Surface-Integral QSPR Models: Local Energy Properties

Bernd Ehresmann,[†] Marcel J. de Groot,[‡] and Timothy Clark*,[†]

Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany, and MISD, Pfizer Ltd., Global Research and Development, Sandwich Laboratories, Sandwich, Kent CT13 9NJ, U.K.

Received January 25, 2005

Surface-integral models based on AM1 semiempirical molecular orbital calculations are presented for the free energies of solvation in water, *n*-octanol, and chloroform and for the enthalpy of solvation in water. A parametrized function of four local properties calculated at the isodensity surface (the molecular electrostatic potential, local ionization energy, electron affinity, and polarizability) is integrated over the triangulated surface area to obtain the target quantity. The resulting models give results only slightly less accurate than those reported for parametrized generalized Born/polar surface area models despite relying only on gas-phase calculations. The water and octanol free-energy models were validated by calculating the water−octanol partition coefficient for a test set of organic compounds with moderate success. The models lead to a local solvation energy, which can be projected onto the molecular isodensity surface and provides insight into "hot" areas for solvation in water or the other solvents.

INTRODUCTION

The need to consider new paradigms (i.e. other than the almost universal atom-based ansatz) for molecular modeling, quantitative structure−activity (QSAR) and structure−property (QSPR) relationships, docking and scoring functions, and other applications in modeling and cheminformatics has been highlighted recently.[1] We have been working toward a new approach based entirely on local properties of the molecule at the molecular surface, which may be defined in any of a number of ways. A further important aim of this work is to make the models as widely applicable as possible by eliminating element-specific descriptors. The ultimate justification for this approach is that current models are almost all local in nature (i.e. they can only be used successfully for a limited interpolation range) and that models based only on global properties are more likely to be generally applicable.

The first stage of this development was to define a series of new local properties[2] in addition to the commonly used molecular electrostatic potential (MEP).[3] These properties include the local ionization energy, originally proposed by Sjoberg et al.[4] and examined in detail by Murray, Politzer, and their co-workers,[5−8] the local electron affinity,[2] and the local polarizability,[2] which is based on our additive distributed polarizability model.[9] This model, in turn, is derived from Rivail's variational treatment[10] parametrized to reproduce experimental molecular electronic polarizabilities accurately.[11] An important step in the development of the new techniques was to define a new set of molecular descriptors derived from the MEP and the new local properties.[12] These descriptors have been shown[12] to be powerful for deriving new QSAR models. The quantum mechanical framework for this work is standard NDDO[13]-based semiempirical molecular orbital (MO) theory, usually AM1.[14] Thus, the only limitation on the use of the model is that the elements present in the molecule must be available for the semiempirical MO-Hamiltonian used. In practice, this limitation only rules out some transition-metal compounds, for which parameters are, however, slowly becoming available. The surface-based technique itself is relatively compute-intensive. Semiempirical geometry optimizations take of the order of seconds to minutes for typical druglike molecules, and the generation of the surface and the local properties may take up to a minute on modern hardware for large (200 atom) molecules. Thus, the techniques discussed here are applicable to tens of thousands of compounds routinely but not to millions.

Two important aims of our work, which are closely connected, are to establish the feasibility and reliability of surface-integral QSPR models using our four local properties and to develop surface-based scoring functions for evaluating docking results. Surface-integral models, in which a physical property is estimated by integrating one or more local properties over the molecular surface, have been used successfully by Brickmann and his group[15−17] within the molecular free energy surface density (MolFESD), and among others by Merz et al.[18] within a generalized Born/surface area (GB/SA) model, and by Cramer, Truhlar, and co-workers[19] within their series of parametrized solvation models. The last two applications are not strictly local because only part of the solvation energy is treated as a pure surface term. The bulk polarization contribution to the solvation energy is clearly not local, but we note that the conductor-like screening model (COSMO) of Klammt and Schüürmann[20] actually uses a local approximation, perfect shielding, to obtain the polarizing virtual charges.

We now report successful pure surface-integral models for free energies of solvation and the enthalpy of solvation in water based only on the four local properties discussed in ref 2. Thus, our treatment is, at least in principle, able to

* Corresponding author e-mail: clark@chemie.uni-erlangen.de.
[†] Universität Erlangen-Nürnberg.
[‡] Pfizer Ltd.

10.1021/ci050025n CCC: $30.25
Published on Web 05/13/2005

deal with donor/acceptor and dispersion interactions between solute and solvent. Merz[21,22] has emphasized the former in a series of papers, and dispersion clearly plays a major role in solvation by nonpolar solvents. As we demonstrate below, we are now able to define local solvation energies (either enthalpy or free energy) at molecular surfaces for use in QSPR solvation models.

METHODS

All geometries were fully optimized using the standard AM1 Hamiltonian[14] with VAMP 8.2.[23] In contrast to our earlier work, isodensity surfaces[24] were used throughout and were generated and triangulated using an in-house version of the marching-cube algorithm[25] implemented in the program ParaSurf.[26] MEPs were calculated using our zero-differential-overlap based atomic multipole technique[27] and the local ionization energy, electron affinity, and polarizability as described in refs 3 and 12. Multiple linear regressions used Tsar 3.3.[28] For standardized models, the variables for the multiple regression were standardized and scaled by mean and standard deviation.

The functions that were precalculated and included in the multiple linear regression are powers, exponentials, etc. of the local properties and combinations (products, etc.) thereof. The complete list of functions included in the regression is given in Table S1 of the Supporting Information.

Surface-Integral Models. Our surface-integral models are defined using the expression

$$P = \sum_{i=1}^{ntri} f(V^i, IE_L^i, EA_L^i, \alpha_L^i, \eta_L^i) \cdot A^i \qquad (1)$$

where P is the target property, usually a free energy, f is a nonlinear function of the electrostatic potential, V, the local ionization energy, IE_L, the local electron affinity, EA_L, the local polarizability, α_L and the local hardness, η_L, which is given by[2]

$$\eta_L = \frac{(IE_L - EA_L)}{2} \qquad (2)$$

The summation runs over the *ntri* triangles that make up the molecular surface, and the superscript i denotes the value of the relevant local property at the center of the surface triangle i of area A^i. The function f is determined by multiple linear regression using precalculated sums of the individual components of the functions given in Table S1. Note that we have previously[2,12] shown the four local properties to be essentially orthogonal, so that the data from which our functions are derived show no linear dependencies. The exception is the local hardness, which correlates weakly with the local ionization energy.[12] However, the local hardness was included here because the terms used in the model are nonlinear functions of the local properties. Functions with a correlation coefficient R greater than 0.9 were excluded from the regression process. In fact, very few of the functions included in the models (less than 4% of the possible correlations in every model) are correlated with $R > 0.7$.

Table 1. Data Sets and Results for the Five Surface-Integral QSPR Models and the Two Validation Models[a]

property	N	MSE	MUE	RMSD	R^2	R^2_{cv}	ref
Surface-Integral Models							
$\Delta G_{solv}(H_2O)_{total}$	387	0.00	1.54	1.96	0.99	0.91	35−37
$\Delta G_{solv}(H_2O)_{neutral}$	362	0.00	0.85	1.09	0.88	0.83	35,36
$\Delta H_{solv}(H_2O)$	93	0.00	2.24	2.87	0.99	0.98	29−34
ΔG_{solv}(octanol)	168	0.00	0.67	0.82	0.90	0.80	35
$\Delta G_{solv}(CHCl_3)$	85	0.01	0.58	0.83	0.74	0.67	38
Validation Model							
$\log P_{OW}$[b]	91	−0.20	0.91	1.24	0.54	0.49	39

[a] N = number of data points, MSE = mean signed error, MUE = mean unsigned error, RMSD = root-mean-square deviation between experiment and prediction, R^2 = correlation coefficient, R^2_{cv} = leave-one-out cross-validated correlation coefficient. [b] Phenylalanine excluded from the statistics as an extreme outlier.

Data Sets. The data set used for enthalpies of hydration was taken from six published data sets[29−34] and includes 93 organic and inorganic species, of which seven are anions and 19 cations.

The data for the free energy of hydration were taken from three sources[35−37] for a total of 387 species including 13 anions and 12 cations.

The data sets for the free energies of solvation in chloroform (88 neutral compounds)[38] and *n*-octanol (168 neutral compounds)[35] were taken directly from the literature.

The $\log P_{OW}$ validation set was that also used for validation in our earlier neural-net based $\log P_{OW}$ work.[39] Of the original 105 compounds, one was removed from the data set because it also occurs in one of the two training sets.

Because of the known problems of AM1 for nitro compounds[40] and because polynitro compounds were always the most significant outliers in our models, we omitted all nitro compounds from the data sets. Similarly, iodine-containing compounds are treated poorly by AM1 and were removed from the data sets. Table 1 gives a summary of the data sets used and the results obtained for the surface-integral QSPR models.

RESULTS

Free Energy of Hydration. The results obtained for the free energy of hydration are shown graphically for the entire data set (neutral compounds, anions, and cations) in Figure 1. A listing of the compounds, the experimental and calculated $\Delta G_{solv}(H_2O)$ values is given in Table S2 of the Supporting Information. Throughout the discussion, all energy values are given in kcal mol^{-1}, and the acronyms MSE and MUE refer to the mean signed error and the mean unsigned error, respectively, between experiment and calculation.

The regression equation using the raw function values is shown in eq 3 and that using standardized input data, which allow us to judge the contributions of the individual terms from the coefficients, in eq 4. To test the statistical significance of the model obtained, the target free energies were randomized and the regression analysis repeated. The resulting model gave $R^2 = 0.13$, $R^2_{cv} = -0.01$, MSE = 0.00, MUE = 2.32, and RMSD = 3.00. The F-probablility of all the models reported here was less than 10^{-27} and in most

cases zero within the precision of the TSAR statistical package.

$$f_{(\Delta G_{solv}(H_2O))}(\mathbf{r}) = 6.1462 \times 10^{-2}\cdot\sqrt{|V(\mathbf{r})|} - 5.0671 \times 10^{-4}\cdot[|V(\mathbf{r})|]^{3/2} + 7.2498 \times 10^{-8}\cdot[V(\mathbf{r})]^3 - 1.9860\cdot[\alpha_L(\mathbf{r})]^3 - 1.5240 \times 10^{-18}\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})|]^{3/2} - 1.3808 \times 10^{-7}\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})|]^{3/2} - 1.5335 \times 10^{-2}\cdot V(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 3.2783 \times 10^{-7}\cdot[V(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 5.1184 \times 10^{-6}\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r}) + 6.6402 \times 10^{-16}\cdot[IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})]^3 + 4.8931 \times 10^{-7}\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^{5/2} + 1.3713 \times 10^{-12}\cdot[|EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{5/2} - 3.3795 \times 10^{-5}\cdot[|EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 1.5082 \times 10^{-4}\cdot\sqrt{|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})|} + 3.6486 \times 10^{-15}\cdot[V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})]^2 + 2.2001 \times 10^{-8}\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r}) + 2.3008 \times 10^{-12}\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 7.6344 \times 10^{-23}\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 7.0707 \times 10^{-14}\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 1.7987 \times 10^{-15}\cdot[IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^2 + 9.0499 \times 10^{-12}\cdot[|IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 1.6015 \times 10^{-14}\cdot[IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 1.3080 \times 10^{-10}\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^2 - 6.1998 \times 10^{-18}[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 0.6803 \quad (3)$$

$$f_{(\Delta G_{solv}(H_2O))}(\mathbf{r}) = 10.7568\cdot\sqrt{|V(\mathbf{r})|} - 15.2318\cdot[|V(\mathbf{r})|]^{3/2} + 6.6229\cdot[V(\mathbf{r})]^3 - 7.7800\cdot[\alpha_L(\mathbf{r})]^3 - 9.1016\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})|]^{3/2} - 10.9507\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})|]^{3/2} - 11.9951\cdot V(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 3.3531\cdot[V(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 12.0307\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r}) + 5.6550\cdot[IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})]^3 + 9.1097\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^{5/2} + 16.5199\cdot[|EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{5/2} - 17.6329\cdot[|EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 5.3916\cdot\sqrt{|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})|} + 7.9905\cdot[V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})]^2 + 8.2607\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r}) + 8.6362\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 3.0000\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 3.7929\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 17.6463\cdot[IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^2 + 13.6101\cdot[|IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 4.0170\cdot[IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 12.2584\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^2 - 2.5834\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 7.3712 \quad (4)$$

The electrostatic potential appears in the form of an $aV + b|V|^{3/2} + cV^3$ expression, but the terms are otherwise difficult to interpret. The terms with the largest coefficients (larger than 20, chosen as an arbitrary cutoff for the discussion) in the standardized equation are the square root of the local polarizability, the 3/2 power is the product of the local ionization energy and the local polarizability and the 3/2 power of the absolute value of the product of the local ionization energy, electron affinity, and hardness. To judge the importance of the purely electrostatic terms (i.e. those that depend only on V), we compared this term with the total calculated free energy of solvation. The mean free energy of solvation given by eq 4 for all the neutral compounds is -2.7 (standard deviation = 2.7) kcal mol^{-1}, compared with -8.3 (standard deviation = 6.7) for the sum of the three terms involving only V. The free energy ranges from -11.8 to $+4.7$ kcal mol^{-1} and the sum of the potential-dependent terms from -40.3 to -0.1. The two terms do not correlate at all ($R^2 = 0.02$). These results do not suggest a dominant role of a purely potential-dependent term, even in water. However, the electrostatic potential does play a significant role in many other terms in combination with one or more of the other properties. At present, there does not seem to be a straightforward interpretation of the regression equation, much as we might like to see one.

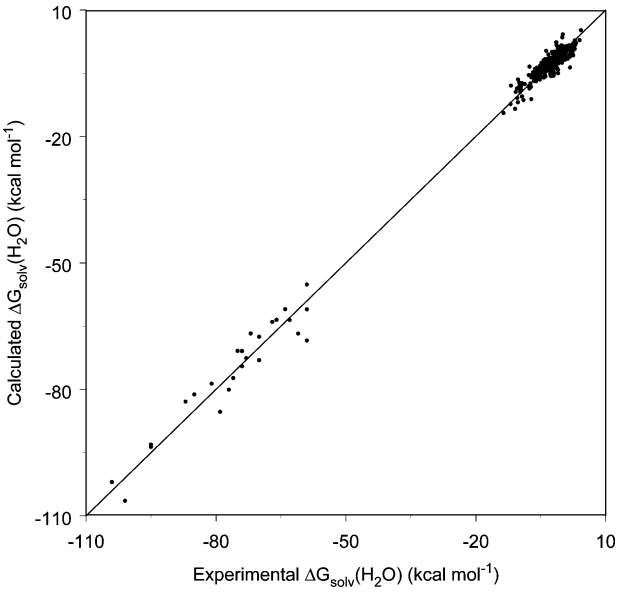


Figure 1. Experimental and calculated free energies of solvation in water for the entire training data set given Table S2. $N = 387$, MSE = 0.00, MUE = 1.54, RMSE = 1.96, $R^2 = 0.989$, $R^2_{cv} = 0.91$, slope = 1.00, intercept = 0.00. The two clusters of points correspond to neutral and charged moieties, as discussed in the text.

The structure of the data shown in Figure 1 (many data points close to each other for the neutral compounds with a second cluster further away for ions) is poorly suited to linear regression, and so we constructed a second model for the free energy of solvation in water using only the neutral compounds. The results are shown in Figure 2. The raw and standardized regression equations are shown as eqs 5 and 6, respectively.

To exclude chance correlations as the basis for the two free-energy models in water, we randomized the free-energy data for the neutral compounds and used the same procedure as above to construct a model. The resulting regression equation consisted of two terms and gave $R^2 = 0.08$, $R^2_{cv} = 0.01$, MSE = 0.0, MUE = 2.32, and RMSD = 3.00. We can therefore exclude chance correlations and conclude that, despite the complexity of the function obtained, the free-energy models for water are significant. Note that only five

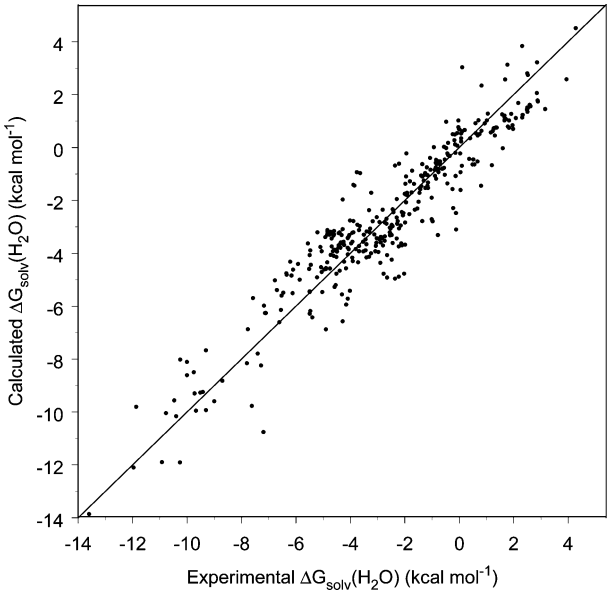


Figure 2. Experimental and calculated free energies of solvation in water for the neutral compounds from the data set given in Table S3. $N = 362$, MSE = 0.00, MUE = 0.85, RMSD = 1.09, $R^2 = 0.88$, $R^2_{cv} = 0.83$, slope = 1.00, intercept = 0.00.

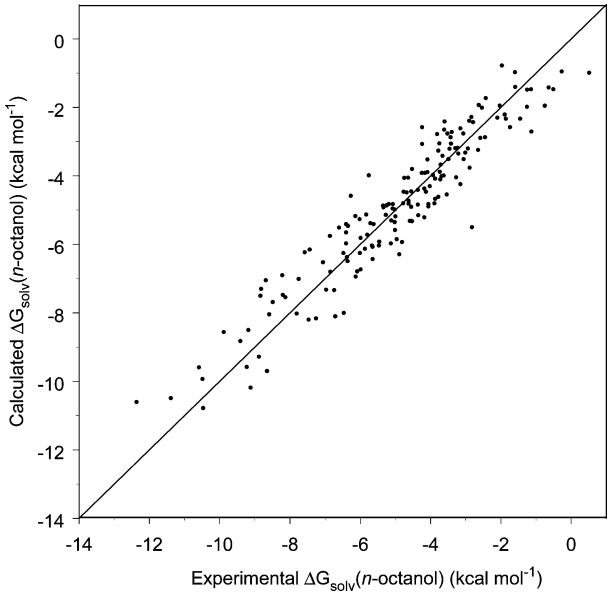


Figure 3. Experimental and calculated free energies of solvation in *n*-octanol for the training data set given in Table S4. $N = 168$, MSE = 0.00, MUE = 0.67, RMSD = 0.82, $R^2 = 0.90$, $R^2_{cv} = 0.80$, slope = 1.00.

real variables used, so that overfitting should not be a problem.

$$f_{(\Delta G_{solv}(H_2O),neutral)}(\mathbf{r}) = 3.2102 \times 10^{-5}\cdot[\eta_L(\mathbf{r})]^{3/2} + 2.4139 \times 10^{-3}\cdot\sqrt{|V(\mathbf{r})\cdot EA_L(\mathbf{r})|} + 1.6708 \times 10^{-2}\cdot V(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 2.6456 \times 10^{-17}\cdot[IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 7.8146 \times 10^{-9}\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 1.5988 \times 10^{-4}\cdot[\cdot\alpha_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^{3/2} + 4.1497 \times 10^{-8}\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r}) + 8.1163 \times 10^{-20}\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{5/2} - 5.8071 \times 10^{-5}\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 1.8980 \times 10^{-10}\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{3/2} + 2.0448 \times 10^{-21}\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 1.6835 \times 10^{-10}\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 1.2508 \times 10^{-12}\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 + 2.2726 \times 10^{-7}\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 0.3840 \quad (5)$$

$$f_{(\Delta G_{solv}(H_2O),neutral)}(\mathbf{r}) = 5.8152\cdot[\eta_L(\mathbf{r})]^{3/2} + 2.7545\cdot\sqrt{|V(\mathbf{r})\cdot EA_L(\mathbf{r})|} + 2.9644\cdot V(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 3.5905\cdot[IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 2.3712\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 5.6818\cdot[\alpha_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^{3/2} + 4.1716\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r}) + 0.7695\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{5/2} - 6.4000\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 5.3263\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{3/2} + 0.3210\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 1.9412\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})|]^{5/2} - 1.2373\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 + 1.6705\cdot V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r}) - 2.6703 \quad (6)$$

Free Energy of Solvation in *n*-Octanol. Figure 3 shows the performance of the model developed for the free energy of solvation in *n*-octanol. The detailed results are shown in Table S4 of the Supporting Information. The regression equations based on the raw (eq 7) and standardized data (eq 8) are shown below.

$$f_{(\Delta G_{solv}(octanol))}(\mathbf{r}) = 3.1998 \times 10^{-5}\cdot V(\mathbf{r})\cdot EA_L(\mathbf{r}) - 8.1936 \times 10^{-7}\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})|]^{3/2} + 4.3643 \times 10^{-9}\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 2.2598 \times 10^{-12}\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{3/2} + 3.0632 \times 10^{-23}\cdot[V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 - 4.3305 \times 10^{-16}\cdot[V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 + 2.0355 \times 10^{-4}\cdot\sqrt{|V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|} + 5.0025 \times 10^{-21}\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 1.9075 \times 10^{-8}\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^2 - 1.2941 \times 10^{-8}\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^{3/2} - 0.0501 \quad (7)$$

$$f_{(\Delta G_{solv}(octanol))}(\mathbf{r}) = 0.94417\cdot V(\mathbf{r})\cdot EA_L(\mathbf{r}) - 4.7881\cdot[|V(\mathbf{r})\cdot EA_L(\mathbf{r})|]^{3/2} + 1.5765\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 - 1.3548\cdot[|V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|]^{3/2} + 1.4514\cdot[V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 - 0.4122\cdot[V(\mathbf{r})\cdot IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^3 + 4.2942\cdot\sqrt{|V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})|} + 0.7293\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^3 + 3.0863\cdot[V(\mathbf{r})\cdot EA_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})]^2 - 5.3489\cdot[IE_L(\mathbf{r})\cdot\alpha_L(\mathbf{r})\cdot\eta_L(\mathbf{r})]^{3/2} - 4.9184 \quad (8)$$

As mentioned above, this function is far simpler than that found for water, although the performance (Figure 3) is almost identical to that of the water model for neutral molecules. It is gratifying to note that the electrostatic potential plays no direct role in the octanol solvation model but occurs exclusively in combination with the local electron affinity, the local ionization energy or both.

Free Energy of Solvation in Chloroform. Figure 4 shows the performance of the model developed for the free energy of solvation in chloroform. The detailed results are shown in Table S5 of the Supporting Information. The regression

equations based on the raw and standardized data are shown in eqs 9 and 10, respectively. As for the *n*-octanol model, the chloroform model consists of seven terms including the constant and is therefore parsimonious and probably quite robust. The models for the two organic solvents are far more satisfactory in this respect than that for water. As suggested above, however, this is probably a consequence of the complexity of the interactions between water solvent and dissolved molecules.

$$f_{(\Delta G_{solv}(CHCl3))}(\mathbf{r}) = -4.1618 \times 10^{-7} \cdot IE_L(\mathbf{r}) \cdot EA_L(\mathbf{r}) - 1.8162 \times 10^{-17} \cdot [IE_L(\mathbf{r}) \cdot \eta_L(\mathbf{r})]^3 - 8.4019 \times 10^{-11} \cdot [|V(\mathbf{r}) \cdot EA_L(\mathbf{r}) \cdot \alpha_L(\mathbf{r})|]^{5/2} - 1.1914 \quad (9)$$

$$f_{(\Delta G_{solv}(CHCl3))}(\mathbf{r}) = -0.6431 \cdot IE_L(\mathbf{r}) \cdot EA_L(\mathbf{r}) - 1.9455 \cdot [IE_L(\mathbf{r}) \cdot \eta_L(\mathbf{r})]^3 - 0.3690 \cdot [|V(\mathbf{r}) \cdot EA_L(\mathbf{r}) \cdot \alpha_L(\mathbf{r})|]^{5/2} - 5.8306 \quad (10)$$

Note at this stage that the model for water is considerably more complex than those given below for *n*-octanol and chloroform. We interpret this as indicating that the model for water must take far more possible types of interaction (Lewis-base, hydrogen-bond donor and acceptor, etc.) than for the less versatile organic solvents.

Enthalpy of Hydration. If we consider the above models to be linear free energy relationships, we should not expect to be able to extend them to the enthalpy of hydration. Nevertheless, we have constructed a model for the enthalpy of hydration in order to test the feasibility of such an approach. The results are shown in Figure 5 and in eqs 11 and 12.

$$f_{(\Delta H_{solv}(H2O))}(\mathbf{r}) = -1.8681 \times 10^{-12} \cdot [|V(\mathbf{r}) \cdot \eta_L(\mathbf{r})|]^{5/2} + 2.9878 \times 10^{-6} \cdot V(\mathbf{r}) \cdot IE_L(\mathbf{r}) \cdot \alpha_L(\mathbf{r}) - 1.5273 \times 10^{-12} \cdot [|IE_L(\mathbf{r}) \cdot EA_L(\mathbf{r}) \cdot \eta_L(\mathbf{r})|]^{3/2} - 2.1388 \quad (11)$$

$$f_{(\Delta H_{solv}(H2O))}(\mathbf{r}) = -27.2662 \cdot [|V(\mathbf{r}) \cdot \eta_L(\mathbf{r})|]^{5/2} + 3.0720 \cdot V(\mathbf{r}) \cdot IE_L(\mathbf{r}) \cdot \alpha_L(\mathbf{r}) - 4.0479 \cdot [|IE_L(\mathbf{r}) \cdot EA_L(\mathbf{r}) \cdot \eta_L(\mathbf{r})|]^{3/2} - 25.7473 \quad (12)$$

Equation 11 shows the regression equation obtained for the raw data and eq 12 that for the standardized data. The model only contains four terms, including the constant, compared with 25 for the free-energy model (eqs 3 and 4).

Validation for logP_{OW}. The models described above for the free energies of solvation in water and *n*-octanol were used to predict the water−octanol partition coefficient, logP_{OW} as

$$\log P_{OW} = \frac{\Delta G_{solv}(H_2O) - \Delta G_{solv}(n\text{-octanol})}{2.303RT} \quad (13)$$

The data set used was taken from our earlier work on logP_{OW}[41] but without nitro- and iodine-containing compounds, as discussed above. The results are shown in Figure 5. The data show a slight systematic trend toward too positive calculated logP_{OW} values (MSE = −0.43) and a mean unsigned error about twice as large as that obtained in our original neural net model.[41] The one very significant outlier

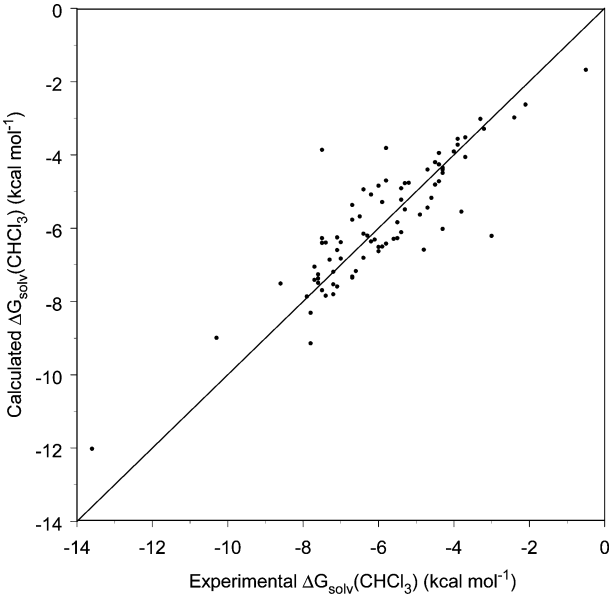


Figure 4. Experimental and calculated free energies of solvation in chloroform for the training data set given in Table S5. $N = 85$, MSE = 0.00, MUE = 0.61, RMSD = 0.81, $R^2 = 0.82$, $R^2_{cv} = 0.62$, slope = 1.00, intercept = 0.00.

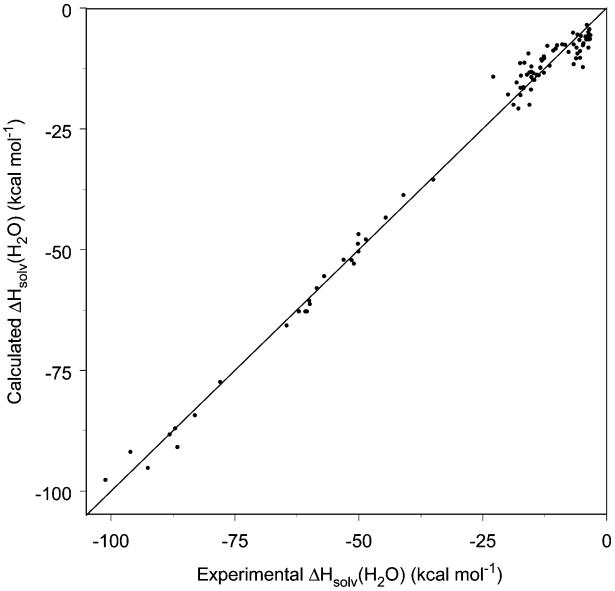


Figure 5. Experimental and calculated enthalpies of solvation in water for training data set given in Table S6. $N = 93$, MSE = 0.00, MUE = 2.24, RMSD = 2.87, $R^2 = 0.99$, $R^2_{cv} = 0.98$, slope = 1.00, intercept = 0.00.

is phenylalanine, which may exist in different equilibrium mixtures (un-ionized, zwitterions, dimer, etc.) in the two different solvents. This situation may, however, also apply to other compounds in the test set and is therefore not sufficient to explain the large deviation found for phenylalanine. Clearly, a model based on eq 11 can only work if the structure calculated is the same in the two solvents.

Equation 11 thus provides a useful independent validation of the models shown in eqs 3 and 7 and underlines the robustness of the surface-integral model approach.

The Solvation Energy as a Local Property. The above results suggest that the approximation of a local free energy of solvation on the molecular surface is relatively accurate. We can therefore consider visualizing this quantity as an aid

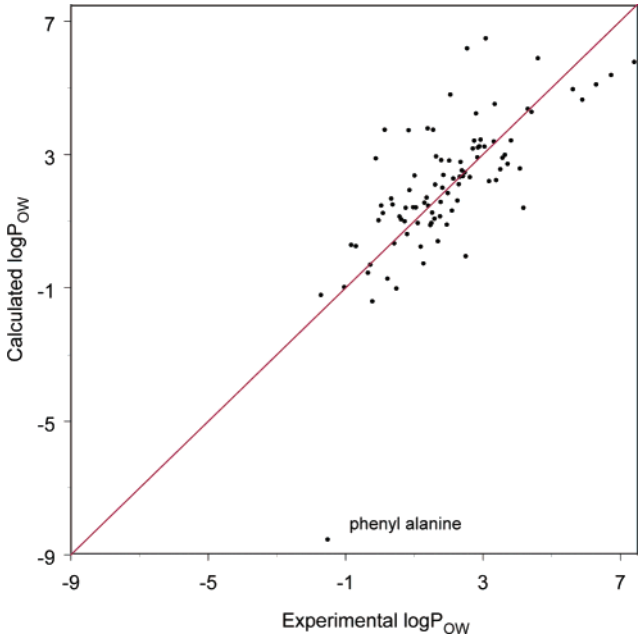


Figure 6. Experimental and calculated values of logP_{OW} for the test data set given in Table S7. $N = 91$, MSE $= -0.20$, MUE $= 0.91$, RMSD $= 1.24$, $R^2 = -0.54$, $R^2_{cv} = 0.49$, slope $= 0.73$, intercept $= 0.39$. The outlier phenylalanine is marked and was excluded from statistics.

to understanding the binding properties of molecules. Figure 7 shows the results for salmeterol, a particularly lipophilic drug. The first striking feature is that the areas of very negative local solvation energy (toward the blue end of the scale and highlighted in the lower picture) are strongly localized and correspond to known molecular binding features such as ether and hydroxyl oxygens, amino nitrogens, and the π-faces of the two aromatic rings. Perhaps surprisingly, areas of quite unfavorable local solvation energy surround the hydroxyl hydrogens, although they can clearly interact favorably with water. This may indicate that the model suffers from some collinearity (hydroxyl hydrogens never occur without their oxygens) or that the compound cannot compete with water as a hydrogen-bond donor, thus making the net effect of breaking up the water hydrogen-bond network to solvate the substrate's hydroxyl hydrogens unfavorable.

However, the local free energy of solvation in water does correspond remarkably well to our ideas of intermolecular interaction sites.

DISCUSSION

The techniques presented here involve two approximations: that the target properties can be treated using local contributions at the surface of the molecule and that gas phase semiempirical MO-electron densities can be used to derive properties that depend on the presence of a polar medium. The results of the surface-integral models suggest that both approximations are relatively well justified. The models are not as accurate as the most reliable approaches currently available (probably the SM*n* solvation models of Cramer and Truhlar,[37,38] with an MUE for the free energy of solvation in water of 0.49 kcal mol^{-1} or the GB/PSA models of Friesner et al.[42] with an MUE of 0.60 kcal mol^{-1}). However, they do give quite reliable estimates of the free energies of solvation within a consistent theoretical framework and are remarkably successful considering that they are based on the gas-phase electron density and can only include solute polarization indirectly via the local polarizability.

The models are also quite robust and can in principle be used to derive further properties. However, as demonstrated above for logP_{OW}, the model obtained is not of useful accuracy, perhaps not surprisingly as it depends on two other models. Merz et al.[18] have pointed out the difficulties encountered when predicting logP_{OW} based solely on models

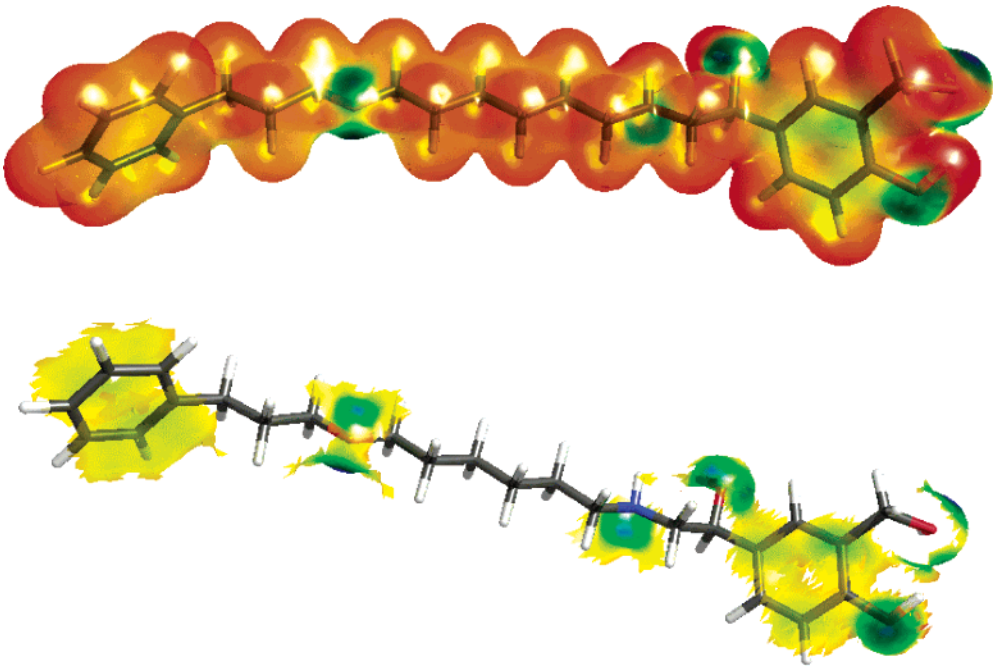
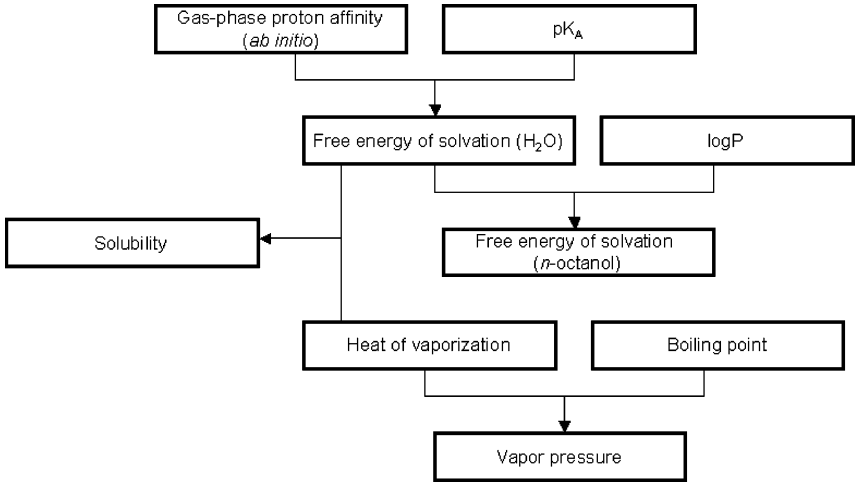


Figure 7. The local free energy of solvation in water, calculated using eq 3 at the isodensity surface (0.008 e$^-$ Å^{-3}) of salmetrol. The color scale ranges from -190 (blue) to $+53$ kcal mol^{-1} Å^{-2} (red). The upper structure shows the entire surface and the lower on only those areas for which the local solvation energy is more negative than -20 kcal mol^{-1} Å^{-2}. The visualization was performed with in-house software.

Chart 1



for free energies of solvation in water and octanol and have highlighted the benefits of data derived both from free-energy and logP_{OW} measurements. Combining data sources in this fashion represents a major opportunity for deriving accurate training data sets with a minimum of inconsistencies. As shown below, a number of different properties can be linked to provide cross checks and even to give more accurate free energies than those measured directly by deriving them from equilibrium data like pK_As or logP_{OW}. An important aspect of this work will be the ability to calculate, for instance, gas-phase proton affinities accurately with extrapolation techniques based on ab initio calculations (see Chart 1). These data will also provide enough resolution in the parametrization data to be able to treat problems such as the effect of conformation on properties. Our current models have relied on a standard protocol to generate conformations because the experimental data are not accurate or well-defined enough to detect conformational influences. The effects of this approximation are, however, within the uncertainties of current models.[43]

One significant disadvantage of the current models is their lack of an obvious physical interpretation. However, this is a general feature of surface-based models, both QSAR and QSPR. Traditionally, the physical interpretability of activity or property models has been placed high on the list of desirable properties. Clearly, in moving away from atomistic modeling, we also depart from the traditional chemists' picture of molecular structure and so must accept models with less direct physical interpretability. There are two possible solutions to this problem. We must either learn to recognize the significance of terms such as those that appear in the models given above or we must develop interrogation software that assesses the impact of structural changes and presents it in terms familiar to chemists. We are pursuing both approaches and hope to be able to recognize some of the important functions as we develop more models. However, we would like to emphasize that the models presented here are based on regression and therefore may not be as amenable to physical interpretation as we might like.

ACKNOWLEDGMENT

We thank Pfizer Ltd for financial support.

Supporting Information Available: Functions included in the regression analysis (Table S1), experimental and calculated free energies of solvation in water for the entire training data set (Table S2), in water for the neutral compounds of the training data set (Table S3), in *n*-octanol for the training data set (Table S4), in chloroform for the training data set (Table S5), and in water for the training data set (Table S6), and experimental and calculated logP_{OW} and calculated free energies of salvation in water ($\Delta G(H_2O)$) and *n*-octanol (ΔG(oct)) for the test set (Table S7). This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES AND NOTES

(1) Clark, T. Modelling the Chemistry: Time to Break the Mould? EuroQSAR 2002: *Designing drugs and crop protectants*; Ford, M., Dearden, J., Eds.; 2003; pp 111−121.
(2) Ehresmann, B.; Martin, B.; Horn, A. H. C.; Clark, T. Local molecular properties and their use in predicting reactivity. *J. Mol. Model.* **2003**, *9*, 342−347.
(3) Murray, J. S.; Politzer, P. Statistical Analysis of the Molecular Surface Electrostatic Potential: An Approach to Describing Noncovalent Interaction in Condensed Phases. *J. Mol. Struct (THEOCHEM)* **1998**, *425*, 107−114. Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grince, M. E.; Politzer, P. Relationships of Critical Constants and Boiling Points to Computed Molecular Surface Properties. *J. Phys. Chem.* **1993**, *97*, 9369−9373.
(4) Sjoberg P.; Murray J. S.; Brinck T.; Politzer P. A. Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity. *Can. J. Chem.* **1990**, *68*, 1440−1443.
(5) Politzer, P.; Murray, J. S.; Grice, M. E.; Brinck, T.; Ranganathan, S. Radial behavior of the average local ionization energies of atoms *J. Chem. Phys.* **1991**, *95*, 6699−6704.
(6) Politzer, P.; Murray, J. S.; Concha, M. C. The complementary roles of molecular surface electrostatic potentials and average local ionization energies with respect to electrophilic processes. *Int. J. Quantum Chem.* **2002**, *88*, 19−27 and references therein.
(7) Hussein, W.; Walker, C. G.; Peralta-Inga, Z.; Murray, J. S. Computed electrostatic potentials and average local ionization energies on the molecular surfaces of some tetracyclines. *Int. J. Quantum Chem.* **2001**, *82*, 160−169.
(8) Murray, J. S.; Abu-Awwad, F.; Politzer, P. Characterization of aromatic hydrocarbons by means of average local ionization energies on their molecular surfaces. *J. Mol. Struct. (THEOCHEM)* **2000**, *501*−*502*, 241−250.
(9) Martin, B.; Gedeck, P.; Clark, T. An Additive NDDO-Based Atomic Polarizability Model. *Int. J. Quantum Chem.* **2000**, *77*, 473−497.
(10) Rinaldi, D.; Rivail, J. L. Calculation of molecular electronic polarizabilities. Comparison of different methods. *Theor. Chim. Acta* **1974**, *32*, 243−251. Rinaldi, D.; Rivail, J. L. Molecular polarisability and dielectric effect of medium in the Liquid phase. Theoretical study of the water molecule and its dimers. *Theor. Chim. Acta* **1973**, *32*, 57−70.
(11) Schürer, G.; Gedeck, P.; Gottschalk, M.; Clark, T. Accurate Parametrized Variational Calculations of the Molecular Electronic Polarizability by NDDO-Based Methods. *Int. J. Quantum Chem.* **1999**, *75*, 17−31.

(12) Ehresmann, B.; de Groot, M. J.; Alex, A.; Clark, T. New Molecular Descriptors Based on Local Properties at the Molecular Surface and a Boiling-Point Model Derived from Them. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 658−668.
(13) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys*. **1965**, *43*, S129−S135.
(14) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. **1985**, *107*, 3902−3909. Holder, A. J. *AM1, Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 8−11.
(15) Pixner, P.; Heiden, W.; Merx, H.; Möller, A.; Moeckel, G.; Brickmann, J. Empirical Method for the Quantification and Localization of Molecular Hydrophobicity. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309−1319.
(16) Jäger, T.; Schmidt, F.; Schilling, B.; Brickmann, J. Localization and quantification of hydrophobicity; The molecular free energy density (MolFESD) concept and ist application to the sweetness recognition. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 631−646.
(17) Jäger, R.; Kast, S. M.; Brickmann, J. Parametrization Strategy for the MolFESD Concept: Quantitative Surface Representation of Local Hydrophobicity. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 237−247.
(18) Best, S. A.; Merz, K. M., Jr.; Reynolds, C. H. GB/SA-Based Continuum Model for Octanol. *J. Phys. Chem. B* **1997**, *101*, 10479−10487. Best, S. A.; Merz, K. M., Jr.; Reynolds, C. H. Free Energy Perturbation Study of Octanol/Water Partition Coefficients: Comparison with Continuum GB/SA Calculations. *J. Phys. Chem. B* **1999**, *103*, 714−726.
(19) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G.: Solvation Model for Chloroform Based on Class IV Atomic Charges. *J. Phys. Chem. B* **1997**, *101*, 2061−2069.
(20) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(21) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M., Jr.. Charge-Transfer Interactions in Macromolecular Systems: A New View of the Protein/Water Interface. *J. Am. Chem. Soc.* **1998**, *120*, 5593−5594.
(22) Van Der Vaart, A.; Merz, K. M., Jr. Charge transfer in biologically important molecules: comparison of high-level ab initio and semi-empirical methods. *Int. J. Quantum Chem.* **2000**, *77*, 27−43.
(23) Horn, A. H. C.; Lin, J.-H.; Clark, T. Multipole electrostatic model for MNDO-like techniques with minimal valence *spd*-basis sets. *Theor. Chem. Acc*. 2005, in press.
(24) Pascual-Ahuir, J. L.; Silla, E.; Tuñon, I. GEPOL: An improved Description of Molecular Surfaces III. A New Algorithm for the Computation of a Solvent-Excluded Surface. *J. Comput. Chem.* **1994**, *15*, 1127−1138.
(25) Heiden, W.; Goetze, T.; Brickmann, J. Fast generation of molecular surfaces from 3D data fields with an enhanced "marching cube" algorithm. *J. Comput. Chem.* **1993**, *14*, 246−50.
(26) Clark, T.; Lin, J.-H.; Horn, A. H. C. ParaSurf 1.0, Computer-Chemie-Centrum, University of Erlangen, Erlangen, 2004.
(27) Rauhut, G.; Clark, T. Multicenter Point Charge Model for High Quality Molecular Electrostatic Potentials from AM1 Calculations. *J. Comput. Chem.* **1993**, *14*, 503−509. Beck, B.; Rauhut, G.; Clark, T. The Natural Atomic Orbital Point Charge Model for PM3: Multipole Moments and Molecular Electrostatic Potentials. *J. Comput. Chem*. **1994**, *15*, 1064−1073.
(28) Tsar 3.3, Oxford Molecular Ltd., Oxford, 2000.
(29) Wilhelm, E.; Battino, R.; Wilcock, R. J. Low-Pressure Solubility of Gases in Liquid Water. *Chem. Rev.* **1977**, *77*, 219−262.
(30) Aue, D. H.; Webb, H. M.; Bowers, M. T. A Thermodynamic Analysis of Solvation Effects on the Basicities of Alkylamines. An Electrostatic Analysis of Substituent Effects. *J. Am. Chem. Soc*. **1976**, *98*, 318−329.
(31) Wilson, B.; Georgiadis, R.; Bartmess, J. E.: Enthalpies of Solvation of Ions. Aliphatic Carboxylyc Acids: Steric Hindrance to Solvation? *J. Am. Chem. Soc*. **1991**, *113*, 1762−1766.
(32) Haberfield, P.; Rakshit, A. K.: Aqueous vs Gas-Phase Acidities of the Haloacetic Acids. Enthalpies of Hydration of Haloacetic Acids and Haloacetate Ions. *J. Am. Chem. Soc*. **1976**, *98*, 4393−4395.
(33) Kubo, M. M.; Gallicchio, E.; Levy, R. M. Thermodynamic Decomposition of Hydration Free Energies by Computer Simulation: Application to Amines, Oxides, and Sulfides. *J. Phys. Chem. B* **1997**, *101*, 10527−10534.
(34) Furuki, T.; Sakurai, M.; Inoue, Y.; Chujo, R.; Harata, K. A semi-empirical methodology applicable to the accurate calculation of hydration enthalpy of organic molecules. *Chem. Phys. Lett*. **1992**, *188*, 584−588.
(35) Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P. A. Solvation Model based on Weighted Solvent Accessible Surface Area. *J. Phys. Chem. B* **2001**, *105*, 5055−5067.
(36) Schüürmann, G. Prediction of Henry's Law Constant of Benzene Derivatives Using Quantum Chemical Continuum-Solvation Models. *J. Comput. Chem*. **2000**, *21*, 17−34.
(37) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. New Universal Solvation Model and Comparison of the Accuracy of the SM5.42R, SM5.43R, C.-PCM, D.-PCM, and IEF-PCM Continuum Solvation Models for Aqueous and Organic Solvation Free Energies and for Vapor Pressures. *J. Phys. Chem. A* **2004**, *108*, 6532−6542.
(38) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G.: Solvation Model for Chloroform Based on Class IV Atomic Charges. *J. Phys. Chem. B* **1997**, *101*, 2061−2069.
(39) Breindl, A.; Beck, B.; Clark, T.; Glen, R. C.; Prediction of the *n*-Octanol/Water Partition Coefficient, logP, Using a Combination of Semiempirical MO−Calculations and a Neural Network. *J. Mol. Model.* **1997**, *3*, 142−155.
(40) Stewart, J. J. P. Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem*. **1989**, *10*, 209−220. Stewart, J. J. P. Optimization of parameters for semiempirical methods. II. Applications. *J. Comput. Chem*. **1989**, *10*, 221−264.
(41) Beck, B.; Breindl, A.; Clark, T. QM/NN QSPR Models with Error Estimation: Vapor Pressure and logP. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1046−1051.
(42) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Honig, B.; Ringnalda, M.; Goddard, W. A. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. *J. Am. Chem. Soc.* **1994**, *116*, 11875−82.
(43) Chalk, A. J.; Beck, B.; Clark, T. A Quantum Mechanical/Neural Net Model for Boiling Points with Error Estimation. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 457−462.

CI050025N