*Journal of Computer-Aided Molecular Design* **17:** 93–102, 2003.



# Ligand binding to domain-3 of human serum albumin: a chemometric analysis

Philip J. Hajduk*, Renaldo Mendoza, Andrew M. Petros, Jeffrey R. Huth, Mark Bures, Stephen W. Fesik & Yvonne C. Martin
*Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, IL 60044*

Received 13 November 2002; Accepted 1 December 2002

*Key words:* Albumin binding, human serum albumin, ligand binding, NMR, chemometric analysis

**Summary**

A detailed chemometric analysis of ligand binding to domain-3A of human serum albumin is described. NMR and fluorescence data on a set of 889 chemically diverse compounds were used to develop a group contribution model based on 74 chemical fragments that is in good agreement with the experimental data ($R^2 = 0.94$, $Q^2 = 0.90$). The structural descriptors used in this analysis comprise a convenient look-up table for quantitatively estimating the effect that a particular group will have on albumin binding. This information can be valuable for optimizing a particular series of compounds for drug development.

## Introduction

Human serum albumin (HSA) is present at high concentrations (approximately 0.6 mM) in serum plasma and is capable of binding to a diverse range of organic molecules, such as fatty acids, metabolites, and drugs [1, 2]. The binding of drugs to albumin can significantly alter their overall activity profile, including *in vivo* distribution, excretion, and toxicity [3]. In general, drugs with high affinity for serum albumin have markedly reduced efficacy *in vivo*, and, therefore, great effort is often exerted to reduce the albumin binding of potential drug candidates while maintaining high affinity for the therapeutic target.

HSA is comprised of three homologous domains, each of which consists of two sub-domains (A and B) that share common structural elements. There are at least two high-affinity drug-binding sites on HSA: the warfarin site (located in domain-2) and the diazepam site, which is located in subsite A on domain-3 (domain-3A or HSA-3A) [2, 4]. The diazepam site is of particular pharmaceutical interest as it has high affinity for small, anionic aromatic compounds that often represent pharmaceutical agents. We have recently demonstrated that compounds with decreased affinity for HSA-3A can be rationally designed using structural information combined with knowledge of the structure-affinity relationships of a lead series to domain-3A of albumin [5]. However, it is not always possible to obtain high-resolution structural information on albumin/drug complexes using NMR or X-ray crystallography. In fact, the only published crystal structures of albumin contain either fatty acids [6], small anesthetics [7], or warfarin [8]. In addition, as there are up to 8 discrete binding sites on HSA for these molecules, the available structural information is of limited value in the rational design of drugs with decreased binding to albumin. Therefore, strategies for reducing drug binding to albumin that do not require high-resolution structural information would be of great utility in the pharmaceutical industry.

In the absence of structural information, medicinal chemists have successfully utilized ligand-based methods for drug design. These strategies rely on the premise that the biological properties of a compound are a function of its physiochemical properties, such as solubility, lipophilicity, polarizability, hydrogen bonding, and steric parameters [9, 10]. As many of these

*To whom correspondence should be addressed. E-mail: philip.hajduk@ abbott.com

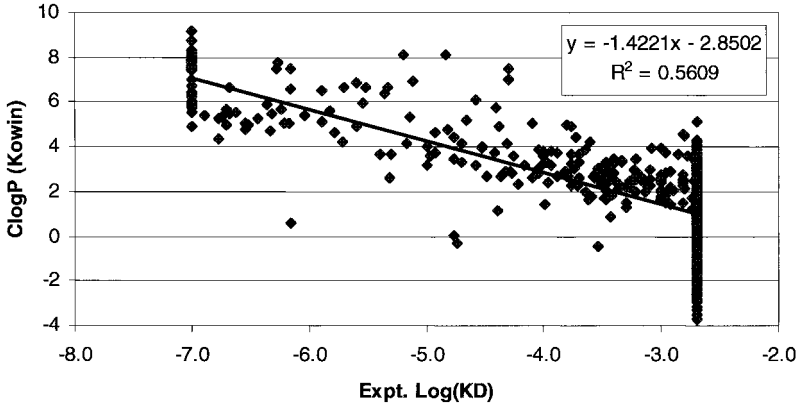


*Figure 1.* Correlation between experimental binding affinities for HSA-3A and (Expt. Log($K_D$)) and Kowin ClogP for the test set of 889 compounds.

properties can be calculated or directly measured, computer models for quantitative structure-activity relationships (QSAR) can be constructed using a number of statistical and machine-learning algorithms. A number of strategies have been described for constructing QSAR models, including the use of bulk or substituent molecular properties (e.g., ClogP, hydrogen bond donors/acceptors, etc.) [10], group contribution models [11], three-dimensional comparative molecular field analyses (CoMFA) [12], and comparative molecular similarity index analysis (CoMSIA) [13]. These models can then be used to predict the activity of proposed molecules before chemistry is initiated. However, the molecular interpretation of these models is critically dependent on the requirement that all of the compounds used in the analysis bind to the same site on the target. This type of data is typically not available for drug binding to albumin. Dissociation constants derived from total protein binding yield no information on individual binding sites, and those obtained from competitive binding experiments are often difficult to interpret.

Here we describe a detailed chemometric analysis of ligand binding to domain-3A of human serum albumin. Binding data on a collection of 889 diverse molecules was obtained using heteronuclear NMR correlation and fluorescence spectroscopy, which enabled the measurement of dissociation constants to domain-3A of HSA. The ability to specifically detect binding to domain-3A allowed a detailed analysis of binding to this critical site on human serum albumin.

## Results and discussion

### *Compound collection*

A total of 1826 compounds were tested for their ability to bind to domain-3A of human serum albumin using NMR and fluorescence spectroscopy. The compounds were chosen to maximize diversity in both structure and physicochemical properties. For construction of the QSAR model, compounds were discarded that were not exhaustively described by the functional groups shown in Tables 2 and 3 (see below) or for which the binding data was ambiguous (see Methods). This process resulted in a final set of 889 compounds that was used for development of the model. The structural diversity of the final compound library was estimated based on Tanimoto coefficients, which indicated that 50% of the compounds in the library are less than 74% similar to any other compound. The structural diversity was further examined using Taylor's clustering algorithm [14], in which at Tanimoto similarities of $\geq$0.74 and $\geq$0.85, the dataset is represented by 562 and 734 clusters, respectively. As shown in Table 1, the compounds also cover wide ranges in various physicochemical properties, such as molecular weight, ClogP, molar refractivity, and polar and non-polar surface area. In addition, the binding affinities vary over 4 orders of magnitude, from $<$0.1 to $>$2000 $\mu$M.

### *QSAR model for HSA-3A binding*

Initial QSAR models were constructed using physicochemical properties. In fact, a simple correlation between ClogP and the observed binding affinity yielded an $R^2$ of 0.56 (see Figure 1). This trend is not

*Table 1.* Physicochemical parameters of the compounds used in the QSAR analysis.

| Parameter[a] | Average | Median | Minimum | Maximum |
|---|---|---|---|---|
| Molecular weight (Da) | 223 | 192 | 31 | 749 |
| ClogP | 1.6 | 1.3 | −8.1 | 9.9 |
| Molar refractivity | 6.0 | 5.3 | 1.0 | 19.6 |
| Log(Solubility) (mg/mL) | 0.17 | 0.65 | −8.9 | 3 |
| Polar surface area (Å$^2$) | 65.9 | 58.4 | 2.8 | 197.2 |
| Nonpolar surface area (Å$^2$) | 217.3 | 195.7 | 9.0 | 666.8 |

[a]All parameters were calculated using SPARK [16,32].

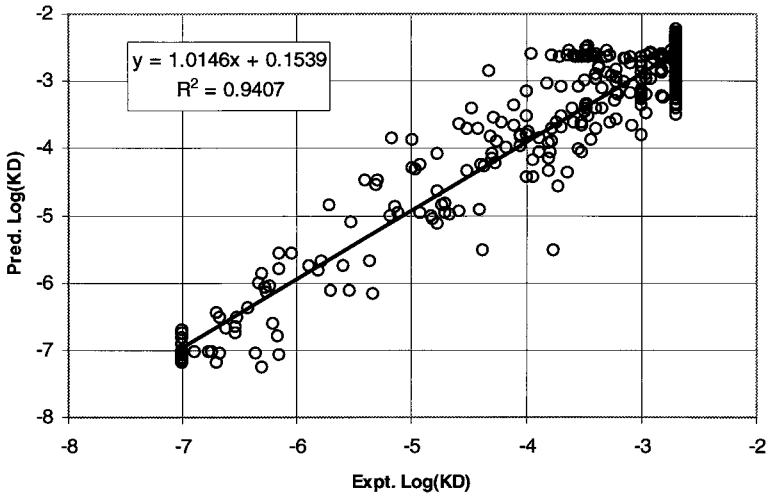


*Figure 2.* Correlation between experimental (Expt. Log($K_D$)) and predicted (Pred. Log($K_D$)) binding affinities for HSA-3A using the group contributions model as describedin the text.

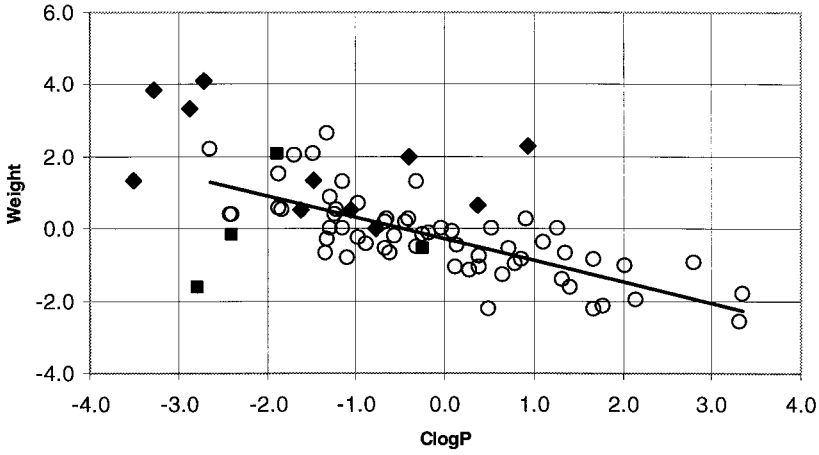


*Figure 3.* Correlation between the ClogP and weighting coefficients of the structural descriptors used in the group contributions model (Tables 2 and 3). Negatively and positively charged fragments are shown with filled squares and diamonds, respectively, while all other fragments are shown with open circles.

unexpected, as lipophilic compounds tend to have high affinity for albumin [15]. However, while this initial correlation was encouraging for the development of a robust QSAR model on the basis of physicochemical properties alone, attempts to improve the model using additional parameters were unsuccessful. Standard properties such as molecular weight [16], ClogP [17], molar refractivity [18], solubility [19, 20], and polar and non-polar surface area [21] were investigated, as well as a number of constitutional [22], topological [23] and charge [24] descriptors, aromaticity indices [25] and WHIM descriptors [26] available using the program DRAGON [27].

A significantly improved QSAR model for ligand binding to domain-3A of albumin was obtained using the group contributions method [11]. In this approach, the compounds were described using structural descriptors or functional groups that each contribute to binding to HSA-3A. Using a set of 74 structural descriptors to describe the 889 compounds in the test set (Tables 2 and 3), a group contribution model with $R^2 = 0.94$ was obtained (Figure 2). The predictive power of this model was validated using the leave-several-out (LSO) method (yielding a cross-validated $Q^2$ of 0.90) and randomization of the binding affinities. The model is remarkably robust, with an average error in the predicted $\text{Log}(K_D)$ of 0.11. The information content of the compounds that do not bind is shown by the fact that the average error increases to 0.36 if only the 232 compounds with $K_D$ values less than 2.0 mM are considered.

The weighting coefficients ($w$) given in Tables 2 and 3 are the contributions (in Log units) that each group makes to HSA-3A binding. Since $\text{Log}(K_D)$ was used, positive values for the weighting coefficients decrease binding to domain 3A (increase $K_D$), while negative values increase binding (decrease $K_D$). Qualitatively, the weighting coefficients are in excellent agreement with what is already widely known about compound binding to serum albumin. For example, cyclic or acyclic amines, that would be positively charged at physiological pH, have positive weighting coefficients (decreased binding to albumin), while negatively charged groups, such as carboxylic and sulfonic acids, have negative weighting coefficients (increased binding to albumin). Also, as shown in Figure 3, there is a general correspondence between the hydrophobicity of the fragment and the weighting coefficient, with more hydrophobic groups tending to increase binding to albumin. However, there are significant and unexpected exceptions to this trend. For example, the benzimidazole (**27**) and allyl (**32**) groups both have a positive contribution to log P but have little effect or even impart decreased binding to HSA-3A. Pyrazole (**56**), triazole (**52**), and imidazole (**49**) all have negative contribution to log P but impart increased binding to domain-3A of albumin. Furthermore, compounds that have similar contributions to log P can have significantly different weighting coefficients. For example, the contributions to log P for carbamate (**4**) and urea (**16**) are essentially identical, but their weighting coefficients differ by more than 1.7 log units. Similarly, the sulfone (**6**) and ketone (**26**) groups have very similar contributions to log P, but differ in their weighting coefficients by more than 1.8 log units.

To reliably calculate an absolute binding affinity, this group contribution QSAR model requires that the compound of interest contain only those fragments that are listed in Tables 2 and 3. Thus, there are many compounds for which absolute binding affinities cannot be predicted by this model. However, even for these compounds, the data contained in Tables 2 and 3 can still be extremely useful in assessing the expected relative change in binding affinity upon modification or substitution. For example, the incorporation of a primary amine (**3**), with a weighting coefficient of 3.33, would be expected to have a significantly larger effect (more than 40-fold) on albumin binding than a pyrrolidine (**18**), which has a weighting coefficient of 0.65. In addition, a decrease in albumin binding of nearly 0.7 log units would be expected by changing a phenyl group (**70**, $w = -1.97$) to a pyridine (**65**, $w = -1.28$), and an additional 0.8 log unit decrease in affinity would be expected upon conversion to a pyrimidine (**48**, $w = -0.49$). The same trend holds true for bicyclic functional groups, with indole (**73**, $w = -2.24$) $\ll$ benzimidazole (**27**, $w = 0.29$) $<$ adenine (**23**, $w = 0.53$) and naphthyl (**74**, $w = -2.57$) $<$ quinoline (**61**, $w = -1.03$). Thus, the descriptors contained in Tables 2 and 3 can be used to quantitatively estimate changes in albumin binding upon chemical modification.

*Comparison to full-length albumin*

A chemometric model to describe ligand binding to full-length albumin has recently been reported [15]. In this study, binding affinities for full-length albumin were derived from affinity chromatography for 95 compounds and used to develop QSAR models based on electronic, topological, spatial, structural, and ther-

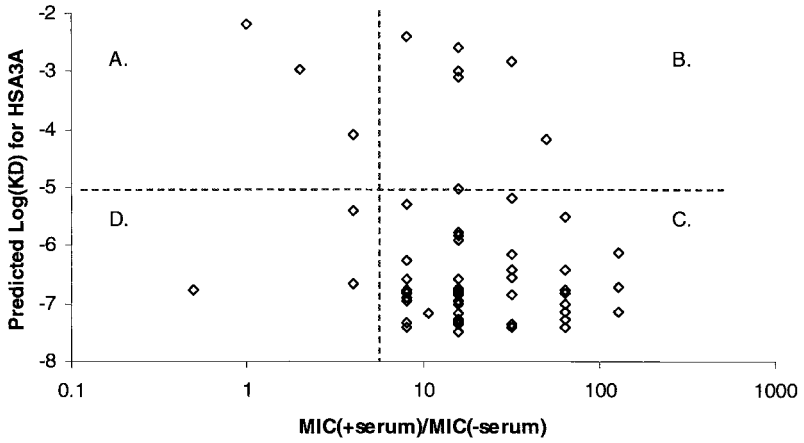


*Figure 4.* Plot of predicted binding affinities for HSA-3A (Pred. Log($K_D$)) using the group contributions model versus fold-reduction in MIC in the presence and absence of serum for a set of 60 compounds. The significant increase in MIC for those compounds that are placed in quadrant C (MIC(+serum)/MIC(-serum) > 8) is expected to be due to high affinity for domain-3A of albumin, while the increase in MIC for those compounds that are placed in quadrant B is expected to be due to binding sites on albumin other than domain-3A or other serum components.

modynamic descriptors. Interestingly, the model did not discriminate between the various binding sites on albumin but instead used the apparent binding affinities for the full-length protein. Thus, a combination of both the full-length and domain-3A QSAR models should allow for a discrimination between compounds that bind to domain-3A of albumin and those that bind elsewhere on the protein. In order to demonstrate the utility of this approach, we predicted and experimentally measured binding affinities to HSA-3A for a subset of the compounds used in the analysis of full-length albumin, and the results are shown in Table 4. For several compounds (e.g., naproxen, clotrimazole, and ketoprofen), the affinity for full-length albumin can be accounted for by the predicted (and observed) affinity for HSA-3A. However, several compounds are predicted to bind with low affinity to domain-3A of albumin, and yet exhibit at least moderate affinity for full-length albumin. For example, norfloxacin, phenytoin, and pindolol all bind to full-length albumin, and yet are predicted to have $K_D$ values for HSA-3A in excess of 2000 μM (the lack of affinity of these compounds for HSA-3A was confirmed experimentally, see Table 4). This result indicates that norfloxacin, phenytoin, and pindolol bind to sites on full-length albumin other than domain-3A. Given the structural similarity between norfloxacin and warfarin, it would not be surprising that the observed affinity of norfloxacin for albumin is due primarily to binding at the warfarin binding site. This type of information can be valuable in the rational design of compounds with reduced affinity for albumin.

*Analysis of serum effects*

In many cases, the first indication of an albumin binding problem with a lead series is a significant reduction in *in vitro* potency in the presence serum. Traditional medicinal chemistry approaches to reduce albumin binding involve the incorporation of polar or positively charged groups at various positions in the lead compound [28]. This process can be facilitated when the serum component responsible for the potency reduction can be identified. In fact, we have recently described a structure-based approach for reducing compound binding to domain-3 of albumin [5] The current chemometric model for binding to HSA-3A can be used in conjunction with serum potency reductions to identify those compounds whose efficacy is modulated primarily through binding to domain-3A. This is demonstrated in Figure 4 for a set of 60 compounds with measured antibacterial activity in the presence and absence of serum. When predicted HSA-3A affinities are plotted against fold-reduction in MIC in the presence of serum, it is observed that the majority of compounds (~75%) that exhibit a >8-fold loss in activity in the presence of serum are also predicted to bind tightly ($K_D < 10^{-5}$M) to domain-3 of albumin (see Figure 4, Quadrant C). Thus, these compounds are candidates for structure-based approaches to reduce binding to HSA-3A, or for incorporating those groups shown in Table 2 that are predicted to have the largest effect on domain-3 binding. Five compounds in this set exhibit a >8-fold loss in activity but would not be expected to bind tightly to domain-3 of albumin

*Table 2.* Structural descriptors that decrease binding to HSA-3A.

| No. | Name | weight[a] | error[b] | #tests[c] | ClogP[d] |
|---|---|---|---|---|---|
| 1 | Secondary amine | 4.12 | 0.12 | 10 | −2.72 |
| 2 | Tertiary amine | 3.85 | 0.02 | 18 | −3.28 |
| 3 | Primary amine | 3.33 | 0.05 | 42 | −1.76 |
| 4 | Carbamate | 2.65 | 0.17 | 24 | −1.32 |
| 5 | Piperidine (1 IN)[e] | 2.28 | 0.02 | 52 | 0.94 |
| 6 | Sulfone | 2.22 | 0.04 | 10 | −2.64 |
| 7 | Amide | 2.09 | 0.03 | 137 | −1.49 |
| 8 | Hydroxamate | 2.07 | 0.04 | 15 | −1.89 |
| 9 | Hydantoin | 2.04 | 1.01 | 36 | −1.69 |
| 10 | Morpholine (1 IN) | 2.00 | 0.06 | 43 | −0.41 |
| 11 | Hydrazide | 1.52 | 0.17 | 18 | −1.88 |
| 12 | Piperazine (2 IN) | 1.35 | 0.06 | 17 | −1.48 |
| 13 | Quaternary amine | 1.33 | 0.22 | 13 | −3.50 |
| 14 | Imine | 1.29 | 0.15 | 7 | −1.15 |
| 15 | 1,3-Dioxane | 1.28 | 0.06 | 18 | −0.32 |
| 16 | Urea | 0.89 | 0.07 | 44 | −1.30 |
| 17 | Formaldehyde | 0.69 | 0.32 | 9 | −0.98 |
| 18 | Pyrollidine (1 IN) | 0.65 | 0.24 | 26 | 0.38 |
| 19 | Alcohol | 0.58 | 0.07 | 128 | −1.86 |
| 20 | Guanidine | 0.53 | 0.06 | 20 | −1.63 |
| 21 | Sulfonamide | 0.53 | 0.04 | 19 | −1.84 |
| 22 | Piperazine (1 IN) | 0.53 | 0.03 | 20 | −1.06 |
| 23 | Adenine | 0.53 | 0.08 | 11 | −1.21 |
| 24 | Aniline | 0.42 | 0.02 | 240 | −1.23 |
| 25 | Ether | 0.41 | 0.05 | 23 | −2.39 |
| 26 | Ketone | 0.39 | 0.03 | 53 | −2.41 |
| 27 | Benzimidazole | 0.29 | 0.06 | 29 | 0.93 |
| 28 | Piperidinone | 0.26 | 0.02 | 8 | −0.41 |
| 29 | Aldehyde | 0.26 | 0.11 | 10 | −0.65 |
| 30 | Thioether | 0.18 | 0.04 | 82 | −0.44 |
| 31 | Phenol | 0.17 | 0.01 | 200 | −0.67 |
| 32 | Allyl | 0.00 | 0.15 | 25 | 1.27 |
| 33 | Ester | 0.00 | 0.13 | 68 | −0.03 |
| 34 | Furanose | 0.00 | 0.46 | 11 | 0.53 |
| 35 | Oxime | 0.00 | 0.06 | 8 | −1.29 |
| 36 | Pyrimidinone | 0.00 | 1.00 | 38 | −1.15 |
| 37 | Pyrroline | 0.00 | 0.11 | 8 | −0.77 |

[a]The weighting coefficient determined using the group contribution method (See text).
[b]The standard deviation in the weighting coefficient calculated from the results of 10 convergent minimzations.
[c]The number of occurrences of the substructure in the test set of 889 compounds.
[d]Calculated ClogP for the fragment using Biobyte as described in the Methods section.
[e]*IN* refers to the number of ionizable nitrogens.

*Table 3.* Structural descriptors that increase binding to HSA-3A.

| No. | Name | weight[a] | error[b] | #tests[c] | ClogP[d] |
|---|---|---|---|---|---|
| 38 | Methylene | −0.10 | 0.01 | 731 | 0.08 |
| 39 | Morpholine (0 IN)[e] | −0.12 | 0.03 | 23 | −0.18 |
| 40 | Nitro | −0.16 | 0.03 | 120 | −0.26 |
| 41 | Acylsulfonamide | −0.18 | 0.08 | 69 | −2.40 |
| 42 | Cyano | −0.22 | 0.12 | 36 | −0.57 |
| 43 | Pyrrolidinone | −0.24 | 0.03 | 26 | −0.97 |
| 44 | Quaternary carbon | −0.31 | 0.08 | 33 | −1.32 |
| 45 | Methyl | −0.37 | 0.01 | 769 | 1.10 |
| 46 | Hydrazine | −0.42 | 0.08 | 18 | −0.88 |
| 47 | Fluorine | −0.45 | 0.03 | 46 | 0.15 |
| 48 | Pyrimidine | −0.49 | 0.03 | 37 | −0.31 |
| 49 | Imidazole | −0.53 | 0.03 | 31 | −0.67 |
| 50 | Carboxylic acid | −0.54 | 0.02 | 130 | −0.26 |
| 51 | Chlorine | −0.57 | 0.04 | 62 | 0.72 |
| 52 | Triazole | −0.69 | 0.02 | 14 | −1.34 |
| 53 | Methine | −0.70 | 0.03 | 90 | −0.62 |
| 54 | Piperidine (0 IN) | −0.70 | 0.05 | 23 | 1.36 |
| 55 | Alkyne | −0.79 | 0.12 | 28 | 0.39 |
| 56 | Pyrazole | −0.83 | 0.02 | 47 | −1.09 |
| 57 | Cyclopropane | −0.86 | 0.27 | 15 | 1.68 |
| 58 | Bromine | −0.86 | 0.02 | 24 | 0.87 |
| 59 | Cyclopentane | −0.92 | 0.02 | 8 | 2.80 |
| 60 | Pyrollidine (0 IN) | −0.97 | 0.03 | 18 | 0.80 |
| 61 | Quinoline | −1.03 | 0.02 | 12 | 2.03 |
| 62 | Thiol | −1.07 | 0.01 | 20 | 0.40 |
| 63 | Isoxazole | −1.09 | 0.03 | 10 | 0.12 |
| 64 | Pyrrole | −1.15 | 0.06 | 8 | 0.29 |
| 65 | Pyridine | −1.28 | 0.04 | 70 | 0.65 |
| 66 | Furan | −1.41 | 0.07 | 49 | 1.32 |
| 67 | Benzoxazole | −1.61 | 0.14 | 11 | 1.42 |
| 68 | Sulfonate | −1.63 | 0.01 | 9 | −2.79 |
| 69 | Cyclohexane | −1.79 | 0.11 | 45 | 3.35 |
| 70 | Phenyl | −1.97 | 0.02 | 678 | 2.14 |
| 71 | Thiophene | −2.13 | 0.04 | 37 | 1.79 |
| 72 | Thiazole | −2.23 | 0.02 | 20 | 0.49 |
| 73 | Indole | −2.24 | 0.10 | 18 | 1.67 |
| 74 | Naphthalene | −2.57 | 0.01 | 16 | 3.32 |

[a]The weighting coefficient determined using the group contribution method (See text).
[b]The standard deviation in the weighting coefficient calculated from the results of 10 convergent minimzations.
[c]The number of occurrences of the substructure in the test set of 889 compounds.
[d]Calculated ClogP for the fragment using Biobyte as described in the Methods Section.
[e]*IN* refers to the number of ionizable nitrogens.

(predicted $K_D > 10^{-4}$ M) (see Figure 4, Quadrant B). The lack of affinity of these compounds for HSA-3A was confirmed experimentally by NMR. Thus, it is likely that other sites on albumin or other serum components, such as α-acid glycoprotein [29], are involved in binding to these compounds and reducing their potency in the presence of serum. This can be readily determined experimentally by measuring the activity in the presence of full-length albumin rather than whole serum. Interestingly, one compound does not exhibit a significant reduction in potency (fold-loss in MIC # 1) but would be expected to bind tightly to domain-3A (predicted $K_D < 10^{-7}$ M) (see Figure 4, Quadrant D). Unfortunately, this compound was not available for experimental confirmation of the binding affinity. However, if in fact this compound does bind tightly to albumin as the model predicts, then the lack of affect on MIC indicates a mechanism of action or access to a target site that is not affected by binding to albumin.

## Conclusions

In summary, we have developed a chemometric model for ligand binding to domain-3 of human serum albumin using a group contributions approach. The fragments and corresponding weighting coefficients comprise a convenient look-up table for quantitatively estimating the effect that a particular group will have on albumin binding- without resorting to complex descriptor calculations not generally available to the typical medicinal chemist. In addition, these data allow for a formulation of general chemical rules for reducing compound affinity for HSA-3A. Application of this QSAR model to compounds for which albumin or serum binding is available can give insight into which serum component or albumin binding site is involved in the interaction. This information can be important in the optimization and development of a particular compound series into a clinical candidate.

## Methods

### *Property calculations*

Tanimoto coefficients were calculated as implemented in the Daylight Fingerprint Toolkit (see <www.daylight.com> for references and documentation). Clustering was performed using an algorithm proposed by Taylor [14, 30, 31]. ClogP [17], molar refractivity [18], polar and non-polar surface area, [21], solubility, [19, 20] and molecular weight [16] were calculated within SPARK [16, 32]. Consitutional descriptors, [22] topological indices, [23] charge descriptors, [24] aromaticity indices, [25] and WHIM descriptors [26] were calculated using the program DRAGON [27].

### *NMR binding data*

Isotopically $^{13}$C(methyl)-labeled [33] HSA-3A was obtained as previously described [34]. The purified protein samples were dissolved in 50 mM sodium phosphate buffer (pH 6.5) containing 150 mM NaCl, 0.02% $NaN_3$, and 0.1 mM EDTA with $H_2O/D_2O$ (95/5). NMR spectra were acquired at 303 K on a Bruker DRX500 spectrometer with xyz-shielded gradient triple resonance probes. Bruker sample changers were used on Bruker DRX-500 MHz spectrometers.

Ligand binding was detected by acquiring $^{13}C/^{1}H$-HSQC spectra on 500 μL of a 50 μM protein solution in the presence and absence of added compound. Compounds were added individually as solutions in perdeuterated DMSO and were initially tested at 0.5 mM each. More than 1800 compounds were tested for their ability to bind to HSA-3A, and the compounds were initially assigned a score ranging from 0 (no observed changes in the spectrum upon the addition of compound) to 5 (significant changes in the spectrum). Compounds that were assigned a score of 2 were not used in the analysis since the single point binding data was considered ambiguous. A total of 657 compounds were assigned a score of 0 or 1 and were estimated to have $K_D$ values greater than 2.0 mM. Dissociation constants were obtained for 232 compounds (score > 3) by monitoring the chemical shift changes as a function of ligand concentration. A least squares grid search was performed by varying the values of $K_D$ and the chemical shift of the fully saturated protein. Data were fit using a single binding site model. The active (232) and inactive (657) compounds comprised the final data set of 889 compounds for chemometric analysis.

### *Fluorescence spectroscopy*

In cases where the NMR data indicated a $K_D$ value of 10 μM or less (73 compounds), a fluorescence competition assay was used to determine dissociation constants for compound binding to domain-3A of

*Table 4.* Comparison of drug molecule binding affinities for full-length and domain-3 of human serum albumin.

| Name | Log K'hsa[a] | Log($K_D$-3A) (pred.)[b] | Log($K_D$-3A) (expt.)[c] |
|---|---|---|---|
| Naproxen | 0.25 | −4.66 | −5.00 |
| Clotrimazole | 1.34 | −7.00[d] | −4.92 |
| Ketoprofen | 0.03 | −5.59 | −4.74 |
| Chlorpropamide | −0.44 | −2.91 | −4.14 |
| Tolbutamide | −0.22 | −3.06 | −4.09 |
| Indomethacin | 0.47 | −3.74 | −3.93 |
| Clofibrate | 0.27 | −3.82 | −3.70 |
| Salicylicacid | −0.66 | −2.70[e] | −3.58 |
| Acetylsalicylicacid | −1.39 | −2.70 | −3.52 |
| Lidocaine | −0.23 | −2.70 | −2.70[f] |
| Pindolol | −0.13 | −2.70 | −2.70 |
| Atenolol | −0.48 | −2.70 | −2.70 |
| Captopril | −2.69 | −2.70 | −2.70 |
| Norfloxacin | 0.14 | −2.70 | −2.70 |
| Phenytoin | 0 | −2.70 | −2.70 |
| Trimethoprim | −0.26 | −2.70 | −2.70 |

[a]Taken from Ref. 15. More positive values indicate longer retention times on an HSA affinity column and thus higher affinity for albumin.
[b]Predicted binding affinity (in log units) for domain-3A of albumin using group contributions model as described in the text.
[c]Experimental binding affinity (in log units) for domain-3A of albumin determined using NMR.
[d]Lower limit of group contributions model.
[e]Upper limit of group contributions model.
[f]Upper limit of NMR estimate.

albumin. Fluorescence intensity measurements were carried out using a SLM8000 fluorimeter and dansyl-sarcosine as the probe [35]. The dissociation constant of dansyl-sarcosine for HSA-3A was found to be 6 μM. Titrations were carried out in a buffer containing 40 mM sodium phosphate (pH 7.2). The protein concentration was 6 μM with the probe concentration at 0.5 μM. Compound of concentrations over the range of 1–50 μM were used. Dissociation constants were determined from titration curves with in-house written software using the analytical expressions of Wang [36]. Under these experimental conditions, it was estimated that $K_D$ values as low as 0.10 μM could be reliably obtained. Calculated $K_D$ values below 0.1 μM were set to 0.1 μM.

*Structural descriptors*

An initial set of structural descriptors was obtained by fragmenting the structures in the database using a modification of the RECAP procedure, [37] in which the bonds of the molecules were recursively cleaved until only the desired types of fragments remained. In addition to the bond cleavage rules employed in RECAP, carbon-halogen bonds and non-ring bonds between aromatic carbon and aliphatic carbon, nitrogen, oxygen and sulfur atoms were also broken. This process resulted in the set of 74 structural descriptors given in Tables 2 and 3 that comprehensively describe final set of 889 compounds used in the analysis. Each descriptor is represented at least 7 times in the data set.

ClogP values for the structural descriptors were calculated using Biobyte [17] with corrections to account for multiple ionization states. For example, the ClogP for morpholine with a charge of +1 (fragment **10** in Table 2) was calculated directly; however, the ClogP for morpholine with a charge of 0 (fragment **39** in Table 3) was obtained by calculating the ClogP of *N*-phenylmorpholine (zero charge) and then subtracting the ClogP of benzene. A similar process was performed for piperidine (fragments **5** and **54**), piperazine (fragments **12** and **22**), and pyrrolidine (fragments **18** and **60**).

*Regression analysis*

The logarithm of the predicted dissociation constant, $Log(K_D)^P$, was expressed as a weighted linear combination of the descriptors,

$$Log(K_D)^P = \sum_{i=0}^{N} w_i x_i$$

where $N$ is the number of descriptors, $w_i$ is the weighting coefficient for the $i$th descriptor, and $x_i$ is the number of times that the $i$th descriptor occurs in the test molecule. The model did not penalize predictions for $Log(K_D) > -2.7$, as this was the experimental limit of detection (see NMR Binding Data). Thus, the maximum allowable predicted value for $Log(K_D)$ was set to $-2.7$ before penalties were calculated. The lower limit of detection was treated similarly, with the lowest allowable predicted value set to $Log(K_D) = -7.0$ (see Fluorescence Spectroscopy). Regression coefficients that minimized the error between predicted and experimental $Log(K_D)$ were generated using the non-linear regression package *Solver* available within Microsoft Excel 2000.

For the group contribution model, uncertainties in the regression coefficients are given as the standard deviations derived from 10 minimizations using randomized initial values. For all 10 solutions, the $R^2$ value was approximately 0.94, with the sum of squares in each case less than 64 (over the set of 889 compounds) and the average error in $Log(K_D)$ less than 0.12. The model was cross-validated with the LSO method using 10 cross-validation groups consisting of approximately 100 compounds each. For each group, the model was trained on the remaining ~800 compounds and binding constants were then predicted for the 100 compounds in the cross-validation group. The cross-validated $Q^2$ from this approach was 0.90. The model was further validated by randomizing the binding affinities with respect to the structural descriptors. All models generated with randomized data sets had $R^2$ values less than 0.23 and average errors in $Log(K_D)$ in excess of 0.40.

## Acknowledgements

The authors would like to thank Bruce Beutel and Theresa Rosenberg for providing the MIC data in the presence and absence of serum.

## References

1. Peters, T.J. *All about albumin: biochemistry, genetics, and medical applications*; Academic Press: San Diego, 1996.
2. Carter, D.C. and Ho, J.X., Adv. Protein. Chem., 45 (1994) 153–203.
3. Russeva, V.N. and Zhivkova, Z.D., Int. J. Pharm., 168 (1998) 23–29.
4. Carter, D.C., He, X.M., Munson, S.H., Twigg, P.D., Gernert, K.M. et al., Science, 244 (1989) 1195–1198.
5. Mao, H., Hajduk, P.J., Craig, R., Bell, R., Borre, T. et al., J. Am. Chem. Soc., 123 (2001) 10429–10435.
6. Curry, S., Mandelkow, H., Brick, P., Franks, N., Nat. Struct. Biol., 5 (1998) 827–835.
7. Bhattacharya, A.A., Curry, S., Franks, N.P., J. Biol. Chem., 49 (2000) 38731–38738.
8. Petitpas, I., Bhattacharya, A.A., Twine, S., East, M., Curry, S., J. Biol. Chem., 276 (2001) 22804–22809.
9. Gensmantel, N.P., Physicochemical Properties and Drug Design. Med. Chem. Principles and Practice, 1994, pp. 98–113.
10. Hansch, C., Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington DC, 1995, p. 557.
11. Gao, C., Govind, R., Tabak, H.H., Env. Tox. Chem., 11 (1992) 631–636.
12. Cramer, R.D., Patterson, D.E., Bunce, J.E., J. Am. Chem. Soc., 110 (1994) 5959–5967.
13. Klebe, G., Abraham, U., Mietzner, T., J. Med. Chem., 37 (1994) 4130–4146.
14. Taylor, R., J. Chem. Inform. Computer Sci., 35 (1995) 59–67.
15. Colmenarejo, G., Alvarez-Pedraglio, A., Lavandera, J.-L., J. Med. Chem., 44, (2001) 4370–4378.
16. Krassavine, A., Jardine, L. *MetaSymphony*; Lion Bioscience: Oxford UK.
17. Leo, A. *CLOGP*, 4.0 ed., Biobyte Corporation: Claremont CA.
18. Leo, A., Weininger, A. *CMR*, 3 ed.; Daylight Chemical Information Systems, Inc.
19. Meylan, W.M., Howard, P.H., Boethling, R.S., Environ. Toxicol. Chem., 15 (1996) 100–106.
20. Meylan, W.M., Howard, P.H., Perspect. Drug Discov. Design (2000)
21. Pearlman, R.S., *SAVOL3*, College of Pharmacy, The University of Texas at Austin, Austin TX. Pearlman@vax.phr.utexas.edu.: Austin TX.
22. Todeschini, R., Consonni, V., *Handbook of Molecular Descriptors*; Wiley-VCH: Wenheim (Germany), 2000.
23. Devliiers, J., Balaban, A.T., *Topological Indices and Related Descriptors in QSAR and Drug Design*; Gordon & Breach: Amsterdam (The Netherlands), 2000.
24. Karelson, M., Lobanov, V.S., Katritzky, A.R., Chem. Rev., 96 (1996) 1027–1043.
25. Jug, K., J. Org. Chem., 48 (1983) 1344.
26. Todeschini, R., Gramatica, P., *3D QSAR in Drug Design*; Kluwer/ESCOM: Dordrecht (The Netherlands), 1998.
27. Todeschini, R., Consonni, V., Pavan, M. *DRAGON*, 2.1 ed.: Milano, Italy.
28. Szczepankiewicz, B.G., Bal, R.B., von Geldern, T.W., Wu-Wong, J.R., Chiou, W.J. et al., Life Sci., 63 (1998) 1905–1912.
29. Israili, Z.H., Dayton, P.G., Drug Metab. Rev., 33 (2001) 161–235.
30. Butina, D., J. Chem. Inform. Computer Sci., 39 (1999) 747–750.

31. Martin, Y.C., Kofron, J.L., Traphagen, L.M., J. Med. Chem., 45 (2002).
32. Martin, Y., Delazzer, J., *SPARK*; Abbott Laboratories.
33. Hajduk, P.J., Augeri, D.A., Mack, J., Mendoza, R., Yang, J. et al., J. Med. Chem., 122 (2002) 7898–7904.
34. Mao, H., Gunasekera, A.H., Fesik, S.W., Prot. Express. Purif., 20 (2000) 492–499.
35. Epps, D.E., Raub, T.J., Hezdy, F.J.A, Anal. Biochem., 227 (1995) 342–350.
36. Wang, Z.-X., FEBS Lett., 360 (1995) 111–114.
37. Lewell, X.Q., Judd, D.B., Watson, S.P., Hann, M.M., J. Chem. Inf. Comput. Sci., 38 (1998) 511–522.