Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 6, Issue 9, 2014

Original Article

IN SILICO QUANTITATIVE STRUCTURE – PHARMACOKINETIC RELATIONSHIP MODELING ON ACIDIC DRUGS: HALF LIFE

ZVETANKA ZHIVKOVA, IRINI DOYTCHINOVA

Department of Chemistry, Faculty of Pharmacy, Medical University –Sofia.
Email: zdzhivkova@pharmfac.acad.bg

Received: 04 Jul 2014 Revised and Accepted: 20 Aug 2014

ABSTRACT

Objective: Drug half-life ($t_{1/2}$) is one of the key pharmacokinetic parameters for establishment of dosing regimen. Surprisingly, the relationship between the chemical structure and $t_{1/2}$ is still poorly explored. The aim of the present study is to derive quantitative structure – pharmacokinetic relationships (QSPkRs) for $t_{1/2}$ of acidic drugs.

Methods: The dataset consisted of 142 molecules which were described with 187 structural and physicochemical descriptors. A three step variable selection procedure was applied to identify the most reliable descriptors. QSPkR modeling was performed using multivariate regression analysis (MLR).

Results: A number of sound and robust QSPkR models were derived. The predictive ability of the models was tested by internal and external validation procedure. The most frequently emerged descriptors were used for construction of a consensus model for $t_{1/2}$ prediction. The model is statistically significant (explained variance 0.688) and predictive (cross validation correlation coefficient 0.600, mean fold error of prediction 2.06, accuracy 61%). It reveals the main structural features affecting $t_{1/2}$. A short check list was proposed determining the cutoff between short half life ($t_{1/2}$ < 1 h) and long half life ($t_{1/2}$ > 24 h) drugs.

Conclusion: The presence of a sulfonyl or phosphonate groups, non-polar substituents at aromatic carbon, 9- or 10-member ring system and donor-acceptor pair separated by 9 skeletal bonds contribute to prolongation of $t_{1/2}$, while the presence of methane group, polar substituents at aromatic carbon and 7-member ring system affect negatively $t_{1/2}$.

Keywords: Computational ADME, Half-life prediction, *In silico* modeling, QSPkR, MLR, Acidic drugs.

INTRODUCTION

The development of a new drug is a long and expensive process. Unfortunately a sizable number of new drug candidates successfully passed the early preclinical trials do not reach the market due to undesirable pharmacokinetic behavior [1]. Hence, in order to be efficient *in vivo*, the drug requires a suitable ADME (absorption, distribution, metabolism and excretion) profile. The recognition of this fact inspired an extensive research aiming to predict the ADME properties in the earliest stages of drug development and to minimize the risk of late stage failures. As a result for the period from 1991 to 2000 years in the late stage candidate attrition due to pharmacokinetics reasons was reduced by approximately 30% [2].

The earliest predictive methods are based on *in vivo* animal pharmacokinetic studies or *in vitro* metabolism data. More recently *in silico* modeling gained increasingly popularity and utility owing to its ability to predict the ADME properties of new drug candidates solely by means of computational techniques, avoiding the need of the time consuming and expensive animal experiments. Besides, *in silico* techniques allow a prediction to be made even for virtual compounds and may provide guidance for targeted synthesis of molecules with desired ADME profile thus accelerating the identification of new drugs and reducing their development costs. One of the most widely used *in silico* approaches is the development of quantitative structure – pharmacokinetics relationships (QSPkR). QSPkR methodology focuses on the development of a mathematical relationship (model) relating the endpoint (pharmacokinetics parameter) to the chemical structure (encoded in structural descriptors) within a group of compounds. The number of reports on successful application of *in silico* methodology for ADME prediction increases, and is a subject of several reviews [3-9].

One of the most important pharmacokinetic parameters is the half-life $t_{1/2}$ as it, together with the therapeutic index, dictates the frequency of dosing.

The maximal dosing interval τ_{max} in which the concentration is maintained within the therapeutic range (between the accepted values of C_{min} and C_{max}) is calculated according to the equation:

$$\tau_{max} = 1.44 * t_{1/2} * ln\frac{C_{max}}{C_{min}}$$

Usually the term half-life refers to the elimination half-life and is determined following *iv* administration in order to avoid the influence of the absorption. Elimination half-life represents the time required for the plasma concentration to reduce by a half after pseudo-equilibrium of distribution is reached between plasma and tissues and the further decrease in plasma concentration is due solely to elimination [10]. For a drug with linear elimination $t_{1/2}$ is calculated from the slope of the terminal linear part of the *lnC/t* curve, corresponding to the rate constant of elimination λ_z:

$$t_{1/2} = \frac{ln\,2}{\lambda_z}$$

Despite the general agreement on the key importance of $t_{1/2}$ for dosing regimen design, there are surprisingly few attempts for its prediction. The early studies are based on *in vivo* animal experiments – allometric scaling or animal versus human cross-drug correlations [11]. The values for the human $t_{1/2}$ are calculated combining the individual predictions of the clearance *CL* and the steady state volume of distribution V^{ss} assuming the following simple relation:

$$t_{1/2} = \frac{ln\,2 * V^{ss}}{CL}$$

It is considered that as $t_{1/2}$ is a composite and dependent parameter, it is more appropriate to use individual models for prediction of *CL* and V^{ss}, and then consider how these factors, acting in a concert, influence $t_{1/2}$ value [12]. However, this simple approach has certain drawbacks. The accuracy of the prediction of $t_{1/2}$ is a function of the accuracy of the prediction of the independent parameters VD^{ss} and *CL*[13]. Another shortcoming is the use of V^{ss}. While perfectly suitable for drugs with one-phase distribution, for drugs with multiphase kinetics the upper relation can result in an under-

prediction of $t_{1/2}$ [11]. For these drugs, the terminal $t_{1/2}$ is related to the terminal volume of distribution V_β which may be much greater than V^{ss} [14]. Both V_β and the slope of the terminal lnC/t phase depend on the rate of drug transfer between plasma and tissues. Therefore for many drugs with multiphase kinetics the observed terminal $t_{1/2}$ differs considerably from the calculated value. For example, the ACE inhibitor Enalapril at shows a biphasic kinetics with a terminal $t_{1/2}$ = 39 h [15]. However, taking into account the reported data for V^{ss} (0.38 L/kg) and CL (6.1 L/h), the calculated $t_{1/2}$ value should be about 3 h. The prolonged terminal phase of Enalapril at is attributed to the slow release of the drug from its complexes with ACE.

We found only two reports on successful *in silico* prediction of $t_{1/2}$ for congeneric series of drugs – fluoroquinolones [16] and antidiabetic agents [17]. Therefore, the prediction of drug half life with *in silico* methodology appears to be a challenging problem. Recently we published a series of reports on the application of *in silico* approach for prediction of key pharmacokinetics properties (steady state volume of distribution, plasma protein binding and unbound clearance) of acidic drugs [18-20]. This study completes our investigations with a modeling of the quantitative relationships between chemical structure and half-life.

EXPERIMENTAL SECTION

Dataset

Success of QSPkR modeling depends crucially on the appropriate selection of the dataset. The dataset used in the present study involves values for $t_{1/2}$ of 142 acidic drugs following *iv* administration, extracted from Obach-Lombardo-Waters database [21]. The drugs are classified as acids, bases, neutral, and zwitterions on the basis of their ionization at physiological pH 7.4. The fractions of the drug ionized as an acid (f_A) and as a base (f_B) are calculated by the equations:

$$f_A = \frac{1}{1+10^{pK_a-7.4}} \quad \text{and} \quad f_B = \frac{1}{1+10^{7.4-pK_a}}$$

The mol files of the drugs are taken from Databank [22] or chemical Book [22]. The pK_a values are calculated using ACD/LogD version 9.08 software (Advanced Chemistry Development Inc., Ontario, Canada). If more than one basic or acidic center present in the molecule, the pK_a of the strongest one is considered. A drug is classified as an acid in two cases: if f_A exceeds 10 % while f_B is negligible or if f_A exceeds f_B and is close to 100%.

The dataset was used for construction of six modeling sets – each one composed by a training set and an external test set. To this end the drugs were arranged in an ascending order with respect to $t_{1/2}$ and were divided to six subsets by allocating one of every six drugs into a different subset. Every subset was used once as a test set for external validation of the models developed by the respective training set comprising the remaining five subsets. For modeling purposes, $t_{1/2}$ was presented as $log\ t_{1/2}$.

Descriptors

The descriptors used in this study were calculated using the software packages ACD/LogD version 9.08 software (Advanced Chemistry Development Inc., Ontario, Canada) and MDL QSAR version 2.2 (MDL Information systems, Inc., San Leonardo, California. Total of 187 descriptors were derived including electro-topological indices, molecular connectivity indices, descriptive properties (the number of atoms of given atom type, rings, hydrogen bond donors and acceptors, etc.), integral 2D (molecular weight, *logP*, $log\ D_{7.4}$, etc.) and 3D (polarizability, surface area, volume, etc.) Properties.

Variable selection

A three step variable selection procedure was performed in order to derive the most relevant descriptors for $t_{1/2}$ prediction. The initial screening reduced the number of descriptors to 145 as descriptors with nonzero values for less than 3 molecules and descriptors correlating to log $t_{1/2}$ with r < 0.1 were excluded. Further selection was performed for everyone training set by applying the genetic algorithm (GA) in order to avoid over-fitting. Selected descriptors entered a step wise linear regression for construction of QSPkRs for $t_{1/2}$.

Using different combinations of descriptors, a number of QSPkR models were derived for each training set. Their performances were assessed by an explained variance (r^2), cross-validated coefficient (q^2), external validation coefficient (r^2_{pred}), accuracy and mean fold error of prediction (*MFEP*) defined in the next section. Descriptors, which emerged in more than 20% of the models, were selected for development of a consensus QSPkR model.

Model assessment and validation

The QSPkR models constructed in the present study were assessed by the explained variance (r^2) given by the equation:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}\left(log\ t_{1/2,obs,i} - log\ t_{1/2,calc,i}\right)}{\sum_{i=1}^{n}\left(log\ t_{1/2,obs,i} - log\ t_{1/2,obs,mean}\right)}$$

Where $t_{1/2,obs,i}$ is the observed $t_{1/2}$ of the ith drug, $t_{1/2,calc,i}$ is the calculated by the model $t_{1/2}$ of the ith drug, and $t_{1/2,obs,mean}$ is the mean value of the observed $t_{1/2}$.

The predictive power of the models was explored by internal and external validation procedures. Internal validation consisted of leave-one-out cross-validation in the training tests (LOO-CV). In this approach each drug is excluded one by one, in turn, from the training set, and a QSPkR model is constructed using the remaining n – 1 compound.

Eventually, the model is used to predict $t_{1/2}$ of the excluded drug. External validation used external test sets of drugs which were not used in any step of model development. The quality of the models was assessed by the coefficients $q^2_{LOO\text{-}CV}$ and r^2_{pred} following the equations:

$$q^2_{LOO-CV} = 1 - \frac{\sum_{i=1}^{n}\left(log\ t_{1/2,obs,i} - log\ t_{1/2,pred,i}\right)}{\sum_{i=1}^{n}\left(log\ t_{1/2,obs,i} - log\ t_{1/2,obs,mean}\right)}$$

$$r^2_{pred} = 1 - \frac{\sum_{i=1}^{n}\left(log\ t_{1/2,obs,i-test} - log\ t_{1/2,pred,i-test}\right)}{\sum_{i=1}^{n}\left(log\ t_{1/2,obs,i-test} - log\ t_{1/2,obs,mean-test}\right)}$$

where $t_{1/2,pred,i}$ is the predicted by the model value of $t_{1/2}$ of the ith drug, and $t_{1/2,obs,i\text{-}test}$, $t_{1/2,obs,i\text{-}mean}$ and $t_{1/2,pre,i\text{-}test}$ are the observed, mean and predicted by the model values of $\mathbf{t_{1/2}}$ for any drug from the external test set, respectively.

The fold error of prediction (*FE*) was calculated as follows:

$$\mathbf{FE} = \mathbf{10}^{\left|\mathbf{log\ t_{1/2,obs,i} - log\ t_{1/2,pre,i}}\right|}$$

The average value of *FE*s represents the mean fold error of prediction (*MFE*). The prediction accuracy is assessed as a percent of the total number of drugs which $t_{1/2}$ is predicted with less than 2-fold error.

RESULTS

Data set analysis

The dataset used in the present study consisted of 142 acidic drugs belonging to different chemical and therapeutic classes. The molecular weight M_w varies between 126 and 1297 g/mol (mean 376.5 g/mol; median 346.8 g/mol). *Log P* ranges between -7.48 and 8.39 (mean 1.49; median 1.52), and $log\ D_{7.4}$ – between -11 and 7.64 (mean -1.5; median -1.36). Most of the drugs (85%) are completely ionized as acids at the physiological pH 7.4 while for only 8% f_a < 0.5. The VD^{ss} is relatively low – it varies between 0.04 and 15L/kg (mean 0.525 L/kg, median 0.220L/kg), and exceeds 0.7L/kg for only 15 drugs (10.5%). The *CL* also varies significantly – between 0.06 and 1070 mL/min/kg (mean 10.82, median 2.10 mL/min/kg). Most of the acids are highly bound to plasma proteins with a free fraction f_u

in the range 0.0004 – 1 (mean 0.3, median 0.15). The values for $t_{1/2}$ vary between 0.12 and 1200 h (mean 17.2, median 1.8 h). 32 drugs (22.5%) have $t_{1/2}$ < 1 h, while for 15 drugs (10.6%) $t_{1/2}$ > 24.

QSPkR models for $t_{1/2}$

In order to derive a robust and predictive QSPkR model for *log* $t_{1/2}$, the whole dataset was divided into six training sets as described in the Experimental section. Four training sets consisted of 118, and two training sets – of 119 molecules. Each training set differed from the others in about 1/6 of the involved drugs. Different combinations of descriptors were used and GA followed by step wise regression was applied for selection the most predictive descriptors. A total of 45 statistically significant QSPR models were derived on the six training sets. The models were validated by LOO-CV and by external validation using six test-sets (four of them containing 24 molecules, and two – 23 molecules). The statistics of the best performing models are given in Table 1.

Table 1: QSPkR models for *log* $t_{1/2}$ constructed for six training sets, validated by external test sets

Training set	Model	r^2	q^2	r^2_{pred}	MFE	Accuracy %
1	$\log t_{1/2} = 0.106(\pm 0.023)SaasC - 0.022(\pm 0.007)SssO - 0.17(\pm 0.044)SdsCH\_acnt + 1.824(\pm 0.179)xc4 - 21.58(\pm 3.49)xvch7 + 71.06(\pm 16.61)xvch10 - 0.013(\pm 0.004)SHBint2 + 0.045(\pm 0.002)SHBint4\_Acnt + 0.043(\pm 0.007)SHBint9 + 0.489$ Outliers of the training set: atovaquone, ceftriaxone, chlorpropamide, losartan, phenobarbital, tesaglitazar, valproic acid Outliers of the test set: 5-fluorouracil, diflunisal, ifetroban, tenoxicam	0.715	0.624	0.414	2.12 ±1.03	46
2	$\log t_{1/2} = -0.065(\pm 0.012)SddssS - 0.119(\pm 0.019)SdsssP - 0.152(\pm 0.048)SdsCH\_acnt - 5.29(\pm 1,28)xvch4 + 0.232(\pm 0.055)xvp7 + 64.42(\pm 17.28)xvch9 - 0.019(\pm 0.004)SHBint2 - 0.420(\pm 0.070)SHCHnX + 0.774$ Outliers of the training set: 5-fluorouracil, acetylsalicylic acid, chlorpropamide, pentobarbital, tesaglitazar Outliers of the test set: artesunate, cefotetan, enalaprilat phenobarbital, quercetin	0.664	0.603	0.671	1.78 ±0.8	58
3	$\log t_{1/2} = 0.358(\pm 0.055)SddssS\_acnt + 0.016(\pm 0.006)SaaCH - 0.208(\pm 0.058)SssssC - 0.131(\pm 0.040)SdsCH\_acnt - 0.027(\pm 0.008)SssO - 20.86(\pm 4.18)xvch7 + 62.16(\pm 19.38)xvch9 + 73.34(\pm 21.1)xvch10 + 0.254$ Outliers of the training set: 5-fluorouracil, atovaquone, chlorpropamide, enalaprilat, ifetroban, phenobarbital, Outliers of the test set: risedronate	0.594	0.539	0.533	2.35 ±2.0	67
4	$\log t_{1/2} = -0.070(\pm 0.0228)SddssS + 0.563(\pm 0.101)SdsssP\_acnt - 0.181(\pm 0.043)SssO\_acnt - 0.154(\pm 0.038)SdsCH\_acnt - 0.087(\pm 0.017)xpc4 - 23.25(\pm 3.96)xvch7 + 16.19(\pm 14.65)xvch9 + 0.064$ Outliers of the training set: 5-fluorouracil, atovaquone, chlorpropamide, enalaprilat, ifetroban, phenobarbital, roqinimex Outliers of the test set: tesaglitazar, warfarin	0.587	0.525	0.832	2.46 ±1.77	42
5	$\log t_{1/2} = -0.096(\pm 0.012)SddssS + 0.609(\pm 0.094)SdsssP\_acnt + 0.072(\pm 0.024)SaasC - 0.123(\pm 0.033)SdsCH\_acnt - 14.64(\pm 3.50)xch7 + 77.75(\pm 14.97)xvch9 + 23.98(\pm 5.39)xch10 + 0.210$ Outliers of the training set: 5-fluorouracil, atovaquone, enalaprilat, phenobarbital, pentobarbital Outliers of the test set: chlorpropamide, roquinimex	0.641	0.590	0.450	2.04 ± 1.44	65
6	$\log t_{1/2} = 0.521(\pm 0.06)SddssS\_acnt - 0.129(\pm 0.022)SdsssP + 0.102(\pm 0.024)SaasC - 0.125(\pm 0.035)SdsCH\_acnt - 14.4(\pm 3.66)xch7 + 77.22(\pm 15.1)xvch9 + 25.31(\pm 4.23)xch10 + 0.177$ Outliers of the training set: 5-fluorouracil, diflunisal, enalaprilat, phenobarbital Outliers of the test set: atovaquone, pentobarbital	0.648	0.560	0.577	2.07 ± 1.06	57

The QSPkR models derived on the six different training sets are quite similar in terms of selected variables, outliers and statistics. The explained variance of the best models r^2 varies between 0.587 and 0.715 (mean 0.642).

The internal $q^2_{LOO\text{-}CV}$ ranges from 0.525 to 0.624 (mean 0.574), and the external r^2_{pred} – between 0.414 and 0.832 (mean 0.580). The *MFEP* varies between 1.78 and 2.46, and the accuracy – between 42 and 47% (mean 56%) Several drugs were identified as outliers by almost all models – for example, *phenobarbital* (100% of the models), *chlorpropamide* (96%), *5-fluorouracil* (96%) and *atovaquone* (73%). Despite some differences, most developed QSPkR models contain common variables. The most frequently emerged descriptors are listed in Table 2.

The 28 most frequently emerged descriptors were used further for development of the final QSPkR model for the whole dataset of 142 acidic drugs. By applying the GA and step wise regression the following consensus model was derived:

$$\log t_{1/2} = -0.116(\pm 0.011)SddssS - 0.140(\pm 0.017)SdsssP + 0.119(\pm 0.022)SaasC - 0.124(\pm 0.040)SdsCH\_acnt - 11.14(\pm 3.47)xvch7 + 48.41(\pm 15.32)xvch9 + 86.39(\pm 16.76)xvch10 - 0.013(\pm 0.004)SHBint2 + 0.031(\pm 0.009)SHBint9 + 0.492$$

N = 133 r^2 0.688 $q^2_{LOO\text{-}CV}$ 0.600 *MFE* 2.06±1.19 Accuracy 61%

Nine drugs were identified as outliers (*5-fluorouracil, losartan, atovaquone, ifetroban diflunisal, chlorpropamide, pentobarbital, phenobarbital and valproic acid*). Their removal resulted in

improved explained variance from r^2 = 0.574 to r^2 = 0.688. The Consensus model predicts $t_{1/2}$ with less than 2-fold error for 61% of the drugs. The plot of calculated by the Consensus model versus experimental values of *log* $t_{1/2}$ is shown in Figure1.

Table 2: The most frequently emerged descriptors in the QSPkR models for log $t_{1/2}$

Descriptor	Encoded structural information	Frequency % of models (training sets)
SdsCH, SdsCH_acnt	Sum of the E-state values or the number of atoms of the type dsCH	76% (6)
SddssS, SddssS_acnt	Sum of the E-state values or the number of atoms of the type ddssS	47% (6)
xch7, xvch7	7-order connectivity index (simple and valence) accounting for a presence of a 7-member ring system	78% (5)
xch9, xvch9	9-order connectivity index (simple and valence) accounting for a presence of a 9-member ring system	51% (5)
SdsssP, SdsssP_acnt	The sum of the E-state or the number of atoms of the type dsssP	42% (5)
xc4	Simple 4-order cluster connectivity index	29% (5)
xch4, xvch4	4-order connectivity index (simple and valence) accounting for a presence of a 4-member ring	24% (5)
SHBint9	Internal hydrogen bond index: the largest product of E-state values for hydrogen acceptor and donor pair separated by 9 skeletal bonds	37% (4)
xch10, xvch10	10-order connectivity index (simple and valence) accounting for a presence of a 10-member ring system	33% (4)
SHBint2	Internal hydrogen bond index: the largest product of E-state values for hydrogen acceptor and donor pair separated by 2 skeletal bonds	31% (4)
SaasC, SaasC_a	The sum of the E-state values or the number of atoms of the type aasC (substituted aromatic carbon atoms)	20% (4)
SssO, SssO_acnt, SaaCH, SaaCH_acnt, xc4, xvpc4, xch8, xvch8, SHBint4_Acnt		Less presented

Table 3: Checklist of criteria for prediction of $t_{1/2}$ for acidic drugs

S. No.	Descriptor	$t_{1/2}$ decreases	$t_{1/2}$ increases
1	***SdsCH_acount*** – presence of methine groups	√	
2	***xvch7*** – presence of a 7-member ring system	√	
3	***SaasC*** – substituted aromatic carbons		
	- with prevalence of polar substituents	√	√
	- with prevalence of non-polar substituents		
4	***SddssS*** – presence of Sulphonyl groups		√
5	***SdsssP*** – presence of Phosphonate groups		√
6	***xvch9*** – presence of a 9-member ring system		√
7	***xvch10*** – presence of a 10-member ring system		√
8	***SHBint_9*** – presence of hydrogen bond donor and acceptor separated by 9 skeletal bonds		√

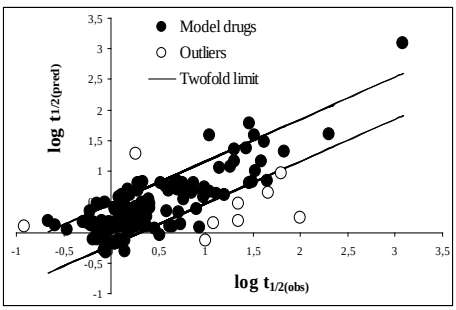

Fig.1: log $t_{1/2}$ predicted by the consensus model versus observed *log* $t_{1/2}$ values for 142 acidic drugs. Nine outliers are shown as blank points. The straight lines represent the 2-fold error limits.

Checklist for prediction of $t_{1/2}$

The descriptors involved in the consensus model describe a number of structural features governing $t_{1/2}$ of the considered drugs. They are given in Table 3 in the form of a checklist of criteria which may be used for prediction of $t_{1/2}$. *SHBint_2* is not included due to its variable effect on $t_{1/2}$.

The checklist was applied to the dataset of the studied drugs, classified into three groups according to the value of $t_{1/2}$. The first group involves 31 compounds with a short $t_{1/2}$ < 1 h. Most of the molecules have structural features affecting negatively $t_{1/2}$. 32% of them contain methine groups, another 32% – a 7-member ring system, and 16% involve aromatic carbons attached mainly to polar substituents. Descriptors with positive impact on $t_{1/2}$ are less represented: no one drug contains neither a sulfonyl nor a phosphonate groups. Only one (*fluvastatin*) has a 9-member ring system and a hydrogen bond donor and acceptor separated by 9 skeletal bonds. In contrast, 58% involve aromatic carbons attached mainly to non-polar substituents and 16% contain a 10-member ring system. Therefore, the difference between the numbers of positively and negatively criteria varies between -2 and 2 with a median value of 0. The second group comprises of 96 molecules with moderate half life (1 h < $t_{1/2}$ < 24 h). Here predominate drugs with positively contributing structural features and 30% meet 2 or 3 positive criteria. Only 35% of the molecules contain a negatively contributing descriptor. The difference between the numbers of positively and negatively criteria varies between -2 and 3 with a median value of 1. The third group consists of 15 drugs with a long half life ($t_{1/2}$ > 24 h). Here emphatically prevail molecules with positively contributing descriptors. Only 1 drug (*epristeride*) contains methine groups, two molecules (*hypericin* and *suramin*) – aromatic carbons attached mainly to polar substituents, and no one – a 7-member ring system. At the same time 47% of the drugs contain either sulfonyl or phosphonate groups, another 47% –a 10-member ring system, and 67% – aromatic carbons attached mainly to non-polar substituents. Thus, the difference between the numbers

of positively and negatively criteria varies between 0 and 3 with a median value of 2. The distribution of the drugs according to the difference between the numbers of positive and negative criteria is shown in Figure 2.

This difference can be used to distinguish between drugs with short and long $t_{1/2}$. Although there are drugs which $t_{1/2}$ does not match exactly to the proposed criteria, the trend is obvious. For 68% of the drugs with a short half-life (< 1 h) the difference is ≤0. At the same time, for 67% of the drugs with a long half life (> 24 h) the difference is ≥ 2. Therefore, a difference between the numbers of positive and negative criteria = 0 can be set as an upper limit for short half life drugs, while a difference = 2 can be set as a lower limit for long half life.

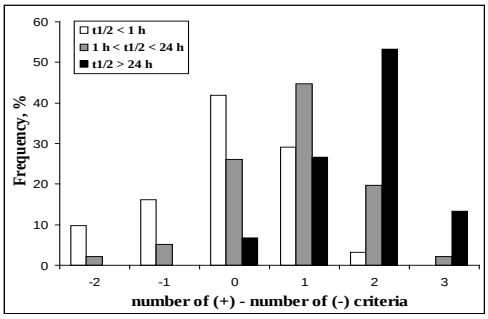

Fig. 2: Distribution of the drugs (in %) according to the difference between the number of positive and negative criteria.

DISCUSSION

Knowing $t_{1/2}$ of new drug candidates in the early stages of drug discovery is of paramount importance as $t_{1/2}$ is a key parameter in determining of dosing regimen. Unfortunately, this parameter is also the most difficult to predict because of the complexity of underlying pharmacokinetic processes. Membrane permeability, plasma and tissue protein binding, affinity to efflux and influx transporters, metabolic ability, etc. are among the numerous factors governing drug half life.

The present work is focused on development of QSPkR for $t_{1/2}$ of acidic drugs. The study was performed on a set of 142 molecules following the conventional workflow. Total of 145 descriptors were used and a three step procedure was applied to identify the most significant variables. MLR was used for model development. In addition to internal validation, a rigorous external validation procedure was performed. To this end, QSPkR models were developed for six different training sets, and were tested on six external test sets. The derived models are very similar in terms of selected variables, outliers and statistics. The most frequently emerged descriptors are used for construction of the final, consensus model. The model is statistically significant (explained variance 69%) and fairly predictive (predicted $t_{1/2}$ for 61% of the drugs with less than 2-fold error). The model is clear and interpretable revealing the most important structural features governing $t_{1/2}$ for acidic drugs.

SddssS represents the sum of the E-state values for all atoms of the type ddssS. It presents in 28 molecules and accounts for the presence of a sulfonyl groups. Its absolute value increases in the order $-SO_2R$ <– SO_2NH_2 < $-SO_2OH$. Especially large is the value for *suramin* containing 4 sulphonate groups. *SdsssP* is equal to the sum of the E-state values for all atoms of the type *dsssP*. It presents in 10 molecules and accounts for the presence of phosphonate groups. Both *SddssS* and *SdsssP* have negative values due to the great number of electronegative atoms. Having negative coefficients in the QSPkR equations they contribute to an increase of $t_{1/2}$.

So *risedronate* ($t_{1/2}$ = 200h) and *suramin* ($t_{1/2}$ = 1200 h) have the longest $t_{1/2}$ in the dataset. *SaasC* represents the sum of E-state values for all atoms of the type aasC (substituted aromatic C-atoms). It is positive for 99 and negative for 18 drugs. The presence of highly electronegative substituents like F, -OH, NO_2, phosphonate or sulphonyl groups results in a lower E-state value (frequently negative), while the prevalence of aliphatic and aromatic substituents determine high positive E-state values. *SaasC* has a positive coefficient in the QSPkR equation. A positive value of *SaasC* contributes positively to $t_{1/2}$ while a negative value of *SaasC* affects negatively $t_{1/2}$. Most of the drugs with SaasC > 1 are highly bound to plasma proteins (more than 99%). *SdsCH_acount* is equal to the number of atoms of the type dsCH (methane groups). This descriptor contributes negatively to $t_{1/2}$. 38% of the drugs containing this structural feature have a short $t_{1/2}$ < 1 h. For only (*epristeride*) $t_{1/2}$ exceeds 5 hours. *Epristeride* has extremely low clearance (0.33 mL/min) which may be due to extensive enterohepatic circulation. It is consistent with the observed second peak in the *C/t* curve following both *iv* and *ev* administration [24]. *Xvch7*, *xvch9* and *xvch 10* represent valence 7th, 9th and 10th order connectivity indices. *Xvch7* accounts for the presence of a 7-member ring system and contributes negatively to $t_{1/2}$. This descriptor presents in 20 molecules: *artesunate* and *chloazepate* containing a seven-member ring, 16 *β-lactam antibiotics* of the penicillin class and the β-lactam inhibitors *sulbactam* and *clavulanic acid* – involving fused β-lactam and five-member rings. All of them have a short $t_{1/2}$ (< 1.4h). *Artesunate* [25] and *chloazepate* [26] are prodrugs, rapidly transformed to active metabolites during absorption. The *penicillins* [27] and *sulbactam* [28] are eliminated almost completely in urine by glomerular filtration and active tubular secretion. *Clavulanic acid* is eliminated almost equally by renal excretion and hepatic metabolism [29]. The predicted values for $t_{1/2}$ are very close to the experimental data with *FE* ranging between 1.03 and 2.83 (*MF*E 1.50±0.40) except for *artesunate* with *FE* = 7. Therefore the presence of a 7-member β-lactam ring system may be considered as favorable for an active transport secretion. *Xvch9* accounts for the presence of a 9-member ring system and contributes positively to $t_{1/2}$. The descriptor presents in 9 molecules with fused five- and six-member rings. The value of *xvch9* is lower for fused aromatic rings, especially those containing N atoms, and higher for saturated systems. Most of the drugs are partially excreted in bile (*epristeride* [24], *fluvastatin* [30], *indomethacin* [31], *Pantoprazole* [32], *telmisartan* [33]. The drugs with the highest values of *xvch9*, therefore the longest $t_{1/2}$, are also extensively distributed in tissues (*telmisartan* [33], *epristeride* [34], *perindoprilat* [35]). The presence of a 9-member ring system may be related to active secretion in bile and to extensive tissue distribution. *Xvch10* accounts for the presence of 10-member ring system and contributes positively to $t_{1/2}$. The value of *xvch10* is lower for aromatic rings containing N or O atoms, connected with =O, -OH, $-NO_2$ or $-NH_2$, and higher for molecules with more than two fused rings. This descriptor presents in 19 molecules. A few drugs show deviations from the positive correlation between *xvch10* and $t_{1/2}$. *Artesunate* has much lower $t_{1/2}$, while *atovaquone* and *suramin* have much longer longer $t_{1/2}$ than expected on the basis of their values of *xvch10. Artesunate* differs from all other structures in that it contains four fused rings: three six-member and one 7-member, which determine a rather high value of *xvch10* leading to overestimation of $t_{1/2}$. Actually the predicted value of 1.54 h is very close to $t_{1/2}$ of dihydroartemisinine [25] – the active metabolite, in which the basic structural elements are preserved. Oppositely, the low value of *xvch10* is inconsistent with its long $t_{1/2}$ which as already suggested is dominated by the large number of sulfonyl groups. *Atovaquone* is identified as an outlier from the model. The drug is highly bound to plasma protein, with negligible metabolism and renal excretion. It is believed that the long $t_{1/2}$ is due to enterohepatic circulation and biliary excretion [36]. *SHBint_2* and *SHBint_9* are internal hydrogen bond indices, indicating the potential for forming an internal hydrogen bond. Their values represent the largest product of E-state value and hydrogen E-state value from all donor-acceptor pairs separated by 2 or 9 skeletal bonds respectively. *SHBint_9* presents in 3 molecules: *fluvastatin* ($t_{1/2}$ 0.7h) with *SHBint_9* = 5.7 characterizing a hydrogen bond between N and OH, and ACE inhibitors *perindoprilat* ($t_{1/2}$ 29 h) and *enalaprilat* ($t_{1/2}$ 39 h) with *SHBint_9* =31.1 corresponding to a hydrogen bond between O and OH. Therefore, a large value of *SHBint_9* contributes to a long $t_{1/2}$. It may be related to the binding to ACE as it is believed that the prolonged elimination of some ACE-inhibitors is due to the slow release of the drug from its complexes

with ACE [35,37]. *SHBint_2* presents in 130 molecules and ranges between 20 and 40. The effect of this descriptor is variable as there are drugs with low value of *SHBint_2* and long $t_{1/2}$, and vice versa. In fact, the presence of hydrogen bond donors and acceptors capable to form hydrogen bonds may have versatile effect on $t_{1/2}$. Hydrogen bonds involved in the binding with plasma and tissue proteins may cause a longer residence of the drug in the body, while those participating in binding with carrier proteins mediating the active secretion in bile and urine facilitate drug elimination. Because of the variable effect of *SHBint_2* it has not been involved in the checklist of predictors.

In summary, the presence of a sulfonyl or phosphonate groups, 9- or 10-member ring system and donor-acceptor pair separated by 9 skeletal bonds contribute to prolongation of $t_{1/2}$, while the presence of methane groups and 7-member ring system affect negatively $t_{1/2}$. The nature of the substituents at the aromatic carbon atoms may have diversely impact on $t_{1/2}$: the prevelance of polar substituents contributes to a short $t_{1/2}$, while the prevalence of non-polar substituents results in a longer $t_{1/2}$. These findings are in accordance with our previous studies. It was found, that the number of phosphonate groups and the presence of a 9-member ring system contribute positively to VD^{ss} [18]. The number of sulfonate groups affects negatively the unbound clearance CL_u [20].

The number of substituted aromatic C-atoms increases plasma protein binding [19] but decreases CL_u [20].

The outliers from the model are shown in Table 4.

The deviations of the outliers may be due either to very different structure or to unusual disposition patterns. The predicted value of $t_{1/2}$ for *5-fluorouracil* and *losartan* is overestimated with about 10-fold error. Both drugs are substrates of substantial hepatic metabolism. *5-fluorouracil* is a prodrug of the active 5-fluoro-5,6-dihydrouridine and is eliminated with a rather high and dose-dependent clearance ranging from 10 to 26 mL/min/kg [38]. *Losartan* is cleared with a very high hepatic extraction ratio as evidenced by the significant first pass effect following *po* administration [39] and the high *CL*, unrestricted by the extensive plasma protein binding. The other 7 outliers are underestimated

Table 4: Outliers from Consensus model together with the major pharmacokinetic parameters (according to Obach, Lombardo and Waters [21])

Ouitliers	$t_{1/2}$, h		Vd	Cl	f_u	criteria
	exp	pred	L/kg	ml/min/kg		
5-fluorouracil	0.12	1.26	0.23	26	0.64	1 negative (=CH-)
Losartan	1.82	19.05	0.37	8.20	0.01	1 positive (SaasC)
Diflunisal	10	0.74	0.097	0.1	0.0016	1 negative (SaasC)
Valproic acid	12	1.40	0.14	0.16	0.08	-
Ifetroban	21.88	3.00	4.40	6.40	-	1 positive (SaasC)
Pentobarbital	21.88	1.52	0.91	0.47	0.39	-
Chlorpropamide	45.7	4.5	0.19	0.045	0.03	2 positive (SaasC, >SO_2
Atovaquone	63.1	9.1	0.6	0.15	0.001	2 positive (xvch10, SaasC)
Phenobarbital	100	1.74	0.54	0.063	0.49	1 positive (SaasC)

Atovaquone and *ifetroban* are cleared exclusively by biliary excretion. The long $t_{1/2}$ of *atovaquone* is attributed to its extensive plasma protein binding, high metabolic stability and enterohepatic cycling [36]. The latter seems to be the main reason for the prolonged half life of *ifetroban* [40]. The long $t_{1/2}$ is consistent with the large V^{ss}, but contradicts to the high *CL*, questionable for a drug eliminated primarily via bile. If the reported values of V^{ss} and *CL* are true, $t_{1/2}$ should be 8 h – closer to the predicted value. *Diflunisal* is eliminated in urine as two glucuronide metabolites, without Phase 1 biotransformation [41]. The extensive plasma protein binding together with the low *CL* may be considered as a main reason for the long $t_{1/2}$. Biliary elimination and enterohepatic circulation are also reported [42]. *Chlorpropamide* [43] and *pentobarbital* [44] are eliminated mainly by hepatic metabolism with a very low extraction ratio as suggested by their low CL. The clearance of chlorpropamide is additionally restricted by the high plasma protein binding. *Phenobarbital* is also metabolized in liver, however about 30% of the drug is cleared unchanged via kidney [45]. The prolonged $t_{1/2}$ may be a result of the low hepatic extraction ratio [46] and significant tubular reabsorption as *phenobarbital* is a liposoluble, weak acid, predominantly non-ionized at physiological conditions. *Valproic acid* drug may be considered as a structural outlier because it differs significantly from most of the compounds in the dataset: it is a small, simple molecule, with molecular weight of about 144 g/mol, containing only a carboxyl group and two propyl residuals.

CONCLUSION

Statistically significant, predictive and interpretable QSPkR model was constructed for $t_{1/2}$ of acidic drugs. The predictive ability was confirmed by internal and external validation procedures. The predicted $t_{1/2}$ values for 61% of the drugs in the dataset are within the 2-fold error. Descriptors involved in the model have clear physical sense and reveal structural features governing $t_{1/2}$ of acidic drugs. The presence of a sulfonyl or phosphonate groups, non-polar substituents at aromatic carbon, 9- or 10-member ring system and donor-acceptor pair separated by 9 skeletal bonds contribute to prolongation of $t_{1/2}$, while the presence of methane groups, polar substituents in aromatic rings and 7-member ring system affect negatively $t_{1/2}$. A short check list was proposed determining the cutoff between short half life ($t_{1/2}$ < 1 h) and long half life ($t_{1/2}$ > 24 h) drugs.

CONFLICT OF INTERESTS

Declared None

ACKNOWLEDGEMENT

The research was supported by the National Science Fund of Ministry of Education and Science, Bulgaria (Grant 02-1/2009).

REFERENCES

1. Van der Waterbeemd H, Gifford E. ADMET in silico modeling: towards prediction paradise? Nat Rev/Drug Discovery 2000;2:192-204.
2. Kola L, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev/Drug Discovery 2004;3:711-5.
3. Ekins S, Waller CI, Swaan PW, Cruciani G, Wrighon SA, Wikel JH. Progress in prediction human ADME parameters in silico. J Pharmacol Toxicol Methods 2000;44:251-72.
4. Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O. In silico prediction of ADME and pharmacokinetics. Report of an expert meeting organized by COST B15. Eur J Pharm Sci 2002;17:183-93.
5. Butina D, Segall MD, Frankcombe K. Predicting ADME properties in silico: methods and models. 2002;DDT7:S83-S88.
6. Yamashita F, Hashida M. In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet 2004;19:327-38.
7. Mager DE. Quantitative structure pharmacokinetic/ pharmacodynamic relationships. Adv Drug Deliv Res 2006;58:1326-56.
8. Chohan KK, Paine SW, Waters NJ. Advancements in predictive in silico models for ADME. Curr Chem Biol 2008;2:215-28.
9. Wang J, Hou T. Recent advances on in silico modeling. In: Wheeler RA, editor. Annual reports in Computational chemistry, Vol. 5. Amsterdam, San Diego: Elsevier; 2009. p. 102-27.

10. Toutain PL, Bousquet-Melou A. Plasma terminal half-life. J Vet Pharmacol Therap 2004;27:427-39.
11. Obach RS, Baxter JG; Liston TE, Silber BM, Jones BC, McIntyre F, *et al*. The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism data. J Pharmacol Exp Therap 1997;283(1):6-58.
12. Madden JC. In silico approaches for predicting ADME properties. In: Puzyn T, Leszczynski J, Cronin MTD, editors. Recent advances in QSAR studies. Dordrecht, Heidelberg, London, New York: Springer Science + Business Media BV; 2010. p. 283-304.
13. Di L, Feng B, Goosen TC, Lay Y, Steyn SJ, Varma MV, *et al*. A perspective on the prediction of drug pharmacokinetics and disposition in drug research and development. Drug Metab Dispos 2013;41(12):1975-93.
14. Berezhkovskiy LM. Prediction of drug terminal half-life and terminal volume of distribution after intravenous dosing based on drug clearance, steady state volume of distribution, and physiological parameters of the body. J Pharm Sci 2013;102 (2):761-71.
15. Hockings N, Ajayi AA, Reid JL. Age and the pharmacokinetics of the angiotensin converting enzyme inhibitors enalapril and enalaprilat. Br J Clin Pharmacol 1986;21:341-8.
16. Paul Y, Parle M, Dhake AS, Singh B. In silico quantitative strructure – pharmacokinetic relationships for elimination half life of fluoroquinolones. Asian J Chem 2009;21 (7):5483-7.
17. Paul Y, Aman Singla P, Singh B. In silico quantitative structure – pharmacokinetic relationship modeling on antidiabetic drugs: half life. Int J Chem Sci 2013;11(1):177-85.
18. Zhivkova Z, Doytchinova I. Prediction of steady state volume of distribution of acidic drugs by quantitative structure-pharmacokinetics relationships. J Pharm Sci 2012;101 (3):1253-66.
19. Zhivkova Z, Doytchinova I. Quantitative structure – plasma protein binding relationships of acidic drugs. J Pharm Sci 2012;101(12):4627-41.
20. Zhivkova Z, Doytchinova I. Quantitative structure – clearance relationships of acidic drugs. Mol Pharmac 2013;10:3758-68.
21. Obach RS, Lombardo F, Waters NJ. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab Dispos 2008;36 (7):1385-405.
22. http: /www. drigbank. ca.
23. http: /www. chemicalbook. com.
24. Davies NM, Takimoto JK, Brocks DR, Yanez JA. Multiple peaking phenomena in pharmacokinetic disposition. Clin Pharmacokinet 2010;49 (6):351-77.
25. Newton PN, Barnes KI, Smith PJ, Evans AC, Chierakul W, Ruangveerayuth R, *et al*. The pharmacokinetics of intravenous artesunate in adults with severe falciparum malaria. Eur J Clin Pharmacol 2006;62:1003-9.
26. Bertler A, Lindgren S, Magnusson J-O, Malmgren H. Pharmacokinetics of chlorazepate after intravenous and intramuscular administration Psychopharmacol 1983;80 (3):236-9.
27. Barza M, Weinstein L. Pharmacokinetics of the penicillins in man. Clin Pharmacokinet 1976;1:297-308.
28. Foulds G, Stankewich JP, Marshall DC, O'Brien MM, Hayes SL, Weidler DJ, *et al*. Pharmacokinetics of sulbactam in humans. Antimicrob Agents Chemother 1983;23 (5):692-9.
29. Jungbluth GL, Cooper DL, Doyle GD, Chrudzik GM, Jusko WJ. Pharmacokinetics of ticarcillin and clavulanic acid (timentin) in relation to renal function. Antimicrob Agents Chemother 1986;30 (6):896-900.
30. Watanabe T, Kusihara H, Maeda K, Kanamuru H, Saito Y, Hu Z, Sugyiama Y. Investigation of the rate-limiting process in the hepatic elimination of HMG-CoA reductase inhibitors in rats and humans. Drug Metab Dispos 2010;38 (2):215-22.
31. Terhaag B, Hermann U. Biliary elimination of indomethacin in man. Eur J Clin Pharmacol 1986;29 (6):691-5.
32. Huber R, Hartmann M, Bliesash H, Luechmann R, Steinijans VW, Zech K. Pharmacokinetics of pantoprazole in man. Int J Clin Pharmacol Ther 1996;34 (1 Supl.):7-16.
33. Wienen W, Entzeroth M, Van Meel JCA, Stangier J, Buscj U, Ebner T, *et al*. A review on telmisartan: a novel, long acting angiotensin II-receptor antagonist. Cardiovasc Ther 2000;18 (2):127-54.
34. Benincosa LJ, Aidet PR, Lundberg D, Zariffa N, Jorkasky DK. Pharmacokinetics and absolute bioavailability of epristeride in healthy male subjects. Biopharm Drug Dispos 1996;17 (3):249-58.
35. Devissaguet JP, Ammoury N, Devissaguet M, Perret L. Pharmacokinetics of perindopril and its metabolites in healthy volunteers. Fundam Clin Pharmacol 1990;4 (2):175-89.
36. Rolan PE, Mercer AJ, Tate E, Benjamin I, Posner J. Disposition of atovaquone in humans. Antimicrob Agents Chemother 1997;41(6):1319-21.
37. McFadyen RJ, Meredith PA, Elliott HL. Enalapril clinical pharmacokinetics and pharmacodynamic relationships. An overview. Clin Pharmacokinet 1993;25 (4):274-82.
38. Port RE, Daniel B, Ding RW, Herrmann R. Relative importance of dose, body surface area, sex, and age for 5-flurouracil clearance. Oncology 1991;48:277-81.
39. Lo MW, Goldberg MR, McCrea JB, Lu H, Furtek CI, Bjornsson TD. Pharmacokinetics of losartan, an angiotensin II receptor antagonist, and its active metabolite EXP3174 in humans. Clin Pharmacol Therap 1995;58:641-9.
40. Dockens RC, Santone KS, Mitroka JG, Morrison RA, Jemal M, Greene DS, *et al*. Disposition of radiolabeled ifetroban in rats, dogs, monkeys, and humans. Drug Metab Dispos 2000;28:973-80.
41. Tempero KF, Cirillo VJ, Steelman SL. Diflunisal: a review of pharmacokinetics and pharmacodynamic properties, drug interactions, and special tolerability studies in humans. Br J Clin Pharmacol 1977;4:31S-6S.
42. Nuernberg B, Koehler G, Brune K. Pharmacokinetics of diflunisal in patients. Clin Pharmacokinet 1991;20 (1):81-9.
43. Shon JH, Yoon YR, Kim MJ, Kim KA, Lim YC, Liu KH, *et al*. Chlorpropamide 2-hydroxylation is catalysed by CYP2C9 and CYP2C19 *in vitro*: chlorpropamide disposition is influenced by CYP2C9, but not by CYP2C19 genetic polymorphism. Br J Clin Pharmacol 2005;59 (5):552–63.
44. Tang BK, Inaba T, Kalow W. N-hydroxylation of pentobarbital in man. Drug Metab Dispos 1977;5:71-4.
45. Nelson E, Powell JR, Conrad K, Likes K, Byers J, Baker S, *et al*. Phenobarbital pharmacokinetics an bioavailability in adults. J Clin Pharmacol 1982;22(2-3):141-8.
46. Jenkings AJ. Pharmacokinetics of specific drugs. In: Karch SB, editor. Pharmacokinetics and pharmacodynamics of abused drugs. CRC Press: Taylor and Francis Group; 2008. p. 26-64.