Quantitative Structure—Plasma Protein Binding Relationships of Acidic Drugs

ZVETANKA ZHIVKOVA, IRINI DOYTCHINOVA

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

Received 21 June 2012; revised 24 July 2012; accepted 2 August 2012

Published online 7 September 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.23303

ABSTRACT: One of the most important factors, affecting significantly the overall pharmacokinetic and pharmacodynamic profile of a drug, is its binding to plasma protein (PPB). In the present study, we focus on a set of 132 diverse acidic drugs binding to plasma proteins to different extent and develop quantitative structure-plasma protein binding relationships (QSPPBR) to predict their unbound fraction in plasma (f_u) using 178 molecular descriptors. QSPPBR models were derived after variable selection by genetic algorithm followed by stepwise regression and tested by cross- and external validation. The final model has r^2 value of 0.771, q^2 value of 0.737, and four outliers. It predicts 57% of the f_u values with less than twofold error. According to the molecular descriptors selected as the most predictive for PPB, the lipophilicity of the drugs, the presence of aromatic rings, cyano groups, and H-bond donor–acceptor pairs increase the PPB, whereas the presence of tertiary carbon atoms, four-member rings, and iodine atoms decrease PPB. These descriptors were summarized into a short seven-item checklist of criteria responsible for PPB. The checklist could be used as a guide for evaluation of PPB of acidic drug candidates, similarly to the Lipinski's rule of five used for evaluation of oral permeability of drugs. © 2012 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 101:4627–4641, 2012

Keywords: protein binding; plasma protein binding (PPB); computational ADME; QSAR; fraction of unbound drug (f_u); f_u prediction; acidic drugs; complexation; drug design

INTRODUCTION

The analysis of the main reasons for attrition in drug development, published in the late 1990s, reveals that almost 39% of the costly late-stage failures are caused by poor pharmacokinetics.[1] The recognized need for estimation of pharmacokinetic behavior of drug candidates as early as possible led to considerable progress in computational techniques and to development of extensively growing number of models for prediction of various pharmacokinetic properties. As a result, by 2000, the negative contribution of pharmacokinetic factors has decreased dramatically to about 10% of the drug failures.[2] The contemporary state of the matter is reviewed in several articles and books.[3–5] These studies highlighted the potential of *in silico* approach to provide key information at early stages in drug discovery prior to chemical synthesis, allowing prediction to be made even on virtual compounds.

One of the most important factors, affecting significantly the overall pharmacokinetic and pharmacodynamic profile, is the binding of drugs to serum proteins. Although the significance of this phenomena remains a source of sheer never-ending debate,[6] it is generally accepted that protein binding is one of the factors controlling the concentration of free, therapeutically active drug in plasma. Drugs with an excessively high affinity for plasma proteins (>95% bound) may require higher doses to achieve the effective concentrations *in vivo*, and usually have restricted distribution to sites of action. Changes in the binding affinity may result in unforeseen changes in the key parameters for setting up dosing regimen—apparent volume of distribution, clearance, half-life, and bioavailability.[6–8]

The development of computational models for the prediction of plasma protein binding (PPB) affinity is an attractive goal from theoretical and practical point of view. From a theoretical point of view, the structure-PPB relationships gain insights into the chemical

Correspondence to: Zvetanka Zhivkova (Telephone: +359-2-9236514; Fax: +359-2-9879874; E-mail: zzhivkova@pharmfac.acad.bg)

Journal of Pharmaceutical Sciences, Vol. 101, 4627–4641 (2012)

nature of drug–protein interactions. The practical applications of these models in the process of drug development allow the design of new compounds with reduced PPB and less variable pharmacokinetic behavior. Recently, this strategy has been successfully applied by Mao et al.[9]

A good number of studies on the analysis and prediction of PPB have been performed during the last 10–15 years. They could be classified according to several criteria: data, descriptors, and methods. According to the data used, studies are based on congeneric sets of drugs[10–13] or on diverse drugs.[14–24] PPB affinity is usually expressed as a percent of protein-bound drug, or as an equilibrium association (dissociation) constant. The descriptors used in the models are constitutional, topological,[21] electrotopological,[10,16] chemical, physicochemical, quantum chemical, molecular docking based,[24] and combinations of them. The repertoire of methods includes multiple linear regression with variable selection,[12–14,19] artificial neural networks,[22,23] support vector machines (SVMs),[11] ant colony optimization,[19] pharmacophoric similarity and fingerprints,[17] cluster analysis,[18] and so on.

All of the above-mentioned studies use ligand-based methodologies. The only structure-based study to our knowledge is an SVM model combined with automated molecular docking calculations that are able to predict whether albumin binds the query ligand, to determine the probable binding site on albumin, to select the albumin X-ray structure complexed with the most similar ligand and to calculate the geometry of the complex ligand-albumin.[25]

Prediction of binding affinity of drugs to plasma proteins is a rather complicated task. The quality of the model depends on the quality of utilized dataset. Like the other pharmacokinetic parameters, binding affinity data vary significantly from report to report as a result of differences in methodology, experimental conditions, and mathematical approaches. This is especially true for the affinity constants K, which are, in general, determined *in vitro*. In addition, published data for K are restricted to affinity toward a specified plasma protein, and even toward a single binding site in the protein.

However, the pharmacokinetic behavior of drugs is controlled by the free, unbound fraction that is determined by the possible interactions with all proteins in plasma: human serum albumin (HSA), alpha-1-acid glycoprotein (AAG), lipoproteins, globulins, erythrocytes, and so on. Although it is accepted that HSA is primarily responsible for the binding of acidic drugs and AAG displays greater affinity for basic drugs, there are several examples for binding of drugs both to HSA and AAG irrespectively on their ionization state.[7,26,27] Lipoproteins also play a significant role in drug binding, especially for hydrophobic drugs.[28] By increasing the ligand concentration, the ligands can occupy all known binding sites, each with different affinity and different pharmacological relevance.[18]

Considering all these facts, it seems that the extent of protein binding (expressed as free, unbound fraction of the drug f_u) is the most reliable parameter characterizing the overall protein-binding behavior *in vivo*.

The binding of drugs to various plasma proteins seems to be governed by different driving forces. Hydrophobic and electrostatic interactions are involved in complexation of drugs with HSA,[29] whereas binding to AAG and lipoproteins is mainly hydrophobic.[30–32] Recently, it was suggested that the influence of the lipophilicity on binding to HSA is larger for acids than for bases.[15] It is clear that the lipophilicity and the ionization state of the ligand are important for the mode of binding to plasma proteins and contribute to the different behavior of acidic and basic drugs in plasma.

In the present study, we focus on a set of 132 diverse acidic drugs binding to plasma proteins and develop QSPPBR models to predict their unbound fraction in plasma (f_u) using 178 molecular descriptors grouped into several groups: molecular connectivity χ (chi) indices; molecular shape indices; electrotopological state (E-state) indices; and two- and three-dimensional molecular properties. Apart from prediction, the QSPPBR models are used to explore the structural features of drugs important for their PPB. The most relevant descriptors are translated into a simple, easy to use, checklist of criteria governing the PPB.

MATERIALS AND METHODS

Dataset

The dataset used in the present study was compiled from Obach's database[33] which is considered as the largest and the best curated database for pharmacokinetic parameters in humans. The mol files of the drugs were retrieved from DrugBank[34] and ChemicalBook[35] and their pK_a values were calculated by ACD/LogD software (Advanced Chemistry Development Inc., Ontario, Canada). The fraction of a drug ionized at pH 7.4 as acid (f_A) or base (f_B) were calculated according to the equations:

$$\text{for acid}: f_A = \frac{1}{1 + 10^{(pK_a - 7.4)}}$$

$$\text{for base}: f_B = \frac{1}{1 + 10^{(7.4 - pK_a)}}$$

In case of more than one acidic or basic center, the pK_a of the strongest one was taken. A drug was considered as an acid if met at least one of the two conditions: (1) f_A exceeded 10% and $f_B \approx 0$; or/and (2) f_A was

considerably higher than f_B and close to 1. One hundred and thirty-two drugs met one or both of these conditions and entered the dataset used in the study.

The data for f_u in Obach's database were used as a quantitative measure for plasma protein binding of drugs. This parameter represents the fraction of unbound drug in plasma and is given by the equation:

$$f_u = \frac{C_u}{C},$$

where C_u is the unbound drug concentration and C is the total drug concentration in plasma. For QSPPBR purposes, it was used as $pf_u = -\log f_u$. Thus, higher pf_u values correspond to higher PPB.

The whole dataset was ranked in an ascending order of pf_u values and was divided into six subsets of 22 drugs each by allocating one of every six drugs into a different subset. Every subset was used as a test set whereas the remaining five subsets were combined into a training set. The training sets were used to develop and cross-validate the QSPPBR models, and the test sets were used for external validation.

Molecular Descriptors

The chemical structure of the drugs used in the present study was described by 178 molecular descriptors computed using the software packages ACD/LogD version 9.08 (Advanced Chemistry Development Inc., Ontario, Canada) and MDL QSAR version 2.2 (MDL Information Systems, Inc., San Leandro, California). The descriptors were grouped into five types:

(i) Molecular connectivity χ (chi) indices,[36] which represent molecular structure by encoding significant topological features of whole molecule. There are five categories of structural information described by χ indices: degree of branching (low-order χ indices), variable branching pattern (path χ indices), position and influence of heteroatoms (valence χ indices), patterns of adjacency (χ cluster and path/cluster indices), and degree of cyclicity (χ chain indices).
(ii) κ shape indices—a family of graph-based structure descriptors that represent shape.[37]
(iii) E-state indices, which are atom-level molecular descriptors computed for each atom in the molecule.[37] They represent the electron density at each atom and the ability of those electrons to participate in intermolecular interactions.
(iv) Molecular properties—weight, log P, log $D_{7.4}$, number of rings, number of hydrogen bond donors and acceptors, and so on.
(v) Three-dimensional molecular properties such as polarizability, surface area, volume, and so on.

Variable Selection

A genetic algorithm (GA),[38] as implemented in the MDL QSAR package, was used as a variable selection procedure in the present study. GA allows one to select a subset of the most significant predictors using two evolutionary operations: random mutation and genetic recombination (crossover). The algorithm was used in the study with default values for the size of initial population (32), choice of parents (tournament selection), types of crossover (uniform crossover) and mutation (one-point mutation), and fitness function (Friedman's lack-of-fit scoring function with two parameters).[39] The selected variables entered a stepwise linear regression.

QSPPBR by Stepwise Linear Regression

The QSPPBR models in the present study were derived by a stepwise linear regression, as implemented in the MDL QSAR package. It was used in a forward mode with default value for F-to-enter (4.00) and F-to-remove (3.99). Final models were assessed by explained variance (r^2) and standard error of estimate (SEE), according to the following equations:

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

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

where $pf_{u,obs,i}$ is the observed pf_u of the ith drug, $pf_{u,calc,i}$ is the calculated by the model pf_u of the ith drug, n is the number of drugs in the dataset, and d is the number of molecular descriptors in the model. Fisher statistics (F) for models was also calculated.

Validation

Two validation procedures were applied in the study to access the predictive ability of the QSPPBR models: leave-one-out cross-validation (LOO-CV) in the training sets and external validation by test sets. The models were assessed by the coefficients $q^2_{LOO\text{-}CV}$ and r^2_{pred}, respectively, according to the following equations:

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

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

where $pf_{\text{u},pred,i}$ is the predicted by the model pf_{u} of the ith drug. In the LOO-CV procedure, one drug was excluded from the training subset, the model was derived based on the remaining $n - 1$ drugs and used to predict the $pf_{\text{u,pred},i}$ of the excluded ith drug. In the external validation, $pf_{\text{u,pred},i_{test}}$ is the pf_{u} value of a drug from the test subset predicted by the model derived on the basis of the training subset.

Fold error of prediction for the test sets was calculated according to the equation:

$$\text{FEP} = 10^{|pf_{\text{u,obs},i} - pf_{\text{u,pred},i}|}$$

and the average value of FEPs for the test drugs is given as mean fold error MFEP.

The *accuracy* of prediction is given as a percent of the total number of drugs predicted with less than twofold error.

RESULTS

Dataset Analysis

The dataset of acidic drugs used in the present study encompasses a wide variety of structures, therapeutic actions, and physicochemical properties. A drug was considered as an acid if it met at least one of the two conditions: (1) f_{A} exceeded 10% and $f_{\text{B}} \approx 0$; or/and (2) f_{A} was considerably higher than f_{B} and close to 1. The dataset consisted of 132 drugs.

The molecular weight (M) of the analyzed drugs varies between 126 and 1297 g/mol with an average value of 379.4 g/mol. M exceeds 500 g/mol for 14 drugs and two of them (micafungin and suramin) have M higher than 1200 g/mol.

The values of lipophilicity parameters log P and log $D_{7.4}$ also vary significantly. Log P ranges between −7.48 (micafungin) and 7.4 (montelukast) and the same drugs have extreme values of log $D_{7.4}$ (−11 and 5.19, respectively).

The values of f_{u} range from 0.0004 ($pf_{\text{u}} = 3.398$) for verlukast to 1 ($pf_{\text{u}} = 0$) for atovaquone, cidofovir, flucytosine, and fosfomycin with an average value of 0.30 and median value of 0.15. The histogram of the data distribution of the experimental f_{u} values is shown in Figure 1.

QSPPBR Models for pf_{u}

A preliminary or an initial model was derived based on the whole set of 132 acidic drugs and using the whole set of 178 descriptors. The descriptors were selected by GA followed by a stepwise linear regression.

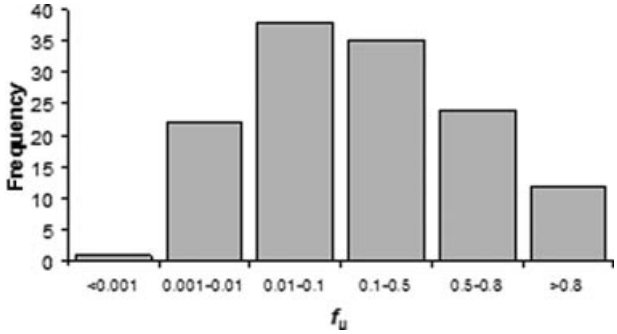

Figure 1. Histogram of the experimental f_{u} values.

This initial model was cross-validated by LOO and it is given below as Model (1):

$$\begin{aligned} pf_u &= 0.177 \log P\,(\pm 0.021) + 0.082\,(\pm 0.020) SaasC\_acnt \\ &+ 0.024(\pm 0.008) SaaCH + 0.057(\pm 0.015) StN \\ &+ 0.04(\pm 0.014) SsCl + 0.071(\pm 0.018) SHBint4\_acnt \\ &+ 0.101(\pm 0.017) SHBint5\_acnt + 0.492 xvp10 \\ &- 0.405(\pm 0.097) SsI\_acnt \\ &- 0.063(\pm 0.013) SHBint6\_acnt - 0.334(\pm 0.068) SssssC\_acnt \\ &- 0.155(\pm 0.050) SsssN + 0.441. \qquad \text{Model,(1)} \end{aligned}$$

$$n = 126,\ r^2 = 0.822,\ SEE = 0.371,\ \text{F} = 43.41\ q^2 = 0.744,$$

$$MFEP = 2.17 \pm 1.33,\ accuracy = 61\%.$$

Six outliers were detected and removed from the dataset (5-aminosalicylic acid, bromfenac, ceftriaxone, telmisartan, tesaglitazar, and verlukast).

To derive a robust model, the whole set of 132 was divided into six training subsets, each consisting of 110 molecules, and six corresponding test subsets, each containing 22 molecules, as described in the section *Materials and Methods*. The training subsets were used for the development and cross-validation of QSPPBR models and the test subsets were used for external validation of the models. The robust modeling of pf_{u} was performed in three steps.

As a first step, variable selection procedure by GA followed by a stepwise linear regression was applied to every training subset using the whole descriptor set. The aim of this step was to select the most predictive descriptors.

Second, from the pool of the most predictive descriptors were selected the most frequently emerged ones. As most frequently emerged descriptors were defined those which appear at least in two of the six models. They were used as a new input descriptor set and underwent variable selection by GA followed by stepwise regression for every training subset.

Finally, a consensus model was derived based on the whole set of 132 acidic drugs using the set of the most frequent descriptors.

If a pair of intercorrelated descriptors ($r > 0.65$) appeared in a model, the less significant descriptor was omitted and the model was rebuild. The results at each step are considered below.

QSPPBR Models for pf_u Using the Whole Descriptor Set

The set of 178 descriptors underwent variable selection by GA followed by a stepwise regression for every training subset of drugs. The derived models were cross-validated and externally validated by the corresponding test sets. Drugs were considered as outliers if the residuals did not obey normal distribution law. They were removed from the training sets, and the models were rebuilt. The statistics of the best performing models is given in Table 1.

The descriptors that have been selected in more than one model were as follows:

- log *P* or log *D*. They intercorrelate with $r^2 = 0.748$ and emerged in all models.
- *SaasC_acnt* or *SaaCH_acnt*. These descriptors count the number of aromatic nonconjugated substituted and nonsubstituted carbon atoms in molecule, respectively. They intercorrelate with $r^2 = 0.512$ and were found in five of the six models.
- *StN_acnt, StN*, or *StsC*. These descriptors account for the presence of cyano (-CN) group. They exist in four models.
- *SsssC_acnt*. This descriptor counts the number of aliphatic tertiary carbon atoms in molecule. It was selected in three models.
- *SsCl_acnt*. This descriptor counts the number of Cl atoms in molecule. It appeared in two models.
- *xch*4. This descriptor accounts for the presence of four-member ring. It was found in two of the models.
- *SsI_acnt* counts the number of I atoms in molecule and it is important descriptor in two of the models.
- *SsssN*. This descriptor gives the sum of all (>N-) E-state values in molecule. It was presented in two of the models.
- *SaaN*. It shows the sum of all E-state values of the aromatic N atoms in molecule and appeared in two models.
- *SHBint 5_acnt* counts the number of pairs of hydrogen bond donors and hydrogen bond acceptors separated by five skeletal bonds (internal hydrogen bonds with five skeletal bonds between donor and acceptor). It is an important descriptor in two models.
- *SHBint4* or *SHBint4_acnt*. These descriptors account for the presence or number of pairs of hydrogen bond donors and hydrogen bond acceptors separated by four skeletal bonds (of internal hydrogen bond with four skeletal bonds between donor and acceptor). They emerged in two models.

These 16 most frequently emerged descriptors form a new descriptor set, which was used in the next step of pf_u modeling.

QSPPBR Models for pf_u Using the Most Frequently Emerged Descriptors

The new set of 16 descriptors underwent variable selection by GA followed by a stepwise regression for every training subset of drugs. The newly derived models were cross-validated and externally validated by the corresponding test sets. The statistics of the best performing models is given in Table 2.

The comparison between the two sets of models shows that reducing the set of descriptors slightly decreases the mean values of r^2 and $q^2_{\text{LOO-CV}}$ but improves the overall prediction by decreasing the MFEP, increasing the *accuracy* and reducing the number of outliers.

Consensus QSPPBR Model for pf_u Using the Most Frequently Emerged Descriptors

At the last step of pf_u modeling, a consensus model was derived using the set of 16 frequently emerged descriptors and based on the whole set of 132 acidic drugs. The derived consensus model is given below as Model (2):

$$\begin{aligned} pf_u &= 0.208(\pm 0.020)\log P + 0.125(\pm 0.019)SaasC\_acnt \\ &+ 0.276(\pm 0.081)StsC + 0.046(\pm 0.013)SHBint4\_acnt \\ &+ 0.074(\pm 0.017)SHBint5\_acnt - 0.193(\pm 0.073)SssssC\_acnt \\ &- 3.104(\pm 0.760)xch4 - 0.498(\pm 0.101)SsI\_acnt \\ &+ 0.419, \qquad \text{Model,(2)} \\ n &= 128, r^2 = 0.771, SEE = 0.410, \mathrm{F} = 49.98, \\ q^2 &= 0.737, MFEP = 2.45 \pm 1.66, accuracy = 57\%. \end{aligned}$$

Model (2) had statistics close to this of Model (1) but with fewer descriptors and only four outliers: bromfenac, suramin, tesaglitazar, and verlukast.

Checklist for PPB Prediction

The descriptors involved in Model (2) account for concrete structural features and reveal their contribution to PPB of acidic drugs. The lipophilicity of molecules (expressed as log *P*) and the presence of aromatic substituted carbon atoms (descriptor *SaasC_acnt*), cyano groups (descriptor *StsC*), and/or H-bond donors and

Table 1. Cross–Validation of the QSPkR Models Based on the Whole Descriptor Set

Training Set	Model	r^2	SEE	F	$q^2_{\text{LOO-CV}}$	r^2_{pred}	MFEP	Accuracy (%)	Outliers
1	$pf_{\text{u}} = 0.169 \log P + 0.107 SaasC\_acnt - 5.312 xch4 - 0.446 SsI\_acnt + 0.433 StN\_acnt + 0.105 SHB\text{int}5\_Acnt - 0.27 SaaaC\_acnt + 0.030 SaaCH - 0.077 SHssNH + 0.576$ Outliers from the training set: 5-aminosalicylic acid, acetylsalicylic acid, ceftriaxone, enalaprilat, and tesaglitazar. Outliers from the test set: cefotetan, glimepiride, irbesartan, and verlukast.	0.847	0.353	58.56	0.811	0.771	2.98	45	9
2	$pf_{\text{u}} = 0.128 \log P + 0.192 SaasC\_acnt - 0.363 SssssC\_acnt - 0.112 SHB\text{int}6\_acnt - 0.0422 SaaN + 0.983 xvp10 + 0.698$ Outliers from the training set: acetylsalicylic acid, amidotrizoate, folinic acid, furosemide, tesaglitazar, and verlukast. Outliers from the test set: fosfluconazole, iothalamic acid, and micafungin.	0.819	0.370	73.01	0.796	0.683	2.47	41	9
3	$pf_{\text{u}} = 0.180 \log P + 0.066 SaaCH\_acnt - 2.550 xch4 + 0.057 StN\_acnt + 0.370 SsCl\_acnt - 0.194 SssssC\_acnt + 0.102 SHB\text{int}8\_Acnt - 0.252 SsssN + 0.590$ Outliers from the training set: acetylsalicylic acid, ceftriaxone, diflunisal, glyburide, tesaglitazar, tolcapone, and verlukast. Outliers from the test set: bromfenac, meloxicam, and tenoxicam.	0.823	0.358	54.55	0.782	0.115	2.84	32	10
4	$pf_{\text{u}} = 0.107 \log D + 0.209 SaasC\_acnt - 0.520 SsI\_acnt + 0.051 StN + 0.007 SHB\text{int}4 - 0.022 SaaN + 45.46 xch10 - 0.074 SsNH_2 - 0.116 SsCH_3 + 0.087 ka3 + 49.08 xvch9 + 0.593266$ Outliers from the training set: bromfenac, ceftriaxone, glimepiride, and suprofen. Outliers from the test set: ibuprofen, suramin, and tesaglitazar.	0.809	0.400	36.15	0.748	0.803	1.94	55	7
5	$pf_{\text{u}} = 0.194 \log D + 0.072 StN + 0.469 SsCl\_acnt + 0.083 SHB\text{int}4\_acnt - 0.217 SssssC\_acnt + 1.346 H_{\max} + 0.042 SHother - 0.026 SHB\text{int}3 - 0.871 SdssS - 14.550 SpcPolarizability + 0.373 H_{\min} - 0.132 SdsCH\_acnt - 2.342$ Outliers from the training set: acetylsalicylic acid, acetazolamide, bromfenac, doxifluridine, and tesafglitazar. Outliers from the test set: 5-aminosalicylic acid, pravastatin, and tolcapone.	0.822	0.387	35.38	0.771	0.698	2.54	45	8
6	$pf_{\text{u}} = 0.083 SaaCH\_acnt + 0.134 \log D + 0.091 SHB\text{int}5\_acnt - 0.248 SssNH\_acnt + 0.380 SsCl\_acnt - 0.186 SsssN + 0.009 SHB\text{int}4 - 0.274 SssssC\_acnt - 0.266 SsNH_2\_acnt + 0.301 StsC + 0.056 k3 + 0.812$ Outliers from the training set: bromfenac, glimepiride, pravastatin, and tesaglitazar. Outliers from the test set: acetylsalicylic acid, ceftriaxone, cephaloridine, epristeride, and folinic acid.	0.834	0.370	43.04	0.781	0.888	2.27	55	9
Mean		0.826	0.373	50.12	0.782	0.660	2.51	46	9

DOI 10.1002/jps

Table 2. Cross-Validation of the QSPkR Models Based on the Most Frequently Emerged Descriptors

Training Set	Model	r^2	SEE	F	$q^2_{\text{LOO-CV}}$	r^2_{pred}	MFEP	Accuracy (%)	Outliers
1	$pf_u = 0.187 \log P + 0.139 SaasC\_acnt - 4.785 xch4 - 0.547 SsI\_acnt + 0.05075 StN + 0.088 SHBint5\_acnt + 0.454$ Outliers from the training set: bromfenac, ceftriaxone, suramin, and tesaglitazar. Outliers from the test set: cefotetan, glimepiride, irbesartan, and verlukast.	0.789	0.395	61.58	0.764	0.775	2.26	45	8
2	$pf_u = 0.197 \log P + 0.102 SaasC\_acnt - 3.203 xch4 - 0.429 SsI\_acnt + 0.009 SHBint4 - 0.215 SssssC\_acnt + 0.589$ Outliers from the training set: bromfenac, tesaglitazar, and verlukast. Outliers from the test set: fosfluconazole and micafungin.	0.751	0.423	50.31	0.719	0.678	2.68	50	5
3	$pf_u = 0.168 \log P + 0.135 SaasC\_acnt - 3.725 xch4 - 0.489 SsI\_acnt + 0.485 StN\_acnt + 0.075 SHBint5\_acnt + 0.250 SsCl\_acnt + 0.426$ Outliers from the training set: suprofen, suramin, tesaglitazar, and verlukast. Outlier from the test set: bromfenac.	0.763	0.418	45.11	0.653	0.514	2.54	50	5
4	$pf_u = 0.192 \log P + 0.138 SaasC\_acn - 4.077 xch4 - 0.5384 SsI\_acnt + 0.053 StN + 0.092 SHBint5\_acnt + 0.418$ Outliers from the training set: bromfenac, ceftriaxone, glimepiride, irbesartan, and verlukast. Outliers from the test set: suramin and tesaglitazar.	0.787	0.403	60.33	0.761	0.731	2.33	50	7
5	$pf_u = 0.192 \log P + 0.132 SaasC\_acnt - 4.01 xch4 - 0.5265 SsI\_acnt + 0.053 StN + 0.099 SHBint5\_acnt + 0.426$ Outliers from the training set: bromfenac, suramin, tesaglitazar, and verlukast. Outlier from the test set: suprofen.	0.745	0.427	48.27	0.710	0.605	2.27	50	5
6	$pf_u = 0.177 \log P + 0.081 SaaCH\_acnt - 2.069 xch4 - 0.32 SsI\_acnt + 0.086 StN + 0.109 SHBint5\_acnt - 0.243 SssssC\_acnt + 0.351 SsCl\_acnt - 0.234 SsssN + 0.522$ Outliers from the training set: 5-aminosalicylic acid, bromfenac, glimepiride, suramin, tesaglitazar, and verlukast. Outliers from the test set: ceftriaxone and diflunisal.	0.818	0.368	52.08	0.777	0.656	2.81	45	8
Mean		0.776	0.406	52.08	0.731	0.660	2.15	48	6

DOI 10.1002/jps

Table 3. Checklist of Criteria for Plasma Protein Binding

Number	Presence of	PPB Increases	PPB Decreases
1	Log $P \geq 3$	√	
2	Aromatic nonconjugated rings ≥2	√	
3	Cyano (CN) group ≥1	√	
4	H-bond donors and acceptors separated by four or five skeletal bonds ≥3	√	
5	Tertiary C atoms ≥1		√
6	Four-member ring ≥1		√
7	I atom ≥1		√

acceptors separated by four or five skeletal bonds (descriptors *SHBint4_acnt* and *SHBint5_acnt*) increase the PPB. The presence of tertiary C atoms (descriptor *SssssC_acnt*), four-member rings (descriptor *xch*4) and/or I atoms (descriptor *SsI_acnt*) decrease the PPB ability of drugs. The above descriptors were summarized into a short seven-item checklist of criteria responsible for PPB (Table 3). Although the number of nonsubstituted aromatic carbon atoms (encoded with the descriptor *SaaCH_acnt*) was not involved in Model (2) it was considered together with *SaasC_acnt* because of its high positive impact on PPB. The presence of aromatic substituted and nonsubstituted carbon atoms was expressed as presence of aromatic nonconjugated rings.

The checklist was applied to the dataset of acidic drugs (Appendix 1). It was found the cutoff for each descriptor distinguishing between high binders ($f_u \leq 0.05$) and the rest of binding drugs and defined as a criterion for PPB. To bind to plasma proteins, an acidic drug needs log P higher than 3, more than two aromatic nonconjugated rings and/or cyano group. Low lipophilicity (log $P < 1.5$) as well as the presence of tertiary carbon atoms, four-member rings, and/or iodine atoms result in decrease in PPB of acidic drugs.

The checklist was applied on the set of acidic drugs and a minimum number of criteria for good PPB was defined (Fig. 2). For PPB above 95% ($f_u = 0.05$, $pf_u = 1.301$), drugs should have difference in the number of positive and negative criteria at least 2. The options for a difference higher or equal to 2 are as follows: four positive criteria (none of the drugs), three positive without any negative (bromfenac, tolcapone, fluvastatin, and glyburide), two positive without any negative (most of the high binders), four positive and two negative (none of the drugs), and three positive and one negative (none of the drugs).

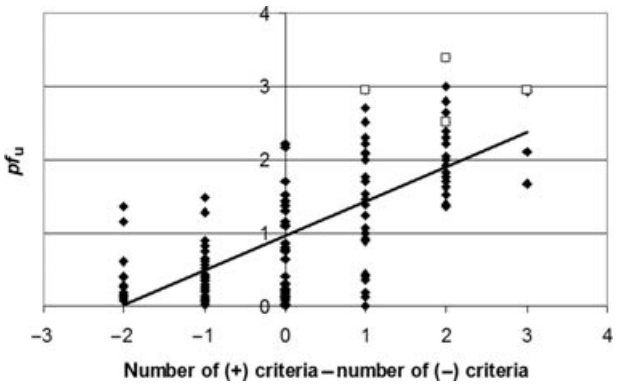

Figure 2. Relationship between pf_u and the difference between the number of positive and negative criteria. The outlies from Model (2) are given as open squares. The correlation coefficient is 0.704 (including outliers).

DISCUSSION

The present study is focused on the quantitative relationships between the molecular structure of 132 acidic drugs and their PPB affinity without considering explicitly any particular plasma protein or binding site. PPB affinity was expressed as pf_u and QSPPBR models were built starting from 178 molecular descriptors. A good QSPPBR model (Model 1) was derived even in the first run on the whole dataset. To avoid any chance correlation and to develop a robust model, the whole dataset was divided systemically into six training and six test subsets, which underwent variable selection by GA followed by a stepwise regression. The derived models were cross- and external validated and showed good predictive ability. A subset of 16 most frequently emerged in the models descriptors was selected and used to derive a consensus model for PPB prediction. The descriptors included in this model were converted into a seven-item checklist of criterion for PPB prediction.

A prerequisite for a good QSAR model is the molecules under study to bind the same binding site on the target biomacromolecule. This is not the case with drug binding in plasma where specific and nonspecific, reversible, and nonreversible interactions with many proteins are possible. The interpretation of models based on such data is further complicated by the presence of numerous secondary binding sites, allosteric interactions between bound drugs, and especially interactions with endogenous ligands arising *in vivo*. In practice, the mechanisms of drug–plasma protein interactions are as diverse, as the drugs themselves, and their binding affinities are determined by different structural features.

In plasma, the acidic drugs bind mainly to albumin. Binding to AAG and lipoproteins is in general low and inconsequential, governed by nonspecific hydrophobic interactions.[30–32] The X-ray structure of HSA in

complexes with several drugs[29,40,41] revealed the architecture of the binding sites and provided an insight into the binding mechanisms. There are at least six binding sites on HSA, most of which are inherent for fatty acids but also accommodate drugs such as azapropazone, indomethacin, diflunisal, ibuprofen, and so on. The complexation of drugs occurs predominantly at two specific sites defined as Site 1 and Site 2.[42] They are topologically similar but not identical. They both contain largely preformed hydrophobic cavities with distinct polar features. Site 1 is larger, with three compartments and two polar clusters, whereas Site 2 is narrower, with a single polar cluster. Hydrophobic and specific interactions with the polar residuals are responsible for the interactions of acidic drugs with albumin. At Site 1 most of the studied drugs have a planar aromatic moiety snugly pinned between the nonpolar residuals at the bottom of the pocket, another hydrophobic part projected in the left or right compartment, and make a number of hydrogen bonds with the polar residuals (Tyr^{150}, His^{242}, Lys^{199}, and Arg^{222}). The presence of a planar hydrophobic structure and two negative centers separated by five or six bonds is a prerequisite for binding of drugs to Site 1. The complexation of drugs at Site 2 occurs through clustering of an aromatic ring in the center of the pocket and hydrogen bonds with Tyr^{411}, Ser^{489}, and Leu^{430}. Obviously, the presence of planar hydrophobic structure and two negative centers separated by a few skeletal bonds is a prerequisite for binding of drugs to both Site 1 and Site 2. These structural features are encoded in our models by the descriptors *SaasC_acnt*, *SaaCH_acnt,* *SHBint4_acnt*, and *SHBint5_acnt*.

The models contain also the lipophilicity descriptors log *P* and log *D*. Higher lipophilicity of drugs corresponds to higher pf_u values and PPB. This is in a good agreement with the general assumption that binding of drugs to plasma proteins is governed by nonspecific hydrophobic interactions. The presence of cyano group encoded by the descriptors *StN_acnt*, *StN*, and *StsC* favors PPB. The positive contribution of this structural feature has been already reported.[43] Another structural feature with positive contribution in PPB is the presence of chlorine atom, encoded by the descriptor *SsCl_acnt*. Among the molecular descriptors with negative impact on the PPB affinity is the simple four-order chain connectivity index *xch*4, reflecting the presence and substitution in four-membered ring. This structure is common in 43 molecules used in the present study, mainly β-lactam antibiotics, and reflects the low (with few exceptions) protein binding of this therapeutic class. The presence and the number of iodine atoms in the molecule reduce the PPB affinity implied by the negative coefficient of *SsI_acnt* in the models. There are only two drugs in the dataset containing I-atoms (iothalamic acid and amidotrizoate). Both show negligible PPB with f_u 0.98 and 1, respectively. Negative influence on the binding affinity has also the number of substituted quaternary aliphatic C atoms in the molecule encoded by the descriptor *SsssC_acnt*. Aromatic and tertiary N atoms, described by *SsssN* and *SaaN*, also decrease PPB.

Many drugs appear as outliers in the models derived in the present study. However, only four drugs do not obey the final consensus Model (2): bromfenac, suramin, tesaglitazar, and verlukast. All of them have f_u values in the lowest range ($f_u < 0.003$) corresponding to PPB higher than 99.7%. Such drugs usually have extremely low plasma levels for the free fraction and require high-sensitive and specific bioassay methods to detect it. Frequently, the free drug plasma concentrations depend on the assay methodology, as it has been demonstrated for ceftriaxone.[44] Therefore, the deviations of the outliers may be as a result of a measurement error in the determination of the unbound fraction, or to saturation of a low-capacity high-affinity binding site in plasma.

The quantitative structure/pf_u relationships derived in the present study allow defining of several clear criteria for binding of acidic drugs to plasma proteins. The lipophilicity of drugs is the most important of them. As more lipophilic is a drug as it binds to plasma proteins in a greater extent (Fig. 3). Among the high binders (PPB $\geq$ 95%, $f_u \leq 0.05$, $pf_u \geq 1.301$), 81% have log *P* higher than 2 and 53% have higher than 3. Micafungin is an outlier with log *P* of −7.48, but its low log *P* value is compensated by another positive criteria. The second very important criterion for good PPB is the presence of at least 2 nonconjugated aromatic rings. The aromatic ring is part of the biophore for Site 1 and Site 2 of HSA.[29] Estrada et al.[21] also have found a high fragment contribution of benzene to HSA binding affinity. The third criterion favoring PPB is the presence of cyano group. The studied set includes five CN-containing drugs: cilomilast ($f_u = 0.006$), entacapone ($f_u = 0.02$), levosimendan ($f_u = 0.02$), milrinone ($f_u = 0.035$), and cefmetazole ($f_u = 0.15$). Four of them are high binders. The fourth criterion for PPB is the presence of at least three pairs of H-bond donor and acceptor separated by four or five skeletal bonds. The two negative centers separated by a few skeletal bonds are parts of the biophore for both Site 1 and Site 2 on HSA.[29] The plot pf_u versus donor–acceptor pairs is quite scattered (Fig. 4). This means, that this criterion is not enough to discriminate between high, moderate, and low binders; and it is recommended to be used only in combination with other positive criteria. Micafungin has 37 such pairs, which compensate its low log *P*. The last three criteria concern the presence of nonfavorite for PPB structural features: tertiary carbon atoms, four-member rings, and iodine atoms. Five drugs (9%)

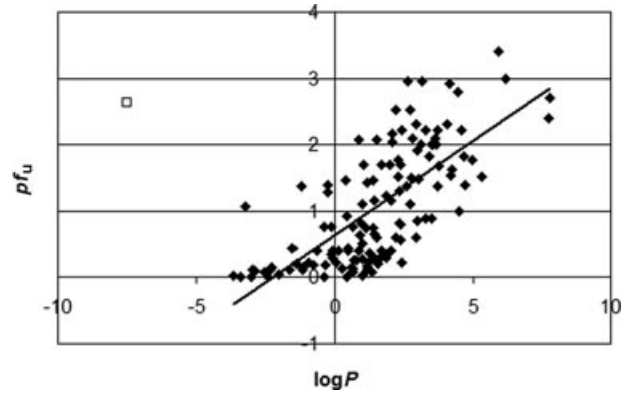

Figure 3. Relationship between pf_u and log P. The outlier micafungin (log $P = -7.48$) is given as an open square. The correlation coefficient is 0.717 (excluding outlier).

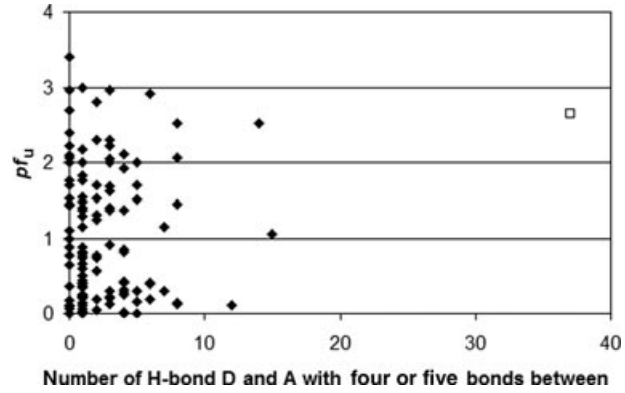

Figure 4. Relationship between pf_u and the number of H-bond donor/acceptor pairs. Micafungin (*SHBint4_acnt*+ *SHBint5_acnt* = 37) is given as an open square. No correlation exists.

among the high binders contain one negative fragment and only two of them (4%) have two negative fragments. The seven-item checklist of criteria derived in the present study could be used as a guide for evaluation of PPB of acidic drug candidates, similarly to the Lipinski's rule of five used for evaluation of oral permeability of drugs.[45]

CONCLUSIONS

The QSPPBR models derived in the present study show that PPB of the acidic drugs depends on several structural features. The lipophilicity, the presence of aromatic rings, cyano groups, and H-bond donor–acceptor pairs increase the PPB. The presence of tertiary carbon atoms, four-member rings, and iodine atoms decrease PPB. The seven-item checklist of criteria could be used as a guide for PPB prediction.

APPENDIX 1

Table A1. The seven-item checklist of criteria for PPB applied to the dataset of acidic drugs.

Table A1. The Seven-Item Checklist of Criteria for Plasma Protein Binding Applied to the Dataset of Acidic Drugs*

			Positive Fragments				Negative Fragments					
Name	pf_u	f_u	Log $P \geq 3$	Aromatic Rings ≥ 2	CN Group	H-Bond Donor–Acceptor ≥ 3	Four-Member Ring	Tertiary C Atom	I Atom	Sum (+) Descriptors	Sum (−) Descriptors	Sum (+) − Sum (−) Descriptors
PPB > 99%												
Verlucast	3.398	0.0004	1	1						2	0	2
Atovaquone	3.000	0.001	1	1						2	0	2
Bromfenac	2.959	0.0011	1	1		1				3	0	3
Tesaglitazar	2.959	0.0011		1						1	0	1
Tolcapone	2.921	0.0012	1	1		1				3	0	3
Diflunisal	2.796	0.0016	1	1						2	0	2
Montelucast	2.699	0.002	1	1				1		2	1	1
Micafungin	2.638	0.0023		1		1				2	0	2
Meloxicam	2.523	0.003				1				1	0	1
Suramin	2.523	0.003		1		1				2	0	2
Telmisartan	2.398	0.004	1	1						2	0	2
Diclofenac	2.301	0.005	1	1						2	0	2
Glimepiride	2.301	0.005				1				1	0	1
Cilomilast	2.222	0.006	1		1			1		2	1	1
Fosinoprilat	2.222	0.006	1			1				2	0	2
Ibuprofen	2.222	0.006	1							1	0	1

(*Continued*)

Table A1. Continued

Name	pf_u	f_u	Positive Fragments: Log $P \geq 3$	Positive Fragments: Aromatic Rings ≥ 2	Positive Fragments: CN Group	Positive Fragments: H-Bond Donor–Acceptor ≥ 3	Negative Fragments: Four-Member Ring	Negative Fragments: Tertiary C Atom	Negative Fragments: I Atom	Sum (+) Descriptors	Sum (−) Descriptors	Sum (+) − Sum (−) Descriptors
Suprofen	2.222	0.006								0	0	0
Ketorolac	2.167	0.0068								0	0	0
Fluvastatin	2.102	0.0079	1	1		1				3	0	3
Ketoprofen	2.097	0.008		1						1	0	1
Tenoxicam	2.071	0.0085				1				1	0	1
Tezosentan	2.071	0.0085		1						1	0	1
Quercetin	2.046	0.009		1		1				2	0	2
Cerivastatin	2.000	0.01	1			1				2	0	2
Chlorambucil	2.000	0.01	1							1	0	1
Indomethacin	2.000	0.01	1	1						2	0	2
Losartan	2.000	0.01	1	1						2	0	2
Torsemide	2.000	0.01	1			1				2	0	2
PPB 95–99%												
Furosemide	1.921	0.012	1			1				2	0	2
Repaglinide	1.824	0.015	1	1						2	0	2
Warfarin	1.824	0.015	1	1						2	0	2
Eprosartan	1.770	0.017	1	1						2	0	2
Sulfinpyrazone	1.770	0.017		1						1	0	1
Entacapone	1.699	0.02			1	1				2	0	2
Glipicide	1.699	0.02								0	0	0
Levosimendan	1.699	0.02			1					1	0	1
Pantoprazole	1.699	0.02								0	0	0
Glyburide	1.678	0.021	1	1		1				3	0	3
Dexloxiglumide	1.620	0.024	1			1				2	0	2
Nateglinide	1.538	0.029	1							1	0	1
Chlorpropamide	1.523	0.03								0	0	0
Epristeride	1.523	0.03	1					1		1	1	0
Bumetadine	1.509	0.031		1		1				2	0	2
Dicloxacillin	1.481	0.033	1				1	1		1	2	−1
Isoxicam	1.456	0.035				1				1	0	1
Milrinone	1.456	0.035			1					1	0	1
Bosentan	1.432	0.037		1				1		1	1	0
Acetazolamide	1.398	0.04				1				1	0	1
Valsartan	1.398	0.04	1	1						2	0	2
Fosfluconazole	1.377	0.042				1		1		1	1	0
Flucloxacillin	1.367	0.043					1	1		0	2	−2
Lovastatin	1.367	0.043	1			1				2	0	2
Tolbutamide	1.301	0.05								0	0	0
PPB 90–95%												
Ceftriaxone	1.284	0.052					1			0	1	−1
Piretanide	1.237	0.058		1						1	0	1
Cefoperazone	1.155	0.07				1	1			1	1	0
Oxacillin	1.155	0.07					1	1		0	2	−2

(*Continued*)

DOI 10.1002/jps

Table A1. Continued

Name	pf_u	f_u	Positive Fragments: Log $P \geq 3$	Positive Fragments: Aromatic Rings ≥ 2	Positive Fragments: CN Group	Positive Fragments: H-Bond Donor–Acceptor ≥ 3	Negative Fragments: Four-Member Ring	Negative Fragments: Tertiary C Atom	Negative Fragments: I Atom	Sum (+) Descriptors	Sum (−) Descriptors	Sum (+) − Sum (−) Descriptors
Sulfisoxazole	1.102	0.079								0	0	0
Valproic acid	1.097	0.08								0	0	0
Folinic acid	1.060	0.087				1				1	0	1
Irbesartan	1.000	0.1	1	1				1		2	1	1
PPB 30–90%												
Rosuvastatin	0.921	0.12				1				1	0	1
Nafcillin	0.886	0.13	1				1	1		1	2	−1
Probenicid	0.886	0.13	1							1	0	1
Thiopental	0.854	0.14				1		1		1	1	0
Cefmetazole	0.824	0.15			1		1	1		1	2	−1
Cefotetan	0.824	0.15				1	1	1		1	2	−1
Perindoprilat	0.796	0.16								0	0	0
Acetylcysteine	0.770	0.17								0	0	0
Betamipron	0.770	0.17								0	0	0
Nitrofurantoin	0.770	0.17								0	0	0
Cefazolin	0.745	0.18					1			0	1	−1
Propylthiouracil	0.745	0.18								0	0	0
Cephalothin	0.658	0.22					1			0	1	−1
Sulfamethoxazole	0.638	0.23								0	0	0
Artesunate	0.602	0.25						1		0	1	−1
Cefamandole	0.602	0.25					1			0	1	−1
Phenethicillin	0.602	0.25					1	1		0	2	−2
Methohexital	0.569	0.27						1		0	1	−1
Cefixime	0.509	0.31					1			0	1	−1
Cefuroxime	0.444	0.36					1			0	1	−1
Methotrexate	0.432	0.37				1				1	0	1
5-Aminosal_acid	0.409	0.39				1				1	0	1
Cefadroxil	0.409	0.39				1	1			1	1	0
Moxalactam	0.409	0.39					1	1		0	2	−2
Pentobarbital	0.409	0.39						1		0	1	−1
Aztreonam	0.398	0.4					1	1		0	2	−2
Cefatrizine	0.398	0.4					1			0	1	−1
Penicillin_G	0.398	0.4					1	1		0	2	−2
Methicillin	0.367	0.43				1	1	1		1	2	−1
Sulfadiazine	0.357	0.44				1				1	0	1
Phenoxymethylpenicillin	0.347	0.45				1	1	1		1	2	−1
Phenobarbital	0.310	0.49				1		1		1	1	0
Piperacillin	0.301	0.5				1	1	1		1	2	−1
Pravastatin	0.301	0.5								0	0	0
Sulbenicillin	0.301	0.5				1	1	1		1	2	−1
Carbenicillin	0.284	0.52					1	1		0	2	−2
Cephapirn	0.260	0.55					1			0	1	−1
Ticarcillin	0.260	0.55					1	1		0	2	−2

(*Continued*)

DOI 10.1002/jps

Table A1. Continued

Name	pf_u	f_u	Positive Fragments				Negative Fragments			Sum (+) Descriptors	Sum (−) Descriptors	Sum (+) − Sum (−) Descriptors
			Log P ≥3	Aromatic Rings ≥2	CN Group	H-Bond Donor–Acceptor ≥3	Four-Member Ring	Tertiary C Atom	I Atom			
Cefprozil	0.229	0.59				1	1			1	1	0
Cefotaxime	0.222	0.6				1	1			1	1	0
Cilastatin	0.222	0.6						1		0	1	−1
Doxifluridine	0.215	0.61								0	0	0
Sulbactam	0.208	0.62				1	1	1		1	2	−1
Enalaprilat	0.208	0.62								0	0	0
5-Fluorouracil	0.194	0.64				1				1	0	1
Azlocillin	0.187	0.65					1	1		0	2	−2
Acetyl_sal_acid	0.167	0.68								0	0	0
Mezlocillin	0.155	0.70					1	1		0	2	−2
PPB <30%												
Carumonam	0.143	0.72				1	1			1	1	0
Ceftizoxime	0.143	0.72					1			0	1	−1
Captopril	0.137	0.73								0	0	0
Cefoxitin	0.137	0.73					1	1		0	2	−2
Fludarabine	0.119	0.762				1				1	0	1
Risedronate	0.119	0.762				1		1		1	1	0
Cefepime	0.108	0.78					1			0	1	−1
Cefetamet	0.108	0.78					1			0	1	−1
Zoledronate	0.108	0.78				1		1		1	1	0
Ceftazidime	0.102	0.79					1	1		0	2	−2
Cephaloridine	0.097	0.8					1			0	1	−1
Amoxicillin	0.071	0.85					1	1		0	2	−2
Cephalexin	0.071	0.85					1			0	1	−1
Foscarnet	0.071	0.85								0	0	0
Ampicilin	0.071	0.85					1	1		0	2	−2
Clavulanic acid	0.041	0.91					1			0	1	−1
Tenofovir	0.032	0.93								0	0	0
Cephradine	0.022	0.95					1			0	1	−1
Iothalamic acid	0.009	0.98				1			1	1	1	0
Amidotrizoate	0.000	1.00				1			1	1	1	0
Cidofovir	0.000	1.00				1				1	0	1
Flucytosine	0.000	1.00								0	0	0
Fosfomycin	0.000	1.00								0	0	0

*The presence of a fragment is denoted by 1.

DOI 10.1002/jps

REFERENCES

1. Kennedy T. 1997. Managing the drug discovery/development interface. Drug Disc Today 2:436–444.
2. Kola I, Landis J. 2004. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3:711–716.
3. Van de Waterbeemd H, Gifford E. 2000. ADMET in silico modeling: Towards prediction paradise? Nat Rev Drug Discov 2:192–204.
4. Chohan KK, Pain SW, Waters, NJ. 2008. Advancements in predictive in silico models for ADME. Curr Chem Biol 2:215–228.
5. Wang J, Hou T. 2009.Recent advances on in silico ADME modeling. Vol. 5. In Annual reports in computational chemistry; Elsevier, Amsterdam, San Diego pp 102–127.
6. Schmidt S, Gonzales D, Derendorf H. 2010. Significance of protein binding in pharmacokinetics and pharmacodynamics. J Pharm Sci 99:1107–1122.
7. Mehvar R. 2005. Role of protein binding in pharmacokinetics. Am J Pharm Edu 69:1–8.
8. Kerns E, Di L. 2008. Drug-like properties: Structures, design and methods. 1st ed. Elsevier, Burlington, San Diego, London pp. 187–196.
9. Mao H, Hajduk PJ, Craig R, Bell R, Bore T, Fesik SW. 2001. Rational design of diflunisal analogues with reduced affinity to human serum albumin. J Am Chem Soc 123:10429–10435.
10. Hall LM, Hall LB, Kier LB. 2003. QSAR modeling of beta-lactam binding to human serum proteins. J Comput Aided Mol Des 17:103–118.
11. Xue CH, Zhang RS, Liu HX, Yao XJ, Liu MC, Hu ZD, Fan BT. 2004. QSAR model for the prediction of binding affinities to human serum albumin using the heuristic method and support vector machine. J Chem Inf Comput Sci 44:1693–1700.
12. Ramamurthi N, Gunturi SB. 2005. In silico ADME modeling: QSPR models for the binding of β-lactams to human serum proteins using genetic algorithms. ARKOVOC XI:102–123.
13. Seedher N, Bhatia S, Singh B. 2008. Quantitative correlation between theoretical molecular descriptors and drug-HSA binding affinities for various cox-2 inhibitors. Chem Biol Drug Des 72:297–302.
14. Colmenarejo G, Alvarez-Pedraglio A, Lavandera J-L. 2001. Cheminformatic models to predict affinities to human serum albumin. J Med Chem 44:4370–4378.
15. Kratochwil AN, Huber W, Mueller F, Kansy M, Gerber PR. 2002. Predicting plasma protein binding of drugs—A new approach. Biochem Pharmacol 64:1355–1374.
16. Hall LM, Hall LH, Kier LB. 2003. Modeling drug-albumin binding affinity with E-state topological structure representation. J Chem Inf Comoput Sci 43:2120–2128.
17. Liu J, Yang L, Li Y, Pan D, Hopfinger AJ. 2005. Prediction of plasma protein binding of drugs using Kier–Hall valence connectivity indices and 4D-fingerprint molecular similarity analyses. J Comput Aided Mol Des 19:567–583.
18. Liu J, Yang L, Li Y, Pan D, Hopfinger AJ. 2006. Constructing plasma protein binding model on a combination of cluster analysis and 4D-fingerprint molecular similarity analyses. Bioorg MedChem 14:611–621.
19. Gunturi SB, Ramamurthi N, Khandelwal A. 2006. In silico ADME modeling. Computational models to predict human serum albumin binding affinity using ant colony systems. Bioorg Med Chem 14:4118–4129.
20. Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S, Tropsha A. 2006. QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation. J Med Chem 49:7169–7181.
21. Estrada E, Uriarte E, Molina E, Simon-Manso Y, Milne GW. 2006. An integrated in silico analysis of drug binding to human serum albumin. J Chem Inf Model 46:2709–2724.
22. Deeb O, Hemmateenehad B. 2007. ANN-QSAR model of drug binding to human serum albumin. Chem Biol Drug Des 70:19–29.
23. Deeb O. 2010. Correlation ranking and stepwise regression procedures in principal components artificial neutral networks modeling with application to predict toxic activity and human serum albumin binding affinity. Chem Intell Lab Sys 104:181–194.
24. Chen L, Chen X. 2012. Results of molecular docking as descriptors to predict human serum albumin binding affinity. J Mol Graph Model 33:35–43.
25. Zsila F, Bikadi Z, Malik D, Hari P, Pechan I, Berces A, Hazai E. 2011. Evaluation of drug-human serum albumin binding interactions with support vector machine aided online automated docking. Bioinformatics 27:1806–1813.
26. Meletiadis J, Chanock S, Walsch T. 2006. Human pharmacogenomic variations and their implication for antifungal efficacy. Clin Microbiol Rev 19:763–787.
27. Martinez-Gomez MA, Sagrado S, Villanueva-Camanas RM, Medina-Hernandez MJ. 2006. Characterization of basic drug-human serum albumin interactions by capillary electrophoresis. Electrophoresis 27:3410–3419.
28. Wasan KM, Brocks DR, Lee SD, Sachs-Barrable K, Thornton SJ. 2008. Impact of lipoproteins on the biological activity and disposition of hydrophobic drugs: Implications for drug discovery. Nat Rev Drug Discov 7:84–99.
29. Ghuman J, Zunzain PA, Petipas I, Bhattacharya AA, Otagiri M, Curry S. 2005. Structural basis of the drug-binding specificity of human serum albumin. J Mol Biol 353:38–52.
30. Otagiri M, Yamamichi R, Maruyama T, Imai T, Suenaga A, Imamura Y, Kimachi K. 1989. Drug binding to alpha-1 acid glycoprotein studied by circular dichroism. Pharm Res 6:156–159.
31. Israili ZH, Dayton PG. 2001. Human alpha-1-glycoprotein and its interactions with drugs. Drug Metab Rev 33:161–235.
32. Yamazaki K, Kanaoka M. 2004. Computational prediction of the plasma protein binding percent of diverse pharmaceutical compounds. J Pharm Sci 93:1480–1494.
33. Obach RS, Lombardo F, Waters NJ. 2008. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab Dispos 36:1385–1405.
34. Drug Bank Accessed June 18, 2012, at: http://www.drugbank.ca.
35. Chemical Book Accessed June 18, 2012, at: http://www.chemical book.com.
36. Hall LH, Kier LB. 1999. Molecular connectivity chi indices for database analysis and structure–property modeling. In Topological indices and related descriptors in QSAR and QSPR; Devillers J, Balaban A, Eds. London: Gordon and Breach, pp 307–360.
37. Hall LH, Kier LB. 1999. Electrotopological state: Structure modeling for QSAR and database analysis. In Topological indices and related descriptors in QSAR and QSPR; Devillers J, Balaban A, Eds. London: Gordon and Breach, pp 491–562.
38. Leardi R, Boggia R, Terrile M. 1992. Genetic algorithms as a strategy for feature selection. J Chemom 6:267–281.
39. Friedman J. 1988. Multivariate adaptive regression splines. Technical report No. 102. Stanford University, Stanford, California: Laboratory for Computational Statistics, Department of Statistics.
40. He XM, Carter DC. 1992. Atomic structure and chemistry of human serum albumin. Nature 358:209–215.
41. Petitpas I, Bhattacharya AA, Twine S, East M, Curry S. 2001. Crystal structure analysis of warfarin binding to human serum albumin. J Biol Chem 276:22804–22809.

42. Sudlow G, Birkett DJ, Wade DH. 1975. Characterization of two specific drug binding sites on human serum albumin. Mol Pharmacol 12:1052–1061.
43. Hajdik PJ, Mendoza R, Petros AM, Huth JR, Bures M, Fesik W, Martin YC. 2003. Ligand binding to domain-3 of human serum albumin: A chemometric analysis. J Comput Aided Mol Des 17:93–102.
44. Kohlhepp SJ, Gilbert DN, Leggett JE. 1988. Influence of assay methodology on the measurement of free serum ceftriaxone concentration. Antimicrob Agents Chemother 42:2259–2261.
45. Lipinski CA, Lombardo F, Dominy BW, Feeney PG. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Del Rev 46:3–26.