Theoretical Prediction of Partition Coefficients via Molecular Electrostatic and Electronic Properties

Olivier Lamarche,[†] James A. Platts,*[,†] and Anne Hersey[‡]

Department of Chemistry, Cardiff University, P.O. Box 912, Cardiff CF10 3TB, U.K., and Medicines Research Centre, GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY, U.K.

Received November 28, 2003

Previously published methods for calculation of Abraham's polarity/polarizability and hydrogen bond acidity and basicity descriptors are validated for their ability to predict the various partition coefficients of 80 challenging molecules. As well as this indirect validation, accurate log *P* predictions are shown to be possible by using directly the fundamental molecular properties used in the calculation of descriptors. From a general point of view, the van der Waals and hydrogen bond interactions present between the solute and the water/solvent system can be represented by charge-based interactions, which are partitioned into a positive and a negative term ($\Sigma^{T}V_{Max}$ and $\Sigma^{T}V_{Min}$), dipole-based interactions (μ) and induced dipole-based interactions (α); further stabilization is possible if solute and solvent densities come into contact and overlap ($\Sigma^{T}E$). A discussion is opened on the possibility to extend this set to describe systems with electron donor/acceptor interactions other than hydrogen bonding.

INTRODUCTION

A great deal of effort has been spent to obtain an accurate description of processes that involve transport of solutes—either in terms of equilibrium partition between phases or rate of transfer from one phase to another. The most commonly studied solvation property is the partition coefficient (log *P*) between phases, being related to $\Delta G°$ for transfer of the solute between phases. Many different approaches, such as atomic[1] or group contribution methods,[2−4] correlation with molecular properties via neural network[5] or multivariate analysis,[6] or Monte Carlo and molecular dynamics methods,[7] have been developed to predict partition coefficients and related solvation properties. Such methods have met with much success and are widely used in a great many fields. However, they are often restricted to a single system (e.g. octanol−water partition) and give few or no physicochemical insights.

A general physicochemical treatment of solvation effects is the Linear Solvation Energy Relationships (LSER) developed by Taft, Kamlet, Abraham, and co-workers.[8] This approach has been successfully used to describe a large number of solvation effects and offered much insight into them. The LSER encodes two major structural contributions, namely a cavity or volume-related term reflecting the energy needed to create a cavity in the solvent (an endoergic term, which cannot be dissociated from exoergic London dispersion forces) and an exoergic interactive term which results from solute−solvent interactions such as electrostatics and hydrogen bonds.[9]

Thus, a list of descriptors was drawn up to quantitatively account for these terms. To separate the general dispersion interaction from the volume term, an excess molar refraction, R_2, was defined. Several scales of H-bond acidity and basicity have been proposed, Taft's pK_{HB} scale being the first example.[10] This opened the way for Abraham et al. to develop more general scales for any ratio of acid and base, resulting in the $\Sigma\alpha_2^{H}$ and $\Sigma\beta_2^{H}$ scales.[11,12] Finally, the polarity and polarizability of a solute is also important. Initial attempts to use the dipole moment or the Kamlet and Taft's solvatochromic parameter π_1* were unsatisfactory, and a new polarity/polarizability descriptor, π_2^{H}, was therefore developed from a large set of gas−liquid chromatography data.[13] These four descriptors were then combined with McGowan's molecular volume (V_x),[14] into a general solvation equation, or LSER, as displayed in eq 1.[15] This methodology has since been applied to a very wide range of solvation properties.

$$\log SP = c + rR_2 + s\pi_2^{H} + a\Sigma\alpha_2^{H} + b\Sigma\beta_2^{H} + vV_x \quad (1)$$

Recently, a fast, automated method has been developed for the estimation of these descriptors based on fragmental contributions,[16] but the generality of this method is limited by the lack of experimental data for important fragments. We have therefore recently explored methods to predict these descriptors directly from the structure, to feed values into the fragmental model.

In previous studies, we showed that several calculated properties correlate with $\Sigma\alpha_2^{H}$, $\Sigma\beta_2^{H}$, and π_2^{H} with reasonable accuracy. Using the partial least-squares method, four properties stand out as predictors of π_2^{H}: the molecular dipole moment, polarizability, CHelpG atomic charges, and frontier molecular orbital energies.[17] Electrostatic descriptors give simple linear correlations with $\Sigma\alpha_2^{H}$ for monofunctional acids,[18,19] while free energies of hydrogen bond formation with HF were used for $\Sigma\beta_2^{H}$ calculations of monofunctional bases.[20,21] These models were then generalized to cope with molecules with any number of acidic and basic sites.

* Corresponding author phone: +44 2920 874950; fax: +44 2920 874030; e-mail: platts@cf.ac.uk.
[†] Cardiff University.
[‡] GlaxoSmithKline.

10.1021/ci034276c CCC: $27.50
Published on Web 02/27/2004

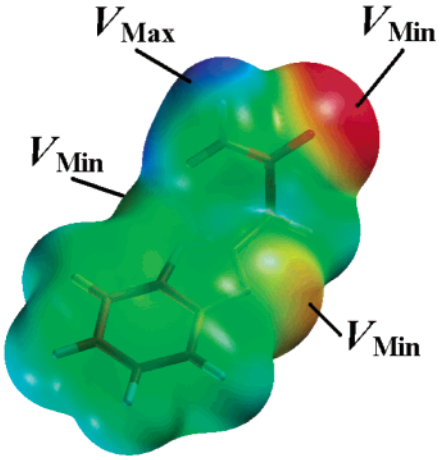

Figure 1. Surface minima and maxima of 2-(carbamylmethyl)-benzotriazole.

Following this work, the $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ scales are regarded as composite descriptors made up of an electrostatic term (by using minimum/maximum electrostatic potential values on the 0.001 au isodensity surface) and of an overlap term (by using energy density values at the bond critical point of 1:1 HF/NCH complexes).[22,23] These models also underlined the complementarity between the two overall hydrogen bond scales. This approach is illustrated in Figure 1, which shows the various surface minima and maxima for a typical organic compound.

As noted earlier, many other methods have been proposed for the estimation of solvation properties. The water/octanol system (log P_{oct}) is widely used, representing one of the most informative physicochemical parameters, with a great deal of reliable experimental data. Among atom/group contribution methods are Klopman's KLOGP, Rekker's fragmentation method, Hansch and Leo's "constitutive" CLOGP, or the computerized ACD/log *P* method.[1−4] These typically estimate log P_{oct} by identifying important substructures and atom types, either predefined or from some fragmentation scheme, and assigning a contribution to log P_{oct} from each fragment. These contributions are found either from multiple linear regressions or by comparison with measured values for fragments. Such methods are applicable to a few systems (e.g. water solubility or log P_{oct}), since large training sets of data are needed. Analyzing log P_{oct} in terms of fragment values is in essence arbitrary and can require a large number of correction factors.

Eisfeld and Maurer used ab initio quantum chemical calculations to develop solvation models based on molecular properties.[6c] Their log P_{oct} model was made of 8/10 descriptors, interpreted as cavity effects, electrostatic interactions, and hydrogen bonding, and gave an excellent standard deviation for relatively simple molecules (0.28). Famini and Wilson were the first workers to calculate solute descriptors for use in LSER, so-called Theoretical Linear Solvation Energy Relationship (TLSER) descriptors. This has been successfully applied to many solvation-related processes, such as Henry's Law constants, solubility in supercritical CO_2, and partition.[6e−g] However, it has recently been suggested that there are difficulties with TLSER descriptors for multifunctional solutes.[6e] Continuum solvation models have also been used to estimate partition coefficients, initially by Tomasi[6h] and more recently in Klamt's COSMO-RS method.[6i]

Our aim is to derive, from the solvatochromic descriptor models (of $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and π_2^H), a theoretical log P_{oct} model for *any* compound, which is made of molecular descriptors with straightforward physical meanings, and which has potential flexibility so it can be extended to any solvation properties (example here with two other partition coefficients: water/chloroform log P_{chl} and water/alkane log P_{alk}) and can incorporate other electron donor/acceptor complexes than the hydrogen bond ones.

CALCULATION METHODS

A set of 80 molecules was selected from the experimental database MedChem02.[24] These molecules were chosen to be chemically diverse and to cover a wide numerical spread of log P_{oct} values. This constitutes a challenging set due to the structural complexity and the numerous multifunctional groups present in some large molecules; many of which are drug-like or common fragments of drugs. The molecules are tabulated in Table 1, along with a sequential number, their experimental log P_{oct}, and, when available, their experimental log P_{chl} and log P_{alk}. While designing the set, one of the main interests was to find molecules with unknown experimental solvatochromic descriptors ($\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and π_2^H; i.e. few known solvation properties), which explains the small number of log P_{chl} and log P_{alk} values present in Table 1.

All ab initio and DFT calculations were performed using GAUSSIAN98[25] running on a Compaq XP1000 workstation. Initially, the geometries of the molecules were optimized at the HF/6-31G(d)[26,27] level, and the resulting structures were confirmed as minima through harmonic frequency calculations. The conformational space of flexible molecules was explored at the same level to ensure that the final optimized structure corresponded to the global minimum—the effect of varying conformation will be discussed in more detail below. Using these HF/6-31G(d) geometries, the magnitude of the dipole moment μ, which describes the charge distribution, and the trace of the molecular polarizability tensor α, defined as the change in dipole moment in response to an applied electromagnetic field, were computed at the B3LYP/6-31++G(d,p) level,[28,29] i.e. consistent with previously published μ and α calculations for predictions of π_2^H.[17]

Following previous publications on H-bond scales,[21−23] the geometries of the isolated molecules were reoptimized at the B3LYP/6-31+G(d,p) level, and their corresponding 1:1 complexes with hydrogen fluoride were also optimized at this level. The resulting structures were confirmed as minima through harmonic frequency calculations. A range of molecular properties, including local minima and maxima on the surface electrostatic potential and energy densities at the H-bond critical point, which were successful in the Atoms In Molecules (AIM) [30] approach for modeling H-bond scales,[22,23] were evaluated for the isolated molecules and for all their corresponding 1:1 HF complexes.

The molecular electrostatic potential created in a space around the isolated molecule has been found to be well suited for the analysis of pK_{HB} and noncovalent interactions in general.[31] B3LYP/6-31+G(d,p) geometries were used to construct regular 6 Å sided cubes (100 × 100 × 100 points) of electronic density and electrostatic potential, centered on the nuclei of interest. From these two grids, an in-house

Table 1. Experimental Partition Coefficient Values from MedChem02[24]

no.[b]	name	log P_{oct}	log P_{chl}	log P_{alk}	no.[b]	name	log P_{oct}	log P_{chl}	log P_{alk}
1	2,4,6-triamino-1,3,5-triazine[a]	−1.37	−1.39		41	6-methyl-2-pyrimidone[a]	−1.45		
2	uracil	−1.07	−1.70	−3.89	42	hypoxanthine[a]	−1.11		
3	methylformamide	−0.97			43	methylacetamide	−1.05		
4	3-carbamoylpyridazine[a]	−0.73			44	methanol	−0.77	−1.36	−2.75
5	2-amino-1,3,4-thiadiazole[a]	−0.57			45	tetrazole[a]	−0.60		
6	methylamine	−0.57	−1.00	−2.04	46	trimethyl phosphate	−0.52	0.76	−2.20
7	1,4-dimethylpiperazine	−0.40	−0.20	−1.71	47	pyrimidine	−0.40	0.32	−1.56
8	dimethylamine	−0.38	−0.44	−1.55	48	purine	−0.37	−1.95	
9	nitromethane	−0.35	0.44	−0.98	49	acetonitrile	−0.34	0.40	−1.24
10	acetone	−0.24	0.59	−1.01	50	ethanol	−0.31	−0.85	−2.14
11	1,3,7-trimethylxanthine[a]	−0.07	1.33	−2.21	51	1-ethylthymine[a]	−0.15		
12	2,5-dimethyl-1,3,4-thiadiazole[a]	0.03			52	ethylamine	−0.13	−0.35	−1.79
13	cyclopropamine[a]	0.07			53	9-methyladenine[a]	−0.03	−0.78	
14	trimethylamine	0.16	0.51	−0.46	54	N-(3-pyridylmethyl)morpholine[a]	0.04		
15	acrylonitrile	0.25			55	3-chloropyridazine[a]	0.10		
16	2-(carbamylmethyl)benzotriazole[a]	0.28			56	1,1,3,3-tetramethylurea	0.19		
17	2-aminomethylfuran[a]	0.37			57	2-fluoropyrazine	0.29	1.07	
18	acetaldehyde	−0.34	0.11	−1.34	58	4-aminopyridine	0.32	−0.71	−1.98
19	*tert*-butylamine	0.40		−0.82	59	1-methyl-4-benzoylpiperazine[a]	0.35		
20	pyrrolidine	0.47			60	2,6-dimethyl-4-pyrimidinamine[a]	0.39		
21	2-aminopyridine	0.49		−1.82	61	chloroacetonitrile	0.45		
22	2-amino-9-oxabenzobicyclo[2,2,1]-heptene[a]	0.75			62	5-ethoxypyrimidine	0.56	1.59	
					63	pyridine	0.65	1.29	−0.31
23	4-aminoacetophenone	0.83	1.20	−0.63	64	2-acetamidothiazole[a]	0.76		
24	3-p-tolyl-5-methyl-1,2,4-oxadiazole[a]	0.89			65	aniline	0.90	1.35	−0.05
25	2-(*N*,*N*-dimethylamino)pyrazine	0.93	1.82	−0.09	66	trimethylpyrazine	0.95	1.93	
26	tetrahydropyran	0.95	1.99	0.78	67	dimethyl sulfide	1.05	0.72	0.98
27	2-(*N*,*N*-dimethylamino)pyrimidine	1.07	1.99		68	4-quinazolinamine[a]	1.28		
28	*N*-methylphthalimide[a]	1.31		0.62	69	phenol	1.47	0.32	−0.90
29	3-chloropyridine	1.33		0.84	70	3-(2-aminoethyl)indole[a]	1.55		
30	4-phenylthioazetidine-2-one[a]	1.58			71	3,4-methylenedioxyamphetamine[a]	1.64		
31	5-phenyltetrazole[a]	1.65			72	2-(*N*,*N*-dimethylamino)pyridine	1.65	2.45	
32	*N*-methylaniline	1.66	2.40	1.10	73	4-fluorophenol	1.77		−0.87
33	4-methylumbelliferone[a]	1.90			74	7-azaindole[a]	1.82		
34	methoxybenzene	2.11	3.12	2.10	75	2-ethylbenzotriazole[a]	2.10		
35	desethyltrietazine[a]	2.14			76	benzene	2.13	2.72	2.17
36	epoxyethylcyclohexane[a]	2.43			77	2-mercaptobenzothiazole[a]	2.41	2.20	
37	hex-1-yne	2.73		2.72	78	1-aminoadamantane[a]	2.44	2.90	1.30
38	3-methyl-5-(2-pyridyl)-1,2,4-oxadiazole[a]	2.74			79	N-(4-bromobutyl)phthalimide[a]	2.96		
39	phenylcyclopropane[a]	3.27			80	2-trifluoromethyl-5,6-dibromo-benzimidazole[a]	4.15		
40	2-phenylbenzothiazole[a]	4.26							

[a] Molecules with no experimentally derived $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and π_2^H. [b] Molecules 1−40 form the training set, while 41−80 form the test set.

C-program extracted the molecular surface, defined as the 0.001 au contour of the electronic density of Bader et al.,[32] and the minimum and maximum electrostatic potential values on this surface were used as descriptors in the following study, similarly to previous work.[19,22,23]

Topological analysis of the electronic density (ρ) is based upon those points where the gradient of the density, $\nabla\rho$, vanishes. In this work we will only consider bond critical points (CPs), wherein one curvature (in the internuclear direction) is positive and two (perpendicular to the bond direction) are negative. Using the AIMPAC suite of programs,[33] in particular EXTREME, the related AIM electronic energy density, E, was computed at the H-bond critical point of the B3LYP/6-31+G(d,p) 1:1 HF complexes and used as descriptor. E reveals whether the accumulation of electronic density is stabilizing ($E < 0$) or destabilizing ($E > 0$).[34]

The computed properties were used as independent variables in a standard least squares regression analysis to directly obtain a correlation for the prediction of log P and therefore to bypass the constraint of using Abraham's experimental LSERs. All multilinear regression analysis (MLRA) employed the JMP discovery software, published by SAS software.[35] The significance or otherwise of calculated properties in the correlation was determined using the standard t-test and the predictive ability of any model tested using the "leave-one-out" cross-validation method. Statistical measures of correlations used are as follows: R^2, the overall correlation coefficient equal to the percentage of variance accounted for in the model; R^2_{CV}, the cross-validated or leave-one-out R^2 value, a measure of the predictive ability of the model; *rms*, the root-mean-square error; and Fischer's F-statistic, a measure of the significance of the model which takes into account the number of variables employed and increases for more significant statistical fits.

RESULTS AND DISCUSSION

A. log *P* Predictions from LSER Molecular Properties. Using models previously published for the estimation of descriptors $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and π_2^H, log P values from Table 1 can be accurately predicted. In the first instance, this is achieved by combining these estimations with exactly calculated values of R_2 and V_x in the appropriate LSER as reported by Abraham et al.[36] This results in encouraging statistics for all three partitions: log P_{oct}: $n = 80$, $sd = 0.61$; log P_{chl}: $n = 36$, $sd = 0.49$; log P_{alk}: $n = 33$, $sd = 0.59$. The quality of these predictions for this challenging data set

provides some validation for the properties used and for our schemes to generalize prediction for multifunctional molecules. We have previously shown that $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and π_2^H can be considered as composite descriptors since they can be expressed by more fundamental properties. Logically, by using these fundamental molecular properties into a multivariate linear analysis, direct prediction of log *P* values could be achieved for any molecule purely from structure; this method being more flexible and easier to retrain than the indirect one used above. Instead of naively injecting all the descriptors into a stepwise regression, and obtaining the best correlation, a methodology will be developed using known properties of solvation to understand the significance of possible descriptors.

To reflect electrostatic and polarization interactions between the solute and the solvent, it seems natural to design descriptors based on the electrostatic potential $V(\mathbf{r})$ as defined by eq 2. $V(\mathbf{r})$, which gives a complete description of molecular charge distribution, may contain more information than is strictly necessary and can be expressed as a multipole expansion of the potential, as displayed in eq 3

$$V(\mathbf{r}) = \sum_I \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \quad (2)$$

$$V(\mathbf{r}) \approx (\sum_I Z_I - \int \rho(\mathbf{r}')\mathrm{d}\mathbf{r}') \cdot \frac{1}{|\mathbf{r}|} - \mathrm{m} \cdot grad\left(\frac{1}{|\mathbf{r}|}\right) + \ldots \quad (3)$$

where Z_I is the charge on nucleus I, $\rho(\mathbf{r})$ is the electronic density, and m is the total electric dipole moment of the molecule. Concerning the multipole expansion, the first term is the potential due to the molecular charge, the second term is the potential due to the molecular total dipole moment, and so on. As we are dealing with neutral species, the molecular charge is zero and, as a start, the potential will be approximated as the potential due to the lowest nonzero multipole: the permanent dipole moment. It should be noted that this multipole expansion is only strictly valid at large intermolecular separations. We do not, however, seek to accurately calculate electrostatic interaction energies but are instead using such ideas as a guide for choosing molecular properties as solvation descriptors.

(i) Total Dipole Moment Solvation Energy. As transport processes are related to Gibbs free energies of solvation, it seems appropriate to use the solvation energy of the total dipole moment as a predictor of log *P*. There are many ways of modeling systems in solution, but a simple and efficient model is the Onsager reaction field model.[37] The solvent is modeled as a continuum of uniform dielectric constant ϵ. A permanent molecular dipole induces a dipole in the medium, and the electric field applied by the solvent dipole in turn interacts with the molecular dipole, leading to net stabilization, the so-called reaction field.

Here, in an electric field, **F**, the total dipole moment (m) is defined as the sum of the permanent and the induced dipole moments (illustrated in eq 4a). The solvation energy of a total dipole moment is proportional to the dot product of the permanent dipole moment (m) and the electric field, which acts upon it in a polarized dielectric. This field may be decomposed into a cavity field **G**, which is proportional to the external field intensity **E**, and a reaction field **R**, which is proportional to the total dipole moment (4b). a_o^3 represents the molecular volume.

$$m = \mu + \alpha\mathbf{F} \quad (4a)$$

$$\mathbf{F} = \mathbf{G} + \mathbf{R} = \mathrm{f}(\epsilon)\mathbf{E} + \mathrm{g}(\epsilon)(m/a_o^3) \quad (4b)$$

As the reaction field **R** depends on the total dipole moment m, this is a convergent iterative process. The point here is not to obtain an exact solution of the solvation energy but to note the form of the parameters, which depend on solute properties. By combining the different parts of (4), one finds that the total dipole moment solvation energy is a function of μ, α , and μ^2/α. Since McGowan's volume V_x is strongly correlated to α ($n = 80$; $R = 0.98$) the molecular volume a_o^3 can be substituted by the polarizability term α. This substitution simplifies greatly our estimation of the total dipole moment solvation energy, since there are no cross-terms such as $\mu\alpha/a_o^3$.

In conclusion, μ is proportional to the solvation energy of a permanent dipole in a nonpolarizable but polarized medium, while μ^2/α describes the solvation of a permanent dipole in a polarizable but nonpolarized medium. Solute polarizability α accounts for the solvation of induced dipoles in both polarizable and nonpolarizable media, since $\alpha^2/a_o^3 \propto \alpha$, as $a_o^3 \propto \alpha$. Since the μ and μ^2/α terms are cross-correlated ($n = 80$; $R = 0.85$), only μ should be used to keep orthogonality between descriptors. Thus, the total dipole moment solvation energy (i.e. dipole and induced-dipole interactions) can be represented as a linear combination of the solute permanent dipole moment and polarizability ($a\mu + b\alpha$). It is worth noting that a strong correlation between the polarizability and the quadrupole moment of a molecule exists (trace of the tensor; $n = 80$; $R = 0.97$), so the polarizability contains the quadrupole information and no higher multipole need be considered.

(ii) Local Charges Solvation Energy and Overlapping Contribution. To reflect hydrogen-bonding interactions, we turn to our previous studies of $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$, which can be regarded as composites of electrostatic and overlap terms.[23] We therefore decompose the scales into their fundamental parts, such that each can be used independently. Therefore, all electrostatic terms used for $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ were summed as well as all overlap terms used for $\Sigma\beta_2^H$. This gave two descriptors for the basicity scale, $\Sigma^T V_{Min}$ and $\Sigma^T E$ and one for the acidity scale, $\Sigma^T V_{Max}$ (the overlap term of the acidity scale can be omitted with little loss of accuracy). Naturally, this summation removes some of the flexibility of previous models of $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ but still yields acceptable correlations of $R^2 \sim 0.86$ and $rms \sim 0.11$.

$\Sigma^T V_{Min}$ is the sum of all electrostatic potential minima on the isosurface with a value less than -0.03 au (one value per basic atom). $\Sigma^T V_{Max}$ is the sum of electrostatic potential maxima on the isosurface due to acidic H's (defined as those with electrostatic potential at the donor H nuclear position greater than -1.025 au—see ref 18), with a limit of one value per group of hydrogen atoms connected to the same nucleus. Together, these properties represent *local* electrostatic interactions with solvent, since although the molecule is neutral from a distance, some charge-based interactions can appear locally. $\Sigma^T E$ is defined as the sum of energy densities at H-bond critical points, for complexes of bases with hydrogen

fluoride (HF). It represents an extra contribution to basicity over the $\Sigma^T V_{Min}$ term, as the densities of solute and solvent molecules can overlap, increasing the stability of this H-bond interaction. No such term is required for H-bond acidity, since the chemical environment of acids is relatively constant. The $\Sigma^T V_{Min}$ and $\Sigma^T V_{Max}$ descriptors can be identified with the first term of (3): along with the dipole, these two terms represent electrostatic interactions. Further stabilization is realized if overlapping occurs (i.e. negative energy density value), otherwise it is no more than a pure electrostatic interaction.

The charge-based interactions, described by $\Sigma^T V_{Max}$ and $\Sigma^T V_{Min}$, are considered as long-range interactions. This is why the interactions are described between small regions (one value per group of hydrogen atoms connected to the same nucleus or per basic atom) rather than between points. On the other hand, overlapping is considered as a short-range, contact-like interaction, and so can be modeled at a single point.

(iii) Comments on Molecular Properties Calculations. Regarding the use of $\Sigma^T V_{Min}$ and $\Sigma^T E$, it can be seen that these properties depend only on the initial state ($\Sigma^T V_{Min}$) and the final state ($\Sigma^T E$) and not potentially important conformational rearrangements occurring during complexation. To obtain a conformational rearrangement scale, the relative pK_{HB} value of each basic site of the molecule was evaluated using calculated Gibbs free energies of complexation with HF, as in ref 21. The H-bond Gibbs free energies contain not only a stabilization term due to formation of a hydrogen bond but also an energetic penalty due to reorganization of geometry and electron density on complexation. A link between estimated pK_{HB} and $\Sigma^T V_{Min}$ and $\Sigma^T E$ systems can be established: *(i)* sites with pK_{HB} value greater than zero contribute to the basicity scale, and $V_{Min} < -0.03$ and/or $E < 0.0$; *(ii)* sites with pK_{HB} value less than zero do not typically contribute to the basicity scale, and $V_{Min} > -0.03$ and $E > 0.0$. Only two structures do not follow this pattern: nitrogen atoms in aniline derivatives or in tertiary amides, which contribute to [$\Sigma^T V_{Min}$-$\Sigma^T E$] but can have negative pK_{HB}. A negative pK_{HB} value indicates that complexation is unfavorable, and the complex will "not appear" in solution if in competition with others. Therefore, for these two types of structure the relative pK_{HB} must be calculated, and if a negative value is obtained and they are in competition with other groups, their contribution to H-bond basicity is not included. Some groups can be always exempt from calculations as they are always negligible. These include π-systems, halides, pyrrole-type nitrogen, and thiophene-type sulfur atoms, and nitrogen atoms in primary or secondary amides.

As the molecular properties are calculated from 3D structures, the current method copes well with 3D structural effects. This is well exemplified with 2-(carbamylmethyl)-benzotriazole. While optimizing the structure, it was found that this compound has a six-membered intramolecular hydrogen bond. This H-bond decreases the strength of the molecule's basicity ($\Sigma^T V_{Min}$), as does the proximity of the nitrogen atoms in the triazole ring (illustrated in Figure 1).

This proximity makes the central nitrogen atom a non-H-bonding site, such that HF slides on the potential surface toward the side nitrogens. The sliding of the HF molecule, due to overlapping of sites' influence sphere, on the surface is not unique in the set. In 2-acetamidothiazole, the sulfur atom and one of the lone pairs of the carbonyl group have no basicity, i.e. they make no contribution to H-bond basicity.

We previously showed that accurate prediction requires that all energetically close conformers of a molecule should be included in the calculation of π_2^{H20} and that all stable tautomers are included in the calculation of $\Sigma\alpha_2^{H}$.[23] However, for the current set of molecules these rules have little impact on the log *P* predictions. Thus, only properties of the most stable conformer appear to be necessary for our purposes.

(iv) Theoretical Linear Solvation Energy Relationship. From the above discussions, we can think of the solvation energy of a molecule as depending on μ, α, $\Sigma^T V_{Max}$, $\Sigma^T V_{Min}$, and $\Sigma^T E$, all of which can be considered as having dimensions related to energy. It is implicit that the presence of the solute will disrupt solvent−solvent interactions, creating an endoergic term. According to the cavity theory, this term is proportional to the solute's volume, and, as seen previously, the solute volume is strongly correlated to polarizability. We therefore denoted α to be a combination of a cavity effect opposing solvation and a London dispersion effect aiding solvation. Therefore, the solvation energy of a solute should be obtainable by multilinear regression of μ, α, $\Sigma^T V_{Max}$, $\Sigma^T V_{Min}$, and $\Sigma^T E$.

To obtain a straightforward interpretation, we limit solute−solvent interactions to charge−charge interactions (with additional overlapping), dipole−dipole interactions, and induced dipole−induced dipole interactions. Other cross interactions present in solution, such as charge-dipole interactions, should be approximated by linear combinations of these. Consequently, for a given solute, the difference in solvation energy between two solvents (i.e. the partition coefficient) depends on the difference in *solvent* descriptors, i.e. μ, α, $\Sigma^T V_{Max}$, $\Sigma^T V_{Min}$, and $\Sigma^T E$ for the *solvent* molecules.

By comparing descriptors for water and octanol, we gain an idea of the descriptors' relative importance for prediction of log P_{oct} values. Water and octanol have almost identical $\Sigma^T V_{Min}$, generated by similar sp^3 oxygen atoms ($\Sigma^T V_{Min}$-water = −0.0589; $\Sigma^T V_{Min}$-octanol = −0.0581), so interactions between a solute's positive charges and the two solvents' negative charges are almost equal, as will be stabilization due to overlapping between solute and solvents. Thus, the solute $\Sigma^T V_{Max}$, representing H-bond acidity, is not likely to be important for log P_{oct}. The difference in solvent dipole moments is small and should have a limited impact (μ-water = 2.18; μ-octanol = 1.90). What really differentiates the two solvents are polarizability (α-water = 21.0; α-octanol = 306.4) and $\Sigma^T V_{Max}$ ($\Sigma^T V_{Max}$-water = 0.0737; $\Sigma^T V_{Max}$-octanol = 0.0658). The strength of the interactions between a solute's negative charges and the two solvents' positive charges will differ, and the solute $\Sigma^T V_{Min}$ is required as well as its corresponding overlap term, $\Sigma^T E$. The sign of the regression coefficients can also be predicted: a positive sign for α, as a polarizable (large) molecule will prefer octanol, a negative sign for μ, as a molecule with a permanent dipole will prefer water, and a positive sign for $\Sigma^T V_{Min}$ and $\Sigma^T E$, as a molecule with negative charges will prefer water (for NB both terms have negative values). We note that these considerations can only act as a guide, since although octanol and water have similar μ and $\Sigma^T V_{Min}$, they have very different densities of H-bonding.

The list of descriptors drawn up to assess the solvation energies of the different interactions taking place in a system

Table 2. Multilinear Regression Analysis (MLRA) Models of log *P*

	regression coefficients						statistics				
model	intercept	μ	α	$\Sigma^T V_{Min}$	$\Sigma^T E$	$\Sigma^T V_{Max}$	set	*n*	R^2	rms	F-stat
log P_{oct} **1**	0.296	−0.213	0.010	11.95	35.90	0	train	40	0.893	0.439	72.8
							test	39	0.862	0.450	
log P_{oct} **2**	0.205	−0.151	0.010	13.44	28.10	0	all	79	0.886	0.431	143.5
log P_{chl}	0.329	0	0.012	9.78	42.03	−18.87	all	36	0.891	0.483	63.1
log P_{alk}	−0.014	0	0.011	25.40	39.68	−19.65	all	33	0.910	0.492	70.9

can be now used in a standard least-squares fitting to experimental log *P* values, the results of which are displayed in Table 2.

The set of log P_{oct} values was further split into training and test sets (see footnote [b] of Table 1). The prediction statistics show that the model is consistent and flexible to the introduction of new chemical structures and other challenging compounds. This robustness is reinforced by the fact that, when combining the training and the test sets to give an overall training set of 79 molecules, all reoptimized regression coefficients are well within the overlap of the model's error limits. log P_{oct} model **2** offers tighter error limits, i.e. higher confidence on the regression coefficients. In the test set, one outlier (1-methyl-4-benzoylpiperazine) was removed. It is noticeable that this is the largest molecule of the set, and it has been seen before that errors in calculations of molecular properties' tend to increase with molecular size. Thus, for the time being, the proposed models are limited to molecules with no more than around 30 atoms. Smaller molecules, i.e. with less than 16 atoms, yield much better statistics ($n = 41$, $rms = 0.33$), but, reassuringly, the new regression coefficients are similar to the ones in Table 2.

To reflect each property's importance, log P_{oct} model **2** was recomputed with scaled and centered variables to give regression coefficients of −0.16 for μ, 0.84 for α, 0.47 for $\Sigma^T V_{Min}$, and 0.30 for $\Sigma^T E$. Not surprisingly, the two main contributions are solute polarizability (0.85) and basicity (0.47 + 0.30). In a trade of speed for accuracy, one could replace the polarizability term by the more readily calculated McGowan's volume ($n = 79$, $R^2 = 0.862$; $rms = 0.474$), since both terms describe similar effects.

Being derived, to some extent, from successful LSER descriptors, these theoretical descriptors should be sufficient to describe as many systems as those modeled by Abraham and co-workers. Regression analyses were made on the log P_{chl} and log P_{alk} values in Table 1 (results are displayed in Table 2), giving similar accuracy to log P_{oct} models. By calculating the descriptors of chloroform, one can conclude that it is less basic, acidic, and dipolar but more polarizable than water ($\Sigma^T V_{Max} = 0.0557$, $\Sigma^T V_{Min} = \Sigma^T E = 0.0$, $\mu = 1.22$, $\alpha = 142.6$). Thus, the regression coefficient of α is positive, as a large, polarizable molecule will prefer chloroform, and those of $\Sigma^T V_{Min}$ and $\Sigma^T E$ positive and $\Sigma^T V_{Max}$ negative as a molecule with *local* charges will prefer water (NB $\Sigma^T V_{Min}$ and $\Sigma^T E$ have negative values, while $\Sigma^T V_{Max}$ has positive values).

These deductions made for the regression coefficients are identical for the water/alkane system, since alkanes have no basicity, acidity, and dipole moments but higher polarizability than water. In these two models, the solute dipole moment is not required to obtain accurate predictions, even though large differences in solvent dipole moments are present (e.g. water vs chloroform: 2.18 vs 1.22). This can be explained by the possibility to roughly correlate the dipole moment with a dipolar descriptor: $\Sigma^T V_{Max} - \Sigma^T V_{Min}$. Although this relation is only approximate, multiplication by a typical μ-regression coefficient suggests an effect smaller than the *rms* error on the prediction of log *P*. Unless the dipole-based interaction gains importance in the model or the data set contains more molecules with large $\Sigma^T V_{Max} - \Sigma^T V_{Min}$ but a small dipole moment, the dipole moment can be approximated by the difference between the two monopoles of the molecule: $\Sigma^T V_{Max}$ and $\Sigma^T V_{Min}$.

Table 2 confirms that the chosen molecular properties successfully predict log *P* values for different systems. These log *P* models can be easily retrained with more experimental log *P* values and offer an alternative to predict log *P* values purely from structure with errors close to experimental ones. A study of the correlation matrix for μ, α, $\Sigma^T V_{Max}$, $\Sigma^T V_{Min}$, and $\Sigma^T E$ shows that the maximum cross-correlation is between $\Sigma^T V_{Min}$ and $\Sigma^T E$, with a value, in terms of R^2, of 0.18. Thus, log *P* values can be explained by the same minimal set of orthogonal descriptors, each of them being well defined and physically understood.

These models show some similarity with the log P_{oct} models previously published by Haeberlein and Brinck or Eisfeld and Maurer:[6] similar accuracy is obtained here with fewer descriptors (4 vs 8/10). Satisfactorily, the accuracy of the log P_{oct} model proposed here is similar: *(i)* to those of commercial quantitative solvation property relationships, based on less fundamental descriptors such as H-bond-scales derived from geometric criteria (BOSS $n = 254$, $rms = 0.61$, 4 descriptors; QikProp: $n = 260$, $rms = 0.65$, 6 descriptors);[38] *(ii)* to those of the fragmental approaches considered by Bodor et al. for complex drugs ($n = 68$; Rekker $sd = 0.83$; KLOGP $sd = 0.73$; CLOGP $sd = 0.67$; and ACD $sd = 0.50$).[39] One clear advantage to use the proposed set of descriptors is the possibility of modeling further solvation properties from small training sets of data, since regression using 4 or 5 descriptors can easily be applied to as little as 25 data.

B. Toward a Generalization to any Electron Donor/Acceptor (EDA) System. Hydrogen bonding can be considered a special case of EDA complexes, where the electron acceptor/Lewis acid is always a hydrogen atom, whether on the solvent (described by $\Sigma^T V_{Min}$-$\Sigma^T E$) or on the solute ($\Sigma^T V_{Max}$). Hydrogen bonded complexes are often thought as precursors of acid−base reactions, while other EDA complexes can be thought of as precursors of electrophile−nucleophile reactions. Thus, a more general model should include any electron acceptor.

The $\Sigma^T V_{Max}$ system is specific to hydrogen bonding because it only contains solute's positive charges generated by acidic hydrogen atoms. A generalization of $\Sigma^T V_{Max}$ is easily achieved by incorporating the maximum potential

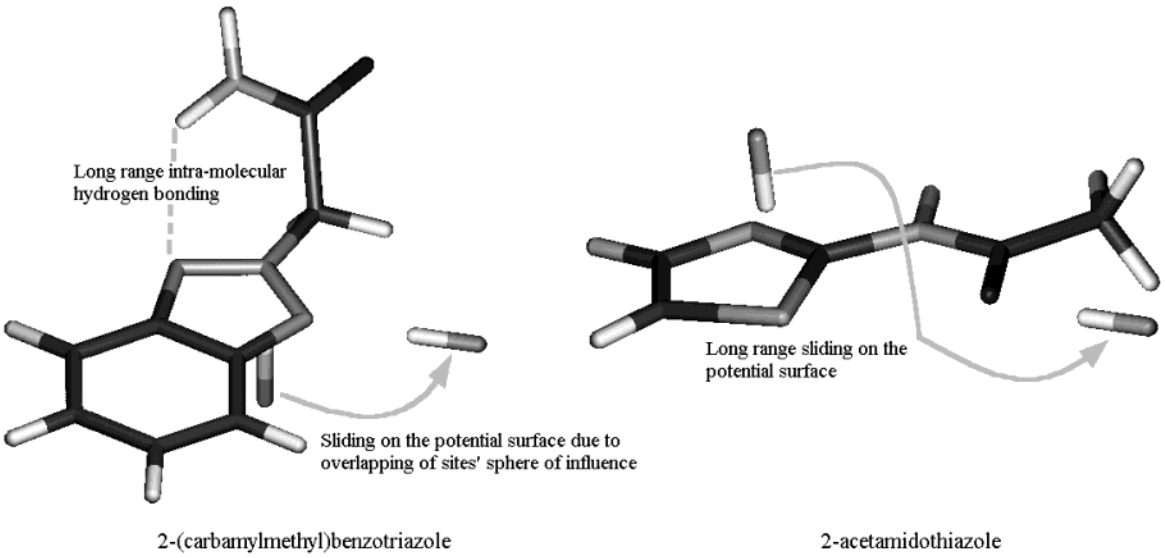

Figure 2. Examples of 3D-structural effects.

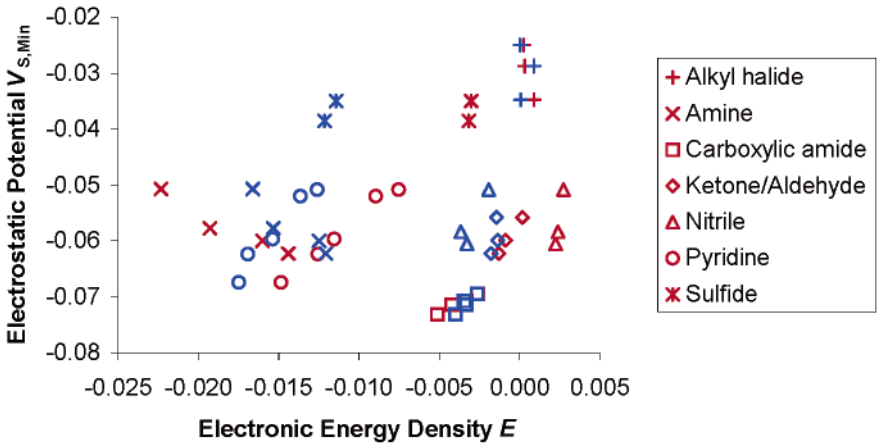

Figure 3. Bivariate plot of $V_{S,Min}$ by E from HF (red) and BrF (blue) complexes.

values of other possible electron acceptors on the solute. Similarly, $\Sigma^T V_{Min}$ already covers a variety of electron donors/Lewis bases and does not need to be extended. It is not specific to hydrogen bonding and can therefore encompass any charge-based interactions, even if the solvent's electron acceptor is not a hydrogen atom.

By using energy densities from HF complexes, the overlap term $\Sigma^T E$ is only relevant to hydrogen bonding and cannot be used to describe other EDA complexes. This is why a combination of $\Sigma^T V_{Min}$ and $\Sigma^T E$ is defined as a H-bond basicity scale. Extra terms from probes other than HF are required to account for overlapping of the solute with other electron acceptors. The necessity of other probes is illustrated by Figure 3, which shows that different families of electron acceptors overlap to very different extents. For example, the overlap in amine−BrF complexes is 0.8 times that in amine−HF complexes. However, overlapping in sulfide−BrF complexes is 3.8 times greater than in sulfide−HF complexes. This is consistent with the picture given by the Hard Soft Acid Base model (HSAB) developed by Pearson et al.[40] HSAB suggests that hard acids (e.g. HF) prefer to associate with hard bases (e.g. amine) and soft acids (e.g. BrF) with soft bases (e.g. sulfide). A similar plot is obtained using H_2O as a reference acid, albeit on a different scale, which explains why HF is capable of describing H_2O and other H-acids.

To account for other electron acceptors, one could customize probes according to the system studied, e.g. by using BrF complexes to describe 1-bromoheptafluoropropane. However, it seems more judicious to design a minimal set of electron acceptor probes capable of reproducing all possible overlap situations. Eight electron acceptor probes (ClF, BrF, BF_3, BH_3, AlH_3, C_2F_4, C_6F_6, SO_3) were complexed with a variety of electron donors (nos. 3,6,10,44,49,63,67, from Table 1 plus diethyl ether, 2-chloropropane, and *N*-methylpyrrole). The two π-systems (C_2F_4, C_6F_6) gave complexes with positive energy densities, so these pure electrostatic complexes do not require an overlap term. A study of the correlation matrix for the probes shows that the minimum cross-correlation is between E_{HF} and E_{AlH3}, with an R^2 of just 0.01. These two overlap terms together can correlate the remaining overlap terms ($n = 10$, $R^2 \sim 0.9-0.8$, $rms < 0.01$), although neither overlap term can be used independently. Thus, it seems possible to use a limited set of orthogonal overlap descriptors to describe a wide range of electron acceptors. This flexibility of the overlap term could also describe some acid−base combinations, which were excluded during the setup of the experimental H-bond scales.[11,12]

By including the maximum potential values of other possible electron acceptors on the solute, $\Sigma^T V_{Max}$ will become

family dependent due to the diversity of atom types, and the introduction of an overlap term will become crucial. This time, a minimal set of electron donor probes should be designed. A real test of the possible extension of the descriptors to any EDA interactions will consist of the prediction of log *P* values of solute and/or solvent systems with other electron acceptors than hydrogen.

CONCLUSION

Through this study, it was possible to directly predict partition coefficients from molecular electrostatic and electronic properties for chemically diverse molecules. Some of these molecules were drugs (with biological activity) or common fragments of drugs and hence represented a challenging test by their structural complexity and variety of sites per molecule. This approach offers a way to study more theoretical cases, such as metal complexes or reactive intermediates, and has also an inherent interest, since it gives an insight into the molecule behavior toward complexation. For example, the 3D structure of the molecule can highlight some long-range intramolecular hydrogen bonds or overlapping of sites' spheres of influence, which partially or totally remove the H-bond ability of some sites.

It was shown that the electrostatic and polarization interactions between the solute and solvent can be developed into local electrostatic interactions, partitioned into positive and negative terms ($\Sigma^T V_{Max}$ and $\Sigma^T V_{Min}$), dipole-based interactions (μ), and induced dipole-based interactions (α). Further stabilization is possible if the molecular densities come into contact and overlap ($\Sigma^T E$). Multilinear regression analysis using these fundamental properties was able to accurately model partition coefficients (*rms* 0.4−0.5). Being derived, to some extent, from successful LSER descriptors, this approach should be sufficient to describe as many systems as those found by Abraham and co-workers.

Hydrogen bonding complexes can be considered as a special case of Electron Donor/Acceptor (EDA) complexes, where the electron acceptor/Lewis acid is always a hydrogen atom, whether on the solvent's molecule (system $\Sigma^T V_{Min}$-$\Sigma^T E$) or on the solute (system $\Sigma^T V_{Max}$). It was found that both systems could be extended so they can describe other EDA interactions.

ACKNOWLEDGMENT

O.L. thanks GlaxoSmithKline for a Ph.D. studentship. The authors are also grateful to Dr. Xiao Qing Lewell for help in the design of the data set.

REFERENCES AND NOTES

(1) Klopman, G.; Li, J.-Y.; Yang, S.; Dimayuga, M. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752.
(2) Rekker, R. F. *The Hydrophobic Fragmental Constant*; Elsevier: New York, 1977.
(3) Leo, A. J. *Chem. Rev.* **1993**, *93*, 1281.
(4) ACD/log*P* DB, Advanced Chemistry Development, Inc., Toronto, 2000.
(5) Bodor, N.; Harget, A.; Huang, M.-J. *J. Am. Chem. Soc*. **1991**, *113*, 9480.
(6) (a) Haeberlein, M.; Brinck, T. *J. Chem. Soc., Perkin Trans. 2* **1997**, *2*, 289. (b) Brinck, T.; Murray, J. S.; Politzer, P. *J. Org. Chem.* **1993**, *58*, 7070. (c) Eisfeld, W.; Maurer, G. *J. Phys. Chem. B* **1999**, *103*, 5716. (d) Nelson, T. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 601. (e) Famini, G.; Benyamin, D.; Kim, C.; Veerawat, R.; Wilson, L. Y. *Collect. Czech. Chem. Commun.* **1999**, *64*, 1727. (f) Famini, G. R.; Wilson, L. Y. *J. Phys. Org. Chem.* **1993**, *6*, 539. (g) Cronce, D. T.; Famini, G. R.; DeSoto, J. A.; Wilson, L. Y. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1293. (h) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. (i) Klamt, A.; Eckert, F.; Hornig, M. *J. Comput.-Aid. Mol. Des.* **2001**, *15*, 355.
(7) Jorgensen, W. L.; Briggs, J. M.; Contreras, M. L. *J. Phys. Chem.* **1990**, *94*, 1683.
(8) (a) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73. (b) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Taft, R. W. *J. Am. Chem. Soc*. **1984**, *106*, 464.
(9) (a) Kamlet, M. J.; Doherty, R. M.; Abboud, J.-L. M.; Abraham, M. H.; Taft, R. W. *Chem. Technol.* **1986**, *16*, 566. (b) Abraham, M. H.; Liszi, J. *J. Chem. Soc., Faraday Trans. 1* **1978**, *74*, 1604.
(10) (a) Gurka, D.; Taft, R. W. *J. Am. Chem. Soc*. **1969**, *91*, 4794. (b) Taft, R. W.; Gurka, D.; Joris, L.; Schleyer, P. v R.; Rakshys, J. W. *J. Am. Chem. Soc.* **1969**, *91*, 4801. (c) Joris, L. J.; Mitsky, L.; Taft, R. W. *J. Am. Chem. Soc.* **1972**, *94*, 3438.
(11) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans. 2* **1989**, *6*, 699.
(12) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, *4*, 521.
(13) Abraham, M. H.; Whiting, G. S.; Doherty, R. M.; Shuely, W. J. *J. Chromatogr*. **1991**, *587*, 213.
(14) (a) Abraham, M. H.; McGowan, J. C. *Chromatographia* **1987**, *23*, 243. (b) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.
(15) See, for example: (a) Gratton, J. A.; Abraham, M. H.; Bradbury, M. W.; Chadha, H. S. *J. Pharm. Pharmacol*. **1997**, *49*, 1211. (b) Abraham, M. H.; Martins, F.; Mitchell, R. C. *J. Pharm. Pharmacol*. **1997**, *49*, 858. (c) Platts, J. A.; Abraham, M. H. *Environ. Sci. Technol*. **2000**, *34*, 318. (d) Poole, S. K.; Poole, C. F. *J. Chromatogr. A* **1999**, *845*, 381.
(16) Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 835.
(17) Lamarche, O.; Platts, J. A.; Hersey, A. *Phys. Chem. Chem. Phys*. **2001**, *3*, 2747.
(18) Platts, J. A. *Phys. Chem. Chem. Phys*. **2000**, *2*, 973.
(19) (a) Hagelin, H.; Murray, J. S.; Brinck, T.; Berthelot, M.; Politzer, P. *Can. J. Chem*. **1995**, *73*, 483. (b) Murray, J. S.; Ranganathan, S.; Politzer, P. *J. Org. Chem.* **1991**, *56*, 3734. (c) Murray, J. S.; Politzer, P. *J. Org. Chem.* **1991**, *56*, 6715. (d) Murray, J. S.; Politzer, P.; *J. Chem. Res. (S)* **1992**, *3*, 110.
(20) Platts, J. A. *Phys. Chem. Chem. Phys*. **2000**, *2*, 3115.
(21) Lamarche, O.; Platts, J. A. *Chem. Eur. J.* **2002**, *8*, 457.
(22) Lamarche, O.; Platts, J. A. *Chem. Phys. Lett.* **2003**, *367*, 123.
(23) Lamarche, O.; Platts, J. A. *Phys. Chem. Chem. Phys*. **2003**, *5*, 677.
(24) MedChem Software, BioByte Corp., Claremont, 2002.
(25) Frisch, M. J.; et al., Gaussian 98, Rev. A.6, Gaussian, Inc., Pittsburgh, PA, U.S.A., 1998.
(26) (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (c) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163. (d) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(27) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265
(28) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(29) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem*. **1983**, *4*, 294.
(30) (a) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, 1990. (b) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893.
(31) Politzer, P.; Truhlar, D. G. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Plenum: New York, 1981.
(32) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. *J. Am. Chem. Soc*. **1987**, *109*, 7968.
(33) Biegler-König, F. W.; Bader, R. F. W.; Tang, T.-H. *J. Comput. Chem.* **1982**, *3*, 317. These programs can be obtained on request from http://www.chemistry.mcmaster.ca/aimpac.
(34) (a) Cremer, D.; Kraka, E. *Angew. Chem., Int. Ed. Engl*. **1984**, *23*, 627. (b) Cremer, D.; Kraka, E. *J. Am. Chem. Soc.* **1985**, *107*, 3800.
(35) JMP, Revision 4, SAS Institute, Inc., Cary, 2000.
(36) (a) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085. (b) Abraham, M. H.; Platts, J. A.; Hersey, A.; Leo, A. J.; Taft, R. W. *J. Pharm. Sci.* **1999**, *88*, 670.
(37) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.
(38) (a) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc*. **2000**, *122*, 2878. (b) Jorgensen, W. L.; Duffy, E. M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155.
(39) Buchwald, P.; Bodor, N. *Curr. Med. Chem.* **1998**, *5*, 353.
(40) Pearson, R. G. *Hard And Soft Acids And Bases*; Dowden, Hutchinson & Ross: Stroudsburg, 1973.

CI034276C