Quantitative Structure – Pharmacokinetics Relationships for Plasma Protein Binding of Basic Drugs

Zvetanka Zhivkova

Faculty of Pharmacy, Medical University of Sofia, Sofia 1000, Bulgaria

ABSTRACT: Purpose. Binding of drugs to plasma proteins is a common physiological occurrence which may have a profound effect on both pharmacokinetics and pharmacodynamics. The early prediction of plasma protein binding (PPB) of new drug candidates is an important step in drug development process. The present study is focused on the development of quantitative structure – pharmacokinetics relationship (QSPkR) for the negative logarithm of the free fraction of the drug in plasma (pf_u) of basic drugs. **Methods.** A dataset includes 220 basic drugs, which chemical structures are encoded by 176 descriptors. Genetic algorithm, stepwise regression and multiple linear regression are used for variable selection and model development. Predictive ability of the model is assessed by internal and external validation. **Results**. A simple, significant, interpretable and predictive QSPkR model is constructed for pf_u of basic drugs. It is able to predict 59% of the drugs from an external validation set within the 2-fold error of the experimental values with squared correlation coefficient of prediction 0.532, geometric mean fold error (GMFE) 1.94 and mean absolute error (MAE) 0.17. **Conclusions.** PPB of basic drugs is favored by the lipophilicity, the presence of aromatic C-atoms (either non-substituted, or involved in bridged aromatic systems) and molecular volume. The fraction ionized as a base f_B and the presence of quaternary C-atoms contribute negatively to PPB. A short checklist of criteria for high PPB is defined, and an empirical rule for distinguishing between low, high and very high plasma protein binders is proposed based. This rule allows correct classification of 69% of the very high binders, 71% of the high binders and 91% of the low binders in plasma.

This article is open to **POST-PUBLICATION REVIEW**. Registered readers (see "For Readers") may **comment** by clicking on ABSTRACT on the issue's contents page.

_____

INTRODUCTION

Most of the drugs bind reversibly to various plasma proteins: human serum albumin (HSA), alpha-1-acid glycoprotein (AGP), lipoproteins, etc. Plasma protein binding (PPB) is a major determinant of both pharmacodynamics (PD) and pharmacokinetics (PK). Only the free fraction of the drug is able to pass across biological membranes and to reach the target, therefore only the free fraction is pharmacologically active (1). Highly bound drugs may require higher doses to achieve effective concentration *in vivo*. PPB may affect considerably key PK parameters such as the apparent volume of distribution (V_d), clearance (CL) and half-life ($t_{1/2}$). V_d is related to the unbound fraction of the drug in plasma f_u and in tissues $f_{u,t}$ as follows:

$$V_d = V + V_t \frac{f_u}{f_{u,t}} \qquad (1)$$

where V and V_t are the volumes of plasma and tissue fluids, respectively (2). The effect of PPB on V_d depends on the relative affinity of the drug for plasma proteins and tissue components. For drugs with high PPB V_d is close to V and is independent of PPB. For drugs with high affinity to tissues V_d increases almost linearly with the increase in f_u. PPB may have either restrictive or permissive effect on drug CL depending on drug's extraction ratio E, which accounts for the fraction of the drug, cleared by given organ. According to the "well-stirred model", the hepatic clearance CL_H is given by the equation:

$$CL_H = \frac{Q_H f_u CL_{int}}{Q_H + f_u CL_{int}} \qquad (2)$$

where Q_H is the liver blood flow, and CL_{int} – the intrinsic clearance of the free drug (2). Drugs with high E (> 0.7) are eliminated with high CL_H, close to

_____

Corresponding Author: Dr. Zvetanka Zhivkova; Associate Professor Faculty of Pharmacy, Medical University of Sofia; 2 Dunav st., 1000 Sofia, Bulgaria; Email: zzhivkova@pharnfac.mu-sofia.bg

Q_H and independent of PPB. In contrast, drugs with low E (< 0.3) have low CL_H, proportional to f_u and restricted by PPB. The same applies to the renal excretion. Drugs that are excreted solely by glomerular filtration have low CL, restricted by PPB, while drugs, substrates of active secretion transporters, are eliminated with high CL, independent of PPB. Consequently, PPB may affect considerably $t_{1/2}$, as it is related with V_d and CL according to the equation:

$$t_{1/2} = \frac{0.693V_d}{CL} \qquad (3)$$

The complex effects of PPB on PK and PD are discussed in several reviews and books (3-9).

The recent advances of combinatorial chemistry and high throughput technologies have led to growing number of structures with drug-like activities. However, the majority of drug candidates fail to become marketable products due to poor PK properties (10). This inspired an intense research focused on the prediction of the PKs of drug candidates at very early stages of discovery process. In the last two decades *in silico* modeling became a powerful strategy for the modeling and prediction of human PK properties. It enables construction of quantitative structure – pharmacokinetics relationships (QSPkRs) based on molecular descriptors. QSPkRs allow prediction of PK properties at early stages of drug discovery, even on virtual compounds, thus reducing time and expense and preventing costly late-stage failures. We have previously reported QSPkR models for V_d (11), pf_u (12), CL (13) and $t_{1/2}$ (14) of acidic drugs, as well as for V_d of basic drugs (15).

Given the importance of PPB for overall drug behavior, a good number of studies on *in silico* modeling and prediction of PPB have been recently published. A few studies concerned congeneric series of drugs – β-lactam antibiotics (16, 17), COX-2 inhibitors (18), etc. In general, QSPkRs derived on congeneric series have higher predictive power because a binding to common binding sites can be expected. However, these QSPkRs are often only valid within the studied series. A good number of studies were focused on the binding of diverse drugs to HSA – the most abundant plasma protein, accounting for 50 – 60% of all proteins in plasma. Most of them were performed on one and the same dataset compiled by Colmenarejo (19), comprising

data for the chromatographic capacity factor on HSA immobilized column for 94 diverse drugs and drug like compounds. A wide array of statistical methods was used to derive the models: multiple linear regression (MLR) (19 – 23), artificial neural networks (ANN) (24), support vector machine (CVM) (22, 23), etc. A few QSPkR models have been proposed based on topological sub-structural molecular descriptors (25, 26). This approach is beneficial as it provides information about the contribution of different groups and fragments to drug binding to HSA. Several QSPR models are based on pharmacophoric similarity principle. It is expected that molecules with similar distribution of the pharmacophoric units (hydrogen bond donors/acceptors or hydrophobic regions) interact in similar manner with HSA and manifest comparable affinity (27 – 29). Only few studies considered PPB to all proteins in plasma and proposed QSPkRs for the %PPB (30 – 32).

Most of the published QSPkRs regarding PPB are developed on diverse datasets, involving molecules of all charged types. There are several clues that acidic and basic drugs follow different PK patterns. With respect to PPB, it is believed that acidic drugs bind primarily, with high affinity and capacity to HSA, while basic drugs complex with high affinity to AGP. Bases are also able to bind with low affinity and high capacity to HSA (33), and a presence of a binding site for acidic ligands in AGP molecule is supposed (34). The binding of drugs to various proteins is governed by different driving forces. Both hydrophobic and electrostatic forces are involved in the complexation of acidic drugs with HSA (35), while binding to AGP and lipoproteins is considered as primarily hydrophobic (28, 36, 37). Although in most cases lipophilicity was identified as the main factor favoring PPB, there are also conflicting findings. Kratochwil didn't find sound correlation between lipophilicity and PPB in a dataset of diverse drugs and stated that bases and acids should be treated separately because the influence of lipophilicity on HSA binding is larger for acids than for bases (27). Yamazaki developed statistically significant non-linear relationship between %PPB and $logD_{7.4}$ for a dataset of 90 basic and neutral drugs, but not for acidic drugs and for diverse dataset (28). Obviously, both lipophilicity and ionization state are important for the mode of drug binding to various plasma proteins and contribute to the different behavior of acidic and basic drugs. Therefore, construction of separate QSPkRs with respect to the ionization type seems

reasonable. A QSPkR model for prediction of PPB of acidic drugs was published recently revealing that the lipophilicity, the presence of aromatic rings, cyano groups and H-bond donor-acceptor pairs increase PPB, whereas the presence of a quaternary C-atoms, four-member ring or iodine atoms are unfavorable for complexation (12).

The present study is focused on the development of QSPkR models for PPB of basic drugs. Presented QSPkR models can be used for predicting PPB of potential drug candidates as well as for searching chemical libraries for drug candidates with desired PPB. Identification of the main molecular features affecting PPB allows structural modifications towards molecules with optimal PK properties. The most relevant descriptors are translated into a simple, easy to use checklist of criteria which enables distinguishing between low, high and very high plasma protein binders.

METHODS

Dataset

The dataset used in the study comprises 220 basic drugs extracted from the database of Obach et al. (38) which resumes the major pharmacokinetic parameters of 669 drugs after *iv* administration in human. A drug was considered as a base if the fraction ionized as a base (f_B) at pH 7.4 exceeded 0.03 and was considerably higher than the fraction ionized as an acid (f_A). The values of f_A and f_B were calculated according to the equations:

$$f_A = \frac{1}{1 + 10^{(pK_a - 7.4)}} \qquad (4)$$

$$f_B = \frac{1}{1 + 10^{(7.4 - pK_a)}} \qquad (5)$$

The pK_a values of the drugs were calculated by ACD/LogD version 9.08 (Advanced Chemistry Development Inc., Ontario, Canada). In the case of more than one acidic or basic center in the molecule, the pK_a of the strongest one was taken into account.

The mol files of the drugs were retrieved from public databases – DrugBank (39) and Chemical Book (40). The fraction of unbound drug in plasma (f_u) was used as a quantitative measure for plasma protein binding. It was transformed to $pf_u = - \log f_u$ in order to approach nearly normal distribution, and for better interpretability. Consequently, higher pf_u values implied higher extent of PPB.

The whole dataset was divided into training set and test set. To this end the drugs were ranked in an ascending order of their pf_u values and one of every five drugs was allocated to a different subset of 44 molecules. The fifth subset was set as external test set, while the other four subsets comprised the training set. The training set was employed for development and cross-validation of the QSPkR models, and the test set was used for external validation.

Molecular descriptors and variable selection

Total of 176 molecular descriptors were computed for description of the chemical structures of the drugs using ACD/LogD version 9.08 (Advanced Chemistry Development Inc., Ontario, Canada) and MDL QSAR version 2.2 (MDL Information Systems Inc., San Leandro, CA). Descriptors encoded various molecular features – constitutional (number of particular atoms and groups, rings, circles, hydrogen bond donors and acceptors), physicochemical (logP, $logD_{7.4}$, PSA, dipole moment, polarizability, etc.), geometrical (volume, surface), electrotopological state and connectivity indices, etc. A three step variable selection procedure was applied to identify the most significant predictors of PPB. Initially, all descriptors with non-zero value for less than five molecules were excluded. The remaining 152 descriptors were subjected to genetic algorithm (GA). Finally, selected descriptors entered a forward stepwise linear regression (SWR) with Fisher's criteria thresholds F-to-enter 4.00 and F-to-remove 3.99. Both GA and SWR algorithms were part of the MDL QSAR package.

Development of QSPkR models

A number of QPkR models for pf_u were constructed by MLR on a training set of 176 molecules using different combinations of descriptor. Drugs which pf_u values were predicted with residuals not obeying normal distribution law were considered as outliers. They were excluded from the training set and the models were rebuilt. The models were assessed by the explained variance r^2, given by the equation:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,calc,i}\right)^2}{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,obs,mean}\right)^2} \qquad (6)$$

where $pf_{u,obs,i}$ and $pf_{u,calc,i}$ are the observed and calculated by the model values of pf_u for the i-th drug,

$pf_{u,obs,mean}$ – the mean pf_u value for the training set. Fisher's statistics of the models was also calculated.

Validation of QSPkR models

QSPkR models were validated by leave-one-out cross validation (LOO-CV) and leave-group-out cross-validation (LGO-CV) in the training set, as well as by external test set. In LOO-CV each drug was excluded successively from the training set, a model was derived on the remaining (n-1) drugs, and was used for prediction of pf_u value of the i^{-th} drug. In LGO-CV one subset of 44 molecules (25%) was excluded successively from the training set, a model was derived on the remaining three subsets, and was used for prediction of pf_u values of excluded drugs. Finally, the predictive ability of the models was proved by the external test set, which had not been involved in any step of model development. The content of the training and test sets is summarized in Table 1.

The model performance was assessed by cross-validated coefficient for the training set (q^2_{LOO-CV}), prediction coefficient for the test set (r^2_{pred}), geometric mean fold error of prediction (GMFEP) and mean absolute error of prediction (MAEP), calculated as follows:

$$q^2_{LOO-CV} = 1 - \frac{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,pred,i}\right)^2}{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,obs,mean}\right)^2} \quad (7)$$

$$r^2_{pred} = 1 - \frac{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,pred,i}\right)^2_{test}}{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,obs,mean}\right)^2_{test}} \quad (8)$$

$$GMFEP = 10^{\left(\sum_{i=1}^{n}\left|pf_{u,obs,i}-pf_{u,pred,i}\right|\right)/n} \quad (9)$$

$$MAEP = \frac{\sum_{i=1}^{n}\left|10^{-pf_{u,obs,i}} - 10^{-pf_{u,pred,i}}\right|}{n} \quad (10)$$

where $pf_{u,obs,i}$ and $pf_{u,pred,i}$ are the observed and calculated by the model values of pf_u for the i^{-th} drug and $pf_{u,obs,mean}$ – the mean pf_u value for the training set or in the test set. The accuracy of prediction was evaluated as a percentage of drugs in the test set, predicted with less than two-fold error of the observed value. The fold error was calculated as follows:

$$FE = 10^{\left|pf_{u,obs,i} - pf_{u,pred,i}\right|} \quad (11)$$

The QSPkR models were considered as well predicting if they met the accepted recently statistical criteria: $q^2_{LOO-CV} > 0.5$ and $r^2_{pred} > 0.5$ (41).

RESULTS

Dataset analysis

The dataset of basic drugs used in the present study comprised 220 basic drugs with diverse chemical structure and therapeutic action. The molecular weight varied between 129 and 1431 (mean 366.55) g/mol. Three drugs – caspofungin, cetrorelix and leuprolide, had molecular weight higher than 1000 g/mol. The values of the lipophilicity parameters also varied significantly. logP ranged between -5.05 (caspofungin) and 8.89 (amiodarone), and the same drugs had extreme values of $logD_{7.4}$. The fraction ionized as a base at pH 7.4 ranged between 0.03 and 1.00 with 60% of the drugs almost completely ionized ($f_B > 0.95$). The values of f_u ranged between 0.0002 (amiodarone) and 1 (gentamicin, metformin, netilmicin and tobramycin) with an average value of 0.37. According to the binding affinity, drugs in the dataset were classified as very high binders ($f_u \leq 0.01$) – 13 drugs, high binders ($0.01 < f_u \leq 0.1$) – 55 drugs, moderate binders ($0.1 < f_u < 0.5$) – 70 drugs, and low binders ($0.5 \leq f_u \leq 1$) – 82 drugs.

Table 1. Training set, LGO-CV and external validation test sets.

	Subsets included	
Training set for model development	1 + 2 + 3 + 4	
LGO-CV Training set A	2+3+4	Test set: 1
LGO-CV Training set B	1+3+4	Test set: 2
LGO-CV Training set C	1+2+4	Test set: 3
LGO-CV Training set D	1+2+3	Test set: 4
External validation test set	5	

QSPkR models for pf$_u$

Numerous significant models were generated on the training set of 176 basic drugs using different combinations of descriptors. The best model in terms of statistics is given below:

$$pf_u = 0.109(\pm 0.021)\log P - 0.428(\pm 0.102)f_B + 0.0022(\pm 0.0004)Vol +$$
$$+ 0.053(\pm 0.013)SaaCH\_acnt - 0.117(\pm 0.04)SssssC\_acnt +$$
$$+ 0.0765(\pm 0.028)SaaaC\_acnt - 0.1454$$

$$n\ 170 \qquad r^2\ 0.601 \qquad F\ 33.33$$

Six drugs were identified as outliers: amlodipine (f$_u$ 0.005), amsalog (f$_u$ 0.0011), caspofungin (f$_u$ 0.035), tamsulosin (f$_u$ 0.01), ziprasidone (f$_u$ 0.0012) and leuprolide (f$_u$ 0.54). They were removed from the training set before building the final model.

The QSPkR model involved six descriptors: log P – lipophilicity parameter, f$_B$ – fraction ionized as a base at physiological pH 7.4, Volume – molecular volume, SaaCH_accnt – number of C-atoms of the type aaCH (aromatic non-substituted C-atoms), SaaaC_acnt – number of C-atoms of the type aaaC (aromatic C-atoms in bridged rings) and SssssC_acnt – number of C-atoms of the type ssssC (quaternary C-atoms). Descriptors with positive coefficients in QSPkR contribute positively to PPB, while descriptors with negative coefficients disfavor PPB.

Validation of QSPkR model

The final QSPkR model was validated by internal LOO-CV and LGO-CV on the training set as it was described in **Methods**. The statistics of LOO-CV (q$^2_{LOO-CV}$ 0.551, GMFEP 2.04 and MAEP 0.21) suggested reasonably goof predictive ability. For 54% of the drugs predicted pf$_u$ value is within the 2-fold error of the observed value, with higher accuracy for moderate and low binders (64%) as compared with high and very high binders (39%).

The results from LGO-CV are presented in Table 2. Four models were built on training sets A, B, C and D – each one containing 132 molecules, or 75% of the whole training set. Each model was tested on the respective test set, involving the rest 44 molecules, as shown in Table 1. Although built on training sets, which content differed by 25%, the generated QSPkRs are extraordinarily similar in terms of selected variables, statistics and outliers. This is a good prove for the robustness of the generated QSPkR model. The explained variance r^2 varied between 0.565 and 0.657 (mean 0.598).The values of q$^2_{LOO-CV}$ in the training sets (mean 0.528), r$^2_{pred}$ in the respective external test sets (mean 0.519) and GMFE of about 2 were indicative for fairly good predictive ability.

The predictive ability of the proposed QSPkR model for pf$_u$ of basic drugs was evaluated using external test set of 44 molecules, not involved in any step of model development. Four drugs were identified as outliers: amiodarone (f$_u$ 0.0002), buprenorphine (f$_u$ 0.04), oxybutinine (f$_u$ 0.0004), and cetrorelix (f$_u$ 0.14). The values of r$^2_{pred}$ = 0.532, GMFEP = 1.94 and MAEP = 0.17 suggested good predictive ability of the model. It was able to predict the pf$_u$ values of 59% of the drugs in the two-fold error of the experimental value. The plot of predicted versus observed values of pf$_u$ for the external test set is presented in in Figure 1.
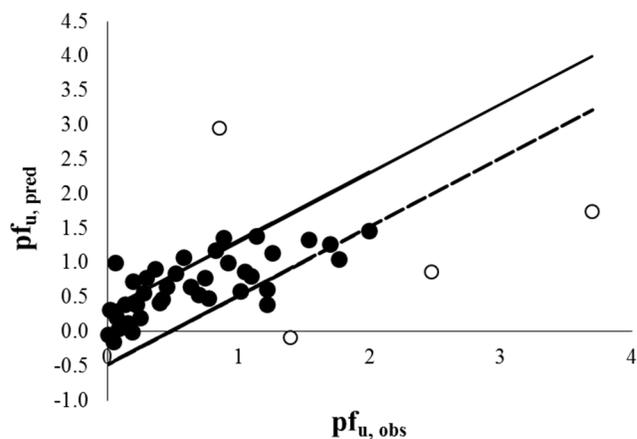

Figure 1. Predicted *vs.* observed pf$_u$ values for the external test set. The four outliers are shown as blank circles. The straight lines represent the twofold error limits.

Table 2. QSPkR models developed in LGO_CV procedure.

Training set	Model	r^2	q^2_{LOO-CV}	r^2_{pred}	GMFEP	Accu-racy, %
A	$pf_u = 0.059 \log P - 0.428 f_B + 0.0012 Vol + 0.089 SssCH\_acnt - 0.171 * SssssC\_acnt + 0.100 SssO\_acnt + 0.0357$	0.567	0.507	0.547	2.00	57
Outliers: amlodipine, amsalog, caspofungin, leuprolide, mibefradil, tubocurarine, ziprasidone Test set: hydroxytaurosporine, clindamycin						
B	$pf_u = 0.132 \log P - 0.399 f_B + 0.0028 Vol + 0.032 SaaCH\_acnt + 0.116 SaaaC\_acnt - 0.334$	0.602	0.544	0.395	2.02	57
Outliers: amlodipine, amsalog, atomoxetine, caspofungin, hydroxychloroquine, reboxetin, telithromycin, ziprasidone Test set: bevantolol, irinotecan, itraconazole, leuprolide, solifenacin, tamsulosin						
C	$pf_u = 0.109 \log P - 0.618 f_B + 0.0015 Vol + 0.066 SaaCH\_acnt - 0.163 SssssC\_acnt + 0.099 SssO\_acnt + 0.067$	0.657	0.606	0.520	2.01	57
Outliers: amsalog, caspofungin, clindamycin, hydroxytaurosporine, leuprolide, mibefradil, tamsulosin Test set: amlodipine, drotaverine, tegaserod, telithromycin, ziprasidone						
D	$pf_u = 0.155 \log P - 0.493 f_B + 0.063 SaaCH\_acnt + 0.108 SssO\_acnt - 0.135 SssssC\_acnt + 0.097 SaaaC\_acnt + 0.2816$	0.565	0.454	0.615	2.04	59
Outliers: amlodipine, clindamycin, drotaverine, tamsulosin, ziprasidone Test set: amsalog, caspofungin, sisomicin						
	Mean	0.598	0.528	0.519	2.02	58

Criteria for pf$_u$ prediction of basic drugs

The descriptors in the generated QSPkR model revealed the main molecular features responsible for PPB of basic drugs. Lipophilicity (expressed as logP), presence of aromatic non-substituted C- atoms (aaCH and aaaC) and molar volume affected positively pf$_u$ (hence, PPB), while basicity (encoded by f$_B$) and presence of quaternary C-atom (ssssC) disfavored PPB. Analysis of the dataset enabled defining some criteria for high PPB, which were summarized in a short checklist (Table 3). Although in general f$_B$ affected negatively PPB, this descriptor was not involved in the checklist because drugs from all PPB affinity groups had relatively high value of f$_B$ > 0.95.

Table 3. Checklist of criteria for PPB of basic drugs

No	Predictor	Increased PPB	Reduced PPB
1	Log P ≥ 3	√	
2	Number of aaCH ≥ 6	√	
3	Presence of aaaC	√	
4	Molar volume > 300 cm^3	√	
5	Presence of ssssC		√

The group of low-PPB drugs involved 82 molecules with very few fulfilled positive criteria. Most of them had low logP values – negative for 15% of the drugs, and exceeding the threshold of 3 for only 10%. The molar volume was less than 300 cm^3 for 75%; only 10% contained six or more aaC-atoms, and 17% involved aaaC atoms. In contrast, 29% possessed one or more ssssC atoms (a negative feature). The group of moderate binders encompased 70 molecules with well-balanced positive and negative features. For 52% logP > 3, 44% contained six or more aaCH atoms, 27% involved aaaC atoms and 24% had molecular volume larger than 300 cm^3. Unfavorable ssssC-atoms presented in 35% of drugs. The high-binders group involved 55 molecules with predominantly positive features. logP > 3 for 64% of the drugs; also 64% involved six or more aaCH-atoms, and 25 % contained aaaC-atoms. The molar volume exceeded 300 cm^3 for 56% of the molecules. Only 25% possessed unfavorable ssssC-atom. Thirteen drugs were classified as very high binders. 92% had molar volume exceeding the threshold of 300 cm^3. For 85% log P > 3, and also 85% contained six or more aaCH groups. 38 % of the drugs contained aaaC-atoms, and only 31% – unfavorable ssssC-atoms.

The difference between the number of positive and the number of negative criteria for high PPB can be used for distinguishing between drugs with different extent of PPB. Based to the analysis of the drugs in the dataset, a simple empiric rule was

proposed. For very high binders the difference between the number of positive and negative criteria should be at least 3, and for high binders – at least 2. In the other extreme, a difference between positive and negative criteria ≤ 1 was considered as an indication for low PPB. The distribution of the drugs according to the difference between the number of positive and negative criteria is shown in Figure 2. Applying the empiric rule, 69% of the very high binders, 71% of the high binders and 91% of the low binders were classified correctly. Seven of the low-binding drugs (8.5%) were incorrectly identified as high-binders (among them – the outlier leuprolide). One of the very high binders and 16 (28%) of the high-binders were incorrectly classified as low-binders (including the outliers buprenorphine, caspofungin and oxybutinine).
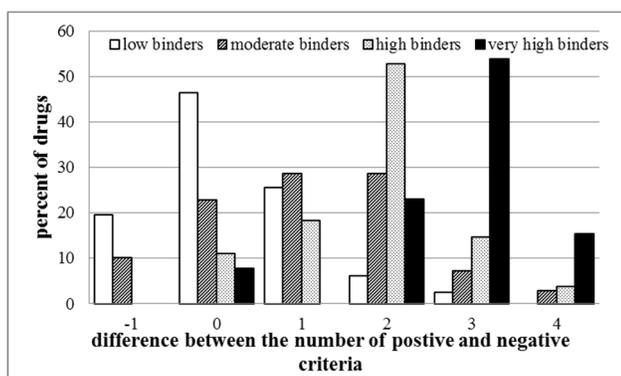

Figure 2. Distribution of drugs with different extent of PPB (given in % on Y-axis) according to the difference between the number of positive and negative criteria (X-axis).

DISCUSSION

The present study is focused on the development of QSPkR for PPB of basic drugs. The extent of PPB is presented as pf_u – the negative logarithm of the free drug fraction in plasma f_u. GA, SWR and MLR are used for model generation on a training set of 176 molecules. The model is validated by LOO-CV and LGO-CV on the training set, and its predictability is evaluated with an external test set of 44 drugs, not involved in any step of model development and validation. The statistics reveals satisfactory performance of the model within the accepted limits: q^2_{LOO-CV} 0.551, mean LGO-CV r^2_{pred} 0.519, external test set r^2_{pred} 0.532, GMFEP 1.94, MAEP 0.17 and accuracy at two-fold level 59%.

Most of the published QSPkRs on PPB refer to the binding of drugs exclusively to HSA. There are very few studies concerning drug binding to all proteins in plasma, generally expressed as % PPB. Votano et al. (30) explored the largest so far dataset of 1008 molecules. Four modeling techniques were used and highly predictive models with external test set r^2_{pred} ranging between 0.59 (MLR, kNN and SVM) and 0.70 (ANN) were developed. However, the large number of descriptors involved in the models (29 – 61) confounded model interpretability. High lipophilicity was identified as the major determinant of PPB. In addition, the importance of the ionization state was suggested as descriptors, encoding acidic function, contributed positively to PPB, while these, signifying bases, had a negative effect. A highly predictive hologram QSPkR was derived on a dataset of 312 molecules with r^2_{pred} 0.86, however no information was gained for the molecular features governing PPB (31). Recently, 794 compounds from Votano's dataset were modeled by linear and non-linear techniques (32). The models with r^2_{pred} between 0.491 and 0.646 confirmed the positive effect of lipophilicity, and the decisive role of the ionization state. Therefore, development of separate QSPkRs for drugs according their ionization type seemed reasonable for gaining more knowledge on the major determinants of PPB of acidic and basic drugs.

In general QSPkR modeling is based on the assumption that all molecules in the dataset have one and the same target and similar mechanism of action. This is not valid for our study focused on PPB to all proteins in plasma, without explicit consideration of any particular protein. The modeling is complicated by the possibility of drug binding to various plasma proteins, to various binding sites – selective or non-selective, reversible or irreversibly, as well as by the occurrence of allosteric interaction and competition with endogenous substances. It is believed that basic drugs bind preferably to AGP. However, given the high concentration of HSA in plasma, it may contribute considerably to the overall binding of basic drugs (33). Also, many lipophilic drugs bind nonspecifically to lipoproteins (42).

The structure of HSA is well characterized, and the architecture of the binding sites for acidic drugs is cleared (35, 43 – 45). There are at least six binding sites primarily inherent for fatty acids, two of which (denoted as Site 1 and Site 2) are specified for binding of acidic drugs. These sites are topologically similar consisting of by predominantly hydrophobic cavities with distinct polar regions. Hydrophobic and electrostatic interactions are considered as responsible for the binding of acidic drugs.

In contrast, the knowledge on the binding sites and mechanisms for basic drugs is rather insufficient. The only crystallographic analysis of HSA complexed with a basic drug, lidocaine, revealed a superficially placed binding site in subdomain IB, different than fatty acids and acidic drugs binding sites. Lidocaine binding is due mainly to cation-π interactions between the phenyl ring and Arg 114, stabilized by electrostatic interactions (46). The presence of a large number of positively charged residuals in the binding site and the lack of hydrophobic interactions is considered as the main reason for the low binding affinity of lidocaine.

Also, little is known about the binding modes of drugs to AGP. Human AGP exists as a mixture of two genetic variants, F1*S and A, which bind drugs with different selectivity (47). The two variants have similar, but not identical topology, which may be a factor for the different substrate selectivity. The F1*S variant possesses a deep and wide branched drug binding pocket consisting of three lobes. The central lobe I is the largest appears to serve as the main hydrophobic drugs binding chamber. Lobes II and III are smaller and negatively charged. Docking studies have shown that the neutral diazepam binds to lobe I, while progesterone – to lobe II. (48). The A variant has the same overall folding as the F1*S variant, but differs in the amino acid sequence and binding site topology. The binding region of A variant is narrower, and involves only lobe I and lobe II, but not lobe III (49). The crystal structures of complexes of a mutant of A variant (with undistinguishable binding affinity from the native type), and three AGP substrates give inside into the binding mode to variant A (49). Disopyramide (DSP) and amitriptyline (AMT), known to be highly selective to variant A, bind in essentially the same manner to the central cavity (lobe I). Both molecules contain two aromatic rings, which are in direct contact with Phe49 and Phe112, resulting in CH-π interactions (edge to face). In addition, van der Waals interactions with Glu64 and Arg90 (DSP) and Leu62 and Arg90 (AMT) are observed. The complex of DSP is further stabilized by hydrogen bonds and van der Waals interactions of the alkyl chain of DSP while the alkyl chain of AMT makes van der Waals contacts with Tyr37 and Val41. Quite different binding mode shows chlorpromazine (CPZ), known to be a nonselective binder to AGP. Its bridged aromatic ring system is involved in π-π stacking interactions with Phe112, and in CH-π interactions with Phe49 and Ala99. Further van der Waals contacts are made with Phe51, Val88, and Arg90.

Obviously, Phe49 and Phe112 are important residuals for the selective binding of drugs to the A variant and the presence of Leu112 in F1*S instead of Phe appear to contribute to the reduced binding affinity of DSP, AMT and other A-variant selective drugs with two aromatic rings and similar structure.

The QSPkR model derived in the present study is well interpretable with respect to the structure of plasma proteins and their binding sites. PPB is favored by lipophilicity, the presence of aromatic C-atoms – both non-substituted and involved in bridged aromatic systems, and molecular volume. Negative impact on PPB the fraction ionized as a base f_B and the presence of quaternary C-atoms. Undoubtedly, lipophilicity is important factor favoring PPB as it is a prerequisite to both selective hydrophobic interactions at the binding sites and non-selective "dissolution" in various plasma proteins. The requirement for six or more aromatic C-atoms means a presence of at least two aromatic rings. The presence of aromatic rings – both separated and bridged, is a prerequisite for the occurrence of specific CH-π, π-π stacking and van der Waals interactions in the binding sites on AGP variants. The large molecular volume ensures a tight fit of the molecule in the binding cavity and closer contact to the suitable hydrophobic amino acid residuals. The negative effect of basicity (encoded by f_B) on PPB is well known. It could be related to the higher tendency of cationic drugs to cross membranes and to distribute into tissues rather than to reside in plasma (50). However, considering basic drugs separately, the negative effect of f_B is not absolute as drugs both with low and very high PPB have $f_B > 0.95$. The negative effect of the presence of quaternary C-atoms may be due to steric hindrances.

The clear physical meaning of the descriptors in the QSPkR model for PPB of basic drugs enabled defining a number of criteria for high PPB. LogP ≥3, number of aromatic non-substituted C-atoms ≥ 6, presence of aromatic bridged ring and molecular volume > 300 L/mol were assigned as positive criteria, and the presence of a quaternary C-atoms – as a negative criterion. The difference between the number of positive and the number of negative criteria allowed distinguishing between drugs with different binding affinity. Applying these criteria, 69% of the very high binders, 71% of the high binders and 91% of the low binders were correctly classified.

Ten drugs (six from the training set and four from the test set) were identified as outliers from the model. PPB was highly underpredicted for the very

high binders amiodarone, amlodipine, amsalog, caspofungin, oxybutinine, tamsulosin and ziprasidone (f_u in the range 0.0002 – 0.01, or 99 – 99.8% PPB) and for buprenorphine (f_u 0.04, 96% PPB). It is generally difficult to estimate correctly the %PPB of the high binders. These drugs usually have very low free plasma levels and require high sensitive assay techniques. In addition, PPB is an equilibrium process, and the free fraction depends crucially on the conditions during the analysis. The deviation of these highly bound drugs may be due to errors in the quantification of the unbound fraction. Alternatively, these drugs may have any structural features favoring PPB, which are not captured in the generated QSPR model. Buprenorphine seems to be a structural outlier from the model as it possesses only one positive feature (logP), and 5 unfavorable quaternery C-atoms. For two drugs – leuprolide (f_u 0.54, 46% PPB) and cetrorelix (f_u 0.14, 86% PPB) PPB was overpredicted. They also could be considered as structural outliers with three positive structural features and no one negative. In addition, the high molecular weight exceeding 1200 g/mol and the low lipophilicity could be restrictive factors for PPB.

CONCLUSIONS

The study presents a significant, predictive and interpretable QSPkR model for PPB of basic drugs. It allows prediction of 59% of the drugs from external validation sets within the 2-fold error of experimental values. The descriptors in the models reveal clear structural features determining PPB of basic drugs. Lipophilicity, the presence of aromatic C-atoms – both non-substituted and involved in bridged aromatic systems, and molecular volume contribute positively to PPB. The fraction ionized as base f_B and the presence of quaternary C-atoms affect negatively PPB. A short checklist of criteria for high PPB is defined, and an empirical rule for distinguishing between low, high and very high plasma protein binders is proposed based on the difference between the number of positive and negative criteria.

REFERENCES

1. Smith DA, Di L, Kerns EH. The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery. Nat Rev Drug Discov, 2010; 9:929-939.

2. Rowland M; Tozer TN. Clinical pharmacokinetics, concepts and applications, 3rd ed., Williams & Williams, Baltimore, Philadelphia, Hong Kong, etc, 1995.

3. Mehvar R. Role of protein binding in pharmacokinetics. Am J Pharm Educ, 2005; 69(5):1-8.

4. Kerns EH; Di L. Drug – like properties. Concepts, structures, design and methods: from ADME to toxicity optimization, Elsevier, Amsterdam, Boston, Heidelberg, etc, pp.187-196, 2008.

5. Schmidt S, Gonzales D, Derendorf H. Significance of protein binding in pharmacokinetics and pharmacodynamics. J Pharm Sci, 2010; 99(3):1107-1122.

6. Bohnert T, Gan L-S. Plasma protein binding: from discovery to development. J Pharm Sci, 2013; 102(9):2953-2994.

7. Yamasaki K, Chuang VTG, Maruyama T, Otagiri M. Albumin – drug interaction and its clinical implication. Biochim Biophys Acta, 2013; 1830:5435-5443.

8. Ascenzi P, Fanali G, Fasano M, Pallotini V, Trezza, V. Clinical relevance of drug binding to plasma proteins. J Mol Struct, 2014; 1077:4-13.

9. Roberts JA, Pea F, Lipman J. The clinical relevance of plasma protein binding changes. Clin Pharmacokinet, 2013; 52(1):1-8.

10. Kennedy T. Managing the drug discovery/development interface. Drug Disc Today, 1997; 2:436-444.

11. Zhivkova Z, Doytchinova I. Prediction of steady-state volume of distribution of acidic drugs by quantitative structure-pharmacokinetic relationships. J Pharm Sci, 2012; 101(3):1253-1266.

12. Zhivkova Z, Doytchinova I. Quantitative structure – plasma protein binding relationships of acidic drugs. J Pharm Sci, 2012; 101(12):4627-4641.

13. Zhivkova Z, Doytchinova I. Quantitative structure – clearance relationships of acidic drugs. Mol Pharmac, 2013; 10:3758-3768.

14. Zhivkova Z, Doytchinova I. In silico guantitative structure – plasma protein binding relationship modeling on acidic drugs: half life. Int J Pharm Pharm Sci, 2014; 6(9):283-289.

15. Zhivkova Z, Mandova Ts, Doytchinova I. Quantitative structure – pharmacokinetics relationships analysis of basic drugs: Volume of distribution. J Pharm Pharm Sci, 2015; 18(3):515-527.

16. Ramamurthi N, Gunturi SB. In silico ADME modeling: QSPR models for the binding of β-lactams to human serum protein using genetic algorithms. ARKOVOC, 2005; XI:102-123.

17. Hall LM, Hall LB, Kier LB. QSAR modeling of beta-lactam binding to human serum proteins. J Comput Aided Mol Des, 2003; 17:103-118.

18. Seedler N, Bhatia S, Singh B. Quantitative correlation between theoretical molecular descriptors and drug-HSA binding affinities for various cox-2 inhibitors. Chem Biol Drug Des, 2008; 72:297-302.

19. Colmenarejo G, Alvarez-Pedraglio A, Lavandera J-L. Cheminformatic models to predict affinities to human serum albumin. J Med Chem, 2001; 44:4370-4378.

20. Hall LM, Hall LH, Kier LB. Modeling drug-albumin binding with E-state topological structure representation. J Chem Inf Comput Sci, 2003; 43:2120-2128.

21. Sitarama BG, Ramamurthi N, Khandelwal A. In silico ADME modeling 2: computational models to predict human serum albumin binding affinity using ant colony systems. Bioorg Med Chem, 2006; 14:4118-4129.

22. Xue CX, Zhang RS, Liu HX, Liu, MC, Hu, ZD, Fan BT. QSAR models for the prediction of binding affinities to human serum albumin using the heuristic method and a support vector machine. J Chem Inf Comput Sci, 2004; 44(5):1693-1700.

23. Chen L, Chen X. Results of molecular docking as descriptors to predict human serum albumin binding affinity. J Mol Graph Model, 2012; 33:35-43.

24. Deeb O. Correlation ranking and stepwise regression procedures in principle components neural networks modeling with application to predict toxic activity and human serum albumin binding affinity. Chem Intel Lab Sys, 2010; 104:181-194.

25. Hajduk PJ, Mendoza R, Petros AM, Huth JR, Bures M, Fesik S, Martin YC. Ligand binding to domain 3 of human serum albumin: a chemometric analysis. J Comput Aid Mol Des, 2003; 17:93-102.

26. Estrada E, Uriarte E, Molina E, Simon-Manso Y, Milne GWA. An integrated in silico analysis of drug binding to human serum albumin. J Chem Inf, 2006; 46:2709-2724.

27. Kratochwil NA, Huber W, Mueller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: a new approach. Biochem Pharmacol, 2002; 64:1355-1374.

28. Yamazaki K, Kanaoka M. Computational prediction of the plasma protein binding percent of diverse pharmaceutical compounds. J Pharm Sci, 2004; 93(6):1480-1494.

29. Liu J, Yang L, li Y, Pan D, Hopfinger AJ. Constructing plasma protein binding model based on combination of cluster analysis and 4D fingerprint molecular similarity analysis. Bioorg Med Chem, 2006; 14:611-621.

30. Votano JR, Parham M, Hall LM, Kier LB, Oloff S, Tropsha A. QSAR modeling of human serum protein binding with several modeling techniques utilizing structure – information representation. J Med Chem, 2006; 49(24):7169-7181.

31. Moda TL, Montanari CA, Andricopulo AD. In silico prediction of human plasma protein binding using

32. Ghafourian T, Amin Z. QSAR models for the prediction of plasma protein binding. Bioimpacts, 2013; 3(1):21-27.

33. Routledge PA. The plasma protein binding of basic drugs. Br J Clin Pharmac, 1986; 22:499-506.

34. Matsumoto K, Sukimoto K, Nishi K, Maruyama T, Suenaga A, Otagiri M. Characterization of ligand binding sites on the alpha 1-acid glycoprotein in humans, bovine and dogs. Drug Metab Pharmacokinet, 2006; 17(4):300-306.

35. Ghuman J, Zunzain PA, Petipas I, Bhattacharya AA, Otagiri M, Curry S. Structural basis of drug binding specificity of human serum albumin. J Mol Biol, 2005; 353:38-52.

36. Otagiri M, Yamamichi R, Maruyama T, Imai T, Suenaga A, Imamura Y, Kimachi K. Drug binding to alpha – 1 acid glycoprotein studied by circular dichroism. Pharm Res, 1989; 6:156-159.

37. Israili ZH, Dayton PG. Human alpha-1-glycoprotein and its interactions with drugs. Drug Metab Rev, 2001; 33:161-235.

38. 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):1385-1405.

39. https://www.drugbank.ca/

40. http://www.chemicalbook.com/

41. Roy K, Kar S, Das RN. Statistical methods in QSAR/QSPR, in Roy K; Kar S; Das RN (eds), A primer on QSAR/QSPR modeling. Fundamental concepts, Springer Cham, Heidelberg, New York, Dordrechs, London, pp 37-59, 2015.

42. Wasan KM, Brooks DR, Lee SD, Sachs-Barrable K, Thornton SJ, Impact of lipoproteins on the biological activity and disposition of hydrophobic drugs: implications for drug discovery. Nat Rev Drug Discov, 2008; 7:84-99.

43. Carter CC, Ho JX. Structure of human serum albumin. Adv Prot Chem, 1994; 45:153-203.

44. Bhattacharia A, Gruene T, Curry S. Crystallographic analysis reveals common modes of binding of medium and long chain fatty acids to human serum albumin. J Mol Biol, 2000; 303(5):721-732.

45. Petipas I, Bhattacharia AA, Twine S, East M, Curry S. Crystal structure analysis of warfarin binding to human serum albumin. Anatomy of drug site I. j boil Chem, 2001; 276(25):22804-22809.

46. Hein KL, Kragh-Hansen U, Morth JP, Jeppesen MD, Otzen D, Moeler JV, Nissen P. Crystallographic analysis reveals a unique lidocaine binding site on human serum albumin. J Struct Biol, 2010; 171:353-360.

47. Herve FO, Caron G, Duch J-C, Gaillard P, Abd Rahman N, Tsantili-Kakoulidou A, Carpurt P-A, D'Athis P, Tillement J-P, Testa, B. Ligand specificity of the genetic variants of human a1-acid

glycoprotein: generation of a three-dimensional quantitative structure-activity relationship model for drug binding to the A-variant. Mol Pharmacol, 1998; 54:129-138.

48. Schoenfeld DL, Ravelli RBG, Mueller U, Skerra A. The 1.8-A crystal structure of α1-acid glycoprotein (orosomucoid) solved by UV RIP reveals the broad drug – binding activity of this human plasma lipocalin. J mol boil, 2008; 384:393-405.

49. Nishi K, Ono T, Nakamura T, Fukunaga N, Izumi M, Watanabe H, Suenaga A, Maruyama T, Yamagata Y, Curry S, Otagiri M. Structural insight between two forms of human α_1-acid glycoprotein genetic variants, the A and F1*S forms. J Biol Chem, 2011; 286:14427-14434.

50. Smith DA, Allerton C, Kalgutkar AS, Waterbeemd H, Walker DK. Distribution, in Smith DA; Allerton C; Kalgutkar AS; Waterbeemd H; Walker DK (eds), Pharmacokinetics and metabolism in drug design, 3rd ed, Wiley – VCH Verlag GmbH&Co. KgaA, Weinheim, pp. 61-79, 2012.