International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 10, Issue 4, 2018

Original Article

# QUANTITATIVE STRUCTURE–PHARMACOKINETICS RELATIONSHIP FOR PLASMA PROTEIN BINDING OF NEUTRAL DRUGS

**ZVETANKA ZHIVKOVA**

**Faculty of Pharmacy, Medical University–Sofia**
**Email: zzhivkova@pharmfac.mu-sofia.bg**

*Received: 30 Jan 2018 Revised and Accepted: 08 Mar 2018*

**ABSTRACT**

**Objective:** Plasma protein binding (PPB) of drugs is important pharmacokinetic (PK) phenomena controlling the free drug concentration in plasma and the overall PK and pharmacodynamic profile. Prediction of PPB at the very early stages of drug development process is of paramount importance for the success of new drug candidates. The study presents a quantitative structure–pharmacokinetics relationship (QSPkR) modelling of PPB for neutral drugs.

**Methods:** The dataset consists of 117 compounds, described by 138 molecular descriptors. Genetic algorithm and stepwise multiple linear regression are used for variable selection and QSPkR models development. The QSPkRs are evaluated by internal and external validation procedures.

**Results**: A robust, significant and predictive QSPkR with explained variance $r^2$ 0.768, cross-validated $q^2_{LOO\text{-}CV}$ 0.731,and geometric mean fold error of prediction (GMFEP) 1.79 is generated, which is able to predict the extent of PPB for 67.6% of the drugs in the dataset within the 2-fold error of experimental values. A simple empiric rule is proposed for distinguishing between drugs with different binding affinity, which allowed correct classification of 78% of the high binders and 87.5% of the low binders.

**Conclusions:** PPB of neutral drugs is favored by lipophilicity, dipole moment, the presence of substituted aromatic and fused rings and a nine-member ring system, and is disfavored by the presence of aromatic N-atoms.

**Keywords:** Plasma protein binding (PPB), Quantitative structure–pharmacokinetics relationship (QSPkR), *In silico* prediction, Human serum albumin (HSA), Alpha-1-acid glycoprotein (AGP).


DOI: http://dx.doi.org/10.22159/ijpps.2018v10i4.24612

## INTRODUCTION

Most of the drugs bind reversibly to various proteins in plasma: human serum albumin (HSA), alpha-1-acid glycoprotein (AGP), lipoproteins, etc. Plasma protein binding (PPB) is a major determinant of the overall pharmacokinetic (PK) and pharmacodynamic (PD) profile. According to the free drug theory, only the free drug is able to be distributed throughout the body and to reach the target receptor–i.e. PPB controls drug potency *in vivo* [1, 2]. Similarly, as only the unbound fraction is available for the clearance organs, PPB may influence drug metabolism, bioavailability and renal excretion [3, 4]. As a result, changes in PPB may have considerable clinical consequences, especially for highly bound drugs (>99%) with narrow therapeutic window, high hepatic extraction ratio (if administered *iv*), and high renal clearance, as well as in critically ill patients [3, 5–7]. The importance of PPB for PK and PD and its clinical relevance have been reviewed recently [1–3, 8–10].

The progress of combinatorial chemistry and high throughput technologies has led to the proposal of extensively growing number of structures with drug-like activities. It has been long recognized that the success of a new drug candidate depends not only on activity and safety but also on its proper absorption–distribution–metabolism–excretion (ADME) characteristics. As PPB appears to be a key determinant for ADME, extensive research has been undertaken targeted to the prediction of PPB at very early stages of drug development process.

In the last two decades *in silico* modeling has established itself as one of the most reliable and promising approaches for early prediction of ADME. It showed the comparable predictive ability to the traditional i*n vivo* and *in vitro* methodologies, while allowing predictions to be made on the basis of theoretical molecular descriptors, even on virtual compounds, thus reducing the investments in technology resources and time [11]. The progress of methodologies for *in silico* prediction of ADME is reviewed in several articles and monographs [11–20]. *In silico* modeling utilizes two general approaches: structure-based and ligand-based. Structure-based methods require knowledge about the 3D structure of both protein and drug and are suitable for modeling specific interactions between drug molecules and proteins. Determination of the X-ray crystallographic structures of the main binding plasma proteins (HSA and AGP) enabled application of structure-based techniques such as molecular dynamics, molecular docking, etc. for more-detailed studying of plasma protein-drug interactions [21–31]. The ligand-based approach is generally used for modeling more complex PK properties, resulting from several underlying processes [18]. One of the most widely used techniques is quantitative structure–pharmacokinetics relationship (QSPkR) modeling. QSPkR models are empirical equations, relating the ADME parameter of interest with a set of molecular descriptors, encoding various aspect of the chemical structure. QSPkR can serve as both predictive and diagnostic tool as it can give insight into the major molecular features governing given ADME property.

Numerous successive QSPkR models have been proposed for analysis and prediction of PPB. Most of them concerned drug binding to HSA–the major plasma protein, accounting for 50–60% of all proteins. A number of QSPkRs were reported for the chromatographic capacity factor on HSA-immobilized stationary phase as a measure for HSA binding affinity, developed on the same dataset of 94 molecules using various statistical tools [32–38]. Several models were based on topological sub-structural descriptors [39, 40] or pharmacophore similarity principle [41–43]. A wide diversity of descriptors have been used, and the repertoire of statistical tools included multiple linear regression (MLR), artificial neural networks (ANN), support vector machines (SVM), pharmacophore similarity and fingerprints, etc.

Although the free fraction of drug in plasma ($f_{u,p}$) was considered as the most reliable parameter characterizing the overall PPB *in vivo*, only a few studies were focused on the prediction of PPB to all

proteins in plasma, without explicit consideration of any particular protein [44–46]. According to all QSPkR models, two factors appeared to be essential for drug binding affinity: lipophilicity and ionization state of the molecule. In general, lipophilicity is favorable for PPB, although there are examples for drugs with equal lipophilicity and quite a different extent of PPB, and *vice versa*. Lexa *et al.* reported a very low correlation between the % HSA binding and log P for the studied dataset with $r^2$ 0.28 and stated that lipophilicity is a necessary but not sufficient requirement for high HSA binding [29]. According to Kratochwil *et al.* the effect of lipophilicity on PPB was larger for acids than for bases [41]. Yamazaki proposed a non-linear relationship between the % PPB and $logD_{7.4}$ for a dataset of 90 basic and neutral drugs, but not for acidic drugs and for the mixed dataset [42]. It is generally accepted, that HSA binds preferably acidic and neutral compounds, while AGP is specialized in complexation of basic and neutral molecules [3, 8]. Therefore different structural requirements for binding to various plasma proteins could be expected, and separate QSPkR analysis according to the ionization state of the molecules seems reasonable.

QSPkR models for PPB of acidic and basic drugs have been published recently [47, 48]. Lipophilicity was identified as a major factor with a positive contribution for both types, while the presence of quaternary C-atom contributed negatively to PPB. PPB of acidic drugs was further favored by the presence of aromatic non-substituted atoms, cyano-groups and a high number of hydrogen bond donor-acceptor pairs, while the presence of a 4-member ring and I-atoms disfavored PPB. PPB of basic drugs was favored by the presence of aromatic non-substituted non-bridged and bridged rings and molecular volume, and disfavored by the ionization. The present study is focused on the development of QSPkR models for PPB of neutral drugs.

## MATERIALS AND METHODS

### Datasets

The dataset used in the study consisted of 117 molecules extracted from the dataset of Obach *et al.*, the largest and best-curated source of data for the key ADME parameters after IV administration [49]. A drug was considered as neutral if the fraction ionized as an acid ($f_A$) or as a base ($f_B$) at physiological pH = 7.4 didn't exceed 3%. The mol files of the drugs were derived from public databases–Drug Bank [50], Chemical books [51], or ChEBI [52]. The value of the free fraction of the drug in plasma ($f_u$) was used as a quantitative measure for PPB. It ranged between 0.0016 and 1 (mean 0.40 ± 0.38, median 0.25), and was logarithmically transformed in order to reach close to normal distribution. For better interpretability, QSPkR models were developed for $pf_u$ =-$logf_u$, so that high value of $pf_u$ implied a high extent of PPB.

The whole dataset was divided into five subsets. To this end, the molecules were arranged in an ascending order according to their $f_u$ values and one of every five drugs was allocated to different subset. Each subset was used once as a test set for validation of the QSPkR model, developed on a training set, consisting of the remaining four subsets. In summary, five training sets (differing in 25% of their content), and five corresponding external test sets were used (table 1).

**Table 1: Training and test sets used for QSPkR development and validation**

| Training set | Subsets included | Test set |
|---|---|---|
| A | 2+3+4+5, n = 93 | 1, n = 24 |
| B | 1+3+4+5, n = 93 | 2, n = 24 |
| C | 1+2+4+5, n = 94 | 3, n = 23 |
| D | 1+2+3+5, n = 94 | 4, n = 23 |
| E | 1+2+3+4, n = 94 | 5, n = 23 |

### Molecular descriptors and variable selection

The chemical structures of the compounds were described with 138 molecular descriptors computed by ACD/logD version 9.08 (Advanced Chemical Development, Inc) and MDL QSAR version 2.2 (MDL Information Systems Inc) software. They included physicochemical (logP, PSA, dipole moment, polarizability), constitutional (number of atoms, groups and bonds of a different type, rings, circles, etc.), geometrical (volume, surface, ovality), electrotopological state and connectivity indices, etc. A three-step variable selection was performed for identification of the more significant predictors: 1. manual rejection of descriptors with a non-zero value for less than 10 molecules; 2. filtering through a genetic algorithm (GA); 3. Stepwise linear regression (SWR) with Fisher criteria F-to-enter 4.00 and F-to-remove 3.99. Both GA and SWR were implemented in MDL QSAR package.

### Generation of QSPkR models for $pf_u$

A number of QSPkR models were constructed for each of the five training sets using MLR and different combinations of descriptors. Drugs for which the values of $pf_u$ were predicted with residuals not following the normal distribution law were considered as outliers. They were removed from the dataset, and the models were rebuilt. The goodness of fit and significance of the models were assessed by:

-Coefficient of determination (explained variance):

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

where $pf_{u,obs,i}$ and $pf_{u,calc,i}$ are the observed and calculated by the model values of $pf_u$ for the $i^{th}$ compound in the training set, and $pf_{u,obs,mean}$ is the mean observed value for $pf_u$.

-Standard error of the estimate:

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,calc,i}\right)^2}{n - p - 1}}$$

Where n is the number of molecules in the train set and p–the number of descriptors in the model.

- Variance ratio, or Fisher statistics:

$$F = \frac{\left(\sum_{i=1}^{n}\left(pf_{u,calc,i} - pf_{u,obs,mean}\right)^2\right)/p}{\left(\sum_{i=1}^{n}\left(pf_{u,obs,i} - pf_{u,calc,i}\right)^2\right)/(n - p - 1)}$$

### QSPkR model validation

The QSPkR models were validated by internal leave-one-out cross-validation (LOO-CV) and external test set validation. Predictive performance was assessed by:

- Cross-validated coefficient for the training set:

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

- Predictive coefficient for the external test set:

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

- Root mean square error of prediction (RMSEP):

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

- Mean fold error of prediction:

$$MFEP = \frac{\sum_{i=1}^{N} 10^{\left|pf_{u,obs,i} - pf_{u,pred,i}\right|}}{n}$$

- Geometric mean fold error of prediction:

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

- Accuracy: the percentage of molecules in the test set which $f_u$ value is predicted within the two-fold or three-fold error of the experimental value.

The QSPkR models were considered as well predictive if they met the proposed recently statistical criteria: $q^2_{LOO\text{-}CV}>0.5$, $r^2_{pred}>0.5$ [53] and GMFEP<2 and accuracy at two-fold error level>60% [54].

**RESULTS**

The QSPkR models developed on the five training sets together with their statistics and outliers are shown in table 2.

**Table 2: QSPkR models for PPB of neutral drugs developed on 5 training sets**

| Train | Model | $r^2$ | $q^2_{LOO\text{-}CV}$ | SEE | F |
|---|---|---|---|---|---|
| 1 | $pf_u = 0.201(\pm0.016)*\log P + 0.065(\pm0.015)*Dipole + 0.593(\pm0.108)*SaaaC + 0.081(\pm0.015)SaasC\_acnt - 0.042(\pm0.011)*SaaN - 0.020$ | 0.781 | 0.732 | 0.309 | 57.2 |
| | Outliers: eltanolone, felodipine, fenoximone, nisoldipine, paclitaxel, tacrolimus, teniposide | | | | |
| 2 | $pf_u = 0.304(\pm0.018)*\log P + 0.140(\pm0.022)*SHBint4\_Acnt + 0.082(\pm0.016)*SaasC\_acnt + 48.54(\pm11.26)*xvch9 - 0.099(\pm0.029)*nrings + 0.093$ | 0.830 | 0.789 | 0.302 | 78.9 |
| | Outliers: diazepam, fenoximone, nisoldipine, paclitaxel, pimobendan, tetrahydrocannabinol | | | | |
| 3 | $pf_u = 0.239(\pm0.016)*\log P + 0.068(\pm0.014)*SaasC\_acnt + 0.058(\pm0.014)*Dipole - 0.253(\pm0.038)*G_{min} - 2.146(\pm0.359)*xc4 + 65.81(\pm11.75)*xvch9 - 0.300$ | 0.838 | 0.801 | 0.280 | 69.6 |
| | Outliers: aminocamptothecin, felodipine, paclitaxel, paricalcitol, pimobendan, Sch34343 | | | | |
| 4 | $pf_u = 0.228(\pm0.016)*\log P + 0.087(\pm0.015)*SaasC\_acnt - 0.204(\pm0.047)*SaaN\_acnt + 0.247(\pm0.082)*SaaaC\_acnt + 2.603(\pm0.891)*xvch5 + 0.043(\pm0.017)*Dipole - 0.023$ | 0.782 | 0.736 | 0.322 | 49.7 |
| | Outliers: aminocamptothecin, nisoldipine, paricalcitol, tetrahydrocannabinol | | | | |
| 5 | $pf_u = 0.314(\pm0.016)*\log P + 0.090(\pm0.023)*SHBint4\_Acnt - 1.973(\pm0.399)*xvc4 + 53.56(\pm9.931)*xvch9 + 0.044(\pm0.015)*SaasC\_acnt - 0.023(\pm0.009)*SaaN + 0.109$ | 0.867 | 0.843 | 0.263 | 87.3 |
| | Outliers: aminocamptothecin, fenoximone, nisoldipine, paclitaxel, Sch 34343, teniposide | | | | |

Predictive ability of developed models was evaluated using five external test sets. Statistical parameters are presented in table 3.

**Table 3: Statistical parameters used for external validation of QSPkR models for neutral drugs**

| Training set | Test set | $r^2_{pred}$ | MFEP | GMFEP | RMSE | Accuracy | Outliers in the test set |
|---|---|---|---|---|---|---|---|
| A | 1 | 0.761 | 2.07 | 1.82 | 0.336 | 63% | paracalcitol |
| B | 2 | 0.785 | 1.81 | 1.73 | 0.273 | 63% | aminocamptothecin, teniposide, Sch 34343 |
| C | 3 | 0.719 | 1.78 | 1.60 | 0.270 | 65% | eltanolone, nisoldipine, doxitaxel, propofol |
| D | 4 | 0.811 | 1.96 | 1.70 | 0.310 | 65% | fenoximone, paclitaxel, tacrolimus |
| E | 5 | 0.765 | 2.14 | 1.89 | 0.347 | 61% | amidulafungin, pimobendan, tetrahydrocanabinol, |
| | Mean | 0.768 | 1.95 | 1.75 | 0.307 | 63.5% | |

The eight most frequently emerging descriptors were used for generation of Consensus model, shown below.

**Consensus model**

$$pf_u = 0.203(\pm0.015)*\log P + 0.046(\pm0.014)*Dipole + 25.5(\pm9.08)*xvch9 + 0.099(\pm0.014)*SaasC\_acnt + 0.208(\pm0.066)*SaaaC\_acnt - 0.038(\pm0.010)*SaaN - 0.038$$

Where logP is the calculated n-octanol/water partition coefficient, Dipole–the dipole moment of the molecule, xvch9–valence 9th order chain connectivity index, SaasC_acnt, SaaaC_acnt and SaaN_acnt–the number of atoms of the type aasC, aaaC and aaN. Six drugs (aminocamptothecin, felodipine, nisoldipine, paclitaxel, paricalcitol, tacrolimus) were identified as outliers and were removed before construction of the final model. Statistic metrics of the model were as follows: $r^2$=0.768, $q^2_{LOO\text{-}CV}$ = 0.731, SEE = 0.323, F = 57.39, MFEP = 1.97, GMFE = 1.79, and accuracy at 2-fold level error = 67.6%.

**DISCUSSION**

The present study was focused on QSPkR modeling of PPB of neutral drugs. The dataset consisted of 117 compounds, described by 138 molecular descriptors. The negative logarithm of the free fraction in plasma ($pf_u$ =-log$f_u$) was used as an end-point variable. A three-step procedure was applied for selection of the most significant variables including manual filtering, GA and SWR. A number of significant QSPkRs were generated by MLR. Although built on training tests which differ in 25% of their content, the models were fairly similar in terms of descriptors involved and the outliers (table 2). This, together with the statistical metrics, suggests the significance robustness of QSPkRs. The predictive performance of the models was assessed by internal and external cross-validation. The values of $q^2_{LOO\text{-}CV}$, $r^2_{pred}$, GMFEP and accuracy are in agreement with the

accepted criteria for good predictive models, namely $q^2_{LOO\text{-}CV}$>0.5, $r^2_{pred}$>0.5 [53], GMFEP<2 and accuracy>60% [54].

The most frequently emerged descriptors were used for the construction of the Consensus model. The model is robust, significant and predictive as proved by the statistical metrics used: $r^2$ = 0.768, $q^2_{LOO\text{-}CV}$ =0.731SEE = 0.323, F = 57.39, MFEP = 1.97 and GMFEP = 1.79. It was able to predict the $f_u$ values of 67.6% of the drugs in the dataset within the 2-fold error of experimental values.

Consensus model involves descriptors with clear physical meaning which reveal the structural features responsible for PPB. The major factor with a positive contribution to PPB is lipophilicity, expressed as logP. It accounts for 55.58% of the explained variance (68.3% without the outliers). The polarity of the molecules (expressed as Dipole), the presence of a nine-member ring system (descriptor xvch9), substituted aromatic C-atoms (SaasC_acnt) and fused rings (SaaaC_acnt) also affect positively PPB, while the presence of aromatic N-atoms (SaaN_acnt) disfavors PPB.

Analysis on the dataset allowed defining a cutoff for each descriptor and criteria for distinguishing between drugs with different extent of PPB. They are summarized in table 4. Although there was a strong positive correlation between $pf_u$ and SaaaC_acnt, the presence of aaaC-atoms was not involved as a criterion for high PPB, because molecules with this structural element were uniformly distributed within all PPB groups.

**Table 4: Checklist of criteria for high PPB of neutral drugs, based on a Consensus model**

| № | Descriptor | Positive effect | Negative effect |
|---|---|---|---|
| 1 | logP>3 | √ | |
| 2 | Dipole>5 | √ | |
| 3 | 9-member ring system | √ | |
| 4 | Number of aasC≥ 3 | √ | |
| 5 | logP<0 | | √ |
| 6 | Presence of aaN | | √ |

The drugs in the dataset were divided into three groups:

- High binders ($f_u \leq 0.1$, i.e. PPB>90%)
- Moderate binders ($0.1 < f_u \leq 0.5$, i.e. 50-90% PPB)
- Low binders ($0.5 < f_u \leq 1.0$, i.e. PPB>50%)

The group of high binders comprised 41 molecules with dominating positive features. For 68% of them logP>3; 49% were fairly polar with Dipole>5; 41% contained 9-member ring system, 60%–at least 4 substituted aromatic C-atoms, and only 24%–aromatic N-atoms. The group of moderate binders involved 36 molecules with well-balanced positive and negative features. logP>3 for only 14%; 46% had Dipole>5; 33% contained 9-member ring system, and 46%–at least 4 aasC-atoms. On the other hand, 30% contained aromatic aaN-atoms, and 11% had logP<0. The low-binders group was represented by 40 molecules with few positive and dominating negative features. No one molecule had log P>0, instead, for 68% logP<0. Only 25% had Dipole>5, 27% contained 9-member ring system, 18%–at least 3 aasC atoms, and 48%–unfavorable aaN-atoms.

The difference between the number of met positive and negative criteria was used for classification of the drugs according to their PPB ability. The following empirical rule was drawn:

Drugs with a difference ≥ 2 were expected to have high PPB ($f_u \leq 0.1$).

Drugs with a difference ≤ 0 were expected to have low PPB ($f_u > 0.5$).

Drugs with difference =1 should have moderate PPB ($0.1 < f_u \leq 0.5$)

Applying this rule, 67% of the drugs were correctly classified, namely 78% of the high binders, 31% of the moderate binders and 87.5% of the low binders. 22% of the high binders were false classified: 5% as low binders, and 17%–as moderate. 12.5% of the low binders were erroneously classified as moderate binders. The prediction for the moderate PPB group was less accurate: 33% of the drugs were incorrectly classified as low binders, and another 36%–as high binders.

Five of the drugs with very high PPB ($f_u \leq 0.01$), although correctly classified as high binders, were identified as outliers, highly under-predicted by the model. These were: paricalcitol ($f_u$0.0016, predicted 0.026), aminocamptothecin ($f_u$0.003, predicted 0.055), nisoldipine ($f_u$0.003, predicted 0.037), felodipine ($f_u$0.0036, predicted 0.031), tacrolimus ($f_u$0.01, predicted 0.092). Similar weakness in the prediction of high plasma protein binders was reported for acidic and basic drugs [45, 46]. Highly bound drugs have very low unbound concentration in plasma which depends crucially on the rate of dissociation of the drug-protein complex. This kinetic factor is not taken into account in QSPkR modeling which assumes rapid dissociation of the complex. In addition, the low free drug concentration in plasma requires highly sensitive analytical techniques and special conditions to preserve the equilibrium state, therefore the possibility for incorrect experimental values for $f_u$ cannot be dismissed.

The developed QSPkR model for neutral drugs is consistent with the structure of the major plasma proteins and their binding sites. Neutral drugs bind with variable affinity to both HSA and AGP [8]. Lipoproteins also contribute to PPB, especially for highly lipophilic compounds [55]. It was suggested that the binding of drugs to HSA occurs at two distinct binding sites defined as Site 1 (warfarin binding site) and Site 2 (benzodiazepine binding site) [56, 57]. The architecture of the binding sites and the modes of complexation were revealed through X-ray crystallographic analysis of HSA complexes with various ligands [58–60]. Both drug binding sites represent large hydrophobic cavities with polar clusters. Site 1 is larger, with three compartments and two polar patches a prerequisite for hydrophobic and electrostatic interactions. Most of Site 1 substrates (warfarin, phenylbutazone, oxyphenbutazone, etc.) are snugly pinned between the non-polar residuals at the bottom of the pocket and make a number of hydrogen bonds with the polar residuals $Tyr^{150}$, $His^{242}$, $Lys^{199}$, $Arg^{222}$. The structure of the pocket and the mode of complexation imply that Site 1 has a preference for molecules with two anionic or electronegative features on the opposite sides of the ligand molecule [59]. Site 2 is narrower, with a single polar cluster, with high affinity for neutral molecules as diazepam, digoxin, clofibrate, 3'-azido-3'-deoxythymidine, acidic ibuprofen, etc. [58, 61]. The presence of only one basic polar patch, located unilaterally in the hydrophobic pocket, determines the specificity of Site 2 for drugs with the peripherally located electronegative group. X-ray analysis revealed that Site 2 drugs are located in the centre of the pocket and able to form hydrogen bonds with $Tyr^{411}$. $Arg^{410}$ and $Ser^{489}$ were also supposed to be involved in salt bridges and hydrogen bonding [59]. Human AGP exists as a mixture of two genetic variants, F1*S and A, which bind drugs with different selectivity [62]. The F1*S variant possesses a deep and wide branched drug binding pocket consisting of three lobes. The central lobe I is the largest and appears to serve as the main hydrophobic drugs binding chamber while lobes II and III are smaller and negatively charged [22]. The binding region of A variant is narrower and involves only lobe I and lobe II [62]. The crystal structures of complexes of a mutant of A variant and three basic AGP substrates (disopyramide, amitriptyline and chlorpromazine) gave inside into the binding mode to variant A [63]. Both disopiramide and amitryptilline contain two aromatic rings, which are in direct contact with $Phe^{49}$ and $Phe^{112}$, resulting in CH-π interactions. Additional van

der Waals interactions with different hydrophobic residuals (Glu[64], Arg[90], Leu[62] and Arg[90]) stabilize the complex. Chlorpromazine has a fused aromatic ring system, involved in π-π stacking interactions with Phe[112], and in CH-π interactions with Phe[49] and Ala[99]. Further van der Waals contacts are made with Phe[51], Val[88], and Arg[90].

According to the Consensus model, the main factor favoring PPB of neutral drugs is lipophilicity. It is a premise for both selective hydrophobic interactions at the binding sites and non-selective "dissolution" in all binding proteins. The presence of aromatic substituted and fused rings (descriptors SaasC_acnt and SaaaC_acnt) and a nine-member ring system (encoded by xvch9, in most cases consisting of fused aromatic six-and five-member rings) is favorable in terms of the possibility of CH-π and π-π stacking at AGP binding sites. The dipole moment of the molecule is a measure of the uneven distribution of the electron density, and it has higher values for extended molecules with distant positive and negative centers. These structural features meet the requirements for binding of drugs at Site 2, as well as for binding at AGP binding site. The negative contribution of the presence of aromatic N-atoms (descriptor SaaN) could be attributed to decrease of lipophilicity and reduced tendency for hydrophobic and Van der Waals interactions in the binding site. Some of the descriptors were suggested to affect positively PPB of acids (logP and SaasC_acnt) and bases (logP and SaaaC_acnt) as well [47, 48].

## CONCLUSION

The present study presents a set of statistically significant, predictive and interpretable models for PPB of neutral drugs. The final Consensus QSPkR allows prediction of 67.6% of the drugs within the two-fold error of experimental values. PPB of neutral drugs is favored by lipophilicity, the presence of aromatic substituted and fused rings, nine-member ring system and high dipole moment of the molecules, while the presence of aromatic N-atoms has a negative effect. A simple rule is proposed for distinguishing between low and high plasma protein binders based on the difference between the number of positive and the number of negative features which allow correct classification of 78% of the high binders and 87.5% of the low binders.

## AUTHOR CONTRIBUTION

All the work has been carried out by the author, Z. Zhivkova.

## CONFLICT OF INTERESTS

Declared none

## REFERENCES

1. Trainor GL. The importance of plasma protein binding in drug discovery. Exp Opin Drug Discovery 2007;2:51-64.
2. Bohnert T, Gan LS. Plasma protein binding: from discovery to development. J Pharm Sci 2013;102:2953-94.
3. Ascenzi P, Fanali G, Fasano M, Pallottini V, Trezza V. Clinical relevance of drug binding to plasma proteins. J Mol Struct 2014;1077:4-13.
4. Zhang F, Xue J, Shao J, Jia L. Compilation of 222 drug's plasma protein binding data and guidance for study design. Drug Discovery Today 2012;17:475-85.
5. Rolan PE. Plasma protein binding displacement interactions–why are they still regarded as clinically important? Br J Clin Pharmacol 1994;37:125-8.
6. Benet LZ, However BA. Changes in plasma protein binding have little clinical relevance. Clin Pharmacol Therap 2002;71:115-21.
7. Roberts JA, Pea F, Lipman J. The clinical relevance of plasma protein binding changes. Clin Pharmacokinet 2013;52:1-8.
8. Schmidt S, Gonzales D, Derendorf H. Significance of protein binding in pharmacokinetics and pharmacodynamics. J Pharm Sci 2010;99:1107-22.
9. Smith DA, Di L, Kerns EH. The effect of plasma protein binding on *in vivo* efficacy: misconceptions in drug discovery. Nat Rev Drug Discovery 2010;9:929-39.
10. Yamasaki K, Chuang VTG, Maruyama T, Otagiri M. Albumin–drug interaction and its clinical implication. Biochim Biophys Acta 2013;1830:5435-43.
11. 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.
12. Van de Waterbeemd H, Giffold E. ADMET *in silico* modeling: towards prediction paradise? Nature 2003;2:192-204.
13. Yamashita F, Hashida M. *In silico* approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet 2004;19:327-38.
14. Butina D, Segall MD, Frankcombe K. Predicting ADME properties *in silico*: methods and models. DDT 2002;7:S83-8.
15. Mager D. Quantitative structure-pharmacokinetic/ pharma-codynamics relationships. Adv Drug Delivery Rev 2006; 58:1326-56.
16. Chohan KK, Paine SW, Waters, S. Advancements in predictive *in silico* models for ADME. Cur Chem Biol 2008;2:215-28.
17. Madden JC. *In silico* approaches for predicting ADME properties. In: Pyzyn T, Leszczynski J, Cronin MTD. editors. Recent advances in QSAR studies. Dordrecht, Heidelberg, London, New York: Springer; 2010. p. 283-304.
18. Xu C, Mager DE. Quantitative structure-pharmacokinetic relationships. Exp Opin Drug Metab Toxicol 2011;7:63-77.
19. Wang J, Hou T. Recent advances on *in silico* ADME modeling. In: Wheeler RA, Spellmeyer D. editors. Annual reports in computational chemistry. Amsterdam, Boston, Heidelberg, etc: Elsevier; 2009. p. 101-27.
20. Ekins S, Waller CL, Qwaan PW, Cruciani G, Wrighton SA, Wikel JH. Progress in predicting human ADME parameters *in silico*. J Pharmacol Toxicol Meth 2000;44:251-72.
21. Huang JH, Cooper MA, Baker MA, Azad MAK, Nation RL, Li J, *et al.* Drug-binding energetics of human α-1-acid glycoprotein assessed by isothermal titration colorimetry and molecular docking simulations. J Mol Recognit 2012;25:642-56.
22. Schoenfeld DL, Ravelli RBG, Mueller U, Skerra A. The 1.8A crystal structure of a-1-acid glycoprotein (orosomucoid) solved by UV RIP reveals the broad drug–binding activity of this human plasma lipocalin. J Mol Biol 2008;384:393-405.
23. Garg A, Manidhar DM, Gokara M, Malleda Ch, Reddy SC, Subramanyam R. Elucidation of the binding mechanism of coumarin derivatives with human serum albumin. PLoS One 2013;8:e63805.
24. Zaidi N, Ajmal MR, Rabbani G, Ahmad E, Khan RH. A comprehensive insight into the binding of hippuric acid to human serum albumin: a study to uncover its impaired elimination through hemodialysis. PLoS One 2013;8:e71422.
25. Rehman MT, Shamsi H, Khan AU. Insight into the binding mechanism of imipenem to human serum albumin by the spectroscopic and computational approach. Mol Pharma 2014;11:1785-97.
26. Dong C, Lu N, Liu Y. Binding of methacycline to human serum albumin at subdomain IIA using multispectroscopic and molecular modeling methods. Luminescence 2013;28:933-41.
27. Yeggoni DP, Gokara M, Manidhar DM, Rachamallu A, Nakka S, Reedy CS, *et al.* Binding and molecular dynamics studies of 7–hydroxycoumarin derivatives with human serum albumin and its pharmacological importance. Mol Pharmacol 2014;11:1117-31.
28. Zhang P, Li Z, Wang X, Shen Z, Wang Y, Yan J, *et al.* Study of the enantioselective interaction of diclofop and human serum albumin by spectroscopic and molecular modeling approaches *in vitro*. Chirality 2013;25:719-25.
29. Lexa KW, Dolghih E, Jacobson MP. A structure-based model for predicting serum albumin binding. PLoS One 2014;9:e93323.
30. Debnath B, Ganguly S. Molecular docking studies and absorption, distribution, metabolism and excretion prediction of novel isatin analogs as human immunodeficiency virus-1-reverse transcriptase inhibitors with broad spectrum chemotherapeutic properties. Asian J Pharm Clin Res 2014;7:186-94.
31. Djajadisastra J, Purnama HD, Yanuar A. *In silico* binding interaction study of mefenamic acid and piroxicam on human serum albumin. Int J Appl Pharmacol 2017;9 Suppl 1:102-6.
32. Colmenarejo G, Alvarez-Pedraglio A, Lavandera JL. Cheminformatic models to predict binding affinities to human serum albumin. J Med Chem 2001;44:4370-8.

33. Hall LM, Hall LH, Kier LB. Modeling drug albumin binding affinity with E-state topological structure representation. J Chem Inf Computer Sci 2003;43:2120-8.
34. Sitarama BG, Ramamurthi N, Akash K. *In silico* ADME modeling 2: computational models to predict serum albumin binding affinity using ant colony system. Bioorg Med Chem 2006;14:4118-29.
35. Xue CX, Zhang RS, Liu HX, Yao XJ, Liu MC, Hu ZD, *et al.* QSAR models for the prediction of binding affinities to human serum albumin using the heuristic method and a support vector machine. J Chem Inf Computer Sci 2004;44:1693-700.
36. Chen L, Chen X. Results of molecular docking as descriptors to predict human serum binding affinity. J Mol Graph Model 2012;33:35-43.
37. Deeb O, Hemmateenehad B. ANN–QSAR model of drug binding to human serum albumin. Chem Biol Drug Res 2007;70:19-29.
38. Deeb O. Correlation ranking and stepwise regression procedures in principal components artificial neural networks modeling with application to predict toxic activity and human serum binding affinity. Chem Intel Lab Syst 2010;104:181-94.
39. Hajduk PJ, Mendoza R, Petros A, Huth JR, Bures M, Fesik W, *et al.* Ligand binding to domain 3 of human serum albumin: a chemometric analysis. J Computer-Aided Mol Des 2003;17:93-102.
40. Estrada E, Uriarte E, Molina E, Simon-Manco Y, Milne GWA. An integrated *in silico* analysis of drug binding to human serum albumin. J Chem Inf Model 2006;46:2709-24.
41. Kratochwil NA, Huber W, Muller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: a new approach. Biochem Pharmacol 2002;64:1355-74.
42. Yamazaki K, Kanaoka M. Computational prediction of the plasma protein binding percent of diverse pharmaceutical compounds. J Pharm Sci 2004;93:1480-94.
43. Liu J, Yang L, Pan D, Hopfinger A. Constructing plasma protein binding model based on a combination of cluster analysis and 4D-fingerprint molecular similarity analyses. Bioorg Med Chem 2006;14:611-21.
44. Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S, *et al.* QSAR modeling of human serum protein binding with several modeling techniques utilizing structure–information representation. J Med Chem 2006;49:7169-81.
45. Moda TL, Montanari CA, Andricopulo AD. *In silico* prediction of human plasma protein binding using hologram QSAR. Lett Drug Des Discovery 2007;4:502-9.
46. Ghafourian T, Amin Z. QSAR models for the prediction of plasma protein binding. BioImpacts 2013;3:21-7.
47. Zhivkova Z, Doytchinova I. Quantitative structure–plasma protein binding relationships of acidic drugs. J Pharm Sci 2012;101:4627-41.
48. Zhivkova Z. Quantitative structure–pharmacokinetics relationships for plasma protein binding of basic drugs. J Pharm Pharm Sci 2017;20:349-59.
49. 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;36:1385-405.
50. https://www.drugbank.ca/. [Last accessed on 01 Dec 2017]
51. http://www.chemicalbook.com/. [Last accessed on 01 Dec 2017].
52. http://www.ebi.ac.uk. [Last accessed on 01 Dec 2017]
53. Roy K, Kar S, Das RN. Statistical methods in QSAR/QSPR. In: Roy K, Kar S, Das RN. editors. A primer on QSAR/QSPR modeling. Fundamental concepts. Heidelberg, New York, Dordrechs, London: Springer Cham; 2015. p. 37-59.
54. Berellini G, Waters NJ, Lombardo F. *In silico* prediction of total plasma clearance. J Chem Inf Model 2012;52:2069-78.
55. Wasan KM, Brocks 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 Discovery 2008;7:84-99.
56. Sudlov G, Birkett DJ, Wade DN. The characterization of two specific drug binding sites on human serum albumin. Mol Pharmacol 1975;11:824-32.
57. Sudlov G, Birkett DJ, Wade DN. Further characterization of two specific drug binding sites on human serum albumin. Mol Pharmacol 1976;12:1052-61.
58. Ghuman J, Zunszain A, 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.
59. He XM, Carter DC. Atomic structure and chemistry of human serum albumin. Nature 1992;358:209-15.
60. Petipas I, Bhattacharya AA, Twine S, East M, Curry S. Crystal structure of warfarin binding to human serum albumin. J Biol Chem 2001;276:22804-9.
61. Carter CC, He XM, Munson SH, Twigg PD, Gernert KM, Broom MB, *et al.* Three-dimensional structure of human serum albumin. Science 1989;244:1195-8.
62. Herve F, Caron G, Duche JC, Gaillard P, Rahman RA, Tsantili-Kakoulidou A, *et al.* Ligand specificity of the genetic variants of human a-1-acid glycoprotein: generation of a three-dimensional quantitative structure-activity relationship model for drug binding to the A variant. Mol Pharmacol 1998;54:129-38.
63. Nishi K, Ono T, Nakamura T, Fukunaga N, Izumi M, Watanabe H, *et al.* Structural insights into differences in drug-binding selectivity between two forms of human alpha1-acid glycoprotein genetic variants, the A and F1*S forms. J Biol Chem 2011;286:14427-34.