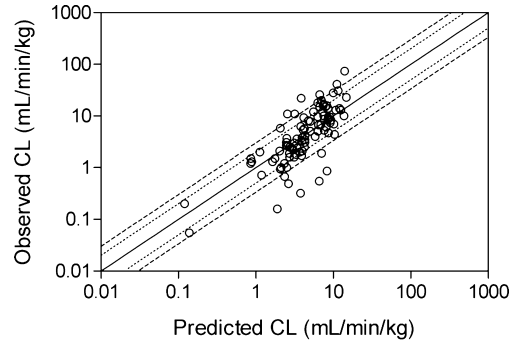

Figure 5. Predicted vs observed clearance (mL/min/kg) for external test set ($n = 92$) using final nonweighted $k = 3$ NN model. The solid line represents unity, the dotted line represents the two-fold error limits, and the dashed line represents the three-fold error limits.

Table 12. Compounds from External Test Set with Greater than Three-Fold Prediction Error[a]

CAS	charge type	predicted clearance[b]	observed clearance[b]	fold error
113-45-1	base	6.6	0.55	12
99-66-1	acid	1.9	0.16	12
95399-71-6	acid	3.8	0.32	12
77-37-2	base	8.3	0.86	9.6
73232-52-7	quaternary ion	3.9	22	5.7
137862-53-4	acid	2.6	0.49	5.4
100-33-4	base	14	74	5.3
287714-41-4	acid	2.5	11	4.3
89565-68-4	base	6.7	26	3.9
1951-25-3	base	7.1	1.9	3.7
42408−82−2	base	11	41	3.7
221877-54-9	acid	2.3	0.67	3.5
55985-32-5	base	3.2	11	3.5

[a] Using final nonweighted $k = 3$ NN Model. [b] Reported as mL/min/kg.

13 and Figure 8). Consistent with the external test set analysis, higher fold error was associated with a smaller number of training set compounds for a particular range of CL values.

Model Validation. Model validation was accomplished using four different methods. The first employed the external test set as described earlier. The second method was a three-fold cross-validation procedure[44] where the 370 compound training set was randomly divided into three groups of 123−124 compounds each. One group was set aside, and the consensus model was built with the remaining two groups. The resulting model was then used to generate predictions for the excluded group. This process was iterated a total of three times so that each of the 370 compounds

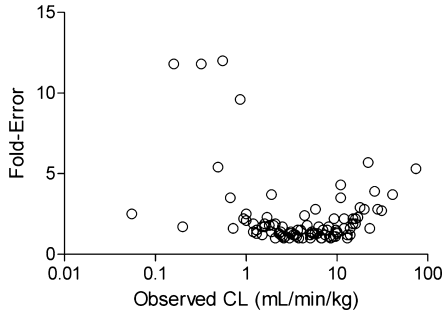

Figure 6. Plot of observed clearance (mL/min/kg) vs fold error for external test set ($n = 92$).

Table 13. Observed and Predicted CL Values for Goteti Set Compounds With >Three-Fold Prediction Error ($n = 20$)[a]

			$k = 3$ NN Model	
name	charge type	observed	predicted	fold error
iloprost	acid	1169	290	4.0
caspefungin	base	11	122	11
acivicin	zwitterion	49	369	7.5
gabapentin	zwitterion	112	373	3.3

[a] Reported as mL/min and assumes a 70 kg person.

was predicted once. As shown in Table 14, model performance was similar across the three groups. Concatenating the predictions for the three groups and calculating the GMFE for the entire training set gave a value of 2.25 ($n = 370$). This is consistent with the final model internal training set LOO GMFE of 2.21 (see Table 6, $k = 3$ unweighted).

To ensure that there were no chance correlations resulting from the data set partitioning, the procedure was repeated using five-fold cross-validation. The 370 compound training set was randomly divided into five groups of 74 compounds each, and the models were built as described above. Concatenating the predictions and calculating the overall GMFE gave a value of 2.17, again consistent with the final model LOO training set GMFE.

The third method of model validation was *y*-randomization.[45] The observed CL values in the training set were randomly shuffled, while leaving the descriptors and the external test set intact. The entire consensus model building procedure was repeated (including automatic descriptor selection), and the resulting system was used to predict values for the external test set. Since statistical significance is typically reported at the 95% confidence level, a total of 25 independent *y*-shuffled runs were examined. As seen in Table 15, the permuted models exhibited poor performance relative to the original.

The fourth method of model validation employed a more demanding variant of *y*-randomization described by Rücker and co-workers in 2007.[46] In this method, models were built using a modified training set comprised of random number pseudodescriptors and original response variables. While the Rücker variation was originally proposed in the context of descriptor selection in multiple linear regression (MLR), the issue they addressed is a general one, i.e., what is the risk associated with selecting a subset of m descriptors from a larger pool of M descriptors (i.e., what is the risk of selection bias).[47] Specifically, will the process utilized by our algorithm

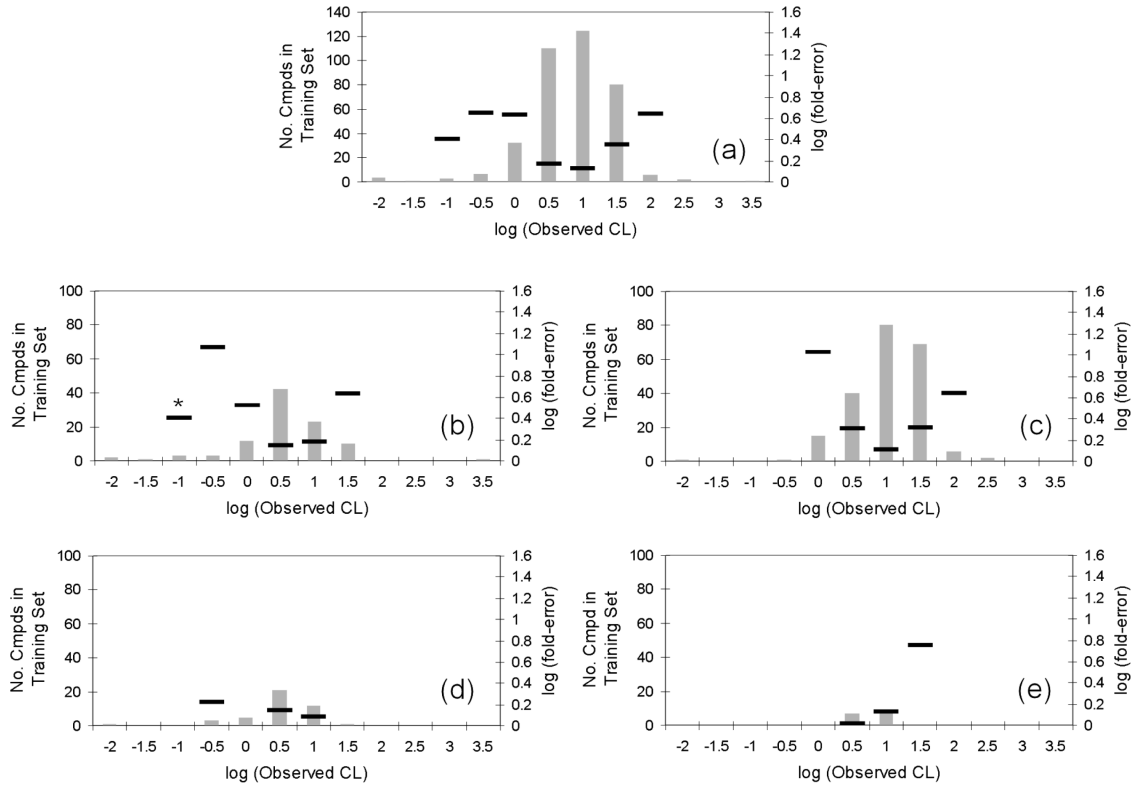

Figure 7. Combined two-axes plots of observed clearance (mL/min/kg) vs number of compounds in training set (vertical histogram bars, left axis) and of observed clearance (mL/min/kg) vs external test set average fold error (heavy line, right axis) for final unweighted $k = 3$ NN model. (a) All compounds in training ($n = 370$) and external test ($n = 92$) sets. (b) Acidic molecules in training ($n = 97$) and external test ($n = 34$) sets. The asterisk marks the fold error symbol for isoglycirrhizinate. (c) Basic molecules in training ($n = 214$) and external test ($n = 47$) sets. (d) Zwitterionic molecules in training ($n = 43$) and external test ($n = 8$) sets. (e) Quaternary ammonium/pyridinium ionic molecules in training ($n = 16$) and external test ($n = 3$) sets.

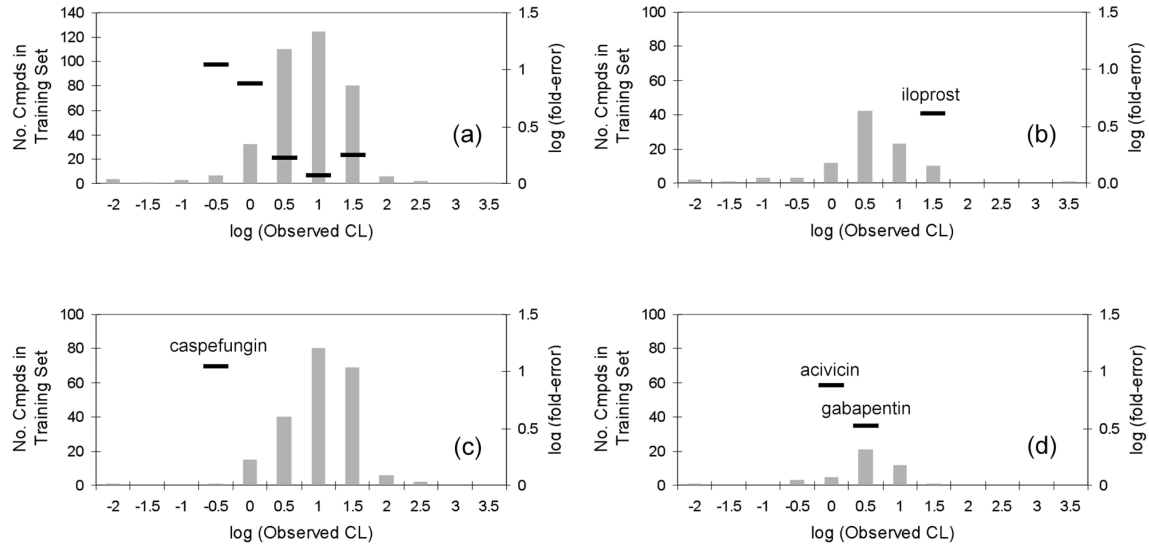

Figure 8. Combined two-axes plots of observed clearance (mL/min/kg) vs number of compounds in training set (vertical histogram bars, left axis) and of observed clearance (mL/min/kg) vs Goteti set average fold error (heavy line, right axis) for final unweighted $k = 3$ NN model. (a) All compounds in training set ($n = 370$) and average fold error for all compounds in Goteti set ($n = 20$). (b) Acidic molecules in training set ($n = 97$) and fold error for iloprost. (c) Basic molecules in training set ($n = 214$) and fold error for caspefungin. (d) Zwitterionic molecules in training set ($n = 43$) and fold error for acivicin and gabapentin.

Table 14. Three- and Five-Fold Cross-Validation Results for Final Clearance Model

	set 1	set 2	set 3	set 4	set 5
GMFE (three-fold CV)	2.20	2.18	2.37	–	–
GMFE (five-fold CV)	2.19	1.95	2.04	2.39	2.34

Table 15. Final Model *y*-Shuffled Results for Internal Training and External Test Sets ($n = 25$)

	LOO training set ($n = 370$)		external test set ($n = 92$)	
model	GMFE	r^2	GMFE	r^2
original	2.18	0.38	1.84	0.57
y-shuffled[a]	3.02 ± 0.07	0.0054 ± 0.0061	2.92 ± 0.2	0.028 ± 0.048

[a] The *y*-shuffled values represent the mean ± standard deviation from 25 independent runs.

to select a smaller number of descriptors from the pool of 12 described in the Methods Section overfit the data and result in a chance correlation.[48] The *y*-shuffling procedure described above broke the connection between the response variable (CL) and the associated molecular descriptors, thereby resulting in very poor model performance. Thus, there is a real connection between the two that cannot be explained by mere chance. However, as argued by Rücker and co-workers, intercorrelations between descriptors in the pool itself may exist that might also give rise to chance correlations. To test this, they replaced the descriptor values used to build the original model with random numbers and compared the resulting mean highest r^2 (mhr) value with that of the original. This set, which lacked the descriptor intercorrelation structure that may have existed in the original descriptor pool, consistently gave higher correlation coefficients than those obtained by traditional *y*-randomization methods. Thus, the pseudodescriptor procedure represents a higher hurdle for comparison.

To determine if intercorrelations in our descriptor set might lead to chance correlations, the original pool of 12 descriptors was replaced with random numbers that spanned the same range as the original. The charge-type designations were also randomly assigned while keeping the original CL values intact. The pseudodescriptors were then normalized, as described in the Methods Section, to provide the modified training set. This was repeated a total of 25 times with fresh random numbers used each time. No modifications were made to the external test set. For each modified descriptor pool, a model was built using the same algorithm as the original, including automatic descriptor selection.

Table 16. Final Model Pseudodescriptor Results for Internal Training and External Test Sets ($n = 25$)[a]

	LOO training set ($n = 370$)		external test set ($n = 92$)	
model	GMFE	r^2	GMFE	r^2
original	2.18	0.38	1.84	0.57
pseudodescriptor	2.90 ± 0.05	0.012 ± 0.015	2.66 ± 0.26	0.12 ± 0.10

[a] The reported pseudodescriptor values represent the mean ± standard deviation from 25 independent runs.

As summarized in Table 16, the pseudodescriptor procedure did provide a higher mean r^2 value relative to that obtained by the traditional *y*-randomization method, but one that was still below the correlation coefficient exhibited by the original model. These studies, therefore, provide additional assurance that the descriptor selection and the model building process utilized by our algorithm was not giving rise to chance correlations and that the final $k = 3$ NN model is valid.

CONCLUSION

A conceptually simple, fully in silico model to predict total clearance in humans is reported. The model was built using a *k*-nearest neighbors approach and validated using an external test set, a combination of three- and five-fold cross-validation and two *y*-randomization procedures. Although currently limited to acidic, basic, zwitterionic, and quaternary

ammonium/pyridinium ionic molecules, the average fold error for the external test set of 92 compounds was <2, with a prediction accuracy rate approaching those reported by Goteti and co-workers for ≥3 species scaling using SA and ME allometric methods. Since the model relies on easily calculated one- and two-dimensional molecular descriptors for input, it is capable of rapidly testing analogue ideas in the form of virtual compounds before chemical synthesis where nothing may be known experimentally beyond a two-dimensional chemical structure. In addition, human clearance projections could be used to supplement in vitro metabolic stability and animal pharmacokinetic data for compound prioritization during lead optimization where sufficient experimental data may not be available to support either interspecies scaling or in vitro−in vivo extrapolation for a large number of derivatives. By maintaining a focus on physical and pharmacokinetic properties in addition to in vitro potency early in the drug discovery process, scientists will be able to make better informed decisions in the design and prioritization of new analogues.

ACKNOWLEDGMENT

The author wishes to thank Dr. Greg Berger for assistance assembling the training set and Dr. Edgar Schuck for helpful discussions. I am also indebted to the reviewers for their invaluable comments and suggestions. Finally, I would like to thank Dr. Scott Obach and the group at Pfizer for publishing the data set of human PK parameters.

Supporting Information Available: Descriptor array for the training and external test sets. Hierarchical chemical charge-type classification scheme. Model development algorithm flowcharts. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Mahmood, I.; Green, M. D. Selection of the First-Time Dose in Humans: Comparison of Different Approaches Based on Interspecies Scaling of Clearance. *J. Clin. Pharmacol.* **2003**, *43*, 692–697.
(2) Reigner, B. G.; Blesch, K. S. Estimating the Starting Dose for Entry into Humans: Principles and Practice. *Eur. J. Clin. Pharmacol.* **2002**, *57*, 835–845.
(3) Mahmood, I.; Balian, J. D. The Pharmacokinetic Principles Behind Scaling from Preclinical Results to Phase I Protocols. *Clin. Pharmacokinet.* **1999**, *36*, 1–11.
(4) Howgate, E. M.; Rowland-Yeo, K.; Proctor, N. J.; Tucker, G. T.; Rostami-Hodjegan, A. Prediction of In Vivo Drug Clearance From In Vitro Data. I: Impact of Inter-Individual Variability. *Xenobiotica* **2006**, *36*, 473–497.
(5) Shiran, M. R.; Proctor, N. J.; Howgate, E. M.; Rowland-Yeo, K.; Tucker, G. T.; Rostami-Hodjegan, A. Prediction of Metabolic Drug Clearance in Humans: In Vitro-In Vivo Extrapolation vs Allometric Scaling. *Xenobiotica* **2006**, *36*, 567–580.
(6) Caldwell, G. W.; Masucci, J. A.; Yan, Z.; Hageman, W. Allometric Scaling of Pharmacokinetic Parameters in Drug Discovery: Can Human CL, Vss and t1/2 be Predicted from In-Vivo Rat Data? *Eur. J. Drug Metab. Pharmacokinet.* **2004**, *29*, 133–143.
(7) Tang, H.; Hussain, A.; Leal, M.; Mayersohn, M.; Fluhler, E. Interspecies Prediction of Human Drug Clearance Based on Scaling Data from One or Two Animal Species. *Drug Metab. Dispos.* **2007**, *35*, 1886–1893.
(8) Wajima, T.; Fukumura, K.; Yano, Y.; Oguma, T. Prediction of Human Clearance from Animal Data and Molecular Structural Parameters using Multivariate Regression Analysis. *J. Pharm. Sci.* **2002**, *91*, 2489–2499.
(9) Mahmood, I. Role of Fixed Coefficients and Exponents in the Prediction of Human Drug Clearance: How Accurate Are the Predictions from One or Two Species. *J. Pharm. Sci.* **2009**, *98*, 2472–2493.
(10) Paixao, P.; Gouveia, L. F.; Morais, J. A. G. Prediction of the In Vitro Intrinsic Clearance Determined in Suspensions of Human Hepatocytes by Using Artificial Neural Networks. *Eur. J. Pharm. Sci.* **2010**, *39*, 310–321.
(11) Nikolic, K.; Agababa, D. Prediction of Hepatic Microsomal Intrinsic Clearance and Human Clearance Values for Drugs. *J. Mol. Graphics Modell.* **2009**, *28*, 245–252.
(12) Li, H.; Sun, J; Sui, X.; Liu, J.; Yan, Z.; Liu, X.; Sun, Y.; He, Z. First Principle, Structure-Based Prediction of Hepatic Metabolic Clearance Values in Human. *Eur. J. Med. Chem.* **2009**, *44*, 1600–1606.
(13) Yang, X.; Gandhi, Y. A.; Duignan, D. B.; Morris, M. E. Prediction of Biliary Excretion in Rats and Humans Using Molecular Weight and Quantitative Structure-Pharmacokinetic Relationships. *AAPS J.* **2009**, *11*, 511–525.
(14) Varma, M. V. S.; Feng, B.; Obach, R. S.; et al. Physicochemical Determinants of Human Renal Clearance. *J. Med. Chem.* **2009**, *52*, 4844–4852.
(15) Yap, C. W.; Li, Z. R.; Chen, Y. Z. Quantitative Structure-Pharmacokinetic Relationships for Drug Clearance by Using Statistical Learning Methods. *J. Mol. Graphics Modell.* **2006**, *24*, 383–395.
(16) Johnson, T. W.; Dress, K. R.; Edwards, M. Using the Golden Triangle to Optimize Clearance and Oral Absorption. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5560–5564.
(17) Waring, M. J. Defining Optimum Lipophilicity and Molecular Weight Ranges for Drug Candidates - Molecular Weight Dependent Lower logD Limits Based on Permeability. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2844–2851.
(18) Rollins, D. E.; Klaassen, C. D. Biliary Excretion of Drugs in Man. *Clin. Pharmacokinet.* **1979**, *4*, 368–379.
(19) Hosea, N. A.; Collard, W. T.; Cole, S.; Maurer, T. S.; Fang, R. X.; Jones, H.; Kakar, S. M.; Nakai, Y.; Smith, B. J.; Webster, R.; Beaumont, K. Prediction of Human Pharmacokinetics from Preclinical Information: Comparative Accuracy of Quantitative Prediction Approaches. *J. Clin. Pharmacol.* **2009**, *49*, 513–533.
(20) van de Waterbeemd, H.; Gifford, E. ADMET In Silico Modelling: Towards Prediction Paradise. *Nat. Rev. Drug Discovery* **2003**, *2*, 192–204.
(21) Bender, A.; Glen, R. C. Molecular Similarity: A Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
(22) Kubinyi, H. Chemical Similarity and Biological Activities. *J. Braz. Chem. Soc.* **2002**, *13*, 717–726.
(23) Nigsch, F.; Bender, A.; van Buuren, B.; Tissen, J.; Nigsch, E.; Mitchell, J. G. O. Melting Point Prediction Employing k-Nearest Neighbor Algorithms and Genetic Parameter Optimization. *J. Chem. Inf. Model.* **2006**, *46*, 2412–2422.
(24) Obach, R. S.; Lombardo, F.; Waters, N. Trend Analysis of a Database of Intravenous Pharmacokinetic Parameters in Humans for 670 Drug Compounds. *J. Drug Metab. Disp.* **2008**, *36*, 1385–1405.
(25) Berellini, G.; Springer, C.; Waters, N. J.; Lombardo, F. In Silico Prediction of Volume of Distribution in Human Using Linear and Nonlinear Models on a 669 Compound Data Set. *J. Med. Chem.* **2009**, *52*, 4488–4495.
(26) Itskowitz, P.; Tropsha, A. k Nearest Neighbors QSAR Modeling as a Variational Problem: Theory and Applications. *J. Chem. Inf. Model.* **2005**, *45*, 777–785.
(27) Ajmani, S.; Jadhav, K.; Kulkarni, S. A. Three-Dimensional QSAR Using the k-Nearest Neighbor Method and Its Interpretation. *J. Chem. Inf. Model.* **2006**, *46*, 24–31.
(28) Neumann, D.; Kohlbacher, O.; Merkwirth, C.; Lengauer, T. Fully Computational Model for Predicting Percutaneous Drug Absorption. *J. Chem. Inf. Model.* **2006**, *46*, 424–429.
(29) Gunturi, S. B.; Archana, K.; Khandelwal, A.; Narayanan, R. Prediction of hERG Potassium Channel Blockade Using kNN-QSAR and Local Lazy Regression Methods. *QSAR Comb. Sci.* **2008**, *27*, 1305–1317.
(30) *Pipeline Pilot*, version 7.5; Accelrys: San Diego, CA, 2008.
(31) Dissociation Constants of Organic Acids and Bases. In *CRC Handbook of Chemistry and Physics*, 87th ed.; Lide, D. R., Ed.; Taylor & Francis Group: Boca Raton, FL, 2006; pp 8−52.
(32) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
(33) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. R., Eds.; Wiley-VCH, Inc., New York, NY, 1991, Vol. 2, pp 367−421.
(34) Goteti, K.; Garner, E.; Mahmood, I. Prediction of Human Drug Clearance from Two Species: A Comparison of Several Allometric Methods. *J. Pharm. Sci.* **2010**, *99*, 1601–1613.
(35) Asikainen, A. H.; Ruuskanen, J.; Tuppurainen, K. A. Performance of (Consensus) kNN QSAR for Predicting Estrogenic Activity in a Large Diverse Set of Organic Compounds. *SAR QSAR Environ. Res.* **2004**, *15*, 19–32.
(36) The Obach and Berellini data set clearance values used as the training set in our models were derived from intravenous administration only.

(37) Goteti, K.; Brassil, P. J.; Good, S. S.; Garner, C. E. Estimation of Human Drug Clearance Using Multiexponential Techniques. *J. Clin. Pharmacol.* **2008**, *48*, 1226–1236.
(38) Mahmood, I.; Balian, J. D. Interspecies Scaling: Predicting Clearance of Drugs in Humans. Three Different Approaches. *Xenobiotica* **1996**, *26*, 887–895.
(39) Mahmood, I. Theoretical Versus Empirical Allometry: Facts Behind Theories and Application to Pharmacokinetics. *J. Pharm. Sci.* **2010**, *99*, 2927–2933.
(40) Mahmood, I. Interspecies Scaling of Biliary Excreted Drugs: A Comparison of Several Methods. *J. Pharm. Sci.* **2005**, *94*, 883–892.
(41) Mahmood, I. Interspecies Scaling of Renally Secreted Drugs. *Life Sci.* **1998**, *63*, 2365–2371.
(42) Fagerholm, U. Prediction of Human Pharmacokinetics - Renal Metabolic and Excretion Clearance. *J. Pharm. Pharmacol.* **2007**, *59*, 1463–1471.
(43) Huang, C.; Zheng, M.; Yang, Z.; Rodrigues, A. D.; Marathe, P. Projection of Exposure and Efficacious Dose Prior to First-in-Human Studies: How Successful Have We Been? *Pharm. Res.* **2008**, *25*, 713–726.
(44) Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing Model Fit by Cross-Validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579–586.
(45) Kubinyi, H. QSAR in Drug Design. In *Handbook of Chemoinformatics*; Gasteiger, J., Ed.; Wiley-VCH: Weinheim, Germany, 2003; Vol. 4, pp 1532−1554.
(46) Rücker, C.; Rücker, G.; Meringer, M. y-Randomization and Its Variants in QSPR/QSAR. *J. Chem. Inf. Model.* **2007**, *47*, 2345–2357.
(47) See for example Livingstone, D. J; Salt, D. W. Judging the Significance of Multiple Linear Regression Models. *J. Med. Chem.* **2005**, *48*, 661–663.
(48) Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.

CI1000295