

# An Integrated in Silico Analysis of Drug-Binding to Human Serum Albumin

Ernesto Estrada,<sup>\*,†</sup> Eugenio Uriarte,<sup>‡</sup> Enrique Molina,<sup>†</sup> Yamil Simón-Manso,<sup>§</sup> and George W. A. Milne<sup>||</sup>

Complex Systems Research Group, X-rays Unit, Edificio CACTUS, Santiago de Compostela 15982, Spain,  
Department of Organic Chemistry, University of Santiago de Compostela,  
Santiago de Compostela 15982, Spain, INEST Group, PM-USA and Center for Theoretical and Computational  
Nanoscience, National Institute of Standards & Technology (NIST), 100 Bureau Drive, Stop 8380,  
Gaithersburg, Maryland 20899-8380, and National Institute of Chemistry, 1000 Ljubljana, Slovenia

Received June 29, 2006

Approaches such as quantitative structure–activity relationships (QSAR) and molecular modeling are integrated with the study of complex networks to understand drug binding to human serum albumin (HSA). A robust QSAR model using the topological substructural molecular descriptors/design (TOPS-MODE) approach has been derived and shows good predictability and interpretability in terms of structural contribution to drug binding to HSA. A perfect agreement exists between the group/fragment contributions found by TOPS-MODE and the specific interactions of drugs with HSA. These results indicate a preponderant contribution of hydrophobic regions of drugs to the specific binding to drug-binding sites 1 and 2 in HSA and specific roles of polar groups which anchor drugs to HSA binding sites. The occurrence of fragments contributing to drug binding to HSA can be represented by complex networks. The fragment-to-fragment complex network displays “small-world” and “scale-free” characteristics and in this way is similar to other complex networks including biological, social, and technological networks. A small number of fragments appear very frequently in most drugs. These molecular “empathic” fragments are good candidates for guiding future drug discovery research.

## INTRODUCTION

The development of computational models for the prediction of absorption, distribution, metabolism, and excretion (ADME) characteristics of drugs and druglike compounds is appealing from both a theoretical and a practical point of view.<sup>1–4</sup> On the theoretical side, the development of these models challenges our current strategies because of the necessity for improving predictability without losing interpretability. The simplest models, with very few variables, are easily interpreted but generally lack prediction power outside the limited scope of the data set used in their development. An improvement of predictability typically requires more variables or more sophisticated statistical techniques which make it difficult to extract structural and biological information from the model.<sup>5</sup> The correct prediction of ADME characteristics at a very early stage of drug development is necessary to avoid costly late-stage failures, which are mainly due to toxicity or poor pharmacokinetics.<sup>6</sup>

An important factor that strongly affects both ADME and toxicity properties of drugs is their interaction with plasma proteins. The majority of drug binding in serum occurs by interactions with two major proteins:<sup>7</sup> human serum albumin (HSA) and  $\alpha_1$ -acid-glycoprotein.<sup>8</sup> HSA, the most abundant protein in plasma, is known to form mainly noncovalent

complexes with exogenous ligands. This relatively large (585 residues) multidomain protein is formed by three homologous helical domains (I–III), which are divided into two subdomains (A and B).<sup>9</sup> Most of the drugs that bind to HSA form complexes in which the drug is located at one of the two main binding sites of HSA, which are located in subdomains IIA and IIIA.<sup>10,11</sup> These two binding sites have been well-characterized in X-ray crystallographic studies of several drug–HSA complexes.<sup>10,12–15</sup>

There have been several attempts to generate models that predict drug binding affinities to HSA.<sup>16,17</sup> These range from quantitative structure–activity relationships (QSAR)<sup>18–21</sup> to the use of molecular modeling.<sup>22,23</sup> Colmenarejo et al.<sup>18</sup> reported the binding affinities to HSA of 95 diverse compounds obtained through high-performance affinity chromatography. A QSAR model with a correlation coefficient  $r^2 = 0.83$  was also obtained in this work. Similar statistical results have been reported by others using different statistical approaches for the same data set.<sup>19–21</sup>

We have used this data set to develop a QSAR model using the topological substructural molecular descriptors/design (TOPS-MODE) approach.<sup>24,25</sup> The novelty of our current approach is that our model provides structural information about the contribution of different groups and fragments to the drug binding to HSA. This information is then compared with the drug–HSA interactions observed by X-ray crystallography. Information from these two sources suggests two pharmacophoric moieties for the binding of drugs at the center of binding site 1 or 2. Another important contribution of the current work is that, for the first time in the context of medicinal chemistry, we have applied the

\* Corresponding author fax: 0034981547077; e-mail: estrada66@yahoo.com.

<sup>†</sup> Complex Systems Research Group.

<sup>‡</sup> Department of Organic Chemistry, University of Santiago de Compostela.

<sup>§</sup> NIST.

<sup>||</sup> National Institute of Chemistry, Slovenia.

techniques of complex networks studies to the drug–HSA binding problem. The complex networks of drug fragments, like other complex systems in chemistry, biology, technology, and society, are “small-world” and “scale-free” networks.

## MATERIALS AND METHODS

**Data Set and Software.** A data set of 95 diverse drugs and druglike compounds reported by Colmenarejo et al.<sup>18</sup> was used in this work. According to these authors, binding constants were determined by high-performance affinity chromatography and recorded on a logarithmic scale such that  $\log K'_{hsa} = \log[(t - t_0)/t_0]$ , where  $t$  and  $t_0$  respectively are the retention time of the drug and of NaNO<sub>3</sub>. These constants are always used in this work when we refer to binding constants.

Three compounds were dropped from the original 95: captopril, for which  $t \approx t_0$ , the Se-containing neuroprotective agent (Se is an element not yet parametrized for TOPS-MODE); ebselen; and quinidine, because it is a stereoisomer of quinine. The remaining 88 compounds (after the exclusion of four outliers, see further) are shown in Table 1. The division of the data set into a training and a prediction set was carried out following strictly the same division reported by Colmenarejo et al.<sup>18</sup> Structures for drug–HSA complexes were obtained from the Protein Data Bank (PDB),<sup>26</sup> and ligand–protein contacts (LPC) were calculated with the LP software.<sup>27</sup> We used here the crystallographic data of nine different drug–HSA complexes as reported in the PDB. They are those for warfarin, phenylbutazone, oxyphenbutazone, 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF), indomethacin, diazepam, ibuprofen, diflunisal, and 3-indoxyl sulfate.

All calculations related to the TOPS-MODE approach were carried out by using the program MODESLAB for Windows ([www.modeslab.com](http://www.modeslab.com)). The input file for these calculations contains the SMILES codes for every compounds in the data set. The program calculates the TOPS-MODE molecular descriptors automatically for the whole file. After the QSAR/quantitative structure–property relationship (QSPR) model is obtained, we use again MODESLAB for calculating the bond contributions of every bond in the molecule, see <http://www.modeslab.com>. The contributions of the different groups/fragments reported in this work are calculated by summing the bond contributions for all bonds in the group/fragment. Such groups/fragments were selected by considering the classical functional groups in organic chemistry as well as other fragments that could be interesting for the purposes of the current work.

Network analysis and drawings were carried out by using the software UCINET for Windows, version 6 (<http://www.analytictech.com/downloaduc6.htm>). We elaborate the input files in the form of an adjacency list for both fragments and molecules. All statistical analyses were carried out using the STATISTICA 6 software (StatSoft Inc.).

**Molecular Descriptors: The TOPS-MODE approach.** TOPS-MODE<sup>28–32</sup> is based on the calculation of spectral moments of molecular bond matrices weighted to account for hydrophobic, electronic, and steric molecular features. Spectral moments are the trace of the  $k$ th power of a matrix, that is, the sum of all of the main diagonal entries of such matrices.<sup>28–30</sup> A bond matrix is a square symmetric matrix

representing the hydrogen-suppressed molecular skeleton in which nondiagonal entries are ones or zeroes if the corresponding bonds have a common atom or not, respectively.<sup>33</sup> Bond weights are placed as diagonal entries of such matrices and represent quantitative contributions to different physicochemical properties. The bond weights currently in use include hydrophobicity (H),<sup>34</sup> polar surface area (PS),<sup>35</sup> polarizability (Pol),<sup>36</sup> molar refractivity (MR),<sup>36</sup> van der Waals radii (vdW),<sup>37</sup> and Gasteiger–Marsilli charges (Ch).<sup>38</sup>

The TOPS-MODE descriptors of the different types, for example, H, PS, Pol, MR, vdW, and Ch, are calculated for the molecules under study. Then, the QSAR model below is developed describing the binding affinity to HSA in terms of the spectral moments  $\mu_j$ :

$$\log K_{hsa} = b_0 + \sum_{j=1}^L b_j \mu_j \quad (1)$$

where  $b_j$  are the coefficients of the QSAR model obtained by linear regression analysis and  $b_0$  is the error. The  $j$ th spectral moment of the bond matrix can be expressed as a sum of bond moments, which are simply the corresponding entries of the  $j$ th power of the bond matrix:

$$\mu_j = \sum_i^m \mu_j(i) \quad (2)$$

where  $\mu_j(i)$  is the bond moment of the  $i$ th bond in a molecule with  $m$  bonds. Then, eq 1 can be written as

$$\log K_{hsa} = b_0 + \sum_{j=1}^L b_j \sum_{i=1}^m \mu_j(i) = b_0 + \sum_{i=1}^m \sum_{j=1}^L b_j \mu_j(i) \quad (3)$$

where the right-hand term in eq 3 represents the contribution of bond  $i$  to  $\log K_{hsa}$  and will be called “bond contribution” and represented by  $\log K_{hsa}(i)$ :

$$\log K_{hsa}(i) = \sum_{j=1}^L b_j \mu_j(i) \quad (4)$$

and  $\log K_{hsa}$  is expressed as an additive function of bond contributions:

$$\log K_{hsa} = \sum_{i=1}^m \log K_{hsa}(i) \quad (5)$$

The general strategy of TOPS-MODE permits the expression of a property, such as  $\log K_{hsa}(i)$ , in terms of fragment or group contributions just by expanding the spectral moments in their linear combination of molecular substructures. These bond contributions are not obtained independently by any different statistical approach, but they correspond to the information encoded in the molecular descriptors used in the QSAR/QSPR model. In general, it has been proved that “if a structure–property data set is sufficiently large to allow building statistically significant models, then any topological index can be replaced with a set of substructural descriptors.”<sup>39,40</sup>

**Orthogonalization of TOPS-MODE Descriptors.** One of the inherent characteristics of the TOPS-MODE approach is that spectral moments are collinear. The main drawback

**Table 1.** Data Set of Compounds Used for the Analysis of HSA Binding, Their CAS Numbers, as Well as Observed and Predicted Binding Affinities

no.	compound	CAS	obsd	calcd
1	cefuroxime	55268-75-2	-1.33	-1.03
2	amoxicillin	61336-70-7	-1.21	-0.78
3	cephalexin	15686-71-2	-1.11	-0.59
4	5-fluorocytosine	2022-85-7	-1.11	-1.25
5	zidovudine	30516-87-1	-1.02	-1.17
6	caffeine	58-08-2	-0.92	-1.12
7	L-tryptophan	73-22-3	-0.78	-0.46
8	methotrexate	59-05-2	-0.77	-0.88
9	antipyrine	60-80-0	-0.69	-0.47
10	phenoxymethylpenicillinic acid	87-08-1	-0.69	-0.74
11	salicylic acid	69-72-7	-0.66	-0.57
12	cefuroxime axetil	64544-07-6	-0.56	-1.03
13	etoposide	33419-42-0	-0.49	-0.04
14	atenolol	29122-68-7	-0.48	-0.54
15	chlorpropamide	94-20-2	-0.44	-0.34
16	sotalol	3930-20-9	-0.44	-0.45
17	hydrochlorothiazide	58-93-5	-0.42	-0.59
18	tolazamide	1156-19-0	-0.42	-0.23
19	hydrocortisone	50-23-7	-0.4	-0.33
20	nadolol	42200-33-9	-0.4	-0.22
21	prednisolone	50-24-8	-0.4	-0.30
22	scopolamine	138-12-5	-0.34	-0.14
23	metoprolol	37350-58-6	-0.29	-0.28
24	trimethoprim	738-70-5	-0.26	-0.44
25	dansylglycine	1091-85-6	-0.26	-0.21
26	lidocaine	137-58-6	-0.23	-0.08
27	methylprednisolone	83-43-2	-0.22	-0.15
28	tolbutamide	64-77-7	-0.22	-0.33
29	sulfaphenazole	526-08-9	-0.21	0.06
30	acebutolol	37517-30-9	-0.21	-0.37
31	procaine	59-46-1	-0.19	-0.09
32	oxprenolol	6452-71-7	-0.15	-0.16
33	lamotrigine	84057-84-1	-0.13	-0.05
34	clonidine	4205-90-7	-0.13	0.02
35	pindolol	13523-86-9	-0.13	-0.22
36	frusemide	54-31-9	-0.13	-0.17
37	carbamazepine	298-46-4	-0.1	0.31
38	ranitidine	66357-35-5	-0.1	-0.37
39	camptothecin	7689-03-4	-0.08	0.24
40	tetracycline	60-54-8	-0.08	-0.17
41	sumatriptan	103628-46-2	-0.05	-0.36
42	warfarin	81-81-2	-0.04	0.43
43	bumetanide	28395-03-1	-0.03	0.16
44	oxyphenbutazone	129-20-4	-0.02	0.20
45	phenytoin	57-41-0	0.00	-0.02
46	doxycycline	564-25-0	0.01	-0.24
47	ketoprofen	22071-15-4	0.03	0.31
48	alprenolol	13655-52-2	0.04	0.01
49	digitoxin	71-63-6	0.13	0.28
50	levofloxacin	100986-85-4	0.14	0.20
51	ciprofloxacin	85721-33-1	0.14	0.23
52	labetalol	36894-69-6	0.14	0.20
53	norfloxacin	70458-96-7	0.14	0.03
54	phenylbutazone	50-33-9	0.19	0.32
55	sancycline	808-26-4	0.21	-0.17
56	minocycline	10118-90-8	0.21	-0.10
57	naproxen	22204-53-1	0.25	0.14
58	propranolol	525-66-6	0.28	0.27
59	tetracaine	94-24-6	0.32	-0.08
60	fusidic acid	03/06/6990	0.33	0.33
61	novobiocin	1476-53-5	0.35	0.34
62	ondansetron	116002-70-1	0.37	0.14
63	droperidol	548-73-2	0.43	0.53
64	indomethacin	53-86-1	0.47	0.51
65	quinine	130-95-0	0.49	0.43
66	sulfasalazine	599-79-1	0.56	0.28
67	progesterone	57-83-0	0.59	0.48
68	desipramine	50-47-5	0.61	0.61
69	estradiol	50-28-2	0.68	0.62
70	glibenclamide	10238-21-8	0.68	0.47
71	testosterone	58-22-0	0.74	0.38
72	imipramine	50-49-7	0.75	0.69
73	ketoconazole	65277-42-1	0.84	0.93
74	promazine	58-40-2	0.92	0.55
75	triflupromazine	146-54-3	1.05	1.05
76	chlorpromazine	50-53-3	1.1	0.78
77	terbinafine	78628-80-5	1.17	1.01
78	clotrimazole	23593-75-1	1.34	1.21

**Table 1.** Continued

no.	compound	CAS	obsd	calcd
External Validation Set				
79	acetylsalicylic acid	50-78-2	-1.39	-0.63
80	acetaminophen	103-90-2	-0.81	-0.72
81	chloramphenicol	56-75-7	-0.46	-0.29
82	timolol	26839-75-8	-0.33	-0.44
83	terazosin	63590-64-7	-0.16	-0.18
84	bupropion	31677-93-7	-0.05	0.35
85	prazosin	19216-56-9	0.06	0.07
86	clofibrate	637-07-0	0.27	0.15
87	verapamil	52-53-9	0.52	1.05
88	itraconazole	84625-61-6	1.04	1.37

of collinearity from the point of view of a QSAR model stems from the stability of the coefficients in the linear regression model. There is difficulty associated with the interpretation of linear models obtained with collinear variables because the sign and magnitude of the coefficients in the regression model can be affected by the removal or introduction to the model of a new variable. In the case of the TOPS-MODE approach, this can lead to a false interpretation of bond contributions because their magnitude and sign can be in error. Consequently, we have implemented the Randić orthogonalization method (ROM)<sup>41-43</sup> to eliminate the collinearities among the TOPS-MODE variables and have developed a new procedure to extract the information contained in these variables after orthogonalization.

ROM has been described in detail,<sup>41-45</sup> and here, we summarize the main steps in the orthogonalization of a QSAR model of the form  $Khsa = a_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3$ , which proceeds as follows:

(i) Orthogonalize  $\mu_1$ :  $\Omega(\mu_1) = \mu_1$

(ii) Orthogonalize  $\mu_2$  with respect to  $\mu_1$ :  $\Omega(\mu_2) = \mu_2 - \hat{\mu}_2$ , where  $\hat{\mu}_2 = b_0 + b_1\mu_1$

(iii) Orthogonalize  $\mu_3$  with respect to  $\Omega(\mu_2)$ :  $\Omega(\mu_3) = \mu_3 - \hat{\mu}_3$ , where  $\hat{\mu}_3 = b'_0 + b'_1\Omega(\mu_2)$

(iv) Orthogonalize  $\Omega(\mu_2)$  with respect to  $\mu_1$ :  ${}^2\Omega(\mu_3) = \Omega(\mu_3) - \hat{\Omega}(\mu_3)$ , where  $\hat{\Omega}(\mu_3) = b''_0 + b''_1\mu_1$ . The orthogonalized variables are  $\mu_1$ ,  $\Omega(\mu_2)$ , and  ${}^2\Omega(\mu_3)$ , and the coefficients in steps ii-iv are obtained by linear regression analysis.

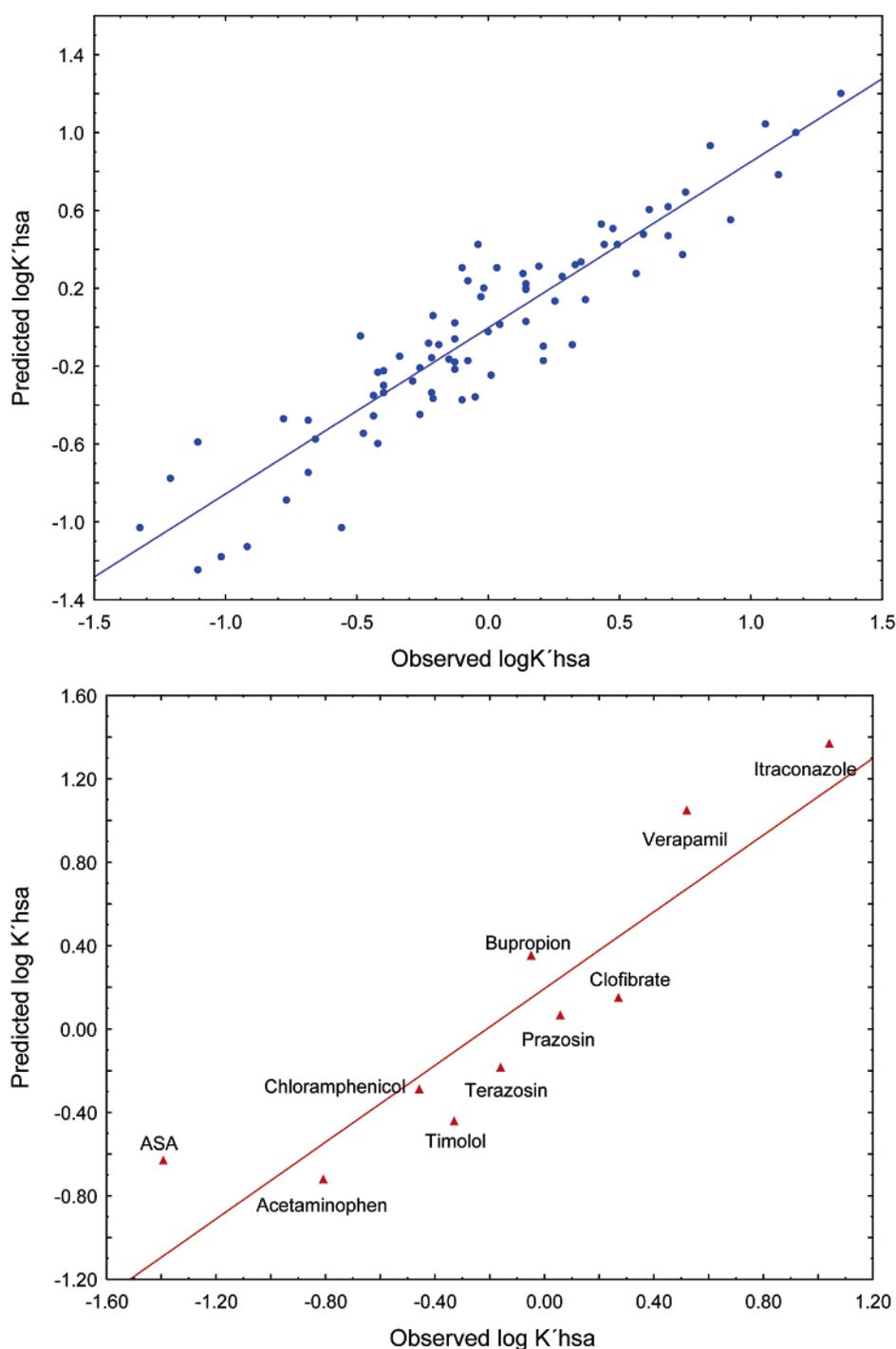
**Calculation of Bond Contributions.** Bond contributions are numeric characterizations of bonds which permit the identification of groups or regions of a molecular framework which may be responsible for a particular property or activity.<sup>46</sup> By analysis of similar regions in different molecules, one can derive general rules about the contributions of molecular fragments to a particular property. This procedure consists of the transformation of a QSAR model into a bond-additive scheme in which the property in question can be calculated as the sum of bond contributions for a molecule.

To calculate the bond contributions in the orthogonalized QSAR obtained with TOPS-MODE, we have implemented the following iterative procedure:<sup>47</sup>

(i) Calculate bond contributions to  $\Omega(\mu_1)$ :  $C[\Omega(\mu_1)] = C(\mu_1)$

(ii) Calculate bond contributions to  $\Omega(\mu_2)$ :  $C[\Omega(\mu_2)] = C(\mu_2) - b_1C(\mu_1)$

(iii) Calculate bond contributions to  $\Omega(\mu_3)$ :  $C[\Omega(\mu_3)] = C(\mu_3) - b'_1C[\Omega(\mu_2)]$



**Figure 1.** Plots of predicted versus observed values of binding affinities of drugs to HAS for the training and prediction sets used in this work (see Table 1).

(iv) Calculate bond contributions to  ${}^2\Omega(\mu_3)$ :  $C[{}^2\Omega(\mu_3)] = C[\Omega(\mu_3)] - b_1' C(\mu_1)$ . The final bond contributions are  $C(\mu_1)$ ,  $C[\Omega(\mu_2)]$ , and  $C[{}^2\Omega(\mu_3)]$ .

This procedure represents the extraction of the information contained into a bond contribution of a variable which is duplicated by the other variables in the model. It is not limited to the training set used to generate the classification model but is also valid for the compounds used in the external prediction set. In the current work, we have orthogonalized the descriptors for the compounds used in the external set using the orthogonalized model obtained for the training

set, and we have also calculated the bond contributions to  $\log K_{hsa}$ .

## RESULTS AND DISCUSSION

**QSAR Modeling.** Our current strategy is based on the consideration of the quantitative relationship between the molecular structure of all drugs and druglike compounds with their binding affinity to HSA without the explicit consideration of their binding sites. Then, we relate these quantitative models with binding sites by using the contribution of the different groups and fragments to the binding and the

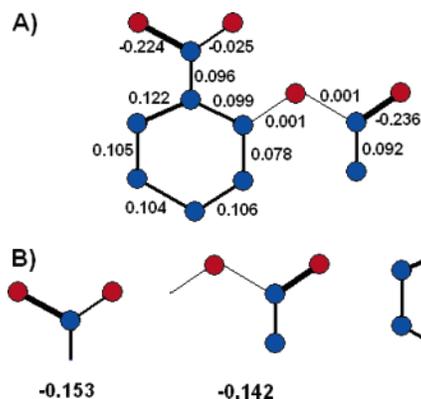
**Table 2.** Coefficients of the QSAR Models Obtained by Using TOPS-MODE Descriptors before and after the Randić Orthogonalization Method (ROM)

variable	coefficient (non-orth.)	error	coefficient (orth.)	error
$\mu_1^H$	0.254	0.026	0.232	0.014
$\mu_1^H$	$-1.320 \times 10^{-6}$	$-2.157 \times 10^{-7}$	$-5.391 \times 10^{-7}$	$1.692 \times 10^{-7}$
$\mu_{10}^{PS}$	$-6.216 \times 10^{-2}$	$-1.393 \times 10^{-2}$	$-4.296 \times 10^{-3}$	$8.717 \times 10^{-4}$
$\mu_3^{PS}$	$1.141 \times 10^{-2}$	$2.937 \times 10^{-3}$	$2.189 \times 10^{-3}$	$1.292 \times 10^{-3}$
$\mu_4^{PS}$	$6.74 \times 10^{-5}$	$1.735 \times 10^{-5}$	$7.472 \times 10^{-7}$	$1.491 \times 10^7$
$\mu_6^{PS}$	$1.959 \times 10^{-7}$	$5.176 \times 10^{-8}$	$5.618 \times 10^{-9}$	$2.587 \times 10^{-9}$
$\mu_8^{PS}$	$2.467 \times 10^{-10}$	$6.989 \times 10^{-11}$	$2.477 \times 10^{-10}$	$6.734 \times 10^{-11}$
$\mu_{10}^{PS}$	$1113 \times 10^{-13}$	$3.453 \times 10^{-14}$	$1.286 \times 10^{-15}$	$7.501 \times 10^{-16}$
$\mu_5^{Ch}$	$1.317 \times 10^{-3}$	$2.167 \times 10^{-4}$	$4.311 \times 10^{-4}$	$5.719 \times 10^{-5}$
intercept	-0.974	0.097	-0.970	0.039

information given by the X-rays structures of the complexes in an a posteriori analysis. The data set we use here displays a large structural variability which makes topological approaches such as TOPS-MODE more appropriate for dealing with the structure–activity relationships. Despite the fact that this method does not consider the three-dimensional structures of molecules explicitly, it describes very well the information contained in their topological structures, which indeed encoded a large part of the geometrical characteristics of molecules.<sup>48</sup> Using TOPS-MODE, we have developed a QSAR model which describes the binding affinities ( $\log K'_{hsa}$ ) of the 78 drugs in the training set in terms of hydrophobicity, polar surface area, and electrostatic charges. Four drugs, cromolyn, propylthiouracil, cimetidine, and acrivastine, were found to be statistical outliers. Our current model, described in Table 1, has the following statistics:  $r^2 = 0.85$ ,  $s = 0.23$ , and  $F(9,68) = 43.6$ . The orthogonalized coefficients and errors for each variable in the model are also given in Table 1. When this model is applied to the prediction of the  $K'_{hsa}$  values for the 10 drugs in the external prediction set, we obtain the following statistics:  $r^2 = 0.83$  and  $s = 0.30$ . Figure 1 shows the plots of observed versus predicted values of  $K'_{hsa}$  for both training and prediction sets, and Table 2 shows the corresponding coefficients.

**Structural Contributions to HSA Binding.** The main advantage of the TOPS-MODE approach is that we can obtain a direct structural interpretation of the results by examination of the bond contributions. The QSAR model given in Table 2 can be transformed into bond contributions for every bond in a molecule. In this way, one can calculate the contributions of all bonds in the 88 molecules studied here (78 in the training set and 10 in the prediction set). The list of bond contributions for the 88 molecules is provided as Supporting Information. All TOPS-MODE contributions are in units of  $\log K'_{hsa}$ . These bond contributions can be transformed into the contributions from fragments or functional groups. The sum of contributions for all bonds forming the fragment is considered to be the fragment's global contribution. Figure 2A, for example, shows the bond contributions for acetylsalicylic acid and the fragment contributions corresponding to three substructures. In this way, the information contained in the 13 bonds of the molecule is reduced to the contribution of three groups.

In this way, we have calculated the contribution of 65 different groups to the drug–HSA binding. The contribution for the same group in different molecules is averaged and

**Figure 2.** Illustration of bond contributions to HAS binding in aspirin (A) and their transformation into fragment/group contributions (B).

reported as the group contribution for this specific fragment independent of the molecule in which it is located. For instance, the contribution of a phenyl ring to  $\log K'_{hsa}$  is  $0.585 \pm 0.045$ —the average of the phenyl ring contribution in the 74 compounds in which this group is present. The maximum value observed for the phenyl group is 0.668 (droperidol), and the lowest is 0.486 (etoposide). In Table 3, the average contributions of the 65 groups selected in this work are shown. A more exhaustive list of contributions can be easily obtained by making use of the bond contributions for all of the drugs studied (see the Supporting Information). This kind of fragment contribution can be implemented in structure-based computer software designed to make predictions of chemical and biological properties of organic molecules on the basis of the identification of structural motifs. Another advantage of group contributions is that they permit the derivation of general rules underlying the role of molecular structure on drug binding to HSA. This kind of general structural interpretation is given below.

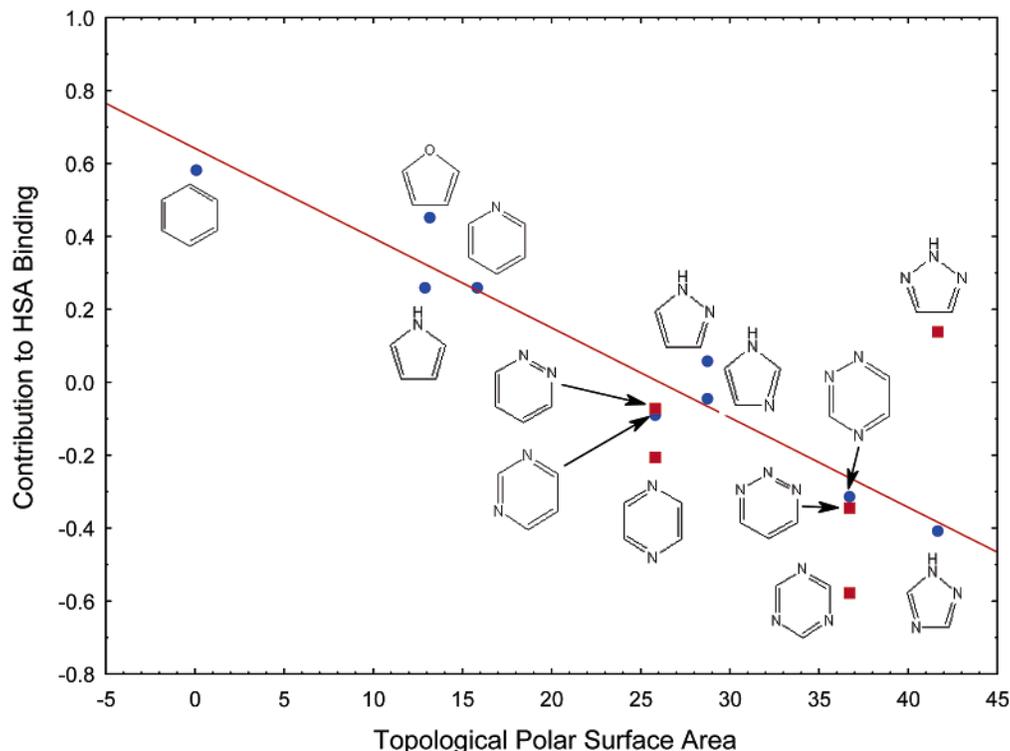
The values of binding constants are expressed here as  $\log K'_{hsa}$ , and so, a positive contribution from a fragment means that it enhances the binding of the drug to HSA by increasing the retention time of the drug with respect to  $\text{NaNO}_3$ . The scale being logarithmic, a negative contribution simply represents a smaller but still significant contribution by a fragment to the binding of the drug to HSA.

There are several general characteristics which can be seen in Table 3. The first is related to the influence of aromaticity on the HSA binding of several groups. It can be seen that functional groups such as an amine bonded directly to aromatic rings display larger contributions to  $\log K'_{hsa}$  than when bonded to an aliphatic framework. Both contributions increase when the hydrogen atoms of the amino group are substituted by methyl or other alkyl groups. In both cases, the same trend of aromatic/aliphatic contributions is observed. For instance, aliphatic dimethylamino groups have contributions of  $-0.026$  and  $0.118$  for the aromatic substituent. Another example is provided by the hydroxyl group, with a low contribution in alcohols ( $-0.178$ ), which is increased in phenols ( $-0.038$ ), and the same trend is observed for methoxyl, chlorine, carboxyl, and carboxamide groups. In the case of the carboxyl group, it is possible to observe the trend when passing from aliphatic ( $-0.192$ ) to  $\alpha,\beta$ -unsaturated ( $-0.150$ ) to aromatic ( $-0.125$ ). These findings suggest that, in groups with p-electron pairs or  $\pi$

**Table 3.** Fragment Contributions to HSA Binding Affinities of Drugs<sup>a</sup>

No.	Fragment	Contrib.									
1		0.150 (± 0.020)	18		0.163 (± 0.002)	35		-0.120 (± 0.001)	51		1.431
2		0.126 (± 0.002)	19		-0.192 (± 0.019)	36		-0.101	52		1.321 (± 0.005)
3		0.157	20		-0.150 (± 0.003)	37		0.260 (± 0.086)	53		0.678 (± 0.073)
4		0.218 (± 0.002)	21		-0.125 (± 0.002)	38		0.060	54		-0.234
5		0.108 (± 0.001)	22		-0.145 (± 0.015)	39		-0.042 (± 0.036)	55		-0.043
6		-0.178 (± 0.012)	23		-0.104 (± 0.003)	40		-0.104	56		-0.227
7		-0.038 (± 0.002)	24		-0.274	41		-0.319 (± 0.001)	57		0.063
8		-0.171	25		-0.258 (± 0.036)	42		-0.489	58		0.177 (± 0.007)
9		-0.082	26		-0.091	43		-0.404	59		0.030 (± 0.003)
10		-0.121 (± 0.008)	27		-0.258	44		-0.350	60		0.873 (± 0.024)
11		-0.032 (± 0.002)	28		-0.186 (± 0.009)	45		0.263 (± 0.024)	61		-0.153 (± 0.035)
12		-0.023 (± 0.016)	29		-0.035 (± 0.029)	46		-0.088 (± 0.026)	62		0.360
13		-0.026 (± 0.003)	30		-0.396 (± 0.026)	47		-0.313	63		0.268
14		0.118	31		-0.287 (± 0.005)	48		-0.227	64		0.148
15		0.189	32		0.585 (± 0.045)	49		-0.178	65		0.146
16		-0.037 (± 0.017)	33		0.453 (± 0.131)	50		1.509			
17		0.141 (± 0.049)	34		0.176						

<sup>a</sup> The contributions reported are the average of the contributions of the corresponding fragment in all molecules having it, with the deviation given in parentheses.



**Figure 3.** Illustration of the relationship existing between group contribution to HAS binding and the TPSA of some aromatic rings. The blue dots corresponds to rings present in the drugs included in this work. Red squares correspond to other aromatic rings with five and six atoms having two or three nitrogens.

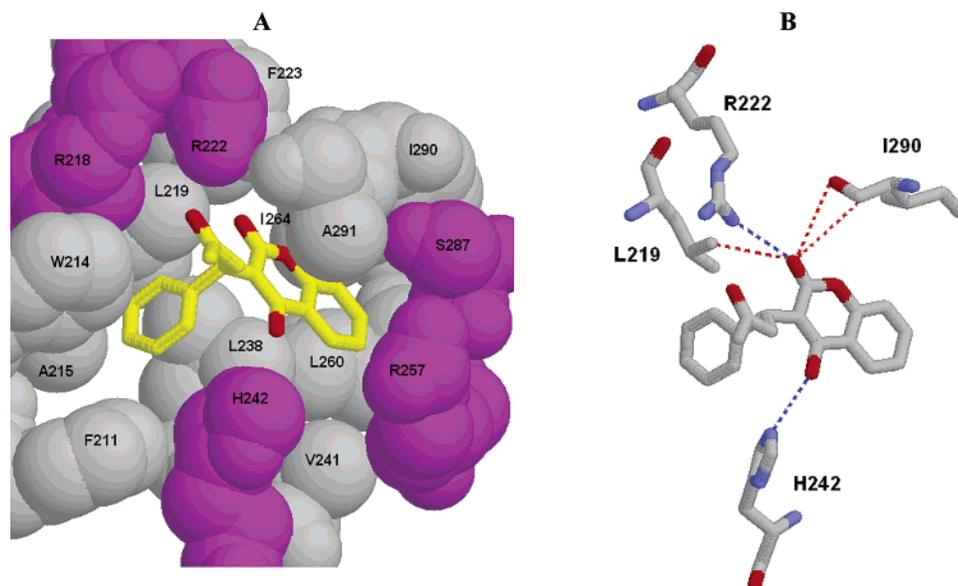
electrons, an increment in the electron delocalization increases the binding to HSA presumably because of the decrease in the polarity of these groups. This effect highlights the roles played by polarity and hydrophobicity in the interaction between drugs and HSA. All of these groups are polar, and the decrease in polarity (or increase in hydrophobicity) produced by electron delocalization associated with aromatic rings increases their binding to nonpolar (or hydrophobic) pockets in HSA.

Another trend observed for the group contributions given in Table 3 concerns aromatic rings. The largest contribution of all aromatic rings studied is 0.585, from the phenyl ring. As heteroatoms are incorporated into the aromatic ring, this contribution decreases dramatically to  $-0.313$  for 1,2,4-triazine. A decrease of ring size also decreases the contribution to drug binding; the lowest contribution of all the aromatic rings studied ( $-0.404$ ) is observed for 1,2,4-triazole. The general trend followed by these aromatic rings is the one of hydrophobicity or polar surface area. There are good correlations ( $r = 0.97$ ) between the contributions of these rings to HSA binding and  $\log P$  as well as the topological polar surface area (TPSA),<sup>35</sup> which can be seen in Figure 3, and this is the main reason that the QSAR model obtained here includes both  $\log P$  and TPSA as molecular descriptors. The order introduced by TPSA and  $\log P$  is not exactly the same as that introduced by the contributions to  $\log K'_{hsa}$ . The differences are related to the dissimilar distributions of hydrophobicity and polar surface area across aromatic rings. This effect is well accounted for by the rest of the TOPS-MODE descriptors and is illustrated by the contributions of pyrazole and imidazole, which have exactly the same global TPSA but a different distribution of the polar regions across the molecule. Pyrazole has a polar region containing both

nitrogen atoms, and the corresponding region in imidazole is distributed over a larger area. If we include the remaining isomers formed by a five- or six-membered ring with two or three nitrogen atoms, the correlation coefficient drops to  $r = 0.82$  (the red squares in Figure 3). The inclusion of the 10 diazaphthalenes in the same plot (not shown) drops the correlation coefficient further, to  $r = 0.63$ , confirming the dissimilar ordering introduced by TOPS-MODE and TPSA for these structural fragments.

**Group Contributions and Specific Drug–HSA Interactions.** Crystallographic studies suggest that HSA has two primary binding sites which are used by different drugs binding to the protein.<sup>10</sup> Drug site 1, in subdomain IIA, has a predominantly apolar interior but also two polar clusters, one formed by H242 and R257 and a second formed mainly by R218 and R222. Drug site 2, in subdomain IIIA, consists of a largely hydrophobic cavity with distinct polar features. This site is the smaller of the sites and includes residues R410, Y411, K414, and S489.

**Binding to Drug Site 1.** Warfarin (**42**) binds at the center of the pocket in drug site 1, as illustrated in Figure 4A.<sup>12,14</sup> This complex is characterized by the location of the coumarin moiety between the apolar side chains of L238 and A291.<sup>14</sup> The bond contributions to the HSA binding affinity indicate a strong positive contribution (0.674) from this group. This location of the coumarin in the complex leaves the hydrophobic phenyl ring in contact with the hydrophobic residues (L260, I264, I290, and the aliphatic portions of R257 and R287) of this pocket. There are also electrostatic interactions between the oxygens of the hydroxyl and carbonyl groups, which form hydrogen bonds with H242 and R222 according to the X-ray structures of the complex,<sup>12,14</sup> respectively. According to TOPS-MODE, however, these two groups



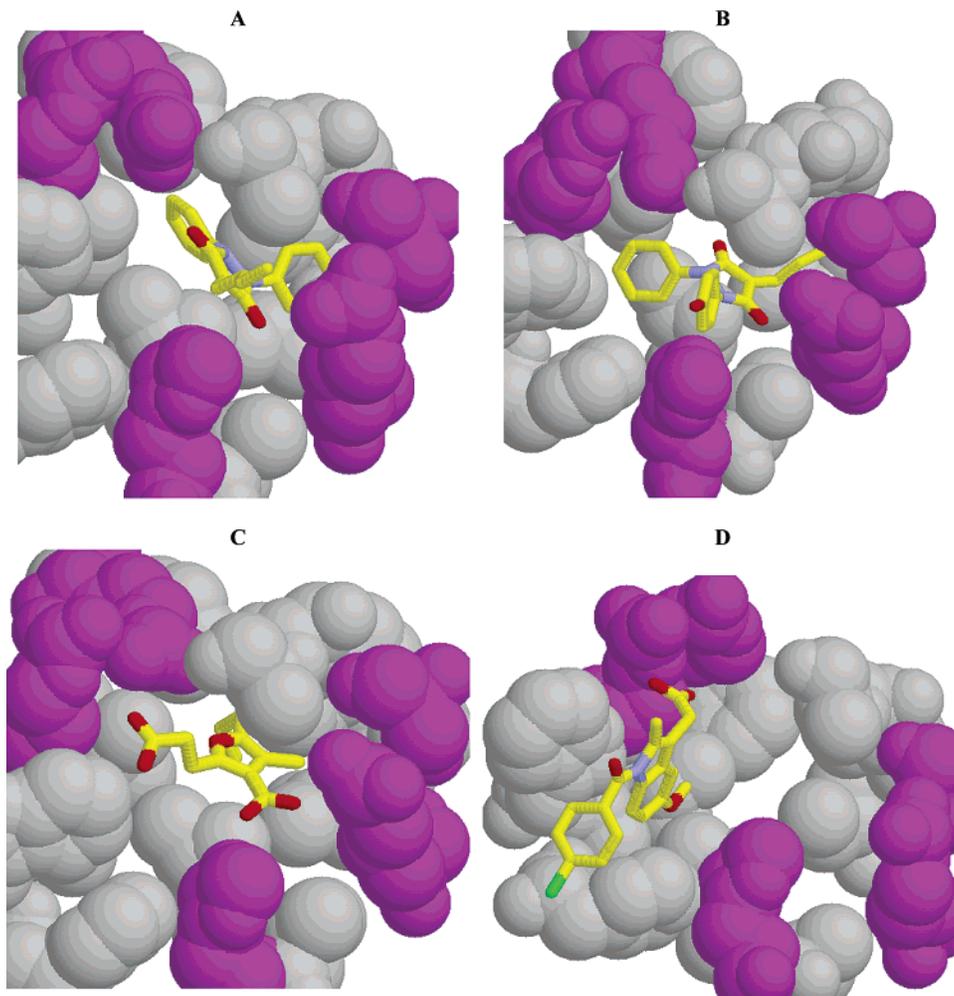
**Figure 4.** Drug binding site 1 (A), in which warfarin is placed at the center of the binding pocket. Interactions of carbonyl and hydroxyl oxygen atoms of warfarin with residues of HSA are shown. In blue are the hydrogen bonds formed with R222 and H242, which stabilize the HSA–warfarin complex. In red are the hydrophobic–hydrophilic and oxygen–oxygen repulsions, which destabilize the warfarin–HSA complex.

make only small contributions to the HSA binding of warfarin. This could be due to the presence of destabilizing interactions between the groups and residues of binding site 1. For example, there is a destabilizing interaction between the coumarin carbonyl group and the carbonyl oxygen of I290, which are 4.29 Å apart. There are also significant destabilizing hydrophobic–hydrophilic interactions between carbons of L219 and I290 and the carbonyl oxygen of the coumarin, which are separated by 4.26 Å and 4.55 Å, respectively (see Figure 4B). The phenyl substituent, with a large contribution to HSA binding (0.646) according to TOPS-MODE calculations, binds in a pocket formed by F211, W214, L219, and L238 with additional aliphatic contacts from R218 and H242. TOPS-MODE finds a small contribution for the acetyl side chain, which is consistent with the experimental observation that this is located at the mouth of the pocket. The destabilizing nature of the interactions experienced by the acetyl group and its small contribution to HSA binding is further confirmed by the fact that the substitution of this group by a propyl moiety in phenprocoumon increases the binding affinity as a result of the formation of hydrophobic contacts between this group and the side chains of W214 and R218.<sup>14</sup>

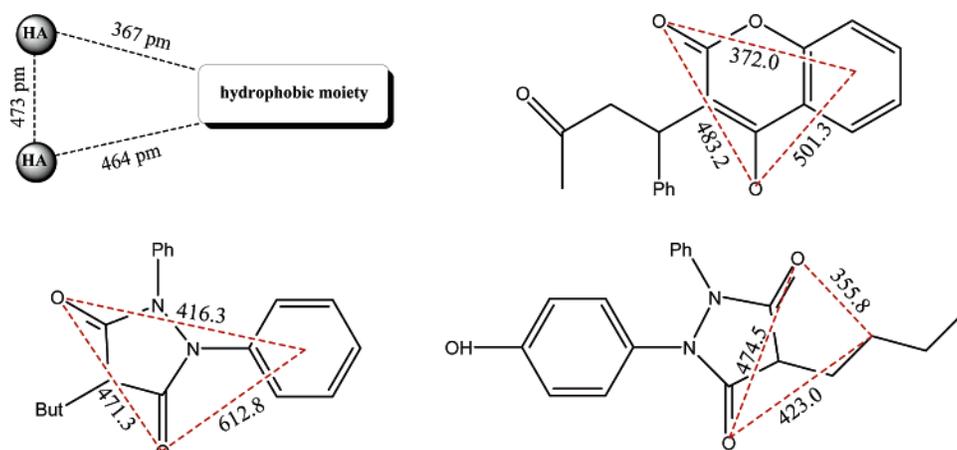
A drug–HSA complex structure is available<sup>14</sup> for three other drugs, oxyphenbutazone (**44**), phenylbutazone (**54**), and indomethacin (**64**) that bind at drug site 1. The structures of the drug–HSA complexes for phenylbutazone and oxyphenbutazone are illustrated in Figure 5A and B. The orientation of phenylbutazone in the binding site is very similar to that of warfarin, but oxyphenbutazone, in the same pocket, is rotated by about 180° relative to the orientation of warfarin and interacts mainly with residues L260, I264, I290, R257, and R287. One of its carbonyl groups forms a hydrogen bond with R222 and the other with H242 and R257. The first also has destabilizing interactions—oxygen–oxygen repulsion and hydrophobic–hydrophilic—with A291. TOPS-MODE predicts a large contribution for phenyl rings (0.580) and small contributions for carbonyl groups (−0.229), and these

coincide with the experimental findings according to which the hydrophobicity of phenyl rings plays a major role in the binding of these compounds to HSA. In the case of oxyphenbutazone, there are two orientations possible for the molecule, both of which place a hydrophobic group in the cavity formed by residues L260, I264, I290, R257, and R287 and allow hydrogen bonds between the carbonyl groups and residues R222, H242, and R257. The first orientation is one in which the phenyl ring is located in the hydrophobic cavity. In this case, the phenol ring has an orientation in which its hydroxyl group is in a cavity formed by residues L219, F223, L234, and I264. These residues are all hydrophobic, and there will be strong destabilizing interactions between the hydroxyl group and these amino acids. The other orientation is one in which the butyl group is located in the hydrophobic chamber formed by residues L260, I264, I290, R257, and R287 with the phenolic moiety directed toward the mouth of the cavity. This is in fact the conformation which is observed for this drug. One of the carbonyl groups forms a hydrogen bond with R222 and the other with R257. The first carbonyl also has destabilizing interactions with L238, L219, I264, and I290 (residues are identified in Figure 4A). The contribution predicted by TOPS-MODE for the butyl group is 0.479. This contribution is lower than the one for the phenyl ring of phenylbutazone. However, it is larger than that for the phenolic moiety (0.465), which is characterized by the small contribution (−0.045) of the hydroxy group. Similar binding characteristics and group contributions are observed for CMPF as depicted in Figure 5C. This compound is not in our original data set, but its complex with HSA has been characterized by X-ray crystallography, which showed that the furan oxygen forms a hydrogen bond with R222 and the carbonyl oxygen of the 3-carboxy moiety forms a hydrogen bond with R257. The hydrophobic methyl and propyl groups are located in the hydrophobic cavity formed by L260, I264, I290, R257, and R287 as well as L238, L219, and F223.

The analysis of the binding characteristics of these four compounds—warfarin, phenylbutazone, oxyphenbutazone,



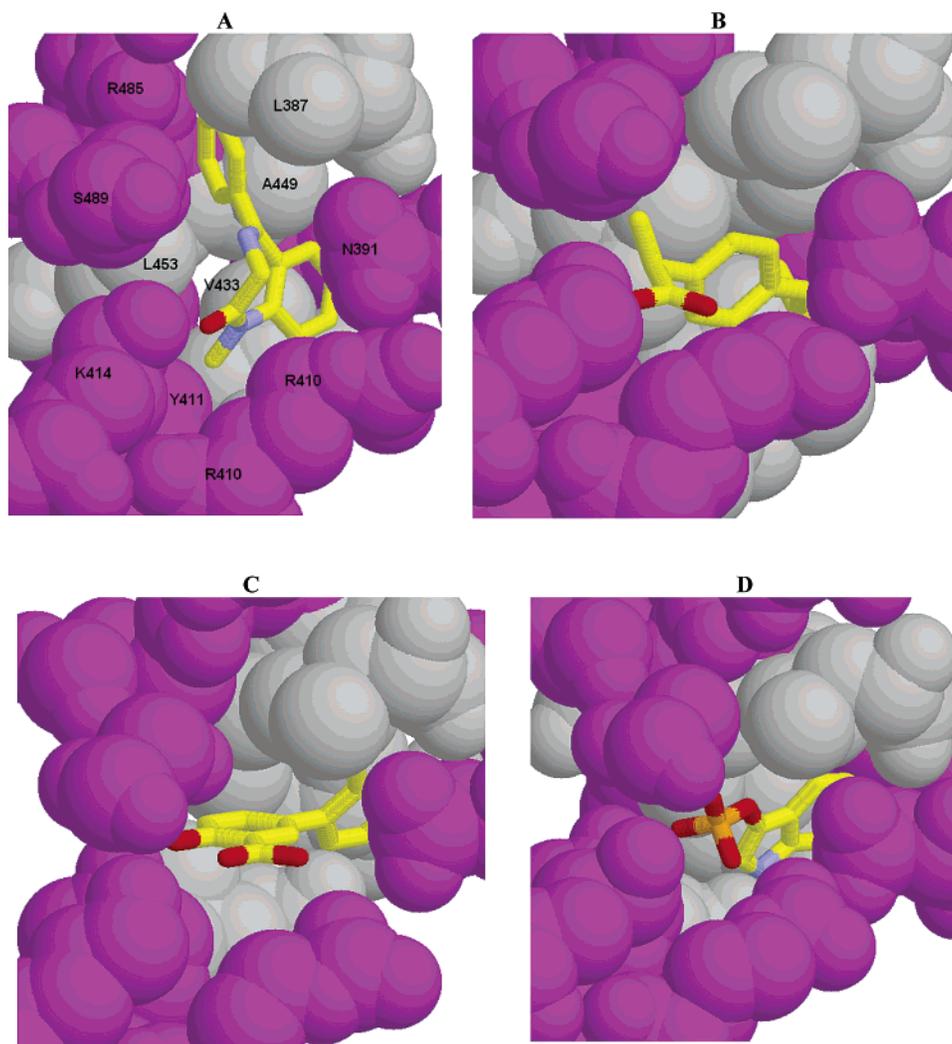
**Figure 5.** Illustration of binding site 1 for the drug–HSA complexes of phenylbutazone (A), oxyphenbutazone (B), CMPF (C), and indomethacin (D).



**Figure 6.** Schematic representation of a possible pharmacophore for the interaction of drugs at the center of binding site 1 of HSA (A). The existence of this pharmacophoric pattern in three drugs studied here is shown: warfarin (B), phenylbutazone (C), and oxyphenbutazone (D).

and CMPF—indicates the presence of a common structural pattern, or pharmacophore, which consists of a triangle formed by a hydrophobic center and two oxygen atoms which form hydrogen bonds with R222 and R257 or H242. These two oxygen atoms are separated by an average distance of 4.73 Å. The other vertex of this triangle is formed by a hydrophobic moiety, a phenyl ring, or an alkyl group. This hydrophobic moiety is separated at an average distance of

3.67 Å from one of the oxygen atoms and 4.64 Å from the other (see Figure 6). Confirmation of this hypothesis comes from the experimental determination of heterocyclic ring binding to HSA.<sup>49</sup> It has been reported that 4-hydroxycoumarin binds more efficiently than warfarin to HSA. With the exception of the benzyl and acetyl groups, 4-hydroxycoumarin has the same template as warfarin, and we hypothesize that this structural pattern improves the binding



**Figure 7.** Illustration of binding site 2 for the drug–HSA complexes of diazepam (A), ibuprofen (B), diflunisal (C), and 3-indoxyl sulfate (D).

at the center of binding site 1, but it is not essential for the binding of a drug to this site. A simple confirmation of this fact is that indomethacin, which does not possess this structural pattern, also binds at drug site 1, but it binds in an unusual way. As can be seen in Figure 5D, indomethacin is located in the cleft formed by L198, F206, A210, F211, W214, L481, V343, and L347. This could be due to the absence of the two oxygen atoms of the putative pharmacophore which, when present, serve to anchor drugs between residues R222 and H242 or R257. According to TOPS-MODE, the chlorophenyl moiety of indomethacin makes a large contribution (0.886) to the HSA binding. This moiety is located in the hydrophobic cavity formed by A210, F211, W214, and L481. The other large contribution comes from the indole ring, located between W214 and residues F211 and K199 and predicted by TOPS-MODE to make a contribution of 0.806. According to TOPS-MODE the carbonyl, carboxyl, and methoxyl groups make only small contributions to the HSA binding. The oxygen of the methoxyl group has only destabilizing hydrophobic–hydrophilic interactions with A215 and L238. The carbonyl has both stabilizing and destabilizing interactions, but the latter, in particular, the hydrophobic–hydrophilic interactions with L198, are dominant. The carboxyl group has mainly stabilizing interactions, appearing to participate in a bidentate salt bridge to R218.

**Binding at Drug Site 2.** The crystallographic structures of several complexes of drugs bound at binding site 2 of HSA have been reported, and information concerning binding site 2 in combination with TOPS-MODE can be derived from these reports. Ghuman et al.<sup>14</sup> reported the structures of diflunisal, diazepam, ibuprofen, and 3-indoxyl sulfate bound to HSA (see Figure 7). These compounds are characterized by a polar group separated from a hydrophobic core consisting mainly of one or more phenyl rings with hydrophobic substituents. These compounds are seen to cluster in the center of the binding pocket of subdomain IIIA, in such a way that the polar groups are oriented to the area surrounded by R410, Y411, and K414. TOPS-MODE predicts small contributions of the polar groups of these drugs interacting with the hydroxyl group of R410. However, according to Ghuman et al., none of the drugs interact with K414, which is also at the polar patch of this binding site. The larger contributions predicted by TOPS-MODE involve the hydrophobic groups. The isobutyl–phenyl moiety of ibuprofen has the largest contribution (1.137); the difluorophenyl moiety of diflunisal has a contribution of 0.898; the chlorophenyl moiety of diazepam has a contribution of 0.822, and the unsubstituted phenyl ring in 3-indoxyl sulfate has a contribution of 0.596.

Clofibrate (*p*-chlorophenoxyisobutyrate; **86**) binds to drug site 2 and is characterized by a polar head formed by the isobutyrate moiety and a hydrophobic tail formed by the *p*-chlorophenyl ring. According to TOPS-MODE, the hydrophobic region has the larger contribution (0.807) to HSA binding, the polar head providing a smaller contribution (0.308). It has been reported<sup>10</sup> that digitoxin (**49**) is also bound to binding site 2, but this site 2 is small and narrow, and large ligands, such as bilirubin, hemin, and hematin, do not bind to it, and digitoxin would appear to be too large to occupy this site. Hage and Sengupta<sup>50</sup> have found that digitoxin does not compete with warfarin or L-tryptophan, and this supports a model in which HSA has a separate binding site for digitoxin-related compounds.

Analysis of binding sites 1 and 2 is useful but does not account for the high-affinity binding of all drugs. Several drugs bind to more than one of these sites, and other sites are known to exist in HSA. Probenecid, amitriptyline, and debrisoquine do not bind significantly to either of the two sites analyzed here,<sup>11</sup> and a site-oriented approach to drug binding to HSA is unable to explain all of the experimental findings concerning drug–HSA interactions. Further, there are several drugs that interact irreversibly with HSA, producing structural modifications of the protein at single amino acids. Examples of this include acylation by aspirin, modification of lysine residues by penicillin, or covalent binding of ethacrynic acid to thioglycolic acid.<sup>51</sup> All of these characteristics make difficult the use of docking approaches based on known X-ray structures for data sets of large structural variability, such as the one used in the current work. However, the TOPS-MODE QSAR model permits the prediction of the binding constants for different drugs, and at the same time, group/fragment contributions permit an understanding of the structural basis of the binding of these drugs to HSA. These contributions can be useful for implementing structure-based algorithms for the fast prediction of HSA–drug binding as well as for understanding experimental findings on the structure of such complexes. In addition, group/fragment contributions could be useful in further studies toward the classification of drugs according to their possible binding sites at HSA.

**Group Contributions and Binding Energetics.** We have used the equation of Andrews et al.<sup>52</sup> as modified by Williams et al.<sup>53,54</sup> to estimate the energy contribution to the binding constant of the different groups. The modification introduced by Williams et al.<sup>53</sup> accounts for the hydrophobic effect, which is assumed to be proportional to the surface area *A* of the hydrocarbon (using Williams et al.<sup>53,54</sup> terminology) that is buried, that is, removed from exposure to water upon formation of the complex. This contribution is estimated in terms of  $\Delta G_h$  per unit area *A* of the buried molecule and is considered to be equal to  $A\Delta G_h$ . The binding constant between HSA and a compound in aqueous solution follows from the equation:<sup>53,54</sup>

$$\Delta G = \Delta G_{t+r} + n\Delta G_r + \Delta G_h + \sum \Delta G_p \quad (6)$$

where  $\Delta G_{t+r}$  and  $\Delta G_r$  respectively are the energy cost of protein–substrate association and that of the restriction of internal rotors, terms that are estimated to contribute approximately +1.29 and +0.33 kcal mol<sup>-1</sup>.<sup>53</sup>  $\Delta G_h$  is the hydrophobic energy per angstrom<sup>2</sup> of buried hydrocarbons,

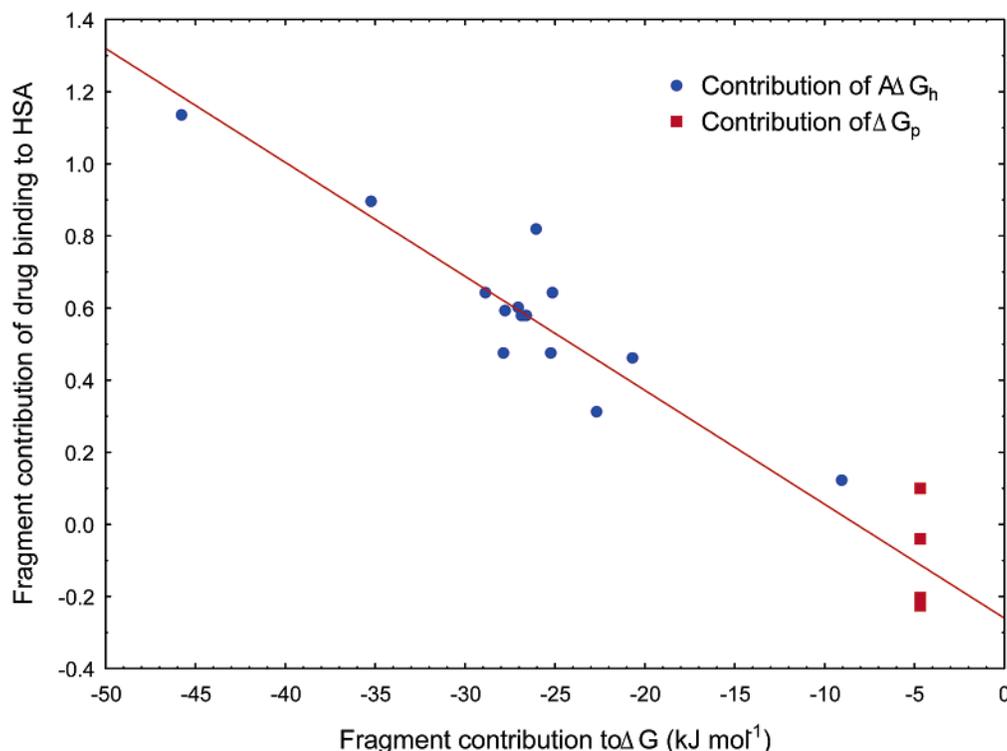
which is estimated at  $-0.04$  kcal mol<sup>-1</sup> Å<sup>-2</sup>.<sup>53</sup> The term  $\sum \Delta G_p$  is the sum of the free energies of binding for all polar interactions between the protein and the substrate. It is comprised of two contributions:<sup>55</sup>  $\Delta G_{\text{ionic}}$ , accounting for ionic interactions, and  $\Delta G_p$ , the energy of a hydrogen bond with optimal geometry. This term contributes  $-1.12$  kcal mol<sup>-1</sup> to the overall binding energy.<sup>55</sup>

Of all of the terms included in eq 6, two are used here to estimate the contribution of a fragment to the energy of drug–HSA binding. These are the energy  $A\Delta G_h$  resulting from hydrophobic interactions and  $\Delta G_p$  from hydrogen bonds. In these calculations, we have considered the interactions of drugs at binding sites 1 and 2 for which X-ray crystallographic structures exist. It is assumed that hydrogen bonds at sites 1 and 2 have “ideal geometry”, a condition which may not be met in all cases. Using the LPC approach (see the Materials and Methods), we have calculated the buried area of hydrophobic substructures interacting with the hydrophobic sites of HSA.

The interaction of warfarin with HSA at binding site 1 is useful for illustrative purposes. The solvent-accessible surface area of free warfarin according to our calculations with LPC is 524.0 Å<sup>2</sup> in the free molecule and only 24.6 Å<sup>2</sup> in warfarin bound to HSA. Thus, the total buried area is 499.4 Å<sup>2</sup>, and of this, the phenyl substituent contributes 148.4 Å<sup>2</sup> and the phenyl ring of the coumarin 159.7 Å<sup>2</sup>. Accordingly, they make contributions, respectively, of  $-6.02$  and  $-6.47$  kcal mol<sup>-1</sup> to the warfarin–HSA binding. Assuming that the hydrogen bonds involving the carbonyl and hydroxyl groups of the coumarinic moiety have ideal geometries, each of them will contribute  $-1.12$  kcal mol<sup>-1</sup> to the binding of warfarin to HSA. We recall that TOPS-MODE contributions for the four mentioned groups are 0.646, 0.602,  $-0.203$ , and  $-0.037$ , respectively.

In Figure 8, we illustrate the linear correlation obtained when the TOPS-MODE estimates of contributions of fragments to HSA binding are plotted against the contributions of the fragments to  $\Delta G$  of binding, as derived above. The correlation coefficient ( $R = 0.96$ ) indicates excellent relationships between both types of fragment contributions. It is clear that there is much oversimplification in the calculation of the group contributions to  $\Delta G$ . We are considering only contributions from hydrophobic interactions or hydrogen bonds, the latter assumed to have an optimal geometry. Despite this oversimplification, however, there is a clear relationship between the fragment contributions to  $\Delta G$  and the data derived by TOPS-MODE, and it appears that TOPS-MODE accounts for a large part of the energy associated with the interaction of drugs with HSA. The major contributions to the binding energy are the hydrophobic effect and hydrogen-bond formation. The first of these contributions appears to play a major role in the binding energetic of drugs to HSA. In other words, we can say that the fragment contributions found by us are mainly due to hydrophobic effects (largest contributions) and to hydrogen-bond formation (lower contributions).

It is known that the binding of drugs at site 1 is controlled by an entropic factor due to the hydrophobic effect.<sup>56,57</sup> However, binding to site 2 has been observed to involve electrostatic and hydrophobic interactions.<sup>56,57</sup> Our finding that hydrophobicity plays a major role in drug binding to HSA might be a consequence of the data set used in this



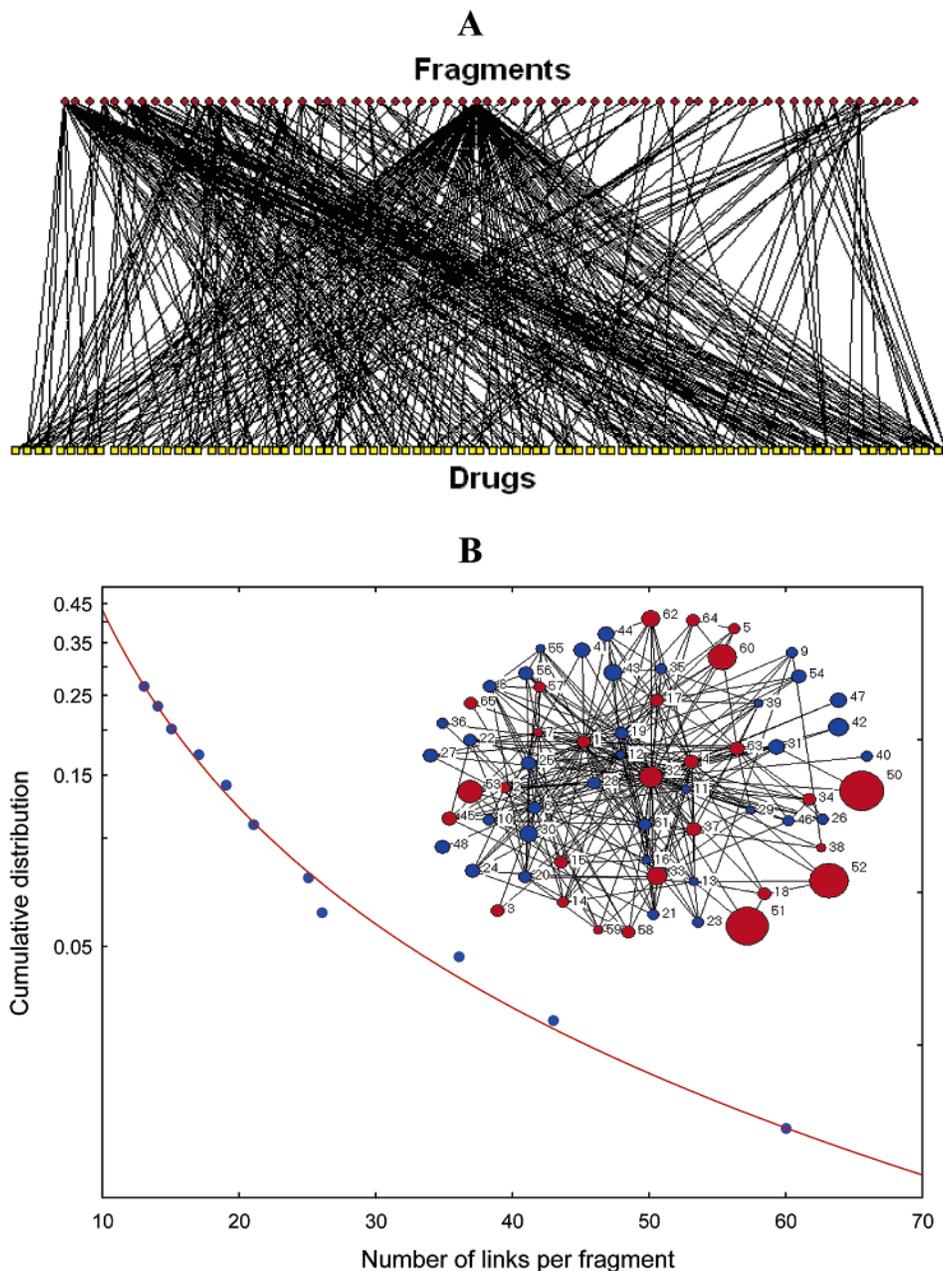
**Figure 8.** Plot of fragment/group contributions to the binding energetic of drug–HSA complexes, restricted to hydrophobic effect and hydrogen bonding, and the contributions to drugs binding to HSA according to TOPS-MODE. The linear plot displays a correlation coefficient of 0.96.

work. Ermondi et al.<sup>23</sup> found different influences of hydrophobicity on the binding affinities determined by ultracentrifugation for neutral, basic, and acidic drugs. While lipophilicity appears to have no effect on the binding of acidic compounds, which should bind at site 2, neutral and basic drugs are significantly influenced by hydrophobic effects. However, in the chromatographic data reported by Colmenarejo et al.,<sup>18</sup> there is a linear relationship between the binding affinity and lipophilicity and no difference between neutral, basic, and acidic drugs.<sup>23</sup> According to these authors, this could be caused by “chromatographic artifacts” because bound HSA is “conformationally modified and is unable to respect its natural three-dimensional structure”.<sup>23</sup>

**Fragment–Fragment Complex Networks.** Networks have attracted considerable recent attention in different scientific disciplines as representations of a variety of complex systems ranging from biological and social to technological and informational networks.<sup>58–60</sup> A study of complex networks can give insights into the organizational and evolutionary principles behind the systems they represent. The nodes of these networks represent the elements of the systems, exemplified by metabolites, proteins, genes, or neurons in biological networks, individuals in social networks, Web pages on the World Wide Web, or scientific papers in informational networks,<sup>58–60</sup> and the edges, which are links between nodes, represent the relationships between the nodes of the system. Two properties of complex networks are common to almost all complex systems. These are the so-called “small-worldness”<sup>61</sup> and “scale-freeness” characteristics.<sup>62</sup> The first is related to the property of a network which has, on average, a small distance between all node pairs in the network but a large clustering, generally represented by the proportion of triangles to connected triples

(pairs of adjacent links) of nodes.<sup>61</sup> The second is related to the statistical distribution of the node degrees. The node degree is the number of ties incident to a particular node. In scale-free networks, this distribution follows a power law scale of the form  $p(k) \sim k^{-\eta}$ , where  $p(k)$  is the probability of finding a node with degree  $k$  in the network.<sup>62</sup> It is in these two properties that real-world networks differ significantly from random networks, which are not “small-worlds”, generally having minimal clustering and a normal distribution of node degrees—they are not in fact “scale-free”.<sup>59–61</sup>

Here, we investigate these properties in the networks of fragments contributing to the binding of drugs to HSA. We first build a bipartite network of fragments-to-molecules as illustrated in Figure 9A. Here, a link exists between a drug and a fragment only if the drug contains that fragment. Thus, each fragment links to many drugs, and each drug links to many fragments. Possible projections of this network give rise to a drug–drug or a fragment–fragment network. In the first case, all nodes represent drugs, and two drugs are connected if they share at least one fragment. In the second case, all nodes represent fragments, and two fragments are connected if they appear simultaneously in at least one drug. We have considered here the fragment–fragment network which can give us some insight into the organizational principles of these fragments in drugs and their binding to HSA. When the 65 fragments selected previously in this work are used, a fragment–fragment network can be built and is immediately seen to have just one isolated node. This corresponds to fragment number 49, which appears only in caffeine. The remaining 64 fragments form a connected network with 343 links. A link means that the two fragments that are linked appear simultaneously in one or more drugs.



**Figure 9.** (A) Bipartite network of fragments to drugs in which nodes at the top represent each of the 65 fragments found in this work to contribute to HSA binding, and the nodes at the bottom represent each of the drugs in our data set. One fragment is connected to a drug if this fragment exists in the drug. (B) Linear-log plot of the cumulative distribution of the number of links *per* fragments in the fragment–fragment network. In this network, nodes represent the fragments found to contribute to HSA binding, and two fragments are connected if they appear simultaneously in at least one drug. The upward curved line is characteristic of power-law (scale-free) distributions.

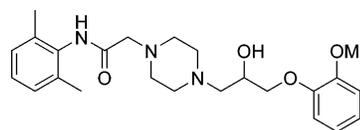
The principal objective of studying this fragment–fragment network is to capture some of the intuition and experience of drug development stored as structural information in successful drugs, such as that included in our data set. The same objective was successfully pursued by Bemis and Murcko,<sup>63,64</sup> who investigated the frequency with which different fragments appear among a large set of drugs. This kind of study is quite important in guiding the selection of fragments for *de novo* design strategies of new drugs or for the combinatorial synthesis of new structures. In the Bemis and Murcko<sup>63,64</sup> strategy, each fragment is represented by means of its frequency of occurrence, or “abundance”, in the set of drugs analyzed. For instance, the phenyl ring appears 433 times in the 5120 drugs analyzed,<sup>63</sup> while the carbonyl group appears 3545 times in 5090 compounds.<sup>64</sup>

Here, we use a different measure: the number of links a fragment has in the fragment network. This number—also known as the node degree—represents the frequency with which a fragment appears together with other fragments in each of the drugs analyzed. Saying in other words, node degree represents the “molecular empathy” of the corresponding fragment to the rest of fragments that can be present in drugs. A fragment with a large node degree has a large “molecular empathy” as it can appear together with many other fragments in a molecule. On the contrary, a fragment with a low node degree is the one that in general appears together with very few other fragments in drugs. A better way of assessing the molecular empathy of a fragment is by dividing its degree by  $N - 1$  so that the empathy ranges from 0 to 1, where  $N$  is the number of fragments considered.

The mean node-to-node distance in this network is  $l = 1.865$ , and the Watt–Strogatz clustering coefficient is  $C^1 = 0.773$ . We have generated 100 random networks having the same average degree and node degree distribution as this fragment–fragment network. For these networks, the mean distance and clustering are  $l = 1.862$  and  $C^1 = 0.590$ . Thus, the fragment–fragment network studied here is a “small-world” network as it has a mean node–node distance similar to that in random networks with the same average degree and node degree distribution, but it is somewhat more clustered than the random networks. The average of 100 Erdős–Rényi random networks with the same average degree (but not the same degree distribution) as our fragment–fragment network has  $l = 1.932$  and  $C^1 = 0.298$ , confirming that the fragment–fragment network is a “small-world” network. We have studied the cumulative distribution of node degrees for the fragment–fragment network, and in Figure 9B, we illustrate the linear log plot of the cumulative distribution of node degrees. Here, it can be seen that for nodes with  $k \geq 6$  a perfect power-law distribution is evident. The power law followed by this network is  $p(k) \sim k^{-1.81}$  with a correlation coefficient of  $R = 0.998$ . In Figure 9B, we show fragments with the largest contributions (positive in log scale) to drug binding to HSA as red nodes and those having the lowest contributions (negative in log scale) as blue nodes. In the case of the larger contributions (red nodes), the larger the area of the node is, the more positive the contribution to drug–HSA binding. In the case of the lowest contributions (blue nodes), the larger the area of the node is, the lower the contribution (most negative) to the drug–HSA binding. The fragment–fragment network studied here is therefore both a “small-world” and “scale-free” network.

The consequences of these network characteristics are evident from the following analysis. First, the small-worldness of the fragment–fragment network indicates that the cliquishness of this network is larger than expected from a random process of assembling fragments into drug molecules which would be expected to give many more outcomes. This process of course is not random but is guided by certain unrelated constraints imposed by the chemist. These include the synthetic feasibility of incorporating these fragments into stable molecules and the necessity to produce good “drug-likeness” in the molecules (e.g., ADME properties such as drug binding to HSA), patent considerations, cost, or a general conservatism. This large clustering indicates that in the fragments studied in drug molecules there is a large transitivity, understood as the property of forming triangles in the network. For instance, if fragment A appears together with fragment B in at least one drug and together with fragment C in at least one drug, then there will be a high probability of finding fragments B and C together in at least one drug. In fact, the number of triangles in the fragment–fragment network is 840, larger than the 682.6 expected in random networks. The “scale-free” properties of this network indicate that there are only a few fragments which appear together with other fragments in a very large number of drugs. These are fragments of large molecular empathy. On the other hand, “scale-freeness” indicates that there are a large number of fragments which appear together with other fragments but in a very low number of drugs. That is the case of fragments with low molecular empathy as, for instance, aliphatic Cl or certain heterocyclic rings,

Scheme 1



such as **36**, **38**, **40**, **41**, **42**, **47**, **48**, and **65** in Table 3. According to this analysis, the most empathetic fragments are (empathy is given in parentheses): phenyl ring (0.94); methyl bonded to aliphatic carbon (0.67); OR bonded to aromatic carbon (0.56); aliphatic OH (0.41); and aromatic chlorine (0.39). Fragment abundance and empathy are related ( $r^2 = 0.83$ ), but they introduced different ordering for the same fragments. For instance, aromatic OR has an empathy of 0.55, and aliphatic OH has an empathy of 0.4, despite the fact that aliphatic OH appears 27 times in the drugs studied and aromatic OR appears only 23 times. The same happens with other fragments, such as aromatic OCH<sub>3</sub> with an empathy of 0.29 and an abundance of 9 and aromatic sulfonamide with an empathy of 0.26 and an abundance of 10. A tertiary aliphatic amine is less empathetic than aliphatic carboxylic acid (empathies of 0.21 and 0.29, respectively) despite the first being more abundant than the second (appearing nine and eight times, respectively).

As an example of the possible utility of this molecular empathy parameter, we studied 13 drugs (lenalidomide, tipranavir, entecavir, ibandronate, ramelteon, nelarabine, sorafenib, conivaptan, ranolazine, lubiprostone, decitabine, sunitinib, and varenicline) approved by the Food and Drug Administration in the period 2005–2006. The objective here was to investigate the frequency with which the most empathetic fragments appear in these drugs, which all have acceptable ADME properties. The fragment which appears more frequently in these 13 drugs is the phenyl ring that appears 14 times and is also the most empathetic fragment according to our network analysis. The second most frequent fragment is the aliphatic OH which appears 11 times and is the fourth most empathetic fragment. In the third place, according to the frequency of appearance, are aliphatic and aromatic methyl groups, both of which appear five times. The aliphatic methyl group is the second most empathetic fragment according to our analysis. The rest of the fragments appear less than four times in all of these drugs. Ranolazine (see Scheme 1), for example, an orally available antianginal compound, contains 10 different fragments from those studied in this paper. These include two phenyl rings, two aromatic methyl groups, one amide, two trisubstituted amino groups (the piperazinyl moiety), one aliphatic OH, one aromatic OR group, and one aromatic OMe. The average molecular empathy of these fragments is 0.413, indicating that this drug is made by linking some of the most empathetic fragments considered here. This suggests that a list of the most empathetic fragments can be useful as a guide for selecting fragments in the design of new druglike compounds or for the modification of existing ones. More importantly, this methodology can guide further studies toward the identification of fragments useful in drug design and combinatorial chemistry, integrating network strategies into medicinal chemistry.

## CONCLUDING REMARKS

We have carried out an analysis of drug binding to HSA by integrating different theoretical strategies and approaches. Starting from the generation of a QSAR model with the use of TOPS-MODE, we have been able to find the contribution of different structural fragments and groups to the binding of drugs to HSA. These fragment/group contributions have been analyzed in the context of the specific binding of drugs at certain sites of HSA, such as drug binding sites 1 and 2. Finally, we have used the set of fragments contributing to the binding affinity of drugs to HSA to build a complex network representing the common occurrence of these fragments in the drugs studied. In this way, we have integrated theoretical approaches of “classical” medicinal chemistry with techniques of statistical physics into a single unified approach to an understanding of drug binding to HSA. We have shown that the QSAR model obtained with TOPS-MODE displays statistical characteristics better than or similar to those of other QSAR models obtained for the same data set but using other molecular descriptors or statistical approaches. Our QSAR model is robust and displays good predictability as illustrated with an external prediction set. Most important is the fact that this QSAR model fulfils the requirements of predictability/interpretability claimed for theoretical models. The fragment/group contributions to binding affinity represent a clear way for interpreting the contribution of the molecular structure to the drug–HSA complexes. We have shown that agreement exists between these contributions and the specific interactions of these groups/fragments with amino acid residues in HSA. Our results indicate a preponderant contribution of hydrophobic regions of drugs to the specific binding to drug binding sites 1 and 2. However, important roles are played by polar groups such as two oxygen atoms separated by four bonds, which form hydrogen bonds at site 1 or polar groups in certain drugs which interact with a polar area in binding site 2. Finally, the study of fragment–fragment complex networks has permitted the understanding of certain organizational principles of fragments contributing to drug binding to HSA. We have seen that very few fragments appear commonly in most drugs showing a large “molecular empathy”. These fragments can be used for the design of new druglike compounds as well as for combinatorial chemistry protocols.

## ACKNOWLEDGMENT

E.E. thanks the program Ramón y Cajal, Spain, for partial financial support. E.M. thanks the University of Santiago de Compostela for a postdoctoral fellowship.

**Supporting Information Available:** A table with the bond contribution for all bonds in the molecules studied is provided (47 pages). This information is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- Butina, D.; Segall, M. D.; Frankcombe, K. Predicting ADME Properties *in Silico*: Methods and Models. *Drug Discovery Today* **2002**, *7*, S83–S88.
- van de Waterbeemd, H.; Gifford, E. ADMET *in Silico* Modeling: Towards Prediction Paradise? *Nat. Rev.* **2003**, *2*, 192–204.
- Delisle, R. K.; Jowrie, J. F.; Hobbs, D. W.; Diller, J. Computational ADME/Tox Modeling: Aiding Understanding and Enhancing Decision Making in Drug Design. *Curr. Comput.-Aided Drug Des.* **2003**, *1*, 325–345.
- Bidault, Y. A Flexible Approach for Optimizing *in Silico* ADME/Tox Characterization of Lead Candidates. *Exp. Opin. Drug Metab. Tox.* **2006**, *2*, 157–168.
- Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361–1375.
- Dimasi, J. A. Risk in New Drug Development Approval Success Rates for Investigational Drugs. *Clin. Pharmacol. Ther.* **2001**, *69*, 297–307.
- Urien, S.; Tillement, J. P.; Barré, J. The Significance of Plasma – Protein Binding in Drug Research. In *Pharmacokinetic Optimization in Drug Research*; Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R. H., Eds.; Wiley VCH: Zürich, Switzerland, 2001; pp 189–197.
- Peters, T., Jr. *All About Albumin, Biochemistry, Genetics, and Medical Applications*; Academic Press: New York, 1996; pp 79–131.
- Carter, D. C.; Ho, J. X. Structure of Serum Albumin. *Adv. Protein Chem.* **1994**, *45*, 152–203.
- Xiao, M. H.; Carter, D. C. Atomic Structure and Chemistry of Human Serum Albumin. *Nature (London)* **1992**, *358*, 209–215.
- Sudlow, G.; Birkett, D. J.; Wade, D. H. Characterization of Two Specific Drug Binding Sites on Human Serum Albumin. *Mol. Pharmacol.* **1975**, *12*, 1052–1061.
- Petitpas, I.; Bhattacharya, A. A.; Twine, S.; East, M.; Curry, S. Crystal Structure Analysis of Warfarin Binding to Human Serum Albumin. *J. Biol. Chem.* **2001**, *276*, 22804–22809.
- Bhattacharya, A. A.; Curry, S.; Franks, N. P. Binding of General Anesthetics Propofol and Halothane to Human Serum Albumin: High-Resolution Crystal Structures. *J. Biol. Chem.* **2000**, *275*, 38731–38738.
- Ghuman, J. J.; Zunsain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. Structural Basis of the Drug-Binding Specificity of Human Serum Albumin. *J. Mol. Biol.* **2005**, *353*, 38–52.
- Kragh-Hansen, U.; Chuang, V. T. G.; Otagiri, M. Practical Aspects of the Ligand-Binding and Enzymatic Properties of Human Serum Albumin. *Biol. Pharm. Bull.* **2002**, *25*, 695–704.
- Colmenarejo, G. *In Silico* Prediction of Drug-Binding Strengths to Human Serum Albumin. *Med. Res. Rev.* **2003**, *23*, 275–301.
- Kratochvil, N. A.; Huber, W.; Müller, F.; Kansy, M.; Gerber, P. R. Predicting Plasma Protein Binding of Drugs: Revisited. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 507–512.
- Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J. L. Cheminformatic Models to Predict Binding Affinities to Human Serum Albumin. *J. Med. Chem.* **2001**, *44*, 4370–4378.
- Hall, L. M.; Hall, L. H.; Kier, L. B. Modeling Drug Binding Affinity with E-State Topological Structure Representation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2120–2128.
- Xue, C. X.; Zhang, H. X.; Liu, X. J.; Yao, M. C.; Liu, Z. D.; Fan, B. T. QSAR Models for the Prediction of Binding Affinities to Human Serum Albumin Using the Heuristic Method and a Support Vector Machine. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1693–1700.
- Gunturi, S. B.; Narayanan, R.; Khandelwal, A. *In Silico* ADME Modelling 2: Computational Models to Predict Human Serum Albumin Binding Affinity Using Ant Colony Systems. *Bioorg. Med. Chem.* **2006**, *14*, 4118–4129.
- Diaz, N.; Suárez, D.; Sordo, T. L.; Merz, K. M., Jr. Molecular Dynamics Study of the IIA Binding Site in Human Serum Albumin: Influence of the Protonation State of Lys195 and Lys199. *J. Med. Chem.* **2001**, *44*, 250–260.
- Ermondi, G.; Lorenti, M.; Caron, G. Contribution of Ionization and Lipophilicity to Drug Binding to Albumin: A Preliminary Step toward Biodistribution Prediction. *J. Med. Chem.* **2004**, *47*, 3949–3961.
- Estrada, E.; Uriarte, E. Recent Advance on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1573–1588.
- Estrada, E. Extracting Structural Information from Molecular Graphs. The TOPS-MODE Approach. *Adv. Quantum Chem.* In press.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhet, J. N.; Weissing, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- Soboler, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated Analysis of Interatomic Contacts in Protein. *Bioinformatics* **1999**, *15*, 327–332.
- Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graph. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844–849.
- Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graph. 2. Molecules Containing Heteroatoms and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320–328.

- (30) Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graph. 3. Molecules Containing Cycles. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23–27.
- (31) Estrada, E.; Peña, A.; García-Domenech, R. Designing Sedative/Hypnotic Compounds from a Novel Substructural Graph – Theoretical Approach. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 583–595.
- (32) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach to the Rational Selection and Design of Anticancer Compounds. *J. Med. Chem.* **2000**, *43*, 1975–1985.
- (33) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
- (34) Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-Additive Method. *Perspect. Drug Discovery Des.* **2000**, *19*, 47–66.
- (35) Ertl, P.; Rhode, P.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Applications to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.
- (36) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure–Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21–35.
- (37) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (38) Gasteiger, J.; Marsilli, M. A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **1978**, *34*, 3181–3184.
- (39) Baskin, I. I.; Skvortsova, M. I.; Stankevich, I. V.; Zefirov, N. S. On the Basis of Invariants of Labeled Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 527–531.
- (40) Zefirov, N. S.; Palyulin, V. A. Fragmental Approach in QSPR. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1112–1122.
- (41) Randić, M. Resolution of Ambiguities in Structure–Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311–320.
- (42) Randić, M. Orthogonal Molecular Descriptors. *New J. Chem.* **1991**, *15*, 517–525.
- (43) Randić, M. Correlation of Enthalpy of Octanes with Orthogonal Connectivity Indices. *THEOCHEM* **1991**, *233*, 45–49.
- (44) Lučić, B.; Nicolčić, S.; Trinajstić, N.; Jurić, D. The Structure–Property Models Can Be Improved Using the Orthogonalized Descriptors. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 532–538.
- (45) Klein, D. J.; Randić, M.; Babić, D.; Lučić, B.; Nicolčić, S.; Trinajstić, N. Hierarchical Orthogonalization of Descriptors. *Int. J. Quantum Chem.* **1997**, *63*, 215–222.
- (46) Estrada, E.; Molina, E. Novel Local (Fragment-Based) Topological Molecular Descriptors for QSPR/QSAR and Molecular Design. *J. Mol. Graphics Modell.* **2001**, *20*, 54–64.
- (47) Estrada, E.; Molina, E. Automatic Extraction of Structural Alerts for Predicting Chromosome Aberrations of Organic Compounds. *J. Mol. Graphics Modell.* **2006**, *25*, 275–288.
- (48) Estrada, E.; Molina, E.; Perdomo-López, I. Can 3D Structural Parameters Be Predicted from 2D (topological) Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1015–1021.
- (49) Zafón, A. M. L.; Villamor, J. P. Study of Heterocycle Rings Binding to Human Serum Albumin. *Chem.-Biol. Interact.* **2000**, *124*, 1–11.
- (50) Hage, D. S.; Sengupta, A. Characterization of the Binding of Digitoxin and Acetyldigitoxin to Human Serum Albumin by High-Performance Affinity Chromatography. *J. Chromatogr., B* **1999**, *724*, 91–100.
- (51) Bertucci, C.; Domenici, E. Reversible and Covalent Binding of Drugs to Human Serum Albumin: Methodological Approaches and Physiological Relevance. *Curr. Med. Chem.* **2002**, *9*, 1463–1481.
- (52) Andrews, P. R.; Craik, D. J.; Martin, J. L. Functional Group Contributions to Drug–Receptor Interactions. *J. Med. Chem.* **1984**, *27*, 1648–1657.
- (53) Williams, D. H.; Cox, J. P. L.; Doig, A. J.; Gardner, M.; Gerhard, U.; Kaye, P. T.; Lal, A. R.; Nicholls, I. A.; Salter, C. J. Mitchell, R. C. Toward the Semiquantitative Estimation of Binding Constants. Guides for Peptide–Peptide Binding in Aqueous Solution. *J. Am. Chem. Soc.* **1991**, *113*, 7020–7030.
- (54) Williams, D. H.; Stephens, E.; O’Brien, D. P.; Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596–6616.
- (55) Böhm, H. J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein–Ligand Complex of Known Three-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243–256.
- (56) Urien, S.; Nguyen, P.; Berilos, S.; Brée, F.; Vacherot, F.; Tillement, J. P. Characterization of Discrete Classes of Binding Sites of Human Serum Albumin by Application of Thermodynamic Principles. *Biochem. J.* **1994**, *302*, 69–72.
- (57) Aki, H.; Yamamoto, M. Thermodynamic Characterization of Drug Binding to Human Serum Albumin by Isothermal Titration Microcalorimetry. *J. Pharmacol. Sci.* **1994**, *83*, 1712–1716.
- (58) Strogatz, S. H. Exploring Complex Networks. *Nature (London)* **2001**, *2*, 268–276.
- (59) Albert, R.; Barabási, A. L. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97.
- (60) Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256.
- (61) Watts, D. J.; Strogatz, S. H. Collective Dynamics of “Small-World” Networks. *Nature (London)* **1998**, *393*, 440–448.
- (62) Barabási, A. L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512.
- (63) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.
- (64) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42*, 5095–5099.

CI600274F