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

ZVETANKA ZHIVKOVA, IRINI DOYTCHINNOVA

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

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 $t_{1/2}$. A short check list was proposed determining the cutoff between short half life ($t_{1/2} < 1 \text{ h}$) and long half life ($t_{1/2} > 24 \text{ h}$) drugs.

Conclusion: The presence of a sulfanyl 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 $s_{max}$ 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:

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

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 IntC/t curve, corresponding to the rate constant of elimination $\lambda$.

$$t_{1/2} = \frac{\ln 2}{\lambda}$$

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 $V^0$ assuming the following simple relation:

$$t_{1/2} = \frac{\ln 2 * V^0}{CL}$$

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 $V^0$, 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 $V^0$ and $CL$[13]. Another shortcoming is the use of $V^0$. 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 \( V_t \), which may be much greater than \( V_d \) [14]. Both \( V_d \) and the slope of the terminal \( InC/t \) phase depend on the rate of drug transfer between plasma and tissues. Therefore for many drugs with multisphere 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 \( V_d = 0.38 \) 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 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 congeners of drugs – fluoroquinolones [16] and antiangiogenic 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 pharmokinetics 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 consists of 142 acidic drugs (10.5%). The \( V_d \) 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 dataset was used for construction of six modeling sets – each belonging to different chemical and therapeutic classes. The dataset used in the present study consisted 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:

\[
f_A = \frac{1}{1 + 10^{pK_a - \text{phys}}}, \quad f_B = \frac{1}{1 + 10^{\text{phys} - pK_a}}.
\]

The mol files of the drugs are taken from Databank [22] or chemical Book [23]. 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 \( f_B \) exceeds \( f_A \) 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 Dc, 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 every 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 QSPkR 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 (\( r_{\text{pred}, \text{ext}} \)) accuracy and mean fold error of prediction (MFE) defined in the next section. Descriptors, which emerged in more than 20% of the models, were selected for development of a consensus QSPK model.

Model assessment and validation

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

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

where \( t_{1/2, \text{i,obs}} \) is the observed \( t_{1/2} \) of the \( i \)th drug, \( t_{1/2, \text{i,calc}} \) is the calculated by the model \( t_{1/2} \) of the \( i \)th drug, and \( t_{1/2, \text{i,mean}} \) 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 QSPK 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 \( q^2_{\text{LOO-CV}} \) and \( r_{\text{pred, ext}} \) following the equations:

\[
q^2_{\text{LOO-CV}} = 1 - \sum_{i=1}^{n \text{min}} \left( \log t_{1/2, \text{i,obs}} - \log t_{1/2, \text{i,calc}} \right)^2 / \sum_{i=1}^{n \text{min}} \left( \log t_{1/2, \text{i,obs}} - \log t_{1/2, \text{i,mean}} \right)^2
\]

\[
r_{\text{pred, ext}} = 1 - \sum_{i=1}^{n \text{mean}} \left( \log t_{1/2, \text{i,obs}} - \log t_{1/2, \text{i,calc}} \right)^2 / \sum_{i=1}^{n \text{mean}} \left( \log t_{1/2, \text{i,obs}} - \log t_{1/2, \text{i,mean}} \right)^2
\]

where \( t_{1/2, \text{i,obs}} \) is the predicted by the model value of \( t_{1/2} \) of the \( i \)th drug, and \( t_{1/2, \text{i,calc}}, t_{1/2, \text{i,mean}} \) and \( t_{1/2, \text{i,obs}} \) are the observed, mean and predicted by the model values of \( t_{1/2} \) for any drug from the external test set, respectively.

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

\[
FE = 10^{\left( \sum_{i=1}^{n \text{mean}} \log t_{1/2, \text{i,obs}} - \log t_{1/2, \text{i,calc}} \right) / n \text{mean}}
\]

The average value of FE(\% 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 \( Dc \) 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(\%) is relatively low – it varies between 0.04 and 15 L/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 \( f_p \).
in the range 0.0004 – 1 (mean 0.3, median 0.15). The values for t1/2 vary between 0.12 and 1200 h (mean 17.2, median 1.8 h). 32 drugs (22.5%) have t1/2 < 1 h, while for 15 drugs (10.6%) t1/2 > 24.

**QSPkR models for t1/2**

In order to derive a robust and predictive QSPkR model for log t1/2, 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 stepwise regression was applied for selection in 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 t1/2, constructed for six training sets, validated by external test sets**

<table>
<thead>
<tr>
<th>Training set</th>
<th>Model</th>
<th>r²</th>
<th>q²</th>
<th>r²pred</th>
<th>MFE</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[ \log t_{1/2} = -0.106(\pm 0.023) SsaCS - 0.022(\pm 0.007) SsoS - 0.17(\pm 0.044) SdsCH _acnt + 1.824(\pm 0.179) xc4 - 21.58(\pm 3.49) xvc7 + 71.06(\pm 16.6) xvc10 - 0.013(\pm 0.004) SHB int 2 + 0.045(\pm 0.002) SHB int 4 , Acnt + 0.043(\pm 0.007) SHdB int 9 + 0.489 ]</td>
<td>0.715</td>
<td>0.624</td>
<td>0.414</td>
<td>2.12</td>
<td>46 ±1.03</td>
</tr>
<tr>
<td>2</td>
<td>[ \log t_{1/2} = -0.065(\pm 0.012) SdsS - 0.119(\pm 0.019) SdsS &lt;p&gt; - 0.152(\pm 0.048) SdsCH _acnt - 5.29(\pm 1.28) xvc4 + 0.232(\pm 0.055) xvp7 + 64.42(\pm 7.28) xvc9 - 0.019(\pm 0.004) SHBD int 2 - 0.420(\pm 0.070) SHCH6nX + 0.774 ]</td>
<td>0.664</td>
<td>0.603</td>
<td>0.671</td>
<td>1.78</td>
<td>58 ±0.8</td>
</tr>
<tr>
<td>3</td>
<td>[ \log t_{1/2} = 0.358(\pm 0.055) SsdSS _acnt + 0.016(\pm 0.006) SaaAC - 0.208(\pm 0.058) SssS &lt;p&gt; - 0.131(\pm 0.040) SdsCH _acnt - 0.027(\pm 0.008) SsoS - 20.86(\pm 4.18) xvc7 + 62.16(\pm 19.38) xvc9 + 73.34(\pm 21.1) xvc10 + 0.254 ]</td>
<td>0.594</td>
<td>0.539</td>
<td>0.533</td>
<td>2.35</td>
<td>67 ±2.0</td>
</tr>
<tr>
<td>4</td>
<td>[ \log t_{1/2} = -0.070(\pm 0.0228) SddSS + 0.563(\pm 0.101) SdssP _acnt - 0.181(\pm 0.043) SssS &lt;p&gt; - 0.154(\pm 0.038) SdsCH _acnt - 0.087(\pm 0.017) xpc4 - 23.25(\pm 3.96) xvc7 + 16.19(\pm 14.65) xvc9 + 0.064 ]</td>
<td>0.587</td>
<td>0.525</td>
<td>0.832</td>
<td>2.46</td>
<td>42 ±1.77</td>
</tr>
<tr>
<td>5</td>
<td>[ \log t_{1/2} = -0.096(\pm 0.012) SddSS + 0.609(\pm 0.094) SdssS _acnt + 0.072(\pm 0.024) SaaAC - 0.123(\pm 0.033) SdsCH _acnt - 14.64(\pm 3.50) xch7 + 77.75(\pm 14.97) xvc9 + 23.98(\pm 25.39) xch10 + 0.210 ]</td>
<td>0.641</td>
<td>0.590</td>
<td>0.450</td>
<td>2.04</td>
<td>65 ±1.44</td>
</tr>
<tr>
<td>6</td>
<td>[ \log t_{1/2} = 0.521(\pm 0.06) SdSS _acnt - 0.129(\pm 0.022) SdsS + 0.102(\pm 0.024) SaaAC - 0.125(\pm 0.035) SdsCH _acnt - 14.4(\pm 3.66) xch7 + 77.22(\pm 15.1) xvc9 + 25.31(\pm 4.23) xch10 + 0.177 ]</td>
<td>0.648</td>
<td>0.560</td>
<td>0.577</td>
<td>2.07</td>
<td>57 ±1.06</td>
</tr>
</tbody>
</table>

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 r² varies between 0.587 and 0.715 (mean 0.642).

The internal q²<sub>LOO</sub>-CV ranges from 0.525 to 0.624 (mean 0.574), and the external r²<sub>pred</sub> between 0.414 and 0.832 (mean 0.580). The MFE<sub>P</sub> 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), florouracil (96%), atovaquone (73%). Despite some differences, the 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 acetic drugs. By applying the GA and stepwise regression the following consensus model was derived:

\[ \log t_{1/2} = -0.116(\pm 0.011) SddSS - 0.140(\pm 0.017) SdsS - 0.119(\pm 0.022) SaaAC - 0.124(\pm 0.040) SdsCH _acnt - 11.14(\pm 5.47) xvc7 + 48.41(\pm 15.32) xvc9 + 8.39(\pm 6.76) xch10 - 0.013(\pm 0.004) SHBD int 2 + 0.03(\pm 0.009) SHBD int 9 + 0.492 \]

N = 133 \( \pm 0.688 \) q²<sub>LOO</sub>-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...
The checklist was applied to the dataset of the studied drugs, classified into three groups according to the value of \(t_{1/2}\) for prediction of experimental values of \(\log t_{1/2}\). The plot of calculated by the Consensus model versus experimental values of \(\log t_{1/2}\) is shown in Figure 1.

Table 2: The most frequently emerged descriptors in the QSPK models for \(\log t_{1/2}\)

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Encoded structural information</th>
<th>Frequency % of models (training sets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SdsCH, SdsCH_acnt</td>
<td>Sum of the E-state values or the number of atoms of the type dsCH</td>
<td>76% (6)</td>
</tr>
<tr>
<td>SdsssS, SdsssS_acnt</td>
<td>Sum of the E-state values or the number of atoms of the type dsssS</td>
<td>37% (5)</td>
</tr>
<tr>
<td>xch7, xvch7</td>
<td>7-order connectivity index (simple and valence) accounting for a presence of a 7-member ring system</td>
<td>78% (5)</td>
</tr>
<tr>
<td>xch9, xvch9</td>
<td>9-order connectivity index (simple and valence) accounting for a presence of a 9-member ring system</td>
<td>51% (5)</td>
</tr>
<tr>
<td>SdssP, SdssP_acnt</td>
<td>The sum of the E-state or the number of atoms of the type dsSP</td>
<td>42% (5)</td>
</tr>
<tr>
<td>xch4</td>
<td>Simple 4-order cluster connectivity index</td>
<td>29% (5)</td>
</tr>
<tr>
<td>xch4, xvch4</td>
<td>4-order connectivity index (simple and valence) accounting for a presence of a 4-member ring pair separated by 9 skeletal bonds</td>
<td>24% (5)</td>
</tr>
<tr>
<td>SHBint9</td>
<td>Internal hydrogen bond index: the largest product of E-state values for hydrogen acceptor and donor pair separated by 9 skeletal bonds</td>
<td>37% (4)</td>
</tr>
<tr>
<td>xch10, xvch10</td>
<td>10-order connectivity index (simple and valence) accounting for a presence of a 10-member ring system</td>
<td>33% (4)</td>
</tr>
<tr>
<td>SHBint2</td>
<td>Internal hydrogen bond index: the largest product of E-state values for hydrogen acceptor and donor pair separated by 2 skeletal bonds</td>
<td>31% (4)</td>
</tr>
<tr>
<td>SaasC, SaasC_a</td>
<td>The sum of the E-state values or the number of atoms of the type aasC (substituted aromatic carbon atoms)</td>
<td>20% (4)</td>
</tr>
<tr>
<td>SssO, SssO_acnt, SaaCH, SaaCH_acnt, xc4, xvpc4, xch8, xvch8, SHBint4_Acct</td>
<td>Less presented</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Descriptor</th>
<th>(t_{1/2}) decreases</th>
<th>(t_{1/2}) increases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(SdsCH), (SdsCH_acnt) – presence of methine groups</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>2</td>
<td>(xvch7) – presence of a 7-member ring system</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>3</td>
<td>(SaasC) – substituted aromatic carbons</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td></td>
<td>- with prevalence of polar substituents</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td></td>
<td>- with prevalence of non-polar substituents</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>4</td>
<td>(SdsssS) – presence of Sulphonyl groups</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>5</td>
<td>(SdssSP) – presence of Phosphonate groups</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>6</td>
<td>(xvch9) – presence of a 9-member ring system</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>7</td>
<td>(xvch10) – presence of a 10-member ring system</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>8</td>
<td>(SHBint_9) – presence of hydrogen bond donor and acceptor separated by 9 skeletal bonds</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
</tbody>
</table>

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 predominates 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 prevalent 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 t1/2. Although there are drugs which t1/2 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 ≥ 2. At the same time, for 62% 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.

The present work is focused on development of QSPkR for drug half life. Metabolic ability, etc. are among the numerous factors governing t1/2. Among the drugs, while a difference = 2 can be set as a lower limit for long half life, the difference can be used to distinguish between drugs with short and long t1/2. Although there are drugs which t1/2 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 ≥ 2. At the same time, for 62% 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.

DISCUSSION

Knowing t1/2 of new drug candidates in the early stages of drug discovery is of paramount importance as t1/2 is a key parameter in determining of dosing regimen. Unfortunately, this parameter is also one of 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 t1/2 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 t1/2 for 61% of the drugs with less than 2-σ deviation error). The model is clear and interpretable revealing the most important structural features governing t1/2 for acidic drugs.

SaasC represents the sum of the E-state values for all atoms of the type dsCH. It presents in 28 molecules and accounts for the presence of a sulfonyl groups. Its absolute value increases in the order -SO2R < -SO2NH2 < -SO2OH. Especially large is the value for suramin containing 4 sulphonate groups. SaassS is equal to the sum of the E-state values for all atoms of the type dsS. It presents in 10 molecules and accounts for the presence of phosphate groups. Both SaassS and SaassP have negative values due to the great number of electronegative atoms. Having negative coefficients in the QSPkR equations they contribute to an increase of t1/2. So risidonate (t1/2 = 200h) and suramin (t1/2 = 1200 h) have the longest t1/2 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 10 drugs. The presence of highly electronegative substituents like F, -OH, NO2, phosphonate or sulphonate 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 t1/2 while a negative value of SaasC affects negatively t1/2. Most of the drugs with SaasC > 1 are highly bound to plasma proteins (more than 99%). SaassCH_acount is equal to the number of atoms of the type dsCH (methane groups). This descriptor contributes negatively to t1/2. 38% of the drugs containing this structural feature have a short half-life. A descriptor epristeride t1/2 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]. Xch7, xch9 and xch 10 represent valence 7, 9, and 10th order connectivity indices. Xch7 accounts for the presence of a 7-member ring system and contributes negatively to t1/2. This descriptor presents in 20 molecules: artesunate and chloroquine 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 t1/2 (< 1.4h). Artesunate [25] and chloroquine [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 t1/2 are very close to the experimental data with FE ranging between 1.03 and 2.48 (Wong, 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. Xch9 accounts for the presence of a 9-member ring system and contributes positively to t1/2. The descriptor presents in 9 molecules with fused five- and six-member rings. The value of xch9 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 xch9, therefore the longest t1/2, 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. Xch10 accounts for the presence of 10-member ring system and contributes positively to t1/2. The value of xch10 is lower for aromatic rings containing N or O atoms, connected with -O- and -OH, and for molecules with more than two fused rings. This descriptor represents in 17% of the drugs showing deviations from the positive correlation between xch10 and t1/2. Artesunate has much lower t1/2, while atovapone and suramin have much longer t1/2 than expected on the basis of their values of xch10. Artesunate differs from all other structures in that it contains four fused rings. FE 3 six-member and one 7-member, which determine a rather high value of xch10 leading to overestimation of t1/2. Actually the predicted value of 1.54 h is very close to 1.52, of dihydroartemisinine [25] – the active metabolite, in which the basic structural elements are preserved. Oppositely, the low value of xch10 is consistent with its long t1/2 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 t1/2 is due to enterohepatic circulation and biliary excretion [36]. ShiBint2 and ShiBint9 are internal hydrocarbon bond indices, indicating the potential for forming an internal hydrogen bond. The value of xch10 represents the largest product of E-state value and hydrogen E-state value from all donor-acceptor pairs separated by 2 or 9 skeletal groups respectively. ShiBint9 presents in 3 molecules: fluvastatin (t1/2 = 0.7h) with ShiBint9 = 5.7 characterizing a hydrogen bond between N and OH, and ACE inhibitors perindoprilat (t1/2 = 29 h) and enalaprilat (t1/2 = 39 h) with ShiBint9 = 31.1 corresponding to a hydrogen bond between O and OH. Therefore, a large value of ShiBint9 contributes to a long t1/2. 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.

![Fig. 2: Distribution of the drugs (in %) according to the difference between the number of positive and negative criteria.](image-url)
Zhivkova et al.  
Int J Pharm Pharm Sci, Vol 6, Issue 9, 283-289

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

The number of substituted aromatic C-atoms increases plasma protein binding [19] but decreases CL\textsubscript{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\textsubscript{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 produg 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\textsubscript{u} 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]) |
|------------------|------------------|------------------|------------------|------------------|
| Outliers         | t1/2, h          | Vd               | CL               | f\textsubscript{0}  |
|                  | exp              | [L/kg]           | [ml/min/kg]      | criteria         |
| 5-fluorouracil   | 0.12             | 1.26             | 0.23             | 26               | 0.64             | 1 negative (=CH-) |
| Losartan         | 1.62             | 19.05            | 0.37             | 8.20             | 0.01             | 1 positive (SaasC) |
| Diflusin         | 10               | 0.74             | 0.016            | 0.12             | 0.0016           | 1 negative (SaasC) |
| Valproic acid    | 12               | 1.40             | 0.14             | 0.16             | 0.08             | -                |
| Iandrobrof       | 21.88            | 3.00             | 4.40             | 6.40             | -                | 1 positive (SaasC) |
| Pentobarbital    | 21.88            | 1.52             | 0.91             | 0.47             | -                | 1 positive (SaasC) |
| Chloropramide    | 45.7             | 4.5              | 0.19             | 0.045            | 0.03             | 2 positive (SaasC >SO\textsubscript{2}) |
| Atovaquone       | 63.1             | 9.1              | 0.6              | 0.15             | 0.001            | 2 positive (xvch18, SaasC) |
| Phenobarbital    | 100              | 1.74             | 0.54             | 0.063            | 0.49             | 1 positive (SaasC) |

Atovaquone and ifetroben are cleared exclusively by biliary excretion. The long t\textsubscript{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 ifetroben [40]. The long t\textsubscript{1/2} is consistent with the large V\textsubscript{d}, but contradicts to the high CL\textsubscript{u}, questionable for a drug may be considered as a main reason for the long t\textsubscript{1/2}. Biliary elimination and enterohepatic circulation are also reported [42]. Chloropramide [43] and pentobarbital [44] are eliminated mainly by hepatic metabolism with a very low extraction ratio as suggested by their low CL\textsubscript{u}. The clearance of chloropramide 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\textsubscript{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 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 QSPr model was constructed for t\textsubscript{1/2} of acidic drugs. The predictive ability was confirmed by internal and external validation procedures. The predicted t\textsubscript{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\textsubscript{1/2} of acidic drugs. The presence of a sulfonyl or phosphate 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\textsubscript{1/2} while the presence of methane groups, polar

REFERENCES


