# Prediction of water–octanol partition coefficients using theoretical descriptors derived from the molecular surface area and the electrostatic potential

**Markus Haeberlein**† **and Tore Brinck** *,†

*Royal Institute of Technology, Department of Chemistry, Physical Chemistry,*
*S-100 44 Stockholm, Sweden*

**Octanol–water partition coefficients are frequently used in quantitative structure–activity relationships. A correlation based on computed theoretical descriptors is presented for the prediction of octanol–water partition coefficients ($P_{ow}$). An *ab initio* SCF approach was used to compute the molecular descriptors at the HF/6-31G* level. It was shown that only three theoretical parameters representing a cavity term, a dipolarity/polarisability term and a hydrogen bonding term were needed for the correlation. The corresponding parameters were deduced from the molecular surface area, the surface electrostatic potential and the spatial minima of the electrostatic potential, respectively. The predictive power of log $P_{ow}$ was demonstrated on a number of molecules which have biological activity.**

Octanol–water partition coefficients, $P_{ow}$, defined as the ratio of the equilibrium concentration of a solute in octanol to that in water, are amongst the first and most frequently used parameters in QSARs (Quantitative structure activity relationships). They have been used in numerous examples for quantitative predictions of the activity of organic molecules in their interactions with proteins.[1] Pharmacological and environmental research is often concerned with molecules which are poorly characterised, or not yet synthesised, and for which the octanol–water partition coefficients are not known. It is therefore of great importance to be able to estimate them accurately. The fragmental method of Leo and Hansch[2] have been used extensively for prediction of octanol–water partition coefficients. However, this requires that parametrisation has been conducted for molecules with the same type of fragments as those of the molecules of interest.

Kamlet and co-workers have shown that many solution properties can be expressed as a function of mainly three types of terms:[3] cavity terms, dipolarity/polarisability terms, and hydrogen-bond acceptor and donor terms. Their LSER (linear solvation energy relationship) for the partitioning of a solute between water and octanol is written as eqn. (1),[4] where $V_I$ is the

$$\log P_{ow} = mV_I + s(\pi^* + d\delta) + b\beta_m + a\alpha_m + C \qquad (1)$$

van der Waals volume of the solute. $\pi^* + d\delta$ takes account of the dipolarity/polarisability interactions, and the parameters $\beta_m$ and $\alpha_m$ are hydrogen-bond acceptor and donor terms, respectively. The $\beta_m$ and $\alpha_m$ that originally were used were solvent parameters. However, they have been shown to function satisfactorily also as solute parameters.[4] Abraham *et al.* have developed solute parameters of the hydrogen-bond basicity,[5,6] $\beta_2^H$, and hydrogen-bond acidity,[7,8] $\alpha_2^H$.

Several molecular orbital (MO) approaches for correlating log $P_{ow}$ have models with parameters which more or less resemble the LSER parameters. Bodor *et al.* used 14 parameters[9] and later 18 parameters[10] in their correlation with log $P_{ow}$. They used several structural descriptors such as the surface area, surface area squared, ovality, and the molecular mass, to describe the cavity term. To describe the hydrogen-bonding, combinations of the atomic charges on oxygen and nitrogen were used. For alkanes an indicator variable was used, which is set to one if the molecule is an alkane and zero otherwise. Sasaki *et al.*[11] have only three parameters in their correlation: a surface tension term, which essentially is a cavity term, together with two interaction terms, one electrostatic term, and one charge transfer term. Famini and co-workers have developed a set of theoretical LSER parameters which they refer to as TLSER.[12] The TLSER consist of six parameters, however, when correlating with the octanol–water partition coefficients only two parameters were found to be significant.[13] These were the molecular volume and the maximum negative charge, the latter representing the hydrogen-bond accepting capability.

We have developed equations based on electrostatic potential properties calculated at the molecular surface.[14] In these equations no site-specific hydrogen-bonding parameters were included.

It has been shown that LSER can be used to predict octanol–water partition coefficients using a limited number of solute parameters, which were not originally designed for this purpose. It is therefore reasonable to believe that a similar equation, based on theoretically computed parameters, could give similar results. Many of the parameters in LSER and similar equations are difficult to determine, which prompts continuing efforts to find improved relationships between log $P_{ow}$ and theoretical descriptors. Compared to fragmental methods, MO approaches have the advantage that they do not need all the special parameters used in the expert systems. The results of the MO approaches reported so far do not have the same accuracy in the correlations with log $P_{ow}$ as the LSER. In recent years progress in computer developments have enabled us to compute descriptors for log $P_{ow}$ correlations at a higher level of theory than previously reported. New correlations with $\alpha_2^H$ and $\beta_2^H$ have shown that the level of theory is important for the accuracy of the correlations.[15] The purpose of the present study was to find theoretical parameters which give a good correlation with log $P_{ow}$. It is also important that the correlation is general and can give reliable predictions for drug molecules which usually have several functional groups and sometimes internal hydrogen-bonding.

## Methodological background

Politzer and co-workers have shown that theoretical descriptors based on the electrostatic potential, $V(\mathbf{r})$, and the average local ionisation energy, $\bar{I}(\mathbf{r})$, are well suited for the description of

---

† *E-mail:* markus@physchem.kth.se and tore@physchem.kth.se

several physicochemical parameters such as critical constants and boiling points,[16] and supercritical solubilities.[17] They have also presented correlations with hydrogen-bond acidity and basicity parameters,[15,18,19] and dipolarity/polarisability parameters.[20,21]

These theoretical descriptors form a good basis for the search of parameters to be used in correlations with log $P_{\mathrm{ow}}$. Below, some previously used theoretical descriptors are discussed in terms of the three types of interaction terms discussed by Kamlet and co-workers.[4]

### Bulk/cavity terms

The molecular volume and the molecular surface area are used mostly as bulk/cavity terms. There is no unique way to define the molecular volume or surface area, but most approaches try to define a surface contour similar to the van der Waals volume.

### Dipolarity/polarisability terms

The dipolarity/polarisability terms reflect electrostatic and polarisation interactions between the solute and the solvent. These interactions are dipole and induced-dipole interactions, together with higher multipole analogues. Therefore, it seems natural to design descriptors based on the electrostatic potential, which gives a complete description of molecular charge distribution, to investigate these properties. The electrostatic potential, $V(\boldsymbol{r})$, at a point $\boldsymbol{r}$ in the space of a molecule is defined by eqn. (2), where $Z_{\mathrm{A}}$ is the charge on nucleus $A$, located at $\boldsymbol{R}_{\mathrm{A}}$,

$$V(r) = \sum_{A} \frac{Z_{\mathrm{A}}}{|\boldsymbol{R}_{\mathrm{A}} - \boldsymbol{r}|} - \int \frac{p(\boldsymbol{r}')d\boldsymbol{r}'}{|\boldsymbol{r}' - \boldsymbol{r}|} \qquad (2)$$

and $p(\boldsymbol{r})$ is the electronic density of the molecule. $V(\boldsymbol{r})$ gives the interaction energy between a positive point charge of unit magnitude located at $\boldsymbol{r}$ and the unperturbed charge distribution of the molecule. For example, an approaching electrophile will initially be attracted to the points where $V(\boldsymbol{r})$ has the most negative values (the local minima $V_{\mathrm{min}}$). Politzer and co-workers have introduced several theoretical descriptors by calculating the electrostatic potential at the surface of the molecules and analysing it statistically.[22] A measurement of local polarity has been defined[20] as eqn. (3), where $V(\boldsymbol{r}_i)$ is the value of $V(\boldsymbol{r})$ at a

$$\Pi = \frac{1}{n} \sum_{i=1}^{n} |V(\boldsymbol{r}_i) - \overline{V}_{\mathrm{S}}| \qquad (3)$$

point $\boldsymbol{r}_i$ on the molecular surface and $\overline{V}_S$ is the average of the potential on the surface, $V(\boldsymbol{r}_i)$. $\Pi$ can be viewed as the average deviation of the electrostatic potential on the molecular surface and has been shown to correlate with the dipolarity/polarisability parameters, $\pi^* + d\delta$, as well as with the relative permittivity.[20] For substituted benzene derivatives a good correlation has also been found between $\Pi$ and Abraham's solute dipolarity/polarisability parameter,[21] $\pi_2^{\mathrm{H}}$. However, in correlations with log $P_{\mathrm{ow}}$ it has been shown that $\Pi$ multiplied by the surface area is more significant than using $\Pi$ alone.[14]

Another parameter which has been shown to be successful in correlations with log $P_{\mathrm{ow}}$[14] is $\sigma^2_{\mathrm{tot}}$ which is defined as eqn. (4),

$$\sigma^2_{\mathrm{tot}} = \sigma^2_{+} + \sigma^2_{-} = \frac{1}{m} \sum_{i=1}^{m} |V^{+}(\boldsymbol{r}_i) - \overline{V}_{\mathrm{S}}^{+}|^2 + \frac{1}{n} \sum_{j=1}^{n} |V^{-}(\boldsymbol{r}_j) - \overline{V}_{\mathrm{S}}^{-}|^2 \qquad (4)$$

where $V^{+}(\boldsymbol{r}_i)$ and $V^{-}(\boldsymbol{r}_i)$ are the positive and negative values, respectively, of the electrostatic potential at the surface. $V_s^{+}$ and $V_s^{-}$ are their averages. $\sigma^2_{\mathrm{tot}}$ reflects the variability of the electrostatic potential over the surface and can be viewed as a measure of the electrostatic interaction tendency. $\sigma^2_{\mathrm{tot}}$ and $\sigma^2_{-}$ have been shown to be highly significant in correlations with log $P_{\mathrm{ow}}$.[14] In these correlations the $\sigma^2$ terms can also be viewed partly as hydrogen-bond terms, since the electrostatic potential has its extremes at hydrogen-bond donor and acceptor sites which results in large contributions to the $\sigma^2$ terms from these sites.

### Hydrogen-bonding terms

Hydrogen-bonding may be divided into an electrostatic term and a polarisation/charge transfer term. The electrostatic term is often dominant.[23,24] For example, the spatial and the surface minima of the electrostatic potential, $V_{\mathrm{min}}$ and $V_{\mathrm{S,min}}$, respectively, have been shown to give good correlations with the solvent hydrogen-bond accepting parameter,[18,19] $\beta$, and the solute hydrogen-bond basicity term,[19] $\beta_2^{\mathrm{H}}$, within families of similar molecules. The hydrogen-bond acidity parameter of Abraham and co-workers, $a_2^{\mathrm{H}}$, has been found to correlate well with the surface maxima of the electrostatic potential, $V_{\mathrm{S,max}}$.[15]

The average local ionisation energy, $\bar{I}(\boldsymbol{r})$, was originally introduced by Sjöberg *et al.*[25] and is defined within the Hartree–Fock theory by eqn. (5), where $\rho_i(\boldsymbol{r})$ is the electronic density of

$$\bar{I}(\boldsymbol{r}) = \sum_{i} \frac{\rho_i(\boldsymbol{r})|\varepsilon_i|}{\rho(\boldsymbol{r})} \qquad (5)$$

the *i*th molecular orbital at the point $\boldsymbol{r}$ and $\varepsilon_i$ is the orbital energy. Since Koopmans' theorem[26] justifies the orbital energies as approximations to the ionisation energies, $\bar{I}(\boldsymbol{r})$ can be interpreted as the average energy required to remove an electron located at the point $\boldsymbol{r}$ from the molecule. Sjöberg *et al.* showed that the positions on the molecular surface where $\bar{I}(\boldsymbol{r})$ has its lowest values ($\bar{I}_{\mathrm{S,min}}$) are indicative of the sites that are most reactive towards electrophiles.[25] For the analysis of electrophilic interactions we have found $\bar{I}_{\mathrm{S,min}}$ to be complementary to the electrostatic parameters, since it reflects charge transfer and polarisation.[27] A relationship including both $V(\boldsymbol{r})$ and $\bar{I}(\boldsymbol{r})$ has been shown to be more general than those using only the electrostatic potential.[27]

In a recent study, it has been shown that gas-phase proton affinities, complexation energies for phenol–base, and iodine–base complexes and shifts in the OH-stretching frequencies for methanol–base complexes can be correlated with the following relationship,[28] eqn. (6), where $P_{\mathrm{Vmin}}$ is the polarisation correc-

$$-\Delta H \text{ (or } \Delta\nu_{\mathrm{OH}}) = \alpha V_{\mathrm{min}} + \beta P_{\mathrm{Vmin}} + \gamma \bar{I}_{\mathrm{S,min}} + \varepsilon \qquad (6)$$

tion of the electrostatic potential at the position of $V_{\mathrm{min}}$. Good relationships have also been found between $\beta_2^{\mathrm{H}}$ and $V_{\mathrm{S,min}}$, $\bar{I}_{\mathrm{S,min}}$ and $\Pi$ taken together.[22]

## Methods and procedure

Geometry optimisations of the solute molecules were carried out at the HF/6-31G* level using GAUSSIAN 94.[29] From these geometries the molecular properties were calculated at the same level of theory. The 6-31G* basis set has been shown to predict accurate geometries[30] and is well suited for the determination of electrostatic potentials, as well as average local ionisation energies.[25,31,32] Previous studies have shown that for some relationships it is necessary to compute descriptors derived from the electrostatic potential at the 6-31G* level, in order to obtain good correlations.[15,33] To confirm that the correct molecular conformations were obtained, all molecules were first optimised with STO-3G* and subsequent frequency calculations were conducted with the same basis set before the 6-31G* optimisation. The tremendous increase in computer performance in recent years has made it possible to compute molecules of the size of most common drug molecules at the HF/6-31G*-level. All properties on the molecular surface, such as the surface area, $\bar{I}_{\mathbf{S,min}}$, and the statistically computed properties of the electrostatic potential, were calculated at surfaces defined by a contour of constant electron density equal to 0.001 electrons/bohr.[3] Bader *et al.* have shown that this contour gives a good representation of the van der Waals surfaces of the molecules.[34,35] The statistical properties obtained from $V(\mathbf{r})$ were computed at a number of points on the molecular surface defined by a 0.2 bohr square grid.

Approximately 100 theoretical descriptors reflecting the three types of terms (bulk/cavity, dipolarity/polarisability, hydrogen bonding) were designed. Correlations of combinations of the descriptors were tested unprejudiced in order to find the most significant parameters.

## Results and discussion

Table 1 lists the experimental and predicted octanol–water partition coefficients for 74 molecules. The compounds were chosen to cover a wide range of log $P_{\mathbf{ow}}$ values. Efforts were also made to select molecules with large variations in size, shape and functional groups in order to obtain a general relationship between log $P_{\mathbf{ow}}$ and computationally calculated descriptors. The set includes among others, heterocycles, amines, amides, alkanes and molecules with more than one functional group.

A large number of theoretical descriptors were tested using different models in order to obtain a correlation with a good predicting power. Eqn. (7) was found to give the best three parameter model, and the best four-parameter model was given by eqn. (8), where $A_{\mathbf{d}}$ is the surface area (defined as the isodensity surface of 0.001 electrons/bohr[3]). $\mu$ is the dipole moment, and $V_{\mathbf{d}}$ is the molecular volume. $\overline{V_-^2}$ is the mean of the squares of the negative potential points on the surface, eqn. (8′). The last term is the sum of the minimum values of the electrostatic potential where only $V_{\mathbf{min}}$ values more negative than −147 kJ mol$^{-1}$ (−35 kcal mol$^{-1}$) are included. If two $V_{\mathbf{min}}$ appeared within 2.1 Å of each other only the minimum having the most negative potential was included in the sum.

$$\log P_{\mathbf{ow}} = \alpha A_{\mathbf{d}} + \beta A_{\mathbf{d}}\overline{V_-^2} + \delta\Sigma V_{\mathbf{min}} + \varepsilon \qquad (7)$$
$$n = 74 \quad s = 0.316 \quad R = 0.979$$

$$\log P_{\mathbf{ow}} = \alpha A_{\mathbf{d}} + \beta A_{\mathbf{d}}\overline{V_-^2} + \gamma\frac{\mu^2}{V_{\mathbf{d}}} + \delta\Sigma V_{\mathbf{min}} + \varepsilon \qquad (8)$$
$$n = 74 \quad s = 0.292 \quad R = 0.983$$

$$\overline{V_-^2} = \frac{1}{n}\sum_{i=1}^{n}[V^-(\mathbf{r}_i)]^2 \qquad (8')$$

The correlation given by eqn. (7) can be interpreted as having one cavity term, one polarity term and one hydrogen-bonding term whereas the correlation given by eqn. (8) has two polarity terms. In all our correlations the most significant parameter was the cavity term. We found that the surface area, $A_{\mathbf{d}}$, gave better correlations than the molecular volume. This agrees with the findings in a study by Hansch and Leo[1] in which they correlated log $P_{\mathbf{ow}}$ of alkanes with both the molecular surface area and the molecular volume; the surface area gave slightly better correlations than the volume. As mentioned by Camillieri *et al.*[36] the surface area might be better suited to taking care of branching and cyclisation effects, than other methods.

Great effort was made in order to find suitable dipolarity/polarisability terms. We found that the terms most significant in describing the dipolarity/polarisability was $\overline{V_-^2}$ multiplied with the surface area. It may seem more natural to use the surface area multiplied with the sum of $\overline{V_-^2}$ and $\overline{V_+^2}$ as a descriptor of the polarity, than only with $\overline{V_-^2}$. However, the hydrogen-bond accepting term has been shown to be dominant over the hydrogen-bond acidity term,[4,13,37] which indicates that the interactions with the positive parts of the solute are of minor importance for the estimation of log $P_{\mathbf{ow}}$. The $A_{\mathbf{d}}\overline{V_-^2}$ term can be interpreted as a factor that describes a decrease in the size of the surrounding cavity of the solute. When comparing the contributions of the different terms of eqn. (7) in Table 1 we can see that the $A_{\mathbf{d}}\overline{V_-^2}$ term gives a large contribution for some molecules, such as the PAH, whereas no contributions from the hydrogen-bonding parameters were found.

In our models the $\Pi$ parameter showed only small significance. Better significance was achieved when $\Pi$ was multiplied by the surface area. $\Pi$ is a size-independent parameter and, as we pointed out[14] in our earlier work, the reason that this parameter is more significant when multiplied with the surface area is that the strength of the interaction with bulk water depends upon the size of the solute.

The $\mu^2/V_{\mathbf{d}}$ term in eqn. (8) has its origin in the reaction field theory of Kirkwood and Onsager.[38] In this model, the free energy of solvation of a dipole inside a spherical cavity is equal to $k(\varepsilon)\mu^2/a^3$, where $k(\varepsilon)$ is a function of the relative permittivity of the solvent and $a^3$ is the radius of the cavity. Our $\mu^2/V_{\mathbf{d}}$ term gives a positive contribution to log $P_{\mathbf{ow}}$. The interpretation of this behaviour is not straightforward. It may be viewed as a correction to the $A_{\mathbf{d}}\overline{V_-^2}$ term. In order to account for the polar interactions of molecules with a zero dipole moment, we also used the analogue of the Kirkwood–Onsager equation for the quadrupolar case.[38] However, no improvements were found using this formula.

LSER includes a polarisability parameter. Famini and coworkers in their TLSER also use a computed polarisability-index defined as the polarisability volume divided by the molecular volume.[13] We have shown that a good correlation exists between the polarisability of a molecule and its volume divided by the average of $\bar{I}(\mathbf{r})$ over the molecular surface,[39] $\bar{I}_{\mathbf{av}}$, which is defined by eqn. (9).

$$\bar{I}_{\mathbf{av}} = \frac{1}{n}\sum_{i=1}^{n}\bar{I}(\mathbf{r}_i) \qquad (9)$$

However, in the correlations with log $P_{\mathbf{ow}}$ this parameter and similar ones, for example $1/\bar{I}_{\mathbf{av}}$, showed no significance.

We found that the electrostatic potential minima, $V_{\mathbf{min}}$, was best suited to describe the hydrogen-bond accepting ability of the solute. The use of multiparameter descriptions of the hydrogen-bonding, including terms such as $\bar{I}_{\mathbf{S,min}}$ and the polarisation correction to the electrostatic potential, $P(\mathbf{r})$ [eqn. (6)] did not result in any improvement. This is in accordance with previous studies which have shown that hydrogen-bonding can often be described by an electrostatic term alone.[15,23,24]

We found that the most effective description of the hydrogen-bonding was obtained when the summation over the $V_{\mathbf{min}}$ was performed using two conditions: first, we added only the minima where $V_{\mathbf{min}}$ was more negative than −147 kJ mol$^{-1}$. Secondly, if two minima were found within 2.1 Å of each other only the most negative one was included in the sum. No improvement was found when adding a parameter for the number of $V_{\mathbf{min}}$. The use of a cut-off value for the hydrogen-bonding parameter is needed since very weak hydrogen-bond interactions are taken care of in the dipolarity/polarisability term.

We found the descriptor for hydrogen-bond acidity, $V_{\mathbf{S,max}}$, to be of no significance in our correlations. This finding is consistent with others who have shown that the hydrogen-bond acidity is of minor importance in relationships with the octanol–water partition coefficient.[4,13,37]

Table 2 lists the three and four-parameter equations together with our previously reported equations computed for this set of molecules. Very good predictions were made, for example, for heterocycles, alcohols and PAH. In our correlations no molecules with very large deviations from the experimental log $P_{\mathbf{ow}}$

**Table 1** Experimental octanol–water partition coefficients and calculated contributions for each parameter in log $P_{ow}$ [eqn. (7)] together with the predicted log $P_{ow}$.

| | | $\beta A_d \overline{V}_-^2$ | $\delta\Sigma V_{min}$ | log $P_{ow}$ (pred.) | log $P_{ow}$ (exp.)[a] | Residual |
|---|---|---|---|---|---|---|
| Propane | 2.85 | −0.01 | 0.00 | 1.95 | 2.36 | 0.41 |
| Pentane | 3.99 | −0.01 | 0.00 | 3.08 | 3.39 | 0.31 |
| Hexane | 4.57 | −0.01 | 0.00 | 3.66 | 3.90 | 0.24 |
| Cyclohexane | 3.89 | −0.01 | 0.00 | 2.99 | 3.44 | 0.45 |
| Benzene | 3.39 | −0.30 | 0.00 | 2.20 | 2.13 | −0.07 |
| Biphenyl | 5.57 | −0.46 | 0.00 | 4.21 | 4.01 | −0.20 |
| Naphthalene | 4.69 | −0.42 | 0.00 | 3.37 | 3.30 | −0.07 |
| Anthracene | 5.97 | −0.53 | 0.00 | 4.54 | 4.45 | −0.09 |
| Phenanthrene | 5.90 | −0.54 | 0.00 | 4.46 | 4.46 | 0.00 |
| Toluene | 3.96 | −0.35 | 0.00 | 2.72 | 2.73 | 0.01 |
| Methanol | 1.91 | −0.67 | −0.99 | −0.66 | −0.77 | −0.11 |
| Ethanol | 2.53 | −0.87 | −1.00 | −0.24 | −0.31 | −0.07 |
| Butan-1-ol | 3.68 | −0.73 | −0.99 | 1.07 | 0.88 | −0.19 |
| *tert*-Butyl alcohol | 3.53 | −0.79 | −0.99 | 0.85 | 0.35 | −0.50 |
| Neopentyl alcohol | 3.98 | −0.35 | −0.93 | 1.81 | 1.31 | −0.50 |
| Octan-1-ol | 5.95 | −0.80 | −0.99 | 3.26 | 3.00 | −0.26 |
| 2,2,2-Trifluoroethanol | 2.87 | −0.45 | −0.76 | 0.77 | 0.41 | −0.36 |
| 2-Naphthol | 4.92 | −0.61 | −0.76 | 2.66 | 2.70 | 0.04 |
| Benzaldehyde | 4.01 | −0.86 | −0.92 | 1.34 | 1.48 | 0.14 |
| Formic acid | 1.97 | −0.92 | −0.91 | −0.76 | −0.54 | 0.22 |
| Acetic acid | 2.58 | −0.85 | −0.89 | −0.04 | −0.17 | −0.13 |
| Butyric acid | 3.72 | −1.09 | −0.89 | 0.84 | 0.79 | −0.05 |
| Benzoic acid | 4.23 | −0.68 | −0.85 | 1.81 | 1.87 | 0.06 |
| Dimethyl ether | 2.54 | −0.63 | −0.93 | 0.09 | 0.10 | 0.01 |
| Diethyl ether | 3.77 | −0.47 | −0.92 | 1.49 | 0.89 | −0.60 |
| Diethyl sulfide | 4.12 | −0.61 | −0.64 | 1.97 | 1.95 | −0.02 |
| Ethylene oxide | 2.23 | −0.87 | −0.88 | −0.42 | −0.30 | 0.12 |
| Tetrahydrofuran | 3.23 | −0.94 | −1.02 | 0.37 | 0.46 | 0.09 |
| Dioxane | 3.40 | −0.84 | −1.76 | −0.09 | −0.27 | −0.18 |
| Anisole | 4.21 | −0.51 | −0.73 | 2.08 | 2.11 | 0.03 |
| Acetone | 2.92 | −1.44 | −0.94 | −0.35 | −0.24 | 0.11 |
| Pentan-2-one | 4.05 | −1.47 | −0.95 | 0.74 | 0.91 | 0.17 |
| Acetophenone | 4.52 | −0.97 | −0.95 | 1.71 | 1.58 | −0.13 |
| Methyl formate | 2.63 | −1.37 | −0.96 | −0.59 | 0.03 | 0.62 |
| Methyl acetate | 3.20 | −1.12 | −0.93 | 0.25 | 0.18 | −0.07 |
| Ethyl propionate | 4.38 | −1.22 | −0.94 | 1.32 | 1.21 | −0.11 |
| Acetonitrile | 2.31 | −1.02 | −0.95 | −0.55 | −0.34 | 0.21 |
| Propionitrile | 2.91 | −1.34 | −0.97 | −0.29 | 0.16 | 0.45 |
| Benzonitrile | 4.06 | −1.13 | −0.95 | 1.08 | 1.56 | 0.48 |
| Acetamide | 2.72 | −1.59 | −1.10 | −0.86 | −1.26 | −0.40 |
| *N*,*N*-Dimethylformamide | 3.27 | −1.91 | −1.11 | −0.64 | −1.01 | −0.37 |
| *N*,*N*-Dimethylacetamide | 3.77 | −1.91 | −1.13 | −0.16 | −0.77 | −0.61 |
| Urea | 2.52 | −1.30 | −2.40 | −2.08 | −2.11 | −0.03 |
| Methylamine | 2.08 | −0.74 | −1.38 | −0.93 | −0.57 | 0.36 |
| Ethylamine | 2.67 | −0.53 | −1.37 | −0.12 | −0.13 | −0.01 |
| Butylamine | 3.82 | −0.92 | −1.38 | 0.63 | 0.97 | 0.34 |
| Dimethylamine | 2.68 | −0.66 | −1.28 | −0.16 | −0.38 | −0.22 |
| Cyclopropylamine | 2.91 | −0.34 | −1.27 | 0.40 | 0.07 | −0.33 |
| 2,2,2-Trifluoroethylamine | 2.99 | −0.28 | −1.10 | 0.73 | 0.24 | −0.49 |
| Piperidine | 3.74 | −0.41 | −1.26 | 1.18 | 0.84 | −0.34 |
| Quinuclidine | 4.27 | −0.70 | −1.26 | 1.42 | 1.38 | −0.04 |
| Aniline | 3.79 | −0.60 | −1.03 | 1.27 | 0.90 | −0.37 |
| *p*-Toluidine | 4.36 | −0.63 | −1.06 | 1.77 | 1.39 | −0.38 |
| *m*-Aminophenol | 4.02 | −0.76 | −1.77 | 0.59 | 0.21 | −0.38 |
| Adenine | 4.26 | −0.85 | −3.16 | −0.65 | −0.09 | 0.56 |
| *p*-Trifluoromethylaniline | 4.68 | −0.49 | −0.83 | 2.47 | 2.39 | −0.08 |
| Pyridine | 3.27 | −0.56 | −1.16 | 0.65 | 0.65 | 0.00 |
| Quinoline | 4.56 | −0.60 | −1.13 | 1.94 | 2.03 | 0.09 |
| Pyrimidine | 3.10 | −0.68 | −2.06 | −0.54 | −0.40 | 0.14 |
| 2-Aminopyridine | 3.66 | −0.68 | −2.05 | 0.04 | 0.32 | 0.28 |
| 4-Methylpyridine | 3.84 | −0.74 | −1.20 | 1.01 | 1.22 | 0.21 |
| 3,5-Dichloropyridine | 4.37 | −0.25 | −0.90 | 2.32 | 2.56 | 0.24 |
| 1-Methylimidazole | 3.42 | −1.25 | −1.35 | −0.08 | −0.06 | 0.02 |
| Pyridine *N*-oxide | 3.52 | −2.48 | −1.26 | −1.11 | −1.20 | −0.09 |
| Furan | 2.81 | −0.27 | −0.60 | 1.04 | 1.34 | 0.30 |
| Thiophene | 3.18 | −0.23 | 0.00 | 2.06 | 1.81 | −0.25 |
| Nitrobenzene | 4.10 | −1.19 | −0.80 | 1.22 | 1.85 | 0.63 |
| *p*-Nitroaniline | 4.49 | −1.33 | −1.53 | 0.74 | 1.39 | 0.65 |
| 2,4-Dinitrophenol | 5.01 | −0.92 | −1.37 | 1.83 | 1.67 | −0.16 |
| Chloroform | 3.31 | −0.06 | 0.00 | 2.35 | 1.97 | −0.38 |
| Pentachlorophenol | 5.90 | −0.21 | 0.00 | 4.80 | 5.12 | 0.32 |
| Dimethyl sulfoxide | 3.16 | −2.48 | −1.26 | −1.48 | −1.35 | 0.13 |
| Methyl *p*-aminophenyl sulfone | 5.36 | −2.44 | −1.72 | 0.30 | −0.12 | −0.42 |
| Phenyl trifluoromethyl sulfone | 5.27 | −1.22 | −0.76 | 2.39 | 3.01 | 0.62 |

[a] Experimental log $P_{ow}$ values are taken from Hansch *et al.*[42] except the log $P_{ow}$ value for aniline which is taken from ref. 43.

**Table 2** Statistics of the correlation equations between log $P_{ow}$ and theoretical descriptors for the molecules given in Table 1

| Equation[a] | $R$ | SD | $F$ |
|---|---|---|---|
| $\log P_{ow} = 0.0278A_d - 9.99 \times 10^{-7} A_d\overline{V^2_-} + 3.92 \times 10^{-3}\Sigma V_{min} - 0.894$ | 0.979 | 0.316 | 545 |
| $\log P_{ow} = 0.0290A_d - 1.22 \times 10^{-6}A_dV^2_- + 2.02\mu^2/V_d + 3.81 \times 10^{-3}\Sigma V_{min} - 1.04$ | 0.983 | 0.292 | 481 |
| $\log P_{ow} = 0.0315A_d - 1.38 \times 10^{-4}A_d\Pi - 4.43 \times 10^{-4}\sigma^2_- - 1.31^b$ | 0.937 | 0.546 | 167 |
| $\log P_{ow} = 0.0340A_d - 1.19 \times 10^{-4}A_d\Pi - 1.32 \times 10^{-4}N\sigma^2_- - 2.13^b$ | 0.931 | 0.571 | 151 |

[a] In units of kJ mol$^{-1}$ for the electrostatic potential, Å$^2$ for the molecular surface area, Å$^3$ for the molecular volume and, debye for the dipole moment. [b] Relationship from ref. 14 computed at the HF/6-31G*-level using the set of molecules from the current study. $N$ is the number of lone pairs from oxygen and nitrogen atoms. However, $N$ is never lower than unity in order to retain the $\sigma^2_-$ term.

**Table 3** Predictions of log $P_{ow}$ of some biologically active molecules using eqn. (7)

| Mo | log $P_{ow}$ (pred.) | log $P_{ow}$ (exp.) | Residual |
|---|---|---|---|
| Caffeine | 0.18 | −0.07 | −0.25 |
| Clonidine | 2.08 | 1.57 | −0.51 |
| Morpholine | −0.42 | −0.86 | −0.44 |
| Nicotine | 1.64 | 1.17 | −0.47 |
| Sulfanilamide | −0.60 | −0.62 | −0.02 |
| Vanillin | 1.41 | 1.37 | −0.04 |

value were encountered. The largest deviations, when using eqn. (7) was found for *p*-nitroaniline with a deviation in log $P_{ow}$ of 0.65. In the four-parameter relation [eqn. (8)] log $P_{ow}$ deviated most for 2,2,2-trifluoroethanol ($\Delta\log P_{ow} = -0.57$). There were only a few systematic deviations from the correlation. The predicted log $P_{ow}$ of the alkanes were found to be lower than the experimental values. This phenomenon is also found in the LSER where the cavity parameter also gives log $P_{ow}$ values that are too low for the alkanes.[4] A possible explanation may be that the strong hydrophobicity of the alkanes induces cage-like clathrates around them, reducing the entropy of the water molecules to a greater extent than for other solutes.[40] Therefore, it can also be interpreted that the surface area is larger, since the mean distance between the solute and the water molecules increases. The predictions for the highly fluorinated compounds give log $P_{ow}$ values that are slightly too high. This may be due to the fact that the fluorine atoms have some hydrogen-bonding tendency, which is not accounted for here since the $V_{min}$ of the fluorines were only about −60 kJ mol$^{-1}$ compared to −170 to −360 kJ mol$^{-1}$ for the oxygen and nitrogen atoms. Table 2 also includes two relationships reported by us previously;[14] $A_d\Pi$ may be interpreted as a polarity term and $\sigma^2_-$ and $N\sigma^2_-$, where $N$ is the number of nitrogen and oxygen lone-pairs in the molecule, are more similar to hydrogen-bonding terms. The correlations are not as good as those reported previously,[14] which may depend on the fact that we used a slightly more diverse set of molecules and that the $\sigma^2$ descriptor is considerably different for some molecules when computing it, using the 6-31G* basis set compared to the STO-5G* basis set.

To test the predictive power of the correlation, given by eqn. (7), a number of biologically active molecules were used as a validation set. Table 3 lists the predicted log $P_{ow}$ values from eqn. (7) together with the experimental values. This set also included vanillin, a molecule having an internal hydrogen-bond. Molecules with internal hydrogen-bonds usually give poor predictions when using methods based on atomic charges.[41] Of the predicted log $P_{ow}$ values some were very close to the experimental (sulfanilamide and vanillin) and the largest deviation was only 0.51 log units (clonidine). These results show that eqn. (7) has a predictive capability even for more complex molecules.

When we used correlations having six to eight parameters the correlation coefficient and the standard deviation were usually better than for eqn. (7). However, the predictions often showed no improvements.

## Conclusions

Following the LSER approach we have presented a correlation between three computed theoretical descriptors and log $P_{ow}$. These descriptors can be interpreted as a cavity term, a dipolarity/polarisability term and a hydrogen-bonding term which give the descriptors a physical meaning. The relationship has been shown to be applicable to a wide range of molecules and successful predictions have been made for some molecules of biological and pharmacological interest which have several functional groups and also internal hydrogen-bonding.

## Acknowledgements

One of the authors (T. B.) thanks the Swedish Natural Science Research Council (NFR) for financial support.

## References

1 C. Hansch and A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, ACS, Washington DC, 1995.

2 A. Leo, P. Y. C. Jow, C. Silipo and C. Hansch, *J. Med. Chem.*, 1975, **18**, 865.

3 M. J. Kamlet, J. L. M. Abboud and R. W. Taft, *Progr. Phys. Org. Chem.*, 1981, **13**, 485.

4 M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. H. Marcus and R. W. Taft, *J. Phys. Chem.*, 1988, **92**, 5244.

5 M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris, P. J. Taylor, C. Laurence and M. Berthelot, *Tetrahedron Lett.*, 1989, **30**, 521.

6 M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1990, 521.

7 M. H. Abraham, P. P. Duce, P. L. Grellier, D. V. Prior, J. J. Morris and P. J. Taylor, *Tetrahedron Lett.*, 1988, **29**, 1587.

8 M. H. Abraham, P. L. Grellier, D. V. Prior, P. P. Duce, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1989, 699.

9 N. Bodor, Z. Gabanyi and C.-K. Wong, *J. Am. Chem. Soc.*, 1989, **111**, 3783.

10 N. Bodor and M.-J. Huang, *J. Pharm. Sci.*, 1992, **81**, 272.

11 Y. Sasaki, H. Kubodera, T. Matsuszaki and H. Umeyama, *J. Pharmacobio-Dyn.*, 1991, **14**, 207.

12 G. R. Famini and L. Y. Wilson, in *Quantitative Treatments of Solute/Solvent Interactions*, eds. P. Politzer and J. S. Murray, Elsevier, Amsterdam, 1994.

13 G. R. Famini, C. A. Penski and L. Y. Wilson, *J. Phys. Org. Chem.*, 1992, **5**, 395.

14 T. Brinck, J. S. Murray and P. Politzer, *J. Org. Chem.*, 1993, **58**, 7070.

15 H. Hagelin, J. S. Murray, T. Brinck, M. Berthelot and P. Politzer, *Can. J. Chem.*, 1995, **73**, 483.

16 J. S. Murray, P. Lane, T. Brinck, K. Paulsen, M. E. Grice and P. Politzer, *J. Phys. Chem.*, 1993, **97**, 9369.

17 P. Politzer, P. Lane, J. S. Murray and T. Brinck, *J. Phys. Chem.*, 1992, **96**, 7938.

18 J. S. Murray, S. Ranganathan and P. Politzer, *J. Org. Chem.*, 1991, **56**, 3734.

19 J. S. Murray and P. Politzer, *J. Chem. Res.* (*S*), 1992, 110.

20 T. Brinck, J. S. Murray and P. Politzer, *Mol. Phys.*, 1992, **76**, 609.

21 J. S. Murray, T. Brinck, P. Lane, K. Paulsen and P. Politzer, *J. Mol. Struct.* (*Theochem*), 1994, **307**, 55.

22 J. S. Murray and P. Politzer, in *Quantitative Treatments of Solute/Solvent Interactions*, eds. P. Politzer and J. S. Murray, Elsevier, Amsterdam, 1994.

23 U. Dinur and A. T. Hagler, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyd, VCH, New York, 1991.

24 S. Scheiner, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyd, VCH, New York, 1991.

25 P. Sjoberg, J. S. Murray, T. Brinck and P. Politzer, *Can. J. Chem.*, 1990, **68**, 1440.
26 T. A. Koopmans, *Physica*, 1933, **1**, 104.
27 T. Brinck, J. S. Murray and P. Politzer, *Int. J. Quantum Chem.*, 1993, **48**, 73.
28 T. Brinck, unpublished results.
29 M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez and J. A. Pople, GAUSSIAN 94 (Revision B.3), Gaussian Inc., Pittsburgh PA, 1995.
30 W. J. Hehre, L. Radom, P. v. R. Schleyer and J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley, New York, 1986.
31 T. Brinck, J. S. Murray and P. Politzer, *J. Org. Chem.*, 1991, **56**, 5012.
32 J. S. Murray, T. Brinck and P. Politzer, *Int. J. Quantum Chem., Quantum Biol. Symp.*, 1991, **18**, 91.
33 M. Haeberlein and T. Brinck, *J. Phys. Chem.*, 1996, **100**, 10116.
34 R. F. W. Bader and H. J. T. Preston, *Theoret. Chim. Acta*, 1970, **17**, 384.
35 R. F. W. Bader, M. T. Carroll, J. R. Cheeseman and C. Chang, *J. Am. Chem. Soc.*, 1987, **109**, 7968.
36 P. Camilleri, S. A. Watts and J. A. Boraston, *J. Chem. Soc., Perkin Trans. 2*, 1988, 1699.
37 D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, *J. Chem. Soc., Perkin Trans. 2*, 1992, 705.
38 B. Y. Simkin and I. I. Sheikhet, *Quantum Chemical and Statistical Theory of Solutions: A Computational Approach*, Ellis Horwood, London, 1995.
39 T. Brinck, J. S. Murray and P. Politzer, *J. Chem. Phys.*, 1993, **98**, 4305.
40 J. Israelachvili, *Intermolecular and Surface Forces*, Academic, London, 2nd edn., 1992.
41 T. Brinck, unpublished results.
42 C. Hansch, A. Leo and D. Hoekman, *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*, ACS, Washington, DC, 1995.
43 C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.

*Paper* 6/04687H
*Received* 4*th July* 1996
*Accepted* 23*rd August* 1996