

Investigation of Factors Affecting the Performance of *In silico* Volume Distribution QSAR Models for Human, Rat, Mouse, Dog & Monkey

Saw Simeon,^[a, b] Dino Montanari,^[c] and Matthew Paul Gleeson^{*,[d, e]}

Abstract: Volume of distribution ($V_{d_{ss}}$) is a measure of how effectively a drug molecule is distributed throughout the body. Along with the clearance, it determines the half-life and therefore the drug dosing interval. A number of different pre-clinical approaches are available to predict the $V_{d_{ss}}$ in human including quantitative structure activity relationship (QSAR) models. $V_{d_{ss}}$ QSAR models have been reported for human and rat, but not important pre-clinical species including dog, mouse and monkey. In this study, we have generated $V_{d_{ss}}$ QSAR model on the human and commonly used pre-clinical species, each of which differs in terms of size, chemical diversity and data quality. We

discuss the model performance by species, assess the effect the domain of applicability and the relative merits of building chemical series-specific models. In addition, we compare the intrinsic variability of the experimental $\log V_{d_{ss}}$ data (~ 1.2 fold error) to in-vivo interspecies differences (~ 2 fold error) and *in silico* based models (~ 3 fold error). This prompted us to explore whether one species could be used to predict another, particularly where little data for that species is available. i.e. does the expansion in domain of applicability prove beneficial over and above any deterioration due to the use of response values from an alternative species.

Keywords: Applicability domain · data mining · quantitative structure-property relationship · volume of distribution

1 Introduction

Compounds in drug discovery must undergo multiobjective optimization^[1] to ensure that a sufficient concentration of compounds is available to act at the required target over the required dosing interval.^[2] This is a complex challenge arising from the competing dependencies that different biological and physiological parameters have on the molecular properties or functional groups of dosed compounds.^[3] Critical decisions regarding compound progression must be made early in the drug discovery process as to whether a chemical series is likely to achieve the necessary properties required for efficacy and safety.

Volume of distribution is a primary pharmacokinetic parameter that describes how effectively a drug molecule is distributed throughout the body. It is a proportionality factor that relates the concentration of drug in the blood to the total amount of drug in the body. Along with clearance, it determines the half-life and therefore the drug dosing interval. Volume of distribution at steady state ($V_{d_{ss}}$) is the volume of distribution measured at equilibrium, e.g. when the free concentration in plasma is equal to the free concentration in the tissue. $V_{d_{ss}}$ can be calculated from either the concentration-time profile generated following intravenous administration of a drug (standard procedure in drug development) or from the tissue/plasma partition coefficient (K_p). A number of experimental pre-clinical approaches are available to predict the $V_{d_{ss}}$ in human including allometric species scaling,^[4] direct scaling from

human *in vitro* tissues,^[5] and from HPLC based measurements.^[6]

In silico quantitative structure activity relationship (QSAR) models offer a means to rapidly benchmark chemical series in terms of their overall pharmacokinetic profiles using chemical structures alone.^[7] These models are derived by determining the underlying relationship between the response values of previously measured


[a] S. Simeon
Interdisciplinary Graduate Program in Bioscience, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

[b] S. Simeon
Center for Advanced Studies in Nanotechnology for Chemical, Food and Agricultural Industries, KU Institute for Advanced Studies, Kasetsart University, Bangkok 10900, Thailand

[c] D. Montanari
DMPK and Bioanalysis, Aptuit, Via Alessandro Fleming, 4, 37135 Verona VR, Italy

[d] M. P. Gleeson
Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
phone: +66 8697 79678
fax: +66 2 329 8346
E-mail: paul.gl@kmitl.ac.th

[e] M. P. Gleeson
Department of Biomedical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/minf.201900059>

compounds and their physico-chemical properties and/or structural features. Models built on diverse datasets of chemicals are considered to have a greater applicability domain,^[8] but they are not necessarily the most predictive.^[9] QSAR models that are limited to structurally related analogs may prove more predictive than global models due to a reduction in confounding effects of trying to describe complex molecular events for multiple chemotypes with situation-independent set of molecular descriptors.^[9a,10] This can occur when the available descriptors do not directly take into account the physical events being modelled (i.e. orientation specific receptor binding) or where no descriptor allows the model to account for the differing dependencies of the different series being modelled.^[11] A further complication in QSAR modelling is that the *in vitro* and *in vivo* response values used to train the model are themselves variable, differing by assay type and from day to day.^[12] In addition, the presence of data posting errors, including assay units or chemical structures, is another factor that complicates the QSAR model building process.^[13]

Global $V_{d_{ss}}$ QSAR models have previously been generated for human^[14] and rat^[14a,b,f,g] but not to our knowledge other species of pre-clinical interest. While the prediction of human $V_{d_{ss}}$ is the ultimate goal in pharmaceutical research, estimation of the parameter in other species are important since these are often used in clinical validation via animal models of efficacy and toxicity.^[15] We have compiled large $V_{d_{ss}}$ datasets on the 5 main clinical/pre-clinical species ranging from 2550 observations for rat, 1442 for human, 644 for dog, 682 for mouse and 318 for monkey. This compares to previously published QSAR models which utilized datasets ranging from 60–500.^[14c,e,f,16] We assessed the model performance by species, the effect of the model applicability domain and the value of building chemical series-specific models for the purpose of comparison. Given the relatively low experimental error and systematic differences we found between the species, we investigated the relative benefits of using scaled predictions from human to predict others species. Finally, we discuss how *in silico* models perform compared to *in vitro* tissue derived predictions of volume of distribution.

2 Methods

2.1 Data Extraction

The $V_{d_{ss}}$ datasets were extracted directly from the ChEMBL database^[17] and gathered from Lombardo *et al.*^[18] Data on 5 species (human, mouse, dog, rat and monkey) was extracted along with the assay description and earliest reported ChEMBL entry date. Observations with the units (L/Kg) and “=” modifiers were considered only. The individual assay description sections of each measurement were assessed to ensure results were from IV administration (Table S1). The IV $V_{d_{ss}}$ values were transformed to the

$\log V_{d_{ss}}$ scale which ranges from approximately -1.52 to $+2.15$ with mean values of 0.01, 0.39, 0.31, 0.37 and 0.16 for human, rat, dog, mouse and monkey, respectively. Compounds with reported $V_{d_{ss}}$ values greater than 4 standard deviations from the mean were subject to removal as were compounds with a molecular mass greater than 800 (Table 1).

Table 1. $V_{d_{ss}}$ data extracted for this study (IV Only).

Species	Compound	Mean	SD	Range
Human	1442	0.01	0.65	($-1.52/1.91$)
Rat	2550	0.39	0.61	($-1.35/2.15$)
Dog	644	0.31	0.60	($-1.22/1.91$)
Mouse	682	0.37	0.55	($-1.01/2.01$)
Monkey	318	0.16	0.60	($-1.41/1.83$)

2.2 In silico Model Building

All species datasets were split approximately (70%/30%) into training and independent test set. We also ensured that observations with data in human and another species were excluded from the training sets to allow for an assessment of the cross predictivity of the models. This meant that the Compounds were standardized using ChemAxon Standardizer^[19] with the following options: Strip Salt, Aromatize, Clean 3D, Tautomerize, Neutralize, and Remove explicit hydrogen. Molecular descriptors were computed in ChemAxon JChem.^[20] acidic and basic pKa, partition coefficients ($\log D$ (pH 7.4), $\log P$), polar surface area (PSA), molecular mass (MWT) and feature counts hydrogen bond donors (HBD) and acceptors (HBA), aliphatic atoms, aromatic and total rings, aromatic atoms, chiral centers, ring counts and rotatable bond counts and ratio of aromatic atoms vs. aliphatic atoms. The descriptors acid-class and base-class were defined from the predicted acidic and basic pKas. The values were binned from 5 to 1, from weak to strong; $ApKa \geq 9$; 8; 7; 6; ≤ 5 ; and $BpKa \leq 5$; 6; 7; 8; ≥ 9 . Additional E-state descriptors and fingerprints were generated using the PaDEL-Descriptors software.^[21]

In silico models were generated using 3 different types of multivariate modelling methods. Partial least squares (PLS) models were built using the *pls* function from the R package *pls*.^[22] Principal component inclusion was decided from 10-fold cross validation. Artificial neural network (ANN) models were built using feed-forward formalism in the R package *nnet*.^[23] Random Forest (RF) models were built using *randomForest* function from R package *randomForest*.^[24] Finally, a consensus model was generated based on the mean prediction from all prediction methods available. Models were built using physico-chemical descriptors (PC), e-state and fingerprint descriptors (ES) or

combinations of physico-chemical descriptors and one of the other two.

All models were assessed using standard statistical parameters including the correlation coefficient squared (R^2), 10-fold cross validation (Q^2), root mean square error (RMSE) in prediction and the mean prediction error, or prediction bias (ME).^[25] Y randomization was also undertaken as an assessment of relationship between the domain of applicability and the error. The domain of applicability was determined as the mean distance to the 5 nearest neighbors in the training set, computed using the QSAR model descriptors or fingerprint space (D2 M).^[8]

Data augmentation and reduction methods were investigated in our efforts to explore the utility of the models generated. The human dataset consists of a diverse set of drug molecules (primarily singletons) while the pre-clinical rat, mouse, dog and monkey datasets consists of primarily of failed R&D compounds (often present as a series), with little overall overlap between the two. We investigated whether the human training set could be used to improve the model performance by expanding the domain of applicability, albeit lacking the species specificity. This $\log V_{d_{ss}}$ prediction was transformed to remove the systematic bias between species before being compared to the original species specific QSAR model prediction.

Finally, we built chemotype specific QSAR models. Chemotypes were identified using Tanimoto-based clustering with a similarity cutoff of 0.80. The *cmp.cluster* function from the R package *ChemmineR* was used to identify chemotypes.^[26] Local QSAR models were developed using in same manner as the global models described above.

3 Results and Discussions

3.1 Human $V_{d_{ss}}$ QSAR Models

Human QSAR models were the focus of our initial model building efforts. Human QSAR models were initially evaluated using physico-chemical descriptors alone in conjunction with different linear (PLS) and non-linear methodologies (RF and ANN). These particular descriptors are known to be critical in describing $V_{d_{ss}}$ from previous research.^[14e,27] We have also investigated the use of consensus predictions which have often appeared to be more effective than predictions from individual models.^[10,28] The results for human are shown in Table 2 with a graphical illustration of their performance in Figure 1.

The PLS, ANN and RF models have moderate predictive performance on human $V_{d_{ss}}$ training set ($N=941$) with R^2 values ranging from ~ 0.47 – 0.70 . The prediction errors as given by the RMSE range from 0.35 to 0.46 (Table 2). The ANN model shows the best performance of any individual model ($R^2=0.70$). The PLS, model which offers the best interpretability, has the lowest R^2 of all (0.47). However, the

Table 2. Summary of QSAR model results using PC descriptors for Human $V_{d_{ss}}$ QSAR Models.

Model	Training Set			Test Set		
	R^2	RMSE	ME	R^2	RMSE	ME
PLS	0.43	0.48	0.00	0.46	0.48	−0.01
RF	0.56	0.42	−0.01	0.61	0.41	−0.03
ANN	0.61	0.40	0.00	0.52	0.46	−0.02
Con	0.58	0.41	0.00	0.57	0.43	−0.02
N	941			412		
SD	0.63			0.65		

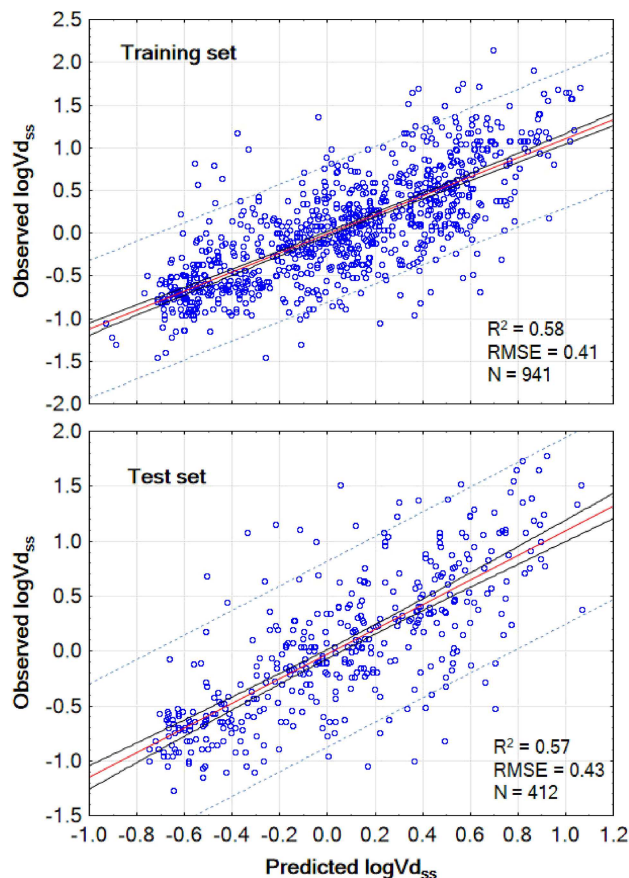


Figure 1. Performance of the human models built using consensus prediction in conjunction with PC descriptors. Dashed lines represent lower and upper prediction interval at 95% confidence level. Black lines represent lower and upper confidence interval at 95% confidence levels.

cross-validated Q^2 values indicates a robust model ($Q^2=0.43$).

The PLS and RF models are found to perform equally well on the test set. In contrast, the ANN model shows a drop in performance from an R^2 of 0.70 to 0.56. The RF model is the best performing model on the test set. The consensus model, which is the average prediction from all 3

different methods, shows reasonable performance in both the training ($R^2=0.64$) and test set ($R^2=0.61$).

3.2 The Effect of Descriptor Type of Vd_{ss} Human Models

It was noted that the ANN model performed marginally better in the initial model building process. In contrast, PLS performed less strongly, but showed less deterioration in performance on the subsequent test set. This, along with the ease of interpretation means PLS models are particularly suited to rational drug design, unless the model is employed in a black-box manner. The descriptor coefficients associated with the human PLS models, computed as the sum of influence over each of the fitted components, are displayed in Figure 2 and Table S2. It can be seen that chemical features that give rise to an increased in $\log Vd_{ss}$ are descriptors encoding acidity and basicity. As the base-class decreases (increasing basic pKa), the $\log Vd_{ss}$ increases (negative coefficient) while as the acid-class increases (decreasing acidic pKa), the $\log Vd_{ss}$ decreases (positive coefficient).^[9a,14b,29] H-bond acceptor features have a negligible effect on the model while H-bond acceptor features lead to a small decrease. Increasing clogP leads to an increase in the $\log Vd_{ss}$ as does MW.^[14g]

3.2.1 The Effect of Using Descriptors Encoding Structural Information

QSAR models were generated using E-state descriptors (ES) topological descriptors to see whether these would lead to an improvement in model performance over simple molecular descriptors. ES descriptors are derived from a graph-based representation of compounds encoding structural information. Among the different topological parameters, the connectivity indices of Kier and Hall^[30] have gained much popularity and have been extensively used in QSAR studies.^[31]

In Table 3 we compare the PLS model results obtained using PC, ES and a combination of both PC+ES. It can be seen that structural descriptors alone are not as useful as PC descriptors alone. However, the combination of both descriptor sets (All) leads to better performance over both the training and test sets. This model displays a slightly lower RMSE on the training set and independent test set than the model built using physicochemical descriptors alone (0.48 vs. 0.57 for E-state).

3.2.2 The Effect of Domain of Applicability on Performance

The relationship between the distance to model and the prediction error is reported for the human consensus model built using all the descriptors (Table 4). The distance was computed as the average Euclidean distance to the 5

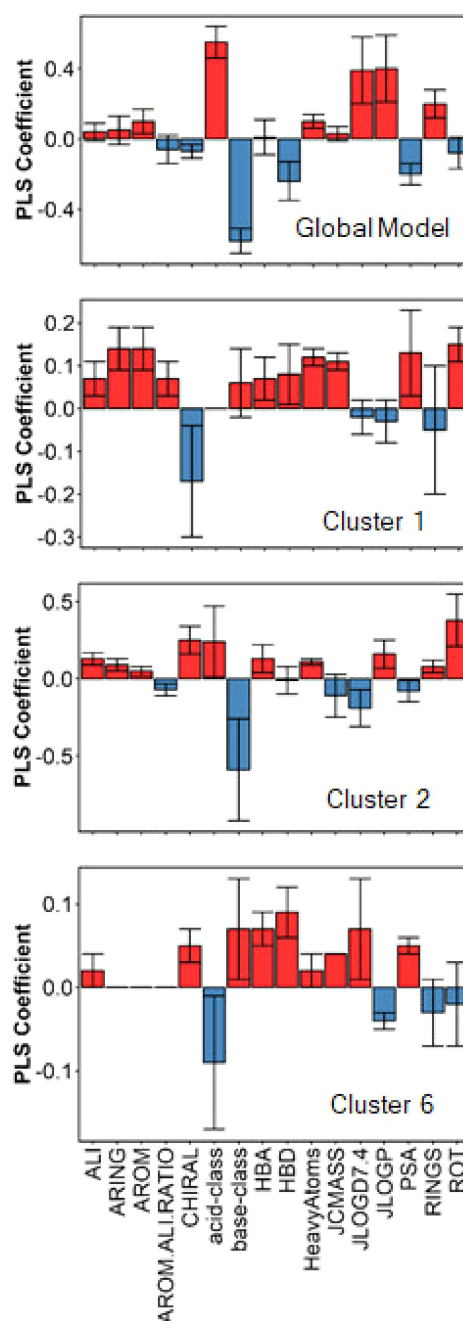


Figure 2. Global and selected cluster PLS model coefficients. Red indicates descriptors positively contribute to $\log Vd_{ss}$ whereas blue represents a negative contribution.

nearest neighbours in the training set for each query within the model descriptor space.^[8f] The distance values were split into 4 equally populated distance bins (Q1–Q4). The mean distance of the training and test sets are reported in Table 4 and pictorially in Figure 3 (Top).

A statistically significant relationship between the distance to model and the error was observed using ANOVA (Kruskal-Wallis, $P < 0.0009$). For the training set, the predic-

Table 3. Summary of the human consensus (CONS) model results (top) and PLS models for comparison using PC, ES or all descriptors.

		Training			Test		
Model	Type	R ²	RMSE	ME	R ²	RMSE	ME
CONS	PC ^{aa}	0.58	0.41	0.00	0.57	0.43	−0.02
	ES ^b	0.59	0.42	0.00	0.50	0.47	−0.02
	All	0.64	0.38	0.00	0.61	0.42	−0.03
PLS	PC ^{aa}	0.43	0.48	0.00	0.46	0.48	−0.01
	ES ^b	0.28	0.53	0.00	0.26	0.57	−0.02
	All	0.47	0.46	0.00	0.50	0.47	−0.02
N		941			0.63		
SD		412			0.65		

^a Physiochemical descriptors. ^b E-state descriptors.

Table 4. Distance to model effect for the human consensus model (all descriptors). Shown are the mean, standard error of the mean and number of observations for training and test set broken down by distance quartile bin

Dataset	Quartile	Mean ^a	SE ^b	N ^c
All	–	0.30	0.01	1353
Training set	1	0.24	0.01	290
Training set	2	0.29	0.01	251
Training set	3	0.33	0.02	220
Training set	4	0.32	0.02	180
Test set	1	0.19	0.03	49
Test set	2	0.28	0.02	85
Test set	3	0.32	0.02	119
Test set	4	0.37	0.02	159

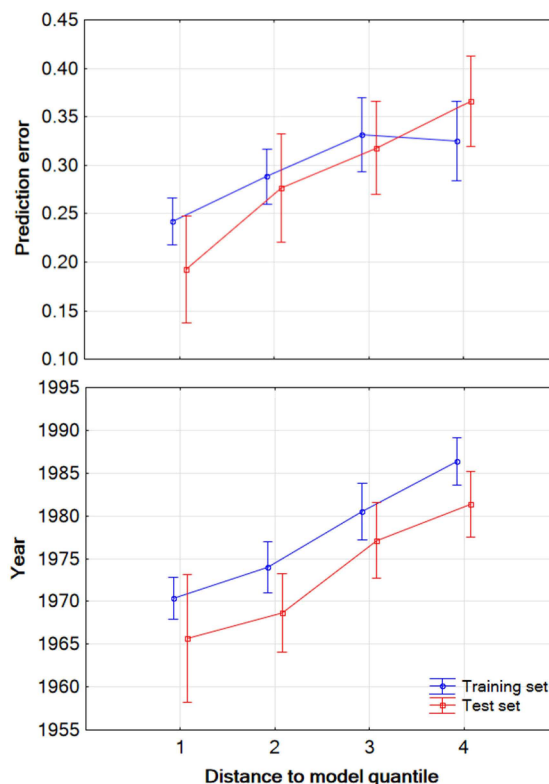
^a Mean absolute error. ^b Standard error mean. ^c Number of observations

tion error increases as the distance bin increases; Bin 1 (0.24), Bin 2 (0.29), Bin 3 (0.33) and Bin 4 (0.32). The same effect was also observed for the test set: Bin 1 (0.19), Bin 2 (0.28), Bin 3 (0.32) and Bin 4 (0.37), indicating that compounds closer to the training set model space are more likely to be predicted better.

It was also observed that the distance to model also shows a strong correlation with the year of first report of compounds. This is consistent with the fact that as time progress, more near neighbours of the compound will appear, a primary reasons given by others to undertake frequent model rebuilding.^[8a,9c]

3.2.3 Benchmarking *in silico* vs. *in vitro* Scaling Methods

Korzekwa *et al.*^[32] recently reported the prediction of human Vd_{ss} for 63 drug molecules using pre-clinical Vd_{ss} species *in vitro* tissue based binding data (microsomes & plasma). To put our model in perspective, we have compared their results to our more conservative PLS model built using PC descriptors only. The *in vitro* model displayed an R² of 0.77 and an RMSE of 0.31 log units (Figure 4, top). In comparison,

**Figure 3.** (Top) Relationship between the distance quartile and the error for the consensus model using all descriptors. Each dataset is denoted with a different color: training (blue) and test set (red). (Bottom) Plot of earliest reported date vs. the distance to model. Each dataset is denoted with a different color: training (blue) and test set (red).

the *in silico* PLS model demonstrates an R² of 0.75 and an RMSE of 0.33.

It should be noted that 51 compounds in the set of Korzekwa *et al.* were also found in the training set of our PLS model. The PLS model was rebuilt excluding these and compounds which led to comparable prediction statistics R² (0.74) and RMSE (0.34).

This suggests that the use of experimentally determined tissue binding data to predict volume of distribution is better than using an *in silico* model. However, it is not clear whether the modest gain in performance can fully justify the cost of determining the *in vitro* measurements.

3.3 Pre-clinical Species Vd_{ss} QSAR Models

In silico models of rat, dog, mouse and monkey were generated using the consensus approach and all descriptors (PC + ES) given that this was the most effective approach found for human data (Table 5). The R² of training sets ranged from 0.38 to 0.47 while the RMSEs ranged from 0.38 to 0.42. Given the different variance of the datasets, a more appropriate measure of the predictive performance is

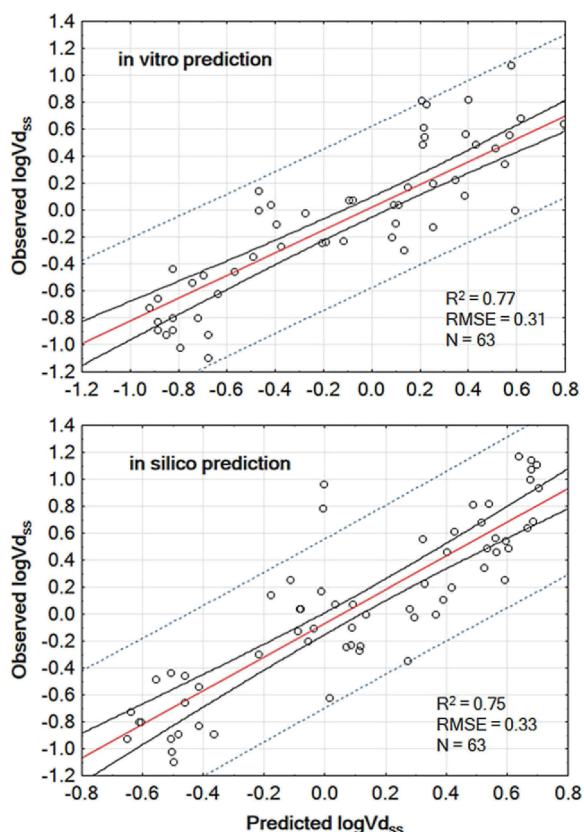


Figure 4. Performance of an *in vitro* derived prediction (using fraction unbound in both plasma and microsomes) for a set of 63 drugs vs that for the most conservative QSAR model generated here (PLS model employing PC descriptors).

Table 5. Consensus model (all descriptors) results obtained from human, dog, monkey, rat and mouse datasets.

Species	Training		R^2	RMSE	ME
	SD	N			
Human	0.63	941	0.64	0.38	0.00
Dog	0.54	448	0.40	0.42	0.00
Monkey	0.52	185	0.47	0.38	0.00
Rat	0.56	1764	0.45	0.42	0.00
Mouse	0.54	474	0.38	0.42	0.99
Species	Test		R^2	RMSE	ME
	SD	N			
Human	0.65	412	0.61	0.42	−0.03
Dog	0.62	194	0.40	0.49	−0.07
Monkey	0.62	124	0.56	0.42	−0.10
Rat	0.57	756	0.48	0.42	−0.01
Mouse	0.53	200	0.33	0.43	0.01

therefore the RMSE. On the independent test set, the models for human, dog and rat have comparable RMSEs, indicating similar predictive performance. Mouse data is somewhat less well predicted at 0.43 followed by dog at 0.49.

All models showed the same dependence on molecular properties (Figure S2). The applicability domain of pre-clinical models were then defined and the relationship between the distance to models and the corresponding prediction error were determined (Figure S1). For dog, rat and mouse, it can be seen that as the distance to model of individual datasets increase, the prediction error increases. However, we do not see any effect for monkey, the smallest dataset of all the species here. This might suggest that the distance to model vs error relationship can only exist with a sizeable, diverse set of compounds in the training set.

3.3.1 Assay Variability & Interspecies Variability

The prediction error of the QSAR models arises from a multitude of sources including structure and data entry errors,^[17,33] descriptor limitations, insufficient data sample to cover the required chemical space and statistical model fitting errors. Another important source of error is with the biological data itself.^[33] A comprehensive analysis on K_i values reported in ChEMBL database estimated experimental uncertainty at ~2.5 fold^[34] and that mixing heterogeneous pK_i and pC_{50} did not lead to a significant loss of quality of models built.^[35]

The volume of distribution data was analyzed to estimate the intrinsic errors associated with the assay. The error between pairs of measurements for any compound reported within a given species via the IV route of administration was assessed. We report the RMSE associated with the experimental $\log V_{dss}$ data and the R^2 derived from a plot of all pairs of measurements (Table 6).

Table 6. The correlation between experimental IV $\log V_{dss}$ values obtained for compounds with multiple repeat measurements.

Species	N	R^2	RMSE ^a	ME	2 fold ^b	4 fold ^c
Human	116	0.77	0.36 (0.05)	−0.03	12	6
Rat	136	0.68	0.42 (0.05)	−0.12	23	11
Dog	41	0.66	0.42 (0.09)	−0.13	32	22
Mouse	35	0.85	0.26 (0.06)	−0.07	1	0
Monkey	12	0.63	0.34 (0.07)	−0.21	42	8

^a Value in parenthesis corresponds to error calculated by excluded all compounds $> \pm 0.2$, ^b Percent of $> / < 0.3$, ^c Percent of $> / < 0.6$

The replicate errors are not normally distributed (Figure 5), unlike those found for the *in silico* model predictions (Figure S1). We therefore report the RMSE calculated as-is and with error data ± 0.2 excluded. The latter results in a considerably lower estimate to the error from a distribution that approaches normality. However, a key take home message is that *in vivo* data can often vary considerably due to different doses and formulations being used, which can be given to diseased and non-diseased animals for example.

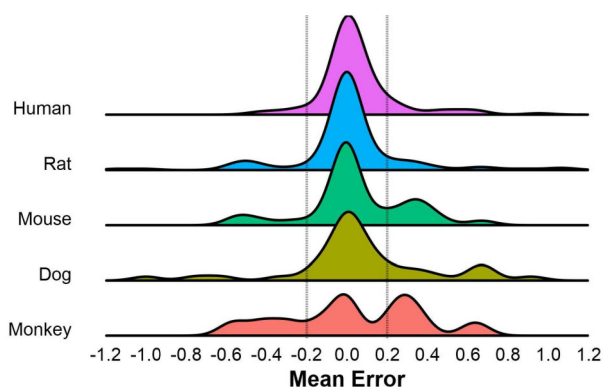


Figure 5. Illustration of the assay replicate errors for IV $V_{d_{ss}}$ data.

We note that of all the $V_{d_{ss}}$ models generated here, the human variants appear the most capable given the R^2 of the training and the test sets (Table 6). Tropsha et al.^[36] have noted that R^2 values of 0.5–0.6 indicate a model of moderate predictive ability. Nevertheless, it is important to note the RMSE also given that some datasets have lower variance. This limits the R^2 due to fact that the species experimental errors are expected to be comparable.^[37] Nevertheless, such relatively weak models could still prove useful in lead generation where even qualitative models can be used to bias chemical libraries to particular outcomes.^[38] The relatively low accuracy of the models could also be managed by focusing on predictions that lie close to the training set model space (domain of applicability) which can often be more reliably predicted than those far from model space.^[39]

The vast majority of the data can be described by a normal distribution with an estimated experimental error of 0.05 log unit or 1.1 fold error. Pre-clinical data in mouse, dog and monkey showed a greater number of extreme errors than rat or human, a reflection of the relatively few measurements available for these species and potentially due to their use in later stage dose ranging studies.^[40]

The $V_{d_{ss}}$ error estimates here are broadly in-line with, those found for in *in vitro* DMPK parameters by Wenlock et al.^[41] (1.4–1.6 fold). These errors of course have implications for QSAR building since the resulting models can be no better than the data it is built from. From the data analysed, models from human and rat data might be expected to result in more reliable fitted relationships than other species. This appears to be a reflection on humans volume being tested late in development with well defined doses and formulations and rat data being obtained using well established protocols and very limited formulations.

3.3.2 Interspecies Extrapolation vs In silico Prediction

Human pharmacokinetic parameters are estimated in pre-clinical species such as rat, mouse, dog and monkey with a

view to extrapolating to human.^[40] Allometric scaling of PK values between species relies on the use of physiological differences in parameters such as body weight, blood flow and tissue binding to derive a prediction in unknown species.^[42] Allometric scaling of $V_{d_{ss}}$ in pre-clinical species can employ corrections for body weight^[43] or for differences in plasma binding between human and the pre-clinical species.^[51] Predictions within 2 fold of the observed value (0.3 log units) are typically considered reasonable.^[42,44] Due to the lack of protein binding data for all compound here, we report the correlation between $\log V_{d_{ss}}$ of pre-clinical species against the values in human as a further means to benchmark the *in silico* models. (Figure 6 and Table 7).

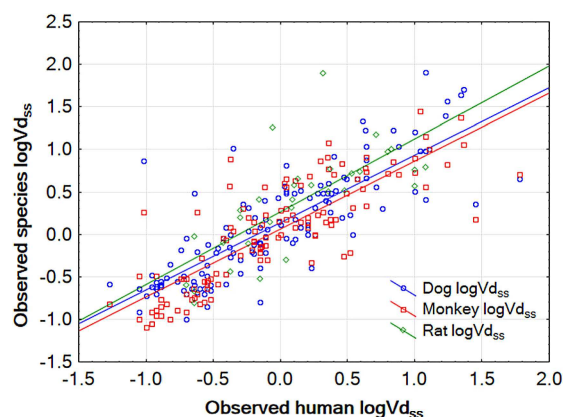


Figure 6. Correlation of human $\log V_{d_{ss}}$ with pre-clinical species.

Table 7. The correlation of human $\log V_{d_{ss}}$ vs pre-clinical species (outliers removed where error $> 4 \times \text{RMSEs}$). Too little data exists for Mouse to ascertain a reliable correlation with human.

Species	B ^a	N	R ²	RMSE	ME	RMSE ^a
Dog	12 Kg	132	0.67	0.41	−0.15	0.36
Monkey	4 Kg	124	0.68	0.38	−0.08	0.35
Rat	0.25 Kg	31	0.52	0.50	−0.25	0.42
Mouse	0.03 Kg	13	0.23	0.63	−0.39	0.42

^a Body weight ^b RMSE is the mean error of the fitted regression for human $V_{d_{ss}}$ as a function of species $V_{d_{ss}}$.

Indeed, it can be seen that a rather strong correlation exists between the human $V_{d_{ss}}$ data and those from dog, rat and monkey. This is of course expected given that each of their QSAR models display the same fundamental dependence on physical properties (Figure S2). Furthermore, the error from the line of unity are lower for species with higher body weight.^[43] The correlation between human and the 4 different pre-clinical species was heavily dependent on the amount of dataset size. The strength of the correlations observed ranged from and R^2 of 0.68 and 0.67 for dog and monkey to 0.68 and 0.23 for mouse and 0.52 to rat.

The strength of correlation found is broadly comparable to our *in silico* models or *in vitro* tissue derived predictions of Korzekwa *et al.*^[32] However, Mc Ginnity show that the use of pre-clinical species $V_{d_{ss}}$ corrected for differences in the free fraction result in much improved predictions – R^2 of 0.93 for prediction human $V_{d_{ss}}$ for a 63 compound set.^[45]

3.3.3 *In silico* Extrapolation of Pre-clinical $V_{d_{ss}}$ using Scaled Prediction from the *in silico* Human Model.

It was observed that certain pre-clinical species had considerably lower numbers of $\log V_{d_{ss}}$ datapoints than others (i.e. Monkey) and that some species showed a tendency towards lower errors (i.e. human). This prompted us to consider whether a prediction of $V_{d_{ss}}$ from the a human *in silico* model, which is build using a diverse datasets of relatively high quality data, could be scaled using the equations in Table 7 to deliver a reasonable estimate of the $V_{d_{ss}}$ for another species.

The training set of a given species here may contain some data already present within the human model dataset. Thus the test set is a more independent measure of the relative performance of *in silico* prediction using a species model or a scaled prediction from the human model. On the training sets, we find in all cases that the explained variance of the scaled human prediction is lower than that from the original model, which is not unexpected given that the former is a fitted relationship (Figure 7). On the two datasets that have the lowest number of observations

(Monkey and Mouse), the scaled human prediction results in improved prediction. On the rat datasets in contrast, the predictions are demonstrably poorer.

The results suggest that where only a small dataset exists for *in vivo* endpoint of a given species, more predictive QSAR models might be possible by adding scaled data derived from another species. This finding in-line with reports of Cortes-Ciriano *et al* who note that data expansion methods to increase its generalizability through an expanded domain of applicability can prove useful in QSAR modelling.^[46] However, it should be noted that for species such as rat which have very large number of measurements available on diverse compounds, the results were in fact poorer suggesting caution is needed.

3.4 Chemotype-specific $V_{d_{ss}}$ QSAR Models

An alternative way of improving the performance of a QSAR model is to limit the chemical diversity of the data used to generate it. This also limits the applicability domain to compounds of the same class which in project specific cases may be beneficial.^[9a,10] While this results in a lower domain of applicability, the lack of confounding descriptor-activity relationships with the diverse compound sets means a potentially more accurate relationship can be determined.^[9a] To explore this further, the human dataset was clustered and the top 6 clusters, having > 19 observations each, were used to generate local models. Model building was limited to PLS and PC descriptors to minimize the possibility of overfitting (Table 8).

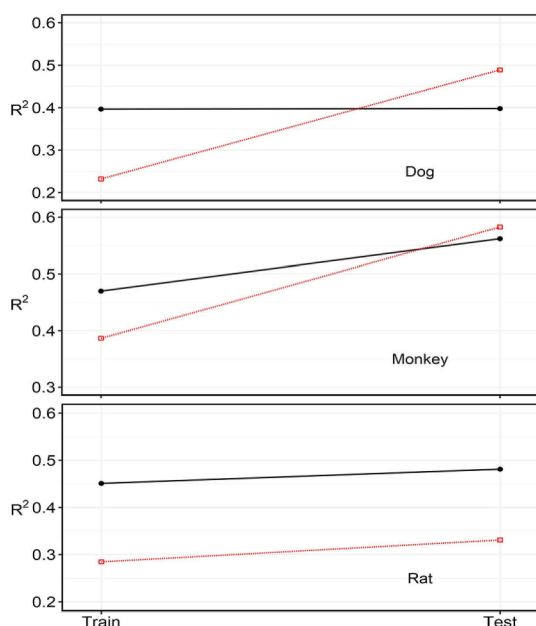


Figure 7. Performance of the scaled human consensus model (all descriptors) predicting non dog, monkey and rat. Predictions were scaled using line of best fit equation reported in Table 7.

Table 8. Global human consensus model results for individual chemotypes identified from clustering.

Cluster	Training		R^2	RMSE (ME)
	N	SD		
1	19	0.39	0.03	0.64 (0.50)
2	39	0.46	0.24	0.40 (0.06)
3	23	0.51	0.22	0.51 (−0.25)
4	29	0.39	0.29	0.35 (0.13)
5	32	0.15	0.00	0.24 (−0.10)
6	19	0.23	0.03	0.44 (−0.34)

Cluster 1 consists of zwitterionic tetracycline like compounds.^[47] the set has a moderate $V_{d_{ss}}$ SD of 0.39, but despite this, the global model is unresponsive of the set ($R^2=0.03$ and t RMSE=0.64) (Table 8). Cluster 2 consists of neutral benzodiazepines^[29,48] and it is again found that the global model performs on this set ($R^2=0.24$). Cluster 3 consists of neutral nicotinic acetylcholine receptor antagonists^[49] having a $V_{d_{ss}}$ SD of 0.51. The global model also predicts this class of compounds poorly (R^2 of 0.22). Cluster 4 contains a set of zwitterionic quinolones^[50] which are predicted with an RMSE equivalent to the global model

(~0.40) and an $R^2=0.29$. The acidic penicillin-like compounds^[51] of cluster 5 show a low variance in $\log V_{d_{ss}}$ (0.15), but also have a very low prediction error (RMSE = 0.24). Thus the low R^2 of the global model ($R^2=0.00$) is expected. Finally, cluster 6 consists of basic aminoglycoside antibiotics^[52] which also have low variance, but are predicted with a larger RMSE (0.44) and the rank ordering of compound is poor ($R^2=0.03$). Our goal was to investigate building local QSAR models on these sets to see if these models could prove any more effective than predictions from a global model.

Local models were built for each of the clusters. The datasets were split at random in an approximately 3:1 ratio into training and test set, respectively. Models were generated using physicochemical descriptors only and the PLS method to minimize the possibility of overfitting the relationship due to the small number of observations (Table 9).

Table 9. Local human PLS model results for individual chemotypes identified from clustering.

Cluster	Training N	R^2 (Q^2)	ME	RMSE
1	14	0.93 (0.65)	0.00	0.11
2	28	0.41 (0.09)	0.00	0.37
3	17	0.45 (0.29)	0.00	0.40
4	21	0.32 (0.17)	0.00	0.37
5	23	0.42 (0.09)	0.00	0.12
6	14	0.76 (0.08)	0.00	0.10
Test				
Cluster	N	R^2 (Q^2)	ME	RMSE
1	5	0.80	0.05	0.18
2	11	0.64	0.22	0.32
3	6	0.00	0.48	-0.27
4	8	0.54	0.15	0.18
5	9	0.41	-0.01	0.09
6	5	0.79	-0.04	0.13

The local models performed well on the training and test set of all six clusters, resulting in either lower prediction errors (RMSE) than the global model, higher R^2 , or both. Cluster 1, 5 and 6 are particularly well predicted by the local models, showing high R^2 values and low RMSEs (Table 8 and Figure S3). Cluster 1 has a training set R^2 of 0.93 and an RMSE of 0.11 while the test set has an R^2 0.80 and RMSE of 0.18. Cluster 6 behaves comparably to cluster 1 while cluster 5 displays RMSE values that are considerably lower than those obtained for the global model. Cluster 5's moderate R^2 (~0.40) is a reflection of the low variance of the dataset. Cluster 2 and cluster 4 are predicted with lower RMSEs than obtained from the global models, however the R^2 values are still relatively low. This is also a reflection of the low variance in the $V_{d_{ss}}$ values in the training and test sets. Finally, the local model built on cluster 3 does not

perform very effectively using either the global or local approach.

These results are consistent with the previous reports that local models on specific series can perform better than predictions derived from global model.^[53] However, a key limitation is that they can only be expected to perform well on compounds from the same series. This can be understood from an analysis of the PLS coefficients of the models (Figure 2). A comparison of clusters 1, 2 and 6 to the global model show that the model coefficients show some degree of similarity. For the zwitterions of cluster 1 for example, the influence of the acid-class and base-class is negligible. In fact base-class shows a positive dependency suggesting lower basicity increases $\log V_{d_{ss}}$. The predominantly neutral molecules of cluster 2 show the same charge class dependency as the global model but the contribution of H-bond donors and acceptors and lipophilicity have inverted. In cluster 6, the basic molecules show increasing $\log V_{d_{ss}}$ with increasing $\log D_{7.4}$, but the effect of increasing acid and basic strength is inverted compared to the global model. In fact, these subtle and not so-subtle differences in the descriptor dependencies helps to explain why global models often struggle to explain SAR of multiple series.

4 Conclusions

In silico models can be used to help make low risk go/no-go decisions early in the drug discovery process or to guide the incorporation of relatively small modifications on later stage lead series. However, the predictive performance of such methods varies considerably depending on the ADMET endpoints in question, species under consideration, sizes and diversities of the available datasets and methodologies employed.

Our QSAR building exercises showed that consensus predictions consisting of PLS, NN and RF models generally performed better than any individual method. We also determined that while physicochemical properties are critical for describing $V_{d_{ss}}$, additional value was obtained by incorporating electrotopological descriptors into the models. Models could be built with prediction errors of ~0.42 log units (~2.6 fold) for most species. *in silico* models were found to be of comparable accuracy to simpler *in vitro* scaling methods that employed *in vitro* binding data.^[43] However, it is important to note that interspecies *in vivo* scaling of species $V_{d_{ss}}$ to human using additional *in vitro* tissue binding data is expected to result in predictions errors below 2 fold.^[45] Nevertheless, while the global *in silico* models can prove advantageous to benchmark chemical series or support early lead optimization efforts due to their low cost and the speed of calculation.^[3a,54]

We investigated methods to augment the training sets using data derived from other species, albeit being suitably scaled. We found that for species with smaller datasets (monkey and mouse), scaling the human QSAR model

prediction to derive a prediction of the independent test set compounds was more effective than using a QSAR model built on small numbers of observations from that species. Finally, we investigated whether reducing the domain of applicability of QSAR models through the generation of chemotype specific models offers any advantages for the 6 largest clusters in the dataset. We found that in a number of cases the local model generated performed better than the global model, either in terms of the absolute prediction error or rank ordering of compounds with the series. This was due to slight differences in the dependencies each series has on the molecular descriptors used to build the model.

Supporting Information

Supporting information consisting of additional tables and figures have been provided. Also provided are all datasets used and QSAR model predictions made.

Conflict of Interest

None declared.

Acknowledgements

M.P.G would like to acknowledge financial support provided by the Thailand Research Fund (RSA6180073) and King Mongkut's Institute of Technology Ladkrabang. S.S is grateful for financial support from the National Research University (NRU) for supporting his Ph.D. studies.

References

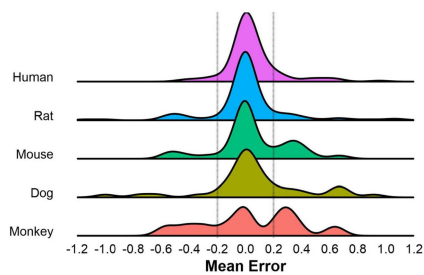
- [1] a) S. Ekins, J. D. Honeycutt, J. T. Metz, *Drug Discovery Today* **2010**, *15*, 451–460; b) M. D. Segall, *Curr. Pharm. Des.* **2012**, *19*, 1292–1310; c) A. T. Chadwick, M. D. Segall, *Drug Discovery Today* **2010**, *15*, 561–569.
- [2] a) D. Montanari, E. Chiarparin, M. P. Gleeson, S. Braggio, R. Longhi, K. Valko, T. Rossi, *Expert Opin. Drug Discovery* **2011**, *6*, 913–920; b) S. Braggio, D. Montanari, T. Rossi, E. Ratti, *Expert Opin. Drug Discovery* **2010**, *5*, 609–618.
- [3] a) M. P. Gleeson, D. Montanari, in *CMC III* (Eds.: D. Rotella, S. E. Ward), Elsevier, Oxford, **2017**, pp. 64–93; b) M. P. Gleeson, A. Hersey, D. Montanari, J. Overington, *Nat. Rev. Drug Discovery* **2011**, *10*, 197–208.
- [4] a) S. Björkman, *J. Pharm. Pharmacol.* **2002**, *54*, 1237–1245; b) R. Nagilla, M. Nord, J. J. McAtee, L. J. Jolivet, *J. Pharm. Sci.* **2011**, *100*, 3862–3874.
- [5] O. Luttringer, F. P. Theil, P. Poulin, A. H. Schmitt-Hoffmann, T. W. Guentert, T. Lavé, *J. Pharm. Sci.* **2003**, *92*, 1990–2007.
- [6] F. Hollosy, K. Valko, A. Hersey, S. Nunhuck, G. Keri, C. Bevan, *J. Med. Chem.* **2006**, *49*, 6958–6971.
- [7] a) K. H. Grime, P. Barton, D. F. McGinnity, *Mol. Pharmaceutics* **2012**, *10*, 1191–1206; b) M. P. Gleeson, D. Montanari, *Expert Opin. Drug Metab. Toxicol.* **2012**, *8*, 1435–1446; c) J. C. Madden, in *Recent Adv. QSAR Studies*, Vol. 8 (Eds.: T. Puzyn, J. Leszczynski, M. T. Cronin), Springer Netherlands, **2010**, pp. 283–304; d) U. Norinder, C. A. Bergstrom, *ChemMedChem* **2006**, *1*, 920–937.
- [8] a) S. L. Rodgers, A. M. Davis, N. P. Tomkinson, H. van de Waterbeemd, *Mol. Inf.* **2011**, *30*, 256–266; b) S. L. Rodgers, A. M. Davis, H. van de Waterbeemd, *QSAR Comb. Sci.* **2007**, *26*, 511–521; c) R. P. Sheridan, *J. Chem. Inf. Model.* **2015**, *55*, 1098–1107; d) A. Tropsha, *Mol. Inf.* **2010**, *29*, 476–488; e) A. Sazonovas, P. Japertas, R. Didziapetris, *SAR QSAR Environ. Res.* **2010**, *21*, 127–148; f) S. Weaver, M. P. Gleeson, *J. Mol. Graphics Modell.* **2008**, *26*, 1315–1326; g) I. V. Tetko, P. Bruneau, H.-W. Mewes, D. C. Rohrer, G. I. Poda, *Drug Discovery Today* **2006**, *11*, 700–707.
- [9] a) M. P. Gleeson, *J. Med. Chem.* **2007**, *50*, 101–112; b) R. P. Sheridan, *J. Chem. Inf. Model.* **2014**, *54*, 1083–1092; c) A. M. Davis, D. J. Wood, *Mol. Pharmaceutics* **2013**, *10*, 1183–1190.
- [10] M. P. Gleeson, N. J. Waters, S. W. Paine, A. M. Davis, *J. Med. Chem.* **2006**, *49*, 1953–1963.
- [11] a) A. M. Davis, in *CMC III* (Eds.: S. Chackalamannil, D. Rotella, S. E. Ward), Elsevier, Oxford, **2017**, pp. 379–392; b) R. Cox, D. V. S. Green, C. N. Luscombe, N. Malcolm, S. D. Pickett, *J. Comp. Aided. Mol. Des.* **2013**, *27*, 321–336.
- [12] a) M. C. Wenlock, L. A. Carlsson, *J. Chem. Inf. Model.* **2015**, *55*, 125–134; b) L. Zhao, W. Wang, A. Sedykh, H. Zhu, *ACS Omega* **2017**, *2*, 2805–2812.
- [13] a) M. Waldman, R. Fraczekiewicz, R. D. Clark, *J. Comp. Aided. Mol. Des.* **2015**, *29*, 897–910; b) D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* **2010**, *50*, 1189–1204; c) D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* **2016**, *56*, 1243–1252.
- [14] a) S. S. De Buck, V. K. Sinha, L. A. Fenu, R. A. Gillissen, C. E. Mackie, M. J. Nijsen, *Drug Metab. Dispos.* **2007**; b) M. P. Gleeson, N. J. Waters, S. W. Paine, A. M. Davis, *J. Med. Chem.* **2006**, *49*, 1953–1963; c) V. K. Gombur, S. D. Hall, *J. Chem. Inf. Model.* **2013**, *53*, 948–957; d) H. Kaneko, K. Funatsu, *J. Chem. Inf. Model.* **2014**, *54*, 2469–2482; e) F. Lombardo, Y. Jing, *J. Chem. Inf. Model.* **2016**, *56*, 2042–2052; f) J. N. O'Donnell, A. Gulati, M. S. Lavhale, S. S. Sharma, A. J. Patel, N. J. Rhodes, M. H. Scheetz, *J. Pharm. Pharmacol.* **2016**, *68*, 56–62; g) P. H. van der Graaf, J. Nilsson, E. A. van Schaick, M. Danhof, *J. Pharm. Sci.* **1999**, *88*, 306–312.
- [15] S. Kar, R. N. Das, K. Roy, J. Leszczynski, *IJQSPR* **2016**, *1*, 23–51.
- [16] U. Norinder, L. Carlsson, S. Boyer, M. Eklund, *J. Chem. Inf. Model.* **2014**, *54*, 1596–1603.
- [17] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, *Nucleic Acids Res.* **2011**, *40*, D1100–D1107.
- [18] F. Lombardo, G. Berellini, R. S. Obach, *Drug Metab. Dispos.* **2018**, *46*, 1466–1477.
- [19] C. Standardizer, *ChemAxon: Budapest, Hungary* **2010**.
- [20] L. Weber, Vol. 5, *Royal Society of Chemistry*, Cambridge, UK, **2008**, pp. 65–66.
- [21] C. W. Yap, *J. Comput. Chem.* **2011**, *32*, 1466–1474.
- [22] R. Wehrens, B.-H. Mevik, *The pls package: principal component and partial least squares regression in R*, **2007**.
- [23] B. Ripley, W. Venables, *nnet: Feed-forward neural networks and multinomial log-linear models*, Vol. 7, **2011**.
- [24] A. Liaw, M. Wiener, *Breiman and Cutler's Random Forests for Classification and Regression Version (4.6-12)*, **2015**.
- [25] a) D. Alexander, A. Tropsha, D. A. Winkler, *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322; b) P. Gramatica, A. Sangion, *J. Chem. Inf.*

- Model.* **2016**, *56*, 1127–1131; c) R. Todeschini, D. Ballabio, F. Grisoni, *J. Chem. Inf. Model.* **2016**, *56*, 1905–1913.
- [26] Y. Cao, A. Charisi, L.-C. Cheng, T. Jiang, T. Girke, *Bioinformatics* **2008**, *24*, 1733–1734.
- [27] a) F. Lombardo, R. S. Obach, M. Y. Shalaeva, F. Gao, *J. Med. Chem.* **2004**, *47*, 1242–1250; b) M. P. Gleeson, *J. Med. Chem.* **2008**, *51*, 817–834.
- [28] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *J. Med. Chem.* **2014**, *57*, 4977–5010.
- [29] G. Berellini, C. Springer, N. J. Waters, F. Lombardo, *J. Med. Chem.* **2009**, *52*, 4488–4495.
- [30] L. H. Hall, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039–1045.
- [31] L. H. Hall, L. B. Kier, *J. Mol. Graphics Modell.* **2001**, *20*, 4–18.
- [32] K. Korzekwa, S. Nagar, *Pharm. Res.* **2017**, *34*, 544–551.
- [33] P. Tiikkainen, L. Bellis, Y. Light, L. Franke, *J. Chem. Inf. Model.* **2013**, *53*, 2499–2505.
- [34] C. Kramer, T. Kalliokoski, P. Gedeck, A. Vulpetti, *J. Med. Chem.* **2012**, *55*, 5165–5173.
- [35] a) T. Kalliokoski, C. Kramer, A. Vulpetti, P. Gedeck, *PLoS One* **2013**, *8*, e61007; b) G. Papadatos, A. Gaulton, A. Hersey, J. P. Overington, *J. Comput.-Aided Mol. Des.* **2015**, *29*, 885–896.
- [36] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2003**, *22*, 69–77.
- [37] D. L. J. Alexander, A. Tropsha, D. A. Winkler, *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322.
- [38] M. P. Gleeson, D. Montanari, *Expt. Opin. Drug Metabol. Toxicol.* **2012**, *8*, 1435–1446.
- [39] S. Weaver, M. P. Gleeson, *J. Mol. Graphics Modell.* **2008**, *26*, 1315–1326.
- [40] Y. Kwon, *Handbook of Essential Pharmacokinetics, Pharmacodynamics and Drug Metabolism for Industrial Scientists*, Springer US, New York, **2002**.
- [41] M. C. Wenlock, L. A. Carlsson, *J. Chem. Inf. Model.* **2014**, *55*, 125–134.
- [42] I. Poggesi, *Curr. Opin. Drug Discov. Devel.* **2004**, *7*, 100–111.
- [43] I. Mahmood, *J. Pharm. Pharmacol.* **1999**, *51*, 905–910.
- [44] N. A. Hosea, W. T. Collard, S. Cole, T. S. Maurer, R. X. Fang, H. Jones, S. M. Kakar, Y. Nakai, B. J. Smith, R. Webster, *J. Clin. Pharmacol.* **2009**, *49*, 513–533.
- [45] D. F. McGinnity, J. Collington, R. P. Austin, R. J. Riley, *Curr. Drug Metab.* **2007**, *8*, 463–479.
- [46] I. Cortes-Ciriano, A. Bender, *J. Chem. Inf. Model.* **2015**, *55*, 2682–2692.
- [47] a) J. J. Stezowski, *J. Am. Chem. Soc.* **1976**, *98*, 6012–6018; b) P.-E. Sum, P. Petersen, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1459–1462.
- [48] C. Martini, T. Gervasio, A. Lucacchini, A. Da Settimo, G. Primofiore, A. M. Marini, *J. Med. Chem.* **1985**, *28*, 506–509.
- [49] Å. Frostell-Karlsson, A. Remaeus, H. Roos, K. Andersson, P. Borg, M. Hämäläinen, R. Karlsson, *J. Med. Chem.* **2000**, *43*, 1986–1992.
- [50] E. L. Ellsworth, T. P. Tran, H. D. Hollis Showalter, J. P. Sanchez, B. M. Watson, M. A. Stier, J. M. Domagala, S. J. Gracheck, E. T. Joannides, M. A. Shapiro, S. A. Dunham, D. L. Hanna, M. D. Huband, J. W. Gage, J. C. Bronstein, J. Y. Liu, D. Q. Nguyen, R. Singh, *J. Med. Chem.* **2006**, *49*, 6435–6438.
- [51] M. F. Brown, M. J. Mitton-Fry, J. T. Arcari, R. Barham, J. Casavant, B. S. Gerstenberger, S. Han, J. R. Hardink, T. M. Harris, T. Hoang, M. D. Huband, M. S. Lall, M. M. Lemmon, C. Li, J. Lin, S. P. McCurdy, E. McElroy, C. McPherson, E. S. Marr, J. P. Mueller, L. Mullins, A. A. Nikitenko, M. C. Noe, J. Penzien, M. S. Plummer, B. P. Schuff, V. Shanmugasundaram, J. T. Starr, J. Sun, A. Tomaras, J. A. Young, R. P. Zaniewski, *J. Med. Chem.* **2013**, *56*, 5541–5552.
- [52] S. Kotretsou, M. P. Mingeot-Leclercq, V. Constantinou-Kokotou, R. Brasseur, M. P. Georgiadis, P. M. Tulkens, *J. Med. Chem.* **1995**, *38*, 4710–4719.
- [53] a) T. Ghafourian, M. Barzegar-Jalali, S. Dastmalchi, T. Khavari-Khorasani, N. Hakimiha, A. Nokhodchi, *Int. J. Pharm.* **2006**, *319*, 82–97; b) T. Ghafourian, M. Barzegar-Jalali, N. Hakimiha, M. T. Cronin, *J. Pharm. Pharmacol.* **2004**, *56*, 339–350.
- [54] M. P. Gleeson, S. Modi, A. Bender, R. L. Marchese Robinson, J. Kirchmair, M. Promkatkaew, S. Hannongbua, R. C. Glen, *Curr. Pharm. Des.* **2012**, *18*, 1266–1291.

Received: May 23, 2019

Accepted: July 3, 2019

Published online on ■■■, ■■■■



S. Simeon, D. Montanari, M. P. Gleeson*

1 – 12

Investigation of Factors Affecting
the Performance of *In silico* Volume
Distribution QSAR Models for
Human, Rat, Mouse, Dog & Monkey

