Original article

Prediction of volume of distribution values in human using immobilized artificial membrane partitioning coefficients, the fraction of compound ionized and plasma protein binding data

Xiaofan Sui [a], Jin Sun [a], Haiyan Li [a], Yongjun Wang [a], Jianfang Liu [b], Xiaohong Liu [a], Wenping Zhang [a], Lijiang Chen [c], Zhonggui He [a,*]

[a] Department of Biopharmaceutics, School of Pharmacy, Shenyang Pharmaceutical University, No. 59 Mailbox, No. 103 of Wenhua Road, Shenyang 110016, China
[b] Department of Pharmacy, Bethune International Peace Hospital, Shijiazhuang 050082, China
[c] School of Pharmacy, Liaoning University, Shenyang 110016, China

ABSTRACT

We developed an improved Lombardeo's method (J. Med. Chem., 2004) for prediction of VD_{ss} in human. With Elog D substituted by $\log k_{IAM}$, together with $f_{i\ (7.4)}$ (the fraction of compound ionized at pH 7.4) and $\log f_u$ (logarithmic fraction of compound unbound in plasma), the predictive equation of f_{ut} (the fraction of the compound unbound in tissues) for the 121 compounds was built, predictive VD_{ss} was further obtained from the Øie–Tozer equation. Our model could be applicable to structurally diverse compounds, including acids, bases, neutrals, and ampholytes. Interior and exterior validation results indicated the model had a robust predictive ability. Compared to the Lombardeo's and Hollósy's (J. Med. Chem., 2006) methods, our model can be generally applicable with better predictive accuracy.

© 2009 Elsevier Masson SAS. All rights reserved.

1. Introduction

With the increasing numbers of potential drug candidates via combinatorial chemistry and computer-aided design and the significant enhancement in the cost of drug development, it is of a great importance to simultaneously optimize human pharmacokinetic behavior of new drug candidates with multiple properties in the drug discovery and development processes. Screening large numbers of compounds for human pharmacokinetics (PKs), especially using animals [1], is a time-consuming, labour-intensive, and ethic-sensitive process, which limits its utility in drug discovery. From the first principle, the ideal situation for the pharmaceutical industry would be that the PKs of a drug candidate could be predicted based on solely its molecular structure and physicochemical properties [2].

One important aspect of PKs of a drug candidate is its distribution performance, which refers to the reversibly transfer capability of drug from the systemic circulation to the peripheral tissues. The extent of distribution is defined as the volume of distribution

(VD at steady state or VD_{ss} throughout this work), a function of the extent of drug partitioning to tissues vs. that remaining within the plasma. It is an important parameter to calculate the mean residence time MRT ($MRT = VD_{ss}/Cl$) and half-life of a drug ($t_{1/2} = 0.693(VD_{ss}/Cl)$), indicative of the drug exposure profile and the dosing frequency for setting up a suitable dosing regimen, which is intrinsically linked to other factors such as the pharmacodynamics and the therapeutical index.

Various methods have been proposed to characterize VD including (i) the extrapolation of animal data [3–8], (ii) physiologically based pharmacokinetic (PBPK) model [6,9–13] and (iii) computational (in silico) approaches [14–20], Additionally, two biomimetic-binding measurement, the phospholipids binding (immobilized artificial membrane IAM) and the plasma protein binding (human serum albumin HSA) have been used to model VD in human [21]. These methods have been discussed in detail in several review papers [22–25]. In brief, interspecies scaling could serve as a guideline for better planning of clinical studies; nevertheless, animal experiments are needed, so, labor, time, resources and even ethical considerations are involved in drug research and development processes. Consequently, the PBPK model and in silico approaches, requiring less or no animal data at all, are recommended. In these methods, a promising approach published by

Lombardo et al. [12,13] used the Øie–Tozer equation to calculate VD_{ss}, which described a relationship between VD_{ss} and f_{ut} (the fraction unbound of the drug in tissues). By measuring in vitro Elog D [26] at pH 7.4, $f_{i\ (7.4)}$ (the fraction ionized at pH 7.4) and f_u (the fraction unbound of the drug in plasma), a good estimate of the *in vivo* VD_{ss} was obtained, yielding an approximate 2-fold error between the predicted and the observed VD_{ss} values in humans. There also existed a very good correlation ($R^2 > 0.86$) between the observed and the predicted fraction unbound in tissue ($\log f_{ut}$) values for a set of 120 structurally diverse drugs. However, the Lombardo's model was applicable only to basic and neutral compounds and was not suitable to acidic and amphoteric drugs because the Elog D can only be accurately determined for basic and neutral drugs by Re-HPLC [26]. Moreover, the octanol/water distribution coefficient at physiological pH ($\log D_{/w\ 7.4}$) was reported as an inadequate parameter to describe the partition of ionized compounds (e.g. acids, bases and ampholytes) into biological membranes and serum albumin [27]. Immobilized artificial membrane chromatography (IAMC) is a good model to determine the lipophilicity equivalent to the liposome–water system [28,29] and is capable of being applied to various types of compounds, including bases, acids, neutrals and ampholytes. To date, IAMC has been applied to predict drug permeability [30,31], VD [21], certain pharmacological potency [32,33] and so on; and its application was reviewed in reference [34].

In this paper, we proposed a method to model the VD_{ss} based on the Lombardo's model. Substituting $\log k_{IAM}$ (logarithmic retention factor in IAM) for Elog D together with $f_{i\ (7.4)}$ and f_u, the VD_{ss} in humans of 121 drugs was modeled. In particular, the developed model acquired the same accurate prediction of acids and ampholytes to those of bases and neutrals. No requirement of *in vivo* experiments and animals, overall faster and cheaper accessibility of parameters used, wide application to various types of compounds, more simply modeling and more accurately predictive ability, make our model highly attractive for prospectively predicting the VD of new chemical entities in human.

2. Materials and methods

2.1. Materials

Most of the drugs (crude drugs) were generously donated by the pharmaceutical laboratories of Shenyang Pharmaceutical University (Shenyang, China). In several cases, they were samples extracted from commercial formulations. The analyte solutions, at concentrations of about 100 $\mu g\ ml^{-1}$, were prepared in acetonitrile or water. All chemicals and solvents were at least of analytical grade.

The chromatographic apparatus (SHIMADZU, Japan) was equipped with a constant flow pump LC-10AT, and an SPD-10A UV–vis detector operating between 200 and 400 nm. The detection was performed at a wavelength of 210 nm. The IAM.PC.DD2 column (15 cm × 4.6 mm) was filled with phosphatidylcholine (PC) residues covalently bonded to silica (Regis Technologies Inc., Morton Grove, IL).

2.2. Determinations of log k_{IAM}

The $\log k_{IAM}$ data were determined using the published method [35], based on a linear regression of capacity factors (as $\log k$) obtained from IAM chromatography determinations and extrapolated to 0% of organic solvent. The mobile phase consisted of 50 mM ammonium acetate and acetonitrile. The pH of the aqueous phase was adjusted to pH 7.0 with concentrated ammonia solution to prevent the loss of phospholipids at a pH of over 7.5. Chromatographic runs were carried out at 30 °C. The mobile phase flow rate

was 1.0 ml min^{-1} and the dead time (t_0) of the system was determined by injecting citric acid solution. For drugs with long retention times, 10, 15, 20, 25, 30, and 35% acetonitrile was used and the retention factor (k) was calculated by the following equation: $k = (t_r - t_0)/t_0$. Back linearly extrapolation of plots of $\log k$ versus at least four percentages of organic modifier with coefficients $r > 0.99$, allowed determination of the retention factor in an entirely aqueous solution. Each value was determined in duplicate. The $\log k_{IAM}$ data spanned over 5 units and were listed in Supplementary material.

2.3. Calculations of $f_{(7.4)}$

The pK_a data were taken from the literatures. When more than one source was available, the pK_a data determined under similar *in vivo* condition was selected. The $f_{i\ (7.4)}$ values were calculated using the following formula [36]: For basic molecules, $f_i = 1/(1 + 10^{7.4 - pK_a})$; for acidic molecules, $f_i = -1/(1 + 10^{pK_a - 7.4})$, and for amphoteric molecules, $f_i = 1/(1 + 10^{7.4 - K_{a(base)}})$. Amphoteric molecules were divided into two categories: (1) compounds with at least one basic $pK_a > 7$, which have been assumed that their distributional behavior has been comparable to base [11]; (2) all other amphoteric molecules which were not included in our drug set. Considering the charge property effect on compound distribution differently, the values of $f_{i\ (7.4)}$ of the acids charged negative at pH 7.4 were taken negative value; the values of $f_{i\ (7.4)}$ of the bases charged positive at pH 7.4 were taken positive value. The pK_a and $f_{i\ (7.4)}$ data were listed in Supplementary material.

2.4. Volume of distribution and plasma protein binding data

VD_{ss} and f_u data of 121 compounds were obtained from the scientific literature after an extensive comparative research. The data and reference sources were compiled in Supplementary material. In few cases, VD data of the compounds used for the calculation of f_{ut} had been reported as VD_{area} values rather than VD_{ss}, and they were used as such because the difference between VD_{area} and VD_{ss} was not usually clinically significant when to make drug dosing decisions [37]. In cases where the VD data were given for the whole body volume in liters, an average body weight of 70 kg for each study subject was assumed, and the VD_{ss} was expressed as L/kg.

2.5. Classification of the training and test sets

The 121 drugs were split into eight classes by cluster analysis (SAS 9.0 software). The physicochemical descriptors used by cluster analysis were 15 terms including calculated *n*-octanol/water partition coefficient (clog P), molecular weight (MW), rotatable bond (RotB), rings, the count of electron pairs on oxygen and nitrogen atoms (ON), the count of hydro-bond acceptor (HBA), the count of OH and NH bonds (OHNH), the count of hydro-bond donor (HBD), polar surface area (PSA) and total surface area (TSA) collected from the literature [38], except for apomorphine, quinine, clonidine, isoniazide, and antipyrine, which descriptors were calculated using Advanced Chemistry Development (ACD/Labs) Software V8.14; *n*-octanol/water partition coefficient at pH 7.0 ($\log D_{(7.0)}$) and molar solubility at pH 7.0 ($\log S$) were calculated using ACD/Labs Software V8.14; The other three parameters were $\log k_{IAM}$, $\log f_u$, and $f_{i\ (7.4)}$. Fourth-fifths of every class was randomly selected as a training set (97 of 121) to build predictive VD_{ss} model. The remaining 24 compounds were used as a test set to validate the predict ability of the developed VD_{ss} model.

2.6. Prediction of VD_ss in humans

The original form of the Øie–Tozer equation [39] (Eq. (1)) was used to calculate the VD_ss for the compounds.

$$VD_{ss} = V_p(1 + R_{E/I}) + f_u V_p(V_E/V_p - R_{E/I}) + \frac{V_R f_u}{f_{ut}} \qquad (1)$$

where VD_ss is the volume of distribution at steady state; f_{ut} and f_u are the fraction unbound in tissues and in plasma, respectively; $R_{E/I}$ is the ratio of binding proteins in extracellular fluid vs. plasma (1.4); V_P, V_E, and V_R are the volumes of plasma, extracellular fluid, and "remainder fluid" with values of 0.0436, 0.151, and 0.380 L/kg body weight in human, respectively. The f_u value could be determined from the in vitro experiment. If the f_{ut} value were known, the VD_ss would be calculated from the Øie–Tozer equation.

Our mechanistic approach to model f_{ut} was based on Brodie theory [40], which highlighted the importance of the proportion of the compound's partition coefficient between the free solution and the plasma as well as the free solution and the tissues. Then the f_{ut} can be determined by the compound's binding to the different components in tissues. The tissues contain a large amount of proteins, some of which are similar to those in plasma, and also contain large amounts of membranes and other lipid-containing components (illustrated by Fig. 1). Then, stated in logarithmic terms, we used f_u to simulate the compound's binding to proteins in tissue, and k_{IAM} to simulate the compound's binding to member/lipid partition in tissue. Considering the different states (ionized or molecular) have the different effects on the compound partition *in vivo*, f_i was included in the model. The log f_{ut} was used as the dependent variable, which was obtained as described in reference [13]. The log k_{IAM}, $f_{i\ (7.4)}$ and log f_u were used as the independent variables. The predictive equation of f_{ut} was built via multiple linear regressions (SAS 9.0 software). Then the predictive VD_ss values were further obtained from the Øie–Tozer equation used the calculated f_{ut} and experimental f_u values.

3. Results

3.1. The new predictive model of VD_ss using the 121 drugs

An equation for f_{ut} of the 121 compounds was built via multiple linear regression (MLR), using log k_{IAM} determined via our published method [35], an experimental determined pK_a for calculation of $f_{i\ (7.4)}$, and the logarithm of f_u values. After the f_{ut} is calculated from Eq. (2), VD_ss could be calculated via the Øie–Tozer equation. Average fold error (AFE) for the prediction of f_{ut} and VD_ss was 1.9273 and 1.6551, respectively. Fig. 2 (A) and (B) shows the plots of
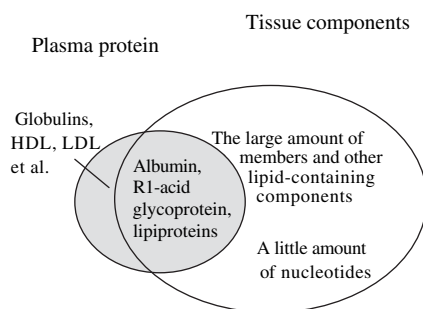

Fig. 1. The schematic diagram of the different components in tissues. The f_{ut} can be integrated by considering the compound's binding to proteins and the binding to member/lipid partition in tissues.
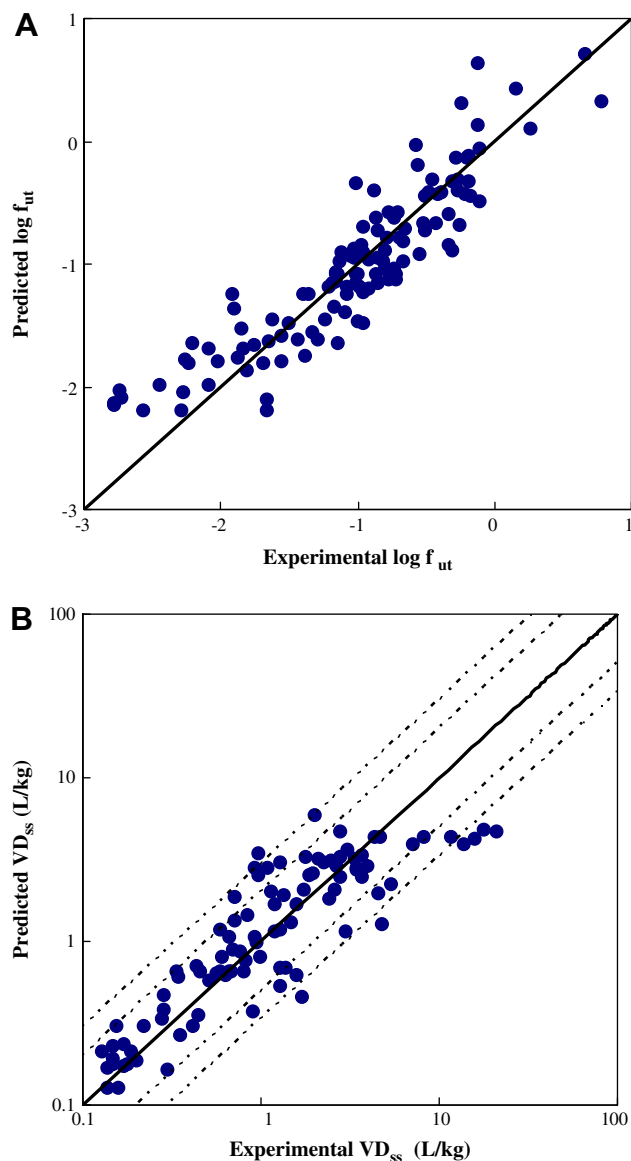

Fig. 2. Plot of calculated log f_{ut} vs. experimental log f_{ut} (A) and plot of predicted VD_ss vs. experimental VD_ss (B) for the 121 compounds. The dotted lines represent the 2-fold and 3-fold error limits.

the predicted vs. experimental log f_{ut} and VD_ss (L/kg) for the 121 drugs, respectively.

$$\log f_{ut} = -0.2197(\pm 0.0489) - 0.3199(\pm 0.0418)\log k_{IAM}$$
$$+ 0.4699(\pm 0.0715)\log f_u - 0.4069(\pm 0.0474)f_{i\ (7.4)}$$
$$(2)$$

$N = 121$; $R^2 = 0.8171$; $Q^2 = 0.8022$; RMSE = 0.3095; $F = 174.24$; $P < 0.0001$.

3.2. The new predictive model of VD_ss using the training and test sets

To further indicate the predictive power and robustness of our method, the data set consisting of 121 drugs was subdivided into a training set and a test set (4:1) via cluster analysis. Ninety-seven drugs in the training set were used to build a predictive equation (Eq. (3)); 24 drugs in the test set as an exterior set to validate the
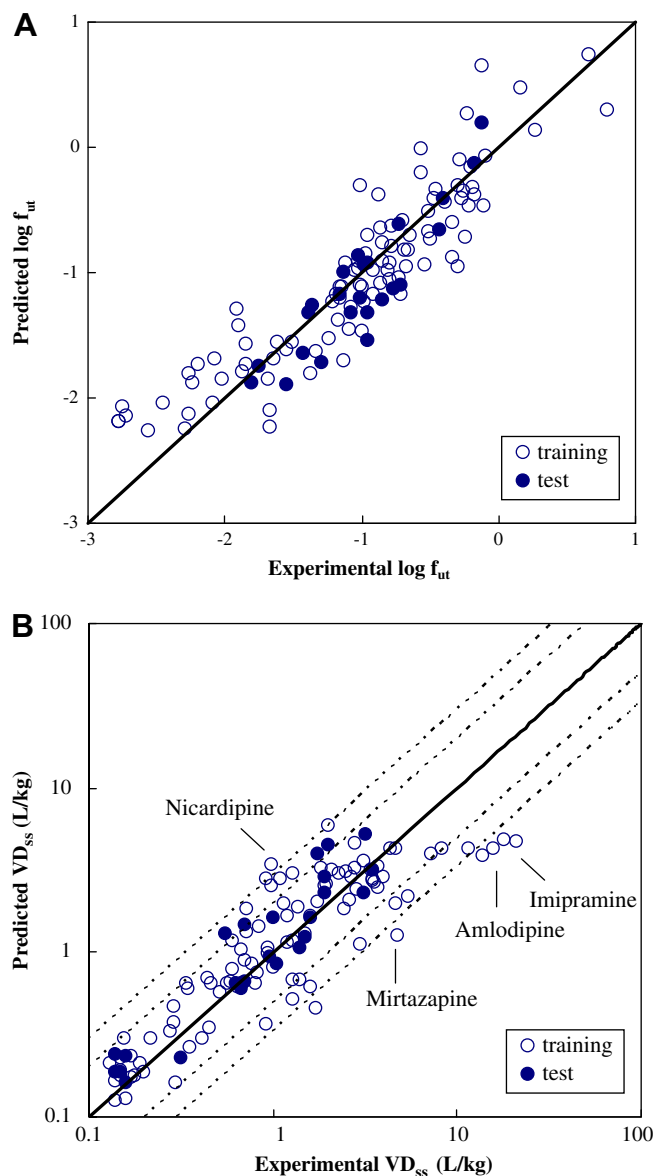
Fig. 3. Plot of calculated $\log f_{ut}$ vs. experimental $\log f_{ut}$ (A) and plot of predicted VD_{ss} vs. experimental VD_{ss} (B) for the 97 compounds in the training set and 24 compounds in the test set. The dotted lines represent the 2-fold and 3-fold error limits.

predictive equation. Fig. 3 (A) and (B) shows the plots of the predicted vs. experimental $\log f_{ut}$ and VD_{ss} (L/kg) for the drugs in the training set and the test set, respectively.

$$\log f_{ut} = -0.2168(\pm 0.0551) - 0.2908(\pm 0.0518)\log k_{IAM}$$
$$+ 0.5577(\pm 0.0956)\log f_u - 0.4497(\pm 0.0594)f_{i(7.4)}$$
$$(3)$$

$N = 97$; $R^2 = 0.8266$; $Q^2 = 0.8083$; RMSE $= 0.3253$; $F = 147.82$; $P < 0.0001$.

Average fold errors (AFEs) for the prediction of f_{ut} and VD_{ss} of 97 compounds in training set and 24 compounds in test set were 2.0012 and 1.7211, 1.6045 and 1.3883, respectively. For the test set, VD_{ss} as the end-point of prediction, 92% and 100% compounds were within 2-fold and 3-fold error of experimental values, respectively. These statistic results indicated the potent predictive power of this method, especially considering the often error

margin for clinical and biological data in general and the heterogeneity of the data.

3.3. The leave-one-class-out

Based on cluster analysis, the 121 compounds were categorized into eight classes; simultaneously, we have identified nine classes of pharmacological analogues. The leave-class-out cross-validations of our predictive model of VD_{ss} in humans were performed according to the above two kinds of classification. Table 1 shows the results of these two tests. The removal of each class generated new equations with similar predictive power and AFEs of f_{ut} and VD_{ss} of every class close to or within 2-fold error (the lowest fold-error of VD_{ss} to 1.3 for cephalosporin's analogues and seventh class), except for psychotolytics analogues and the second cluster class. In which amlodipinel, mirtazapine and imipramine are underestimated by 4-fold factor; nicardipine is overestimated by 3-fold factor.

4. Discussions

4.1. The representative of the data set

The compounds investigated in this study were all available commercially and selected to cover a wide variety of chemical structure, pharmacological activity and PK characteristic. The data for constructing the predictive model should be representative, which is the first premise to construct the predictive model. Normal test was processed for the $\log f_{ut}$ dataset of 121 drugs. The plot of number of observations vs. the $\log f_{ut}$ values showed that the data were normally distributed (see Fig. 4) (Shapiro–Wilk Normal $W = 0.9722$; $Pr < W$, 0.0131; Skewness $= -0.4159$; and Kurtosis $= 0.1807$). Principal component analysis (PCA) of the three variables for the whole dataset was performed. The scores of plot for the first two principal components (PCs) explained about 93.74% of the total variance of the matrix, and the drugs in the training set and those in test set were overlapping in the score plot and greatly

Table 1
Leave-class-out statistics.

	No. of compounds in class	R^2 [a]	AFE of f_{ut}	AFE of VD_{ss}
Class by pharmacological analogues				
Hormones	13	0.8261	1.9900	1.4846
Quinolones	5	0.8202	1.6409	1.5822
Cephalosporins	5	0.8217	3.0667	1.3229
Drugs of digestive system	7	0.8250	2.0349	1.8616
Hypotensive drugs	7	0.8226	2.5421	2.3020
β-Blockers	5	0.8183	1.7498	1.6378
Antipyretic analgesics and anti-inflammatories	12	0.8143	1.9884	1.3430
Psychotolytics	13	0.8150	3.0673	2.8829[b]
Pain-killers	5	0.8178	1.5153	1.4853
Class by cluster analysis				
1	21	0.7982	1.6213	1.4264
2	23	0.7977	3.2314	3.0023[b]
3	22	0.8378	1.8442	1.7080
4	19	0.8254	1.6178	1.4265
5	12	0.8181	2.2911	1.4043
6	16	0.8284	2.5613	2.0525
7	7	0.8108	2.8843	1.3026
8	1	0.8170	1.2633	1.2452

[a] The correlation coefficient of experimental and predictive values of the $\log f_{ut}$ regression as in Eq.(2), based on a data set where this class compounds were left out from 121 compounds.
[b] Mirtazapine, imipramine in this class were underestimated over 4-fold error.
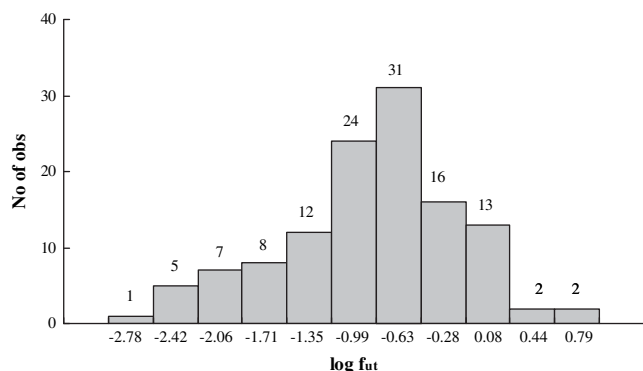
Fig. 4. Histogram of the data distribution of the experimental $\log f_{ut}$ values. Shapiro–Wilk Normal $W = 0.9722$; Pr < W, 0.0131; Skewness $= -0.4159$; and Kurtosis $= 0.1807$.

representative to each other (see Fig. 5). Therefore, we considered this date set being a reasonably large, representative set of drugs to develop a general predictive model.

4.2. The new predictive model of VD_{ss}

In view of Eqs. (2) and (3), three independent variables were all statistically significant and were not correlated each other, the greatest correlation coefficient (-0.55) between $\log k_{IAM}$ and $\log f_u$. We had already tried to refine the model through exploration of computed parameters used in cluster analysis, but no statistically significant term was introduced to the final predictive model by the stepwise regression.

The coefficients of $\log k_{IAM}$, $\log f_u$ and $f_{i\,(7.4)}$ for the centered and scaled data were -0.3347, 0.2919 and -0.3340, respectively. So the influence of the three variables on the $\log f_{ut}$ followed the sequence: $f_{i\,(7.4)} \approx \log k_{IAM} > \log f_u$. Compared to the plasma protein binding (f_u), phospholipid membrane (mimic biomembrane) binding ($\log k_{IAM}$) and the charge property of compound ($f_{i\,(7.4)}$) were more principal factors on tissue distribution ($\log f_{ut}$ and further VD_{ss}). It was generally accepted that the fraction ionized or the electronic state of the compound affected both f_u and $\log k_{IAM}$ [35]. The log-k_{IAM} had a higher value for its coefficient compared to that of Elog D.
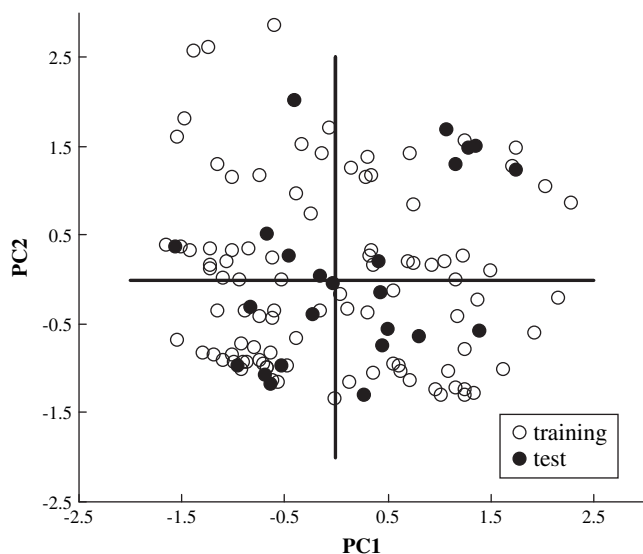

Fig. 5. Score plot of PC1 vs. PC2 for the PCA analysis. The score plot just obtained from $\log k_{IAM}$, $\log f_u$ and $f_{i\,(7.4)}$.

in the Lombardo's equation [13], however, the $\log f_u$ and $f_{i\,(7.4)}$ had lower values for their coefficient. Simultaneously, we noted that for basic and neutral compounds, there was co-linearity between $\log f_u$ and $\log k_{IAM}$ ($r = -0.76$), similar to the relationship between $\log f_u$ and Elog D ($r = 0.80$) [12]. But for all compounds or only for acidic compounds, correlation did not exist ($r = -0.55$, -0.51, respectively). Maybe the reason was that $\log k_{IAM}$ represented not only hydrophobicity [21] but also electronic factor [35] in distribution or partition process between aqueous and biomembrane phases.

The signs of the coefficients of our equation were physically reasonable and agreed with those in Lombardo's equation. The negative coefficient of the $f_{i\,(7.4)}$ term conformed to the general considering that cations tend to bind anionic cellular and tissue components represented greatly by membrane phospholipids; anions to albumin in plasma, so increasing $f_{i\,(7.4)}$ of basic drugs and decreasing $f_{i\,(7.4)}$ of acidic drugs would improve tissue binding ($\log f_{ut}$ decreased). The negative coefficient of the $\log k_{IAM}$ term meant that more lipophilic compounds would have less free fraction in the tissues resulting in the larger VD_{ss} values. The positive coefficient of $\log f_u$ implied that the lower plasma protein binding would increase f_{ut} values, which is reasonable due to the large amount of extravascular proteins presenting in interstitial fluids and in cell membranes.

Compared to Eq. (2) for the 121 drugs, the similar correlation coefficient ($R^2 = 0.82$) and greater predictive results were achieved as seen from Eq. (3). AFEs for f_{ut} of the compounds in training set and in test set were both lower than 2-fold error; AFEs for VD_{ss} were both lower than 1.72-fold error. The coefficients of $\log k_{IAM}$, $\log f_u$ and $f_{i\,(7.4)}$ were comparable to those of Eq. (2). The different size of data might cause some change in the coefficients, but the overall statistical quality and predictive power were not changed, demonstrating the ruggedness of the method. This method and Øie–Tozer equation have the same fundamental assumption that passive diffusion is the only or predominant mechanism of tissue penetration, a purely equilibrium (diffusion) distribution and average binding in all tissues relied heavily on the physicochemical properties of the compound. However, some compounds such as nicardipine, mirtazapine, imipramine, and amlodipine in training set had larger predictive deviation (>3-fold error) as shown in Fig. 3 (B). The reasons may be explained as follows: mirtazapine, imipramine whose VD_{ss} were underestimated have specific affinity for lysosomes and aggregate within lysosomes, as explained in literature [13]. However, we had not the clear reason for overestimated nicardipine. In addition, two compounds loratadine and trifluoperazine which had significant predictive deviation (≫5-fold error) were not included in 121 dataset because the large VD_{ss} (120 and 160 L/kg, (http://www.druginfosys.com) respectively), implied that the two compounds had tight and specific binding to certain constituents of the tissues and organs, which did not agree with the premise of Øie–Tizer equation.

The leave-class-out approach is significant in the investigation of new compound that a predictive modeling based on known classes of drugs would be applicable to the novel classes. Good results of validation of two leave-one-class approaches in drugs set (AFEs of f_{ut} and VD_{ss} of every class close to or within 2-fold error) explore the usefulness and ruggedness of our method for predicting VD_{ss}.

4.3. Compared with two published models

We have also compared our model with other published models. Among the 121 compounds in our dataset, 44 compounds overlapped with Lombardo's [13] dataset and 57 compounds overlapped with Hollósy's [21] dataset. Using VD_{ss} as the end-point of the prediction, we listed the comparative result in Table 2. The AFEs of prediction and accuracy based on Lombardo's model were

Table 2
Comparison of the different predictive models of VD_{ss} for the overlapping drugs.

	No. of compounds overlapped	AFE	FE < 2[a]	FE < 3[a]
Equation published by Lombardo et al.	44	1.6243	0.8409	0.9318
Eq. (2)	44	1.8293	0.6818	0.9091
Eq. (4) [b]	44	1.6913	0.7273	0.9091
Equation published by Hollósy et al.	57	1.9614	0.6897	0.8448
Eq. (2)	57	1.5516	0.8596	0.9474

[a] Percentage of compounds predicted to have a VD_{ss} value within a 2-fold error and 3-fold error from experiment value, respectively.
[b] The equation is constructed only using basic and neutral compounds.

$$\log f_{ut} = -0.1398 - 0.3038 \log k_{IAM} + 0.6189 \log f_u - 0.4921 f_{i(7.4)}. \quad (4)$$

$$n = 78 \ R^2 = 0.8412 \ Q^2 = 0.8230 \ RMSE = 0.2869 \ F = 130.66.$$

slightly better than those obtained from the models of the whole dataset with 121 compounds and of the basic and neutral dataset with 78 compounds. However, our model can be applied to structurally diverse compounds while the Lombardo's model could only be applied to basic and neutral compounds. Further, for 57 compounds overlapped with Hollósy's dataset, our model yielded the same accurate predictive results as Lombardo's model, significantly better than Hollósy's model.

5. Conclusion

We successfully developed a method to model the VD_{ss} for acids, bases, neutrals and ampholytes using the $\log k_{IAM}$ (logarithmic retention factor in IAM), the fraction of compound ionized at pH 7.4, and the fraction of unbound drug in plasma. The VD_{ss} dataset in humans of 121 known drugs was modeled which was internally and externally validated. The method yielded average fold error for VD_{ss} lower than 2-fold. Compared to literature methods, our model could be applied to acids and ampholytes which were not included in Lombardo's model; and it yielded the similar accurate predictive results to Lombardo's model, but significantly better than Hollósy's model.

This method does not require *in vivo* animal experiments, costs less time and resource, is suitable for structure diverse compounds and has better predictive accuracy, all of which make it as an attractive and promising approach for predicting VD_{ss} of new drug candidates. With the in-depth study of quantitative structure retention relationship of immobilized artificial membrane chromatography [35,41,42], and the high-throughout determination of the fraction of unbound drug in plasma [43,44] and the ionized fraction of compound [45], our method would have the higher speed for screening candidate drugs and would be accepted by world-wide pharmaceutical researchers.

Acknowledgments

Appendix. Supplementary information

References

[1] T. Wajima, K. Umura, Y. Yano, T. Oguma, J. Pharm. Pharmacol. 55 (2003) 939–949.
[2] A. Avdeef (Ed.), Absorption and Drug Development, Wiley, Hoboken, NJ, 2003.
[3] H. Boxenbaum, J. Pharmacokinet. Biopharm. 10 (1982) 411–426.
[4] Y. Sawada, M. Hanano, Y. Sugiyama, H. Harashima, T. Iga, J. Pharmacokinet. Biopharm. 12 (1984) 587–596.
[5] Y. Sawada, M. Hanano, Y. Sugiyama, T. Iga, J. Pharmacokinet. Biopharm. 13 (1985) 477–492.
[6] R.S. Obach, J.G. Baxter, T.E. Liston, B.M. Silber, B.C. Jones, F. MacIntyre, D.J. Rance, P. Wastall, J. Pharmacol. Exp. Ther. 283 (1997) 46–58.
[7] V. Karalis, P. Macheras, Pharm. Res. 19 (2002) 697–704.
[8] S. Woo, W.J. Jusko, Drug Metab. Dispos. 35 (2007) 1672–1678.
[9] P. Poulin, F.P. Theil, J. Pharm. Sci. 91 (2002) 129–156.
[10] H.M. Jones, N. Parrott, K. Jorga, T. Lavé, Clin. Pharmacokinet. 45 (2006) 511–542.
[11] T. Rodgers, M. Rowland, J. Pharm. Sci. 95 (2006) 1238–1257.
[12] F. Lombardo, R.S. Obach, M.Y. Shalaeva, F. Gao, J. Med. Chem. 45 (2002) 2867–2876.
[13] F. Lombardo, R.S. Obach, M.Y. Shalaeva, F. Gao, J. Med. Chem. 47 (2004) 1242–1250.
[14] H. Van de Waterbeemd, E. Gifford, Nat. Rev. Drug Discovery 2 (2003) 192–204.
[15] V. Karalis, A. Tsantili-Kakoulidou, P. Macheras, Eur. J. Pharm. Sci. 20 (2003) 115–123.
[16] C. Ng, Y. Xiao, W. Putnam, B. Lum, A. Tropsha, J. Pharm. Sci. 93 (2004) 2535–2544.
[17] J.V. Turner, D.J. Maddalena, D.J. Cutler, Int. J. Pharm. 270 (2004) 209–219.
[18] T. Ghafourian, M. Barzegar-Jalali, S. Dastmalchi, T. Khavari-Khorasani, N. Hakimiha, A. Nokhodchi, Int. J. Pharm. 319 (2006) 82–97.
[19] P. Gleeson, N.J. Waters, S.W. Paine, A.M. Davis, J. Med. Chem. 49 (2006) 1953–1963.
[20] F. Lombardo, R.S. Obach, F.M. DiCapua, G.A. Bakken, J. Lu, D.M. Potter, F. Gao, M.D. Miller, Y. Zhang, J. Med. Chem. 49 (2006) 2262–2267.
[21] F. Hollósy, K. Valkó, A. Hersey, S. Nunhuck, G. Kéri, C. Bevan, J. Med. Chem. 49 (2006) 6858–6971.
[22] J. Keldenich, Drug Discov. Today Technol. 1 (2004) 389–395.
[23] U. Fagerholm, J. Pharm. Pharmacol. 59 (2007) 1181–1190.
[24] D.E. Mager, Adv. Drug Delivery Rev. 58 (2006) 1326–1356.
[25] X.F. Sui, J. Sun, X. Wu, H.Y. Li, J.F. Liu, Z.G. He, Curr. Drug Metab. 9 (2008) 574–580.
[26] F. Lombardo, M.Y. Shalaeva, K.A. Tupper, F. Gao, J. Med. Chem. 44 (2001) 2490–2497.
[27] K. Valko, in: R.T. Borchardt, E.H. Kerns, C.A. Lipinski, D.R. Thakker, B. Wang (Eds.), AAPS, Washington DC, 2004, pp. 127–182.
[28] J. Sun, Y. Deguchi, J.M. Chen, K. Morimoto, Acta Pharmacol. Sin. 23 (2003) 430–438.
[29] J. Sun, Y. Deguchi, J.M. Chen, R.H. Zhang, K. Morimoto, Die Pharmazie 57 (2002) 497–498.
[30] J. Sun, Y. Deguchi, Y. Tauchi, Z. He, G. Cheng, K. Morimoto, Eur. J. Pharm. Biopharm. 64 (2006) 238–245.
[31] Y. Deguchi, J. Sun, Y. Tauchi, S. Sakai, K. Morimoto, Drug Metab. Pharmacokinet. 18 (2003) 319–326.
[32] F. Barbato, M.I. La Rotonda, F. Quaglia, Pharm. Res. 14 (1997) 1699–1705.
[33] F. Barbato, M.I. La Rotonda, F. Quaglia, J. Pharm. Sci. 86 (1997) 225–229.
[34] J. Sun, X. Wu, R. Lu, J. Liu, Y. Wang, Z. He, Curr. Drug Metab. 9 (2008) 152–166.
[35] J. Li, J. Sun, S.M. Cui, Z.G. He, J. Chromatogr. A 1132 (2006) 174–182.
[36] A.N. Martin, J. Swarbrick, A.Cammarata (Eds.), Physical Pharmacy, Physical Chemical Principles in the Pharmaceutical Sciences, Lea & Febiger, Philadelphia, PA, 1969.
[37] G.R. Wilkinson, in: J.G. Hardman, L.E. Limbird, A.G. Gillman (Eds.), Goodman & Gillman's the Pharmacological Bases of Therapeutics, 10th ed, McGraw-Hill, New York, 2001, pp. 20–22.
[38] M. Vieth, M.G. Siegel, R.E. Higgs, I.A. Watson, D.H. Robertson, K.A. Savin, G.L. Durst, P.A. Hipskind, J. Med. Chem. 47 (2004) 224–232.
[39] S. Øie, T.N. Tozer, J. Pharm. Sci. 68 (1979) 1203–1205.
[40] B.B. Brodie, H. Kurtz, L.J. Schanker, J. Pharmacol. Exp. Ther. 130 (1960) 20–25.
[41] J. Li, J. Sun, Z.G. He, J. Chromatogr. A 1140 (2006) 174–179.
[42] L. Sprunger, B.H. Blake-Taylor, A. Wairegi, W.E. Acree Jr., M.H. Abraham, J. Chromatogr. A 1160 (2007) 235–245.
[43] M.J. Banker, T.H. Clark, J.A. Williams, J. Pharm. Sci. 92 (2003) 967–974.
[44] H. Hong Wan, M. Rehngren, J. Chromatogr. A 1102 (2006) 125–134.
[45] Z.J. Jia, Curr. Pharmaceut. Anal. 1 (2005) 41–56.