

# QSPR Studies on Vapor Pressure, Aqueous Solubility, and the Prediction of Water–Air Partition Coefficients

Alan R. Katritzky,\* Yilin Wang, Sulev Sild, and Tarmo Tamm

Center of Heterocyclic Compounds, Department of Chemistry, University of Florida,  
Gainesville, Florida 32611-7200

Mati Karelson

Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu EE 2400, Estonia

Received March 13, 1998

The vapor pressures and the aqueous solubilities of 411 compounds with a large structural diversity were investigated using a quantitative structure–property relationship (QSPR) approach. A five-descriptor equation with the squared correlation coefficient ( $R^2$ ) of 0.949 for vapor pressure and a six-descriptor equation with  $R^2$  of 0.879 for aqueous solubility were obtained. All descriptors were derived solely from the chemical structure of the compounds. The QSPR correlation equations for vapor pressure and aqueous solubility allow the reliable prediction of water–air partition coefficients.

## INTRODUCTION

The assessment of the transport, distribution, and fate of chemicals in air, water, and soil is very important in environmental science. The distribution of a chemical among air, water, and soil is largely dependent on some key physical-chemical properties which include the vapor pressure (VP), water solubility ( $S_w$ ), water–air partition coefficient ( $L_w$ ), octanol–water partition coefficient ( $K_{ow}$ ), distribution coefficient for adsorption or desorption on soil ( $K_p$ ), and Henry's law constant ( $H$ ).<sup>1–3</sup> To study the behavior of chemicals in the environment, much effort is being directed toward developing models based on their physical and chemical properties.

Vapor pressure determines the volatility of a chemical, so it governs the exchange rate of a chemical across an air–water interface through Henry's Law Constant.<sup>1b,4</sup> Accurate vapor pressures of low-volatility chemicals are often not available due to analytical difficulties. In such cases, the vapor pressure may be predicted using either the Clapeyron–Clausius equation and known values of enthalpy of vaporization and the respective compressibility factor or a group-contribution method.<sup>5</sup> Alternatively, the quantitative structure–property relationship (QSPR) approach is highly promising to estimate vapor pressures from descriptors derived solely from the molecular structure to fit experimental data. This method requires only the knowledge of the chemical structure; therefore, it is the most general method and is particularly suitable for new chemical products. Burkhard et al.<sup>6</sup> developed QSPR models for *n*-alkanes, biphenyls, and all polychlorinated biphenyl (PCB) congeners with new variables obtained from a principal component analysis of molecular connectivity indices (MCI). These new variables have a similar meaning for all the compounds and reflect physical characteristics of the molecules. Eleven methods were used by the same group<sup>7</sup> to predict vapor pressures at 25 °C for 15 PCBs. Two of the methods employed linear

**Table 1.** The QSPR Models for VP from Different References<sup>b</sup>

structures	<i>n</i>	$R^2$	<i>s</i>	$N_d$	ref
polychlorinated biphenyls	15	0.953		1	7
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	53	0.960		3	9 <sup>a</sup>
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	479	0.960	0.534	7	28
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	411	0.949	0.331	5	this work

<sup>a</sup> This model contains an empirical descriptor–melting point. <sup>b</sup> *n* is the number of structures,  $R^2$  is the squared correlation coefficient, *s* is the standard error, and  $N_d$  is the number of descriptors in the model.

correlations with MCI and the molecular topological indices of Kier and Hall.<sup>8</sup> Other methods were based on the Clapeyron–Clausius equation, the group-contribution method, or Gibbs' free energy of vaporization. Banerjee et al.<sup>9</sup> found that the vapor pressures of 53 compounds could be correlated with the Kamlet's empirical solvatochromic parameters. The results of relevant work on the QSPR treatment of vapor pressures are summarized in Table 1.

In recent years a general QSPR prediction methodology, which was coded into the CODESSA program,<sup>10</sup> has been developed by our group and successfully applied to the prediction of a variety of physical properties of compounds.<sup>11</sup> The first goal of the present work is to develop a general QSPR model of vapor pressure for diverse compounds using the CODESSA program.

The aqueous solubilities of organic compounds  $S_w$  are very important in many research areas, such as pharmaceutical or environmental science. A confident prediction of the aqueous solubility of a compound could greatly assist drug design by avoiding the synthesis of unsuitable compounds. Many different methods to predict aqueous solubility  $S_w$  have

**Table 2.** The QSPR Models for  $S_w$  from Different References<sup>a</sup>

structures	<i>n</i>	$R^2$	<i>s</i>	$N_d$	ref
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	331	0.965	0.299	18	13
hydrocarbons, halogenated hydrocarbons, O containing compounds	258	0.937	0.374	9	15
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	123	0.980	0.277	9	16
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	170	0.918	0.588	4	14
hydrocarbons, halogenated hydrocarbons	241	0.959	0.386	3	17
hydrocarbons, halogenated hydrocarbons, O and N containing compounds	411	0.879	0.573	6	this work

<sup>a</sup> *n* is the number of structures,  $R^2$  is the squared correlation coefficient, *s* is the standard error, and  $N_d$  is the number of descriptors in the model.

been considered in the literature, as summarized by Yalkowsky and Banerjee.<sup>12</sup> These methods can be categorized into the following types: (1) group contribution methods derived from measured aqueous solubility; (2) correlations with experimentally determined physicochemical properties such as boiling point, molecular surface area, molar volume, chromatographic retention time, and others; and (3) correlations with descriptors calculated only from molecular structure. The third approach has been proven to be particularly successful for the prediction of molecular properties without the need of experiments to obtain empirical correlative parameters. The results of QSPR studies of aqueous solubilities by Bodor and Huang,<sup>13</sup> Suzuki et al.,<sup>14</sup> Nelson Jurs,<sup>15</sup> and Sutter and Jurs,<sup>16</sup> and by our group<sup>17</sup> are summarized in Table 2. The multiple linear regression model previously derived in our group<sup>17</sup> included just three geometrical, topological, and constitutional descriptors; however, the structures correlated were limited to hydrocarbons and halogenated hydrocarbons. To develop a general QSPR model for calculating the aqueous solubilities of diverse organic compounds, it is necessary to enlarge the data set of  $S_w$ .

Vapor pressure and water solubility are fundamental physical parameters, and many other properties can be derived from them using suitable physical or correlation equations.<sup>18</sup> Thus VP and  $S_w$  can be combined to calculate water–air partition coefficients  $L_w$ .<sup>19</sup> The relationship between aqueous solubility  $S_w$  and octanol/water partition coefficient  $K_{ow}$  has been widely studied,<sup>18b,18c,20</sup> but few investigations have attempted to establish the theoretical basis of relationships between  $L_w$  and  $S_w$ . The  $S_w$  and VP are both “saturation” properties, i.e., they are measurements of the maximum capacity which a phase has for the dissolved compound in solution. Although water/air partition coefficients ( $L_w$ ) are not constant over the whole concentration range of aqueous solution, here  $L_w$  means the water/air partition coefficient for a saturated solution. Parameter  $L_w$ , also named the Ostwald solubility coefficient, is defined as the ratio of the solubility of a compound in the aqueous solution to its equilibrium concentration in the gas phase (eq 1). The necessity of the measurement of the solute concen-

tration in both phases raises the difficulty of accurate analytical determination that often also requires different techniques. Therefore it is of great practical importance to develop theoretical approaches for the accurate prediction of  $L_w$ , either directly from the structure or by correlation with VP and/or  $S_w$ , which are more easily measured.

$$L_w = \frac{\text{solubility of solute in aqueous solution}}{\text{equilibrium concn of solute in the gas phase}} \quad (1)$$

$$H = \frac{\text{fugacity of solute in the gas phase}}{\text{activity of solute in aqueous solution}} \quad (2)$$

Another commonly used equilibrium parameter is Henry's law constant  $H$ , which is essentially an air–water partition coefficient (eq 2). Many studies have estimated Henry's law constants using the VP/ $S_w$  method,<sup>3,19b</sup> which is based on the following two key assumptions: (1) the vapor pressure of the water-saturated compound is approximately equal to that of the pure compound and (2) the activity coefficient of the solute does not vary appreciably with concentration and can be neglected. Given these two assumptions,  $H$  is equal to  $L_w - 1$ . By definition,  $L_w = S_w/S_g$ . The equilibrium concentration of solute in the gas-phase  $S_g$  is proportional to the solute-saturated vapor pressure VP. According to the ideal gas law,  $S_g = VP/RT$ . Therefore,  $S_g$  at 25 °C is given by  $S_g = VP/24.45$ , i.e.,  $L_w = 24.45S_w/VP$ , in which VP is the saturated vapor pressure of solute expressed in atm, both  $L_w$  and  $S_w$  are expressed in mol L<sup>-1</sup>.<sup>19a</sup> Meylan and Howard<sup>21</sup> concluded that the VP/ $S_w$  estimation routine was “an excellent estimation method when reliable data are available”. They noted that although the VP/ $S_w$  method is often recommended for compounds with aqueous solubility less than 1 mol L<sup>-1</sup>, compounds with aqueous solubilities as high as 10–30% (by weight) were also found to be adequately estimated by the VP/ $S_w$  approach.

The primary objective of this work is to obtain QSPR models for vapor pressures, aqueous solubilities and the relationship between water–air partition coefficients  $L_w$  and aqueous solubilities,  $S_w$ . A total of 411 compounds with diverse structures were chosen as the data set. It is expected that the QSPR equations derived should provide useful information about the physical mechanisms determining VP,  $S_w$ , and  $L_w$ . The direct relationship connections among VP,  $S_w$ , and  $L_w$  encouraged us to establish a quantitative structure–property relationship (QSPR) model for the relationship between  $S_w$  and  $L_w$  by QSPR treatment for VP. The model can then be used to predict  $L_w$  by VP and  $S_w$  so as to supplement experimental data.

## DATA AND METHODOLOGY

**Data Set.** The data sets of vapor pressure and aqueous solubility at 25 °C were chosen from the PHYSPROP database assembled by Syracuse Research Corporation which was used in ref 21. This database contains 1151 compounds and includes chemical names and structures, molecular weights, aqueous solubilities, vapor pressures, Henry's law constants, and other properties. A total of 411 structurally diverse compounds were chosen from the database. Those included saturated, unsaturated and halogenated hydrocarbons, and compounds containing alcoholic and phenolic hydroxyl, amino, cyano, nitro, thio, ester, ether, carboxyl,

sulfide, and sulfoxide groups, ketones, aldehydes, furans, pyrans, biphenyls, naphthalenes, pyridines, pyrazines, and multifunctional compounds. The data set contains gases, liquids, and solids at normal condition. Table 3 shows the compounds and their corresponding experimental vapor pressures and aqueous solubilities as  $\log(\text{VP})$  and  $\log(S_w)$ , where VP is saturated vapor pressure in atmosphere (atm) and  $S_w$  is the solubility in moles per liter (mol/L). Among the 411 data points, our checks indicated that the solubility data for a total of 11 compounds from the database differ significantly from the values from other references. In these cases, average values from the available references were used instead of the values given in the PHYSPROP database.

The experimental error magnitude in VP and  $S_w$  is difficult to determine, because measurements are often reported in the literature without an error estimation. Even with published error estimations, the variation between different references can be large.<sup>1a</sup> In the present work, the experimental standard deviations  $s$  for VP and  $S_w$  were estimated using eq 3, similar to the method of ref 17. In eq 3,  $X_i$  are the data from different references and  $X_{av}$  is their mean value. The average standard deviations  $s$  for vapor pressures and aqueous solubilities were obtained as 0.32 and 0.58 log units, respectively.

$$s = \sqrt{\frac{1}{N} \sum (X_i - X_{av})^2} \quad (3)$$

**Computational Methods.** The quantitative structure–property relationships were developed using the Microsoft Windows version of the CODESSA program,<sup>10</sup> which is a chemical multipurpose quantitative–activity and structure–property (QSAR/QSPR) statistical analysis and prediction program. The program can perform the calculation of around 1000 descriptors, and the search for the best statistical relationship between calculated descriptors and experimental property data can be carried out using several techniques, which have been described elsewhere.<sup>11a,22</sup> The structures were drawn and preoptimized by the MMX molecular mechanics method using PCMODEL program.<sup>23</sup> The final geometry optimization of compounds was performed using the semiempirical quantum-chemical AM1 parametrization<sup>24</sup> with the MOPAC 6.0 program.<sup>25</sup> The MOPAC output files of individual compounds were loaded into the CODESSA program along with the experimental vapor pressures and aqueous solubilities to calculate five types of molecular descriptors: constitutional, topological, geometrical, electrostatic, and quantum-chemical.<sup>11a</sup> The correlation analysis to find the best QSPR model was carried out using a heuristic method or a forward selection method<sup>26,27</sup> based on the linear regression technique.

## RESULTS AND DISCUSSION

**QSPR Model of Vapor Pressure.** The CODESSA program provides a large variety of nonempirical molecular descriptors which encode the essential structural features of the molecules<sup>11</sup> and that are derived solely from the molecular structure, without requiring any experimental information. In the present treatment, the preliminary regression analysis was carried out using all the original CODESSA descriptors. Altogether more than 800 molecular descriptors

**Table 4.** The Best Five-Descriptor Correlation Model of Vapor Pressure (VP) for the Data Set of 411 Diverse Structures

<i>n</i>	descriptor	$X + \Delta X$	t-test	$R^2$	$R_{cv}^2$	<i>F</i>
0	intercept	$2.30 \pm 0.06$	41.48			
1	$G_1$	$-0.00618 \pm 0.00008$	−76.56	0.699	0.696	950.0
2	HDCA(2)	$-4.02 \pm 0.10$	−38.73	0.879	0.877	1483.4
3	SA-2(F)	$0.129 \pm 0.006$	20.15	0.929	0.927	1766.8
4	MNAC(Cl)	$6.02 \pm 0.574$	10.48	0.939	0.937	1552.2
5	SA(N)	$-0.0143 \pm 0.0017$	−8.44	0.949	0.947	1511.4

$$^a R^2 = 0.949, F = 1511.4, s = 0.331, R_{cv}^2 = 0.947.$$

were calculated for all 411 compounds. The heuristic method of selection of the best regression model<sup>27</sup> was then applied to the complete set of vapor pressures at normal conditions.

The best linear five-descriptor correlation model of vapor pressure developed with 411 compounds is shown in detail in Table 4. The most important descriptor in this regression is the gravitational index ( $G_1$ ) over all bonded atoms  $i, j$  in the molecule, defined by eq 4,<sup>26</sup> where  $m_i$  and  $m_j$  are the atomic masses of the bonded atoms and  $r_{ij}$  denote the respective bond lengths. The gravitational index ( $G_1$ ) reflects the effective mass distribution in the molecule and effectively describes the molecular dispersion forces in the bulk liquid media.<sup>27</sup>

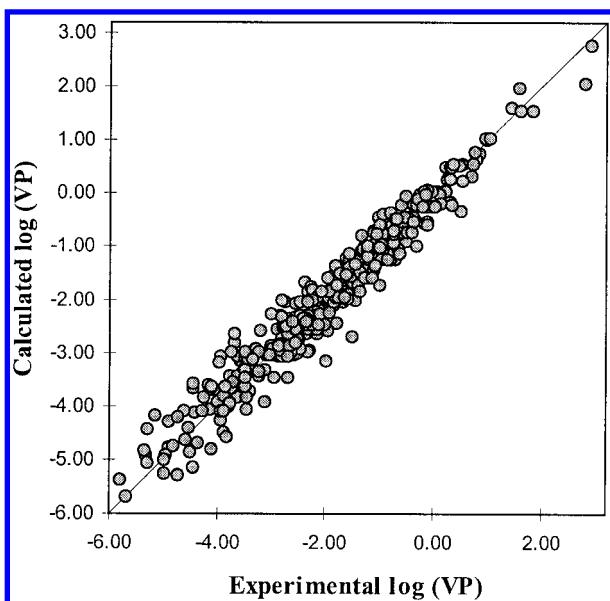
$$G_1 = \sum_{ij}^{\text{all bonded atoms}} (m_i m_j / r_{ij}^2) \quad (4)$$

$$\text{HDCA (2)} = \sum \frac{q_D \sqrt{S_D}}{\sqrt{S_{\text{tot}}}} \quad (5)$$

The second most important descriptor is the hydrogen-bonding donor charged surface area, HDCA (2), defined as in eq 5, where  $q_D$  is the partial charge on the hydrogen bonding donor (H) atom (s),  $S_D$  denotes the exposed surface area of this atom, and  $S_{\text{tot}}$  is the total molecular surface area calculated from the van der Waals' radii of the atoms (overlapping spheres). The summation in the foundation of HDCA (2) (eq 5) is performed over all simultaneously possible hydrogen bonding donor and acceptor pairs per solute molecule. Also, hydrogen atoms attached to carbons connected directly to carbonyl or cyano group were included as possible hydrogen bonding donor centers (their effectiveness is, of course, much smaller because of the smaller partial charge on them).

The combination of the two descriptors  $G_1$  and HDCA (2) adequately represents the forces of intermolecular attraction;  $G_1$  is connected with the dispersion and cavity-formation effects in liquid, and HDCA (2) is connected with the hydrogen-bonding ability of compounds. This relationship indicates that vapor pressure has the similar structure-governing factors as previously found for the boiling point.<sup>27</sup> In the model of boiling point, the most important descriptor is the cubic root of gravitational index ( $G_1^{1/3}$ ); however, when forced into the correlation result of vapor pressure,  $G_1^{1/3}$  gave a poorer correlation. This observation suggests the size of molecule is a key factor for vapor pressure, so the lower dimensional descriptor  $G_1^{1/3}$  cannot give a better description for vapor pressure.





**Figure 1.** Scatter plot of the calculated versus experimental log-(vapor pressure) values for 411 compounds.

The three additional descriptors in the five-parameter model (Table 4) are as follows: the sum of the surface area of fluorine atoms SA-2 (F), the maximum net atomic charge for a chlorine atom MNAC (Cl), and the surface area of nitrogen atoms SA (N). The inclusion of these three further descriptors in the model shows that the  $G_1$  and HDCA (2) could not describe adequately the intermolecular interactions with solute molecules containing fluorine, chlorine, or nitrogen atoms, possibly because of the deficiency in AM1 calculation of partial charges in those atoms or because of the requirement of corrected "effective" atomic masses in the formulation of the gravitation index.

For our five-descriptor model (Table 4), the cross-validated correlation coefficient  $R_{cv}^2 = 0.947$ , as compared to the correlation coefficient  $R^2 = 0.949$ , indicates the high stability of the regression equation. The scatter plot of the logarithm of vapor pressure calculated using the model of Table 4 versus experimental log(VP) is presented in Figure 1.

Most reported quantitative structure–property relationship models of vapor pressure contain at least one or two empirical data, such as boiling points, critical pressures, or critical temperatures. This limits their usefulness in various fields. Recently, Liang and Gallagher<sup>28</sup> developed a seven-descriptor model (eq 6) employing polarizability and six functional group counts. The data set used by them has almost the same structural diversity of compounds as the set employed in this work. However, there are no gaseous compounds in their data set, and their standard error of correlation (0.534) is larger than the standard error of our treatment (0.331, Table 4) which approaches the experimental error 0.32.

$$\begin{aligned} \log(\text{VP}) = & -0.432\alpha - 1.382(\text{OH}) - 0.482(\text{C=O}) - \\ & 0.416(\text{NH}) - 2.197(\text{COOH}) - 1.383(\text{NO}_2) - \\ & 1.101(\text{C}\equiv\text{N}) + 4.610 \quad (6) \end{aligned}$$

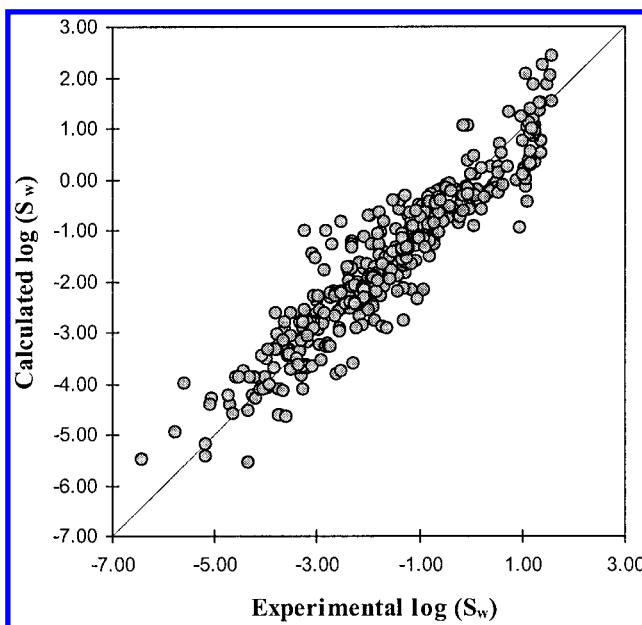
$$R^2 = 0.960, \quad s = 0.534, \quad n = 479, \quad R_{cv}^2 = 0.957$$

**QSPR Model of Aqueous Solubility.** The best six-descriptor model for aqueous solubility with 411 compounds

**Table 5.** The Best Six-Descriptor Correlation Model of Aqueous Solubility ( $S_w$ ) for the Data Set of 411 Diverse Structures

<i>n</i>	descriptor	$X + \Delta X$	t-test	$R^2$	$R_{cv}^2$	<i>F</i>
0	intercept	$-0.882 \pm 0.138$	-6.39			
1	$Q_{\min}$	$-16.1 \pm 0.7$	-24.33	0.348	0.343	218.4
2	$N_{\text{el}}$	$-0.113 \pm 0.005$	-25.12	0.664	0.660	403.4
3	FHDSA(2)	$2.55 \pm 0.22$	11.55	0.776	0.771	469.1
4	ABO(N)	$0.781 \pm 0.064$	12.26	0.842	0.838	541.7
5	$^0\text{SIC}$	$0.328 \pm 0.037$	8.91	0.867	0.862	526.9
6	RNCS	$-0.0143 \pm 0.0022$	-6.32	0.879	0.874	487.9

$$^a R^2 = 0.879, F = 487.9, s = 0.573, R_{cv}^2 = 0.874.$$



**Figure 2.** Scatter plot of the calculated versus experimental log-(aqueous solubility) values for 411 compounds.

was selected by the CODESSA forward selection method using all original descriptors. This model is expressed in Table 5, where  $Q_{\min}$  and RNCS are both electrostatic descriptors,  $N_{\text{el}}$  and ABO (N) are two quantum-chemical descriptors, and  $^0\text{SIC}$  is a topological descriptor. The two most important descriptors are the most negative partial charges in the molecule ( $Q_{\min}$ ) and the number of electrons ( $N_{\text{el}}$ ).<sup>10</sup> The number of electrons is roughly related to the polarizability of the molecule as well as to the cavity-size effects (dispersion and cavity formation). Solute–solvent interaction is a major determining factor for the aqueous solubilities of compounds, and the most negative partial charges  $Q_{\min}$  can be related to one specific type of solute–solvent interaction.

The fractional hydrogen donor surface area FHDSA (2) is defined as HDCA (2)/SQRT (TMSA), where TMSA is the total molecular surface area and HDCA (2)<sup>27</sup> has a similar definition to HDCA (2) introduced in the vapor pressure model.

$$\text{HDCA (2)} = \sum \frac{q_D \sqrt{S_D}}{S_{\text{tot}}} \quad (7)$$

Descriptor RNCS (relative negative charged surface area) is defined as eq 8, where  $q_{\min}$  is the charge on the most negative atom,  $S_{\min}$  is the surface area of the most negative atom, and  $\sum q_i$  is the total negative charges. RNCS is

sensitive to both the size and the charge of the molecule and is responsible for polar interactions between molecules. Both RNCS and FHDSA (2) as well as  $Q_{\min}$  can be also related to hydrogen bonding, and they adjust the model of Table 5 to account for intermolecular association. Hydrogen bonding and other intermolecular interactions obviously have a significant effect on aqueous solubility.

$$\text{RNCS} = \frac{q_{\min} S_{\min}}{\sum q_i} \quad (8)$$

Descriptor  $^0\text{SIC}$  is the structural information content of zeroth order<sup>8,10</sup> and is defined by eq 9, where  $n_i$  is the number of atoms in the  $i$ th class,  $n$  is the total number of atoms in the molecule, and  $^0\text{IC}$  is the information content of zeroth order itself defined by eq 10. Descriptor  $^0\text{SIC}$  describes the atomic connectivity in the molecule and encodes the size and the degree of branching in the compound. The size and the shape of the molecule also directly affect the intermolecular interaction.

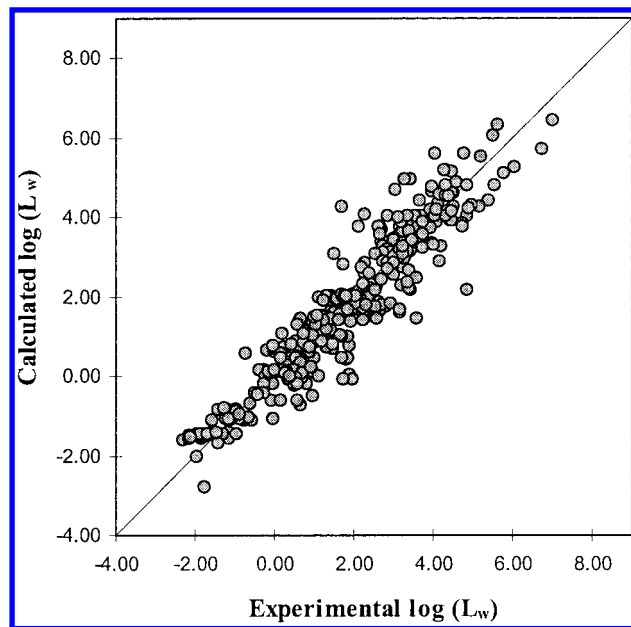
$$^0\text{SIC} = ^0\text{IC}/\log_2 n \quad (9)$$

$$^0\text{IC} = -\sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n} \quad (10)$$

Descriptor ABO (N), the average bond order of nitrogen atom, is apparently a term to correct the deficiency of the electrostatic or hydrogen-bonding parameter for the N-atom containing compounds.

The plot of the calculated versus experimental solubility of the complete set using the six-descriptor model is shown in Figure 2. The standard error is 0.573 log units which is within the estimated experimental error of 0.58 (see above). The cross-validated correlation coefficient  $R_{\text{cv}}^2$  is 0.874, which approaches the correlation coefficient ( $R^2 = 0.879$ ), indicating the stability of the model of Table 5. A comparison of the present result with other models is given in Table 2. Some of the previous models<sup>15–17</sup> are derived from very limited data sets. The model from ref 14 corresponds to a comparatively diverse data set but not containing any carboxylic acids, pyridines, and pyrazines. The data set of ref 13 is very diverse in structures, but the model includes 18 structural descriptors. Thus, compared to the previous models, the correlation model developed in the present work treats a greater variety of structures than those of refs 14–17 and is much more economical in its descriptors than the model of ref 13.

**Prediction of Water–Air Partition Coefficients.** The values for vapor pressure VP and aqueous solubility  $S_w$  predicted by the above models are listed in Table 3. The predicted values of VP and  $S_w$  and the relationship  $L_w = 24.45S_w/\text{VP}$  were used to calculate the water–air partition coefficients  $L_w$ . The calculated  $\log(L_w)$  values are presented in Table 3 along with the experimental data on  $\log(L_w)$ . The latter was obtained by the transformation of Henry's law constants given in the database according to the approximate relationship  $L_w = H^{-1}$ . The plot of the predicted  $\log(L_w)$  versus experimental  $\log(L_w)$  is shown in Figure 3. The mean standard error of this prediction is 0.63 log units, which is close to the standard error of  $L_w$  (0.52) predicted using the



**Figure 3.** Scatter plot of the calculated vs experimental  $\log$ -(partition coefficient) values for 411 compounds.

equation derived directly from the experimental values of  $L_w$ .<sup>26</sup> We conclude that hence this procedure is a valid approach to calculate  $L_w$  by using QSPR predicted values of VP and  $S_w$ .

A direct QSPR equation for  $L_w$ , which was obtained in our previous study,<sup>26</sup> is shown in eq 11, where  $N(\text{O})$  and  $N(\text{N})$  are the number of nitrogen and oxygen atoms in the molecule,  $E_{\text{HOMO}} - E_{\text{LUMO}}$  is the energy gap between HOMO and LUMO,  $\text{PCWT}^E$  is the most negative partial charge weighted topological electronic index, and  $N_{\text{rings}}$  is the number of rings. The most important descriptor in that model was the hydrogen-bonding related descriptor HDCA (2). It is apparent from the present results that  $S_w$  and VP have similar leading structural determining factors. Hydrogen-bonding has a large effect on these physical properties.

$$\begin{aligned} \log(L_w) = & (2.66 \pm 0.22) + (42.37 \pm 1.11)\text{HDCA}(2) + \\ & (0.65 \pm 0.02)[N(\text{O}) + 2*N(\text{N})] + \\ & (-0.16 \pm 0.02)(E_{\text{HOMO}} - E_{\text{LUMO}}) + \\ & (0.12 \pm 0.01)\text{PCWT}^E + (0.82 \pm 0.06)N_{\text{rings}} \quad (11) \end{aligned}$$

## CONCLUSION

The vapor pressures and the aqueous solubilities of 411 compounds with a wide diversity of structure have been satisfactorily described using QSPR equations with only five and six theoretical molecular descriptors, respectively. All the descriptors involved are calculated solely from the chemical structures of compounds and have definite physical meaning corresponding to different intermolecular interactions. As compared to the previous reported models, the models developed by us will be applicable to a wider variety of structures with less molecular descriptors. The predicted values of vapor pressure and aqueous solubility by the QSPR correlation equations also allow an accurate prediction of the water–air partition coefficients for structurally diverse compounds. The QSPR equation obtained provide also significant insight into the relationships of the structure with vapor pressure, aqueous solubility, and water–air partition coefficient.

**Supporting Information Available:** Table 3 gives the experimental and predicted vapor pressure, aqueous solubility, and water–air partition coefficient for the set of 411 compounds (8 pages). See any current masthead page for ordering and Internet access information.

## REFERENCES AND NOTES

- (1) (a) Mackay, D.; Shiu, W. Y.; Ma, K. C. In *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; Lewis Publishers: Boca Raton, FL, 1992; Vol. 1–4. (b) Lyman, W. J. In *Environmental Exposure from Chemicals*; Neely, W. B., Blau, G. E., Eds.; Chemical Rubber: Boca Raton, FL, 1985; Vol. 1.
- (2) (a) Mackay, D. Finding Fugacity Feasible. *Environ. Sci. Technol.* **1979**, *13*, 1218–1223. (b) Mackay, D.; Paterson, S. Calculating Fugacity. *Environ. Sci. Technol.* **1981**, *15*, 1006–1014. (c) Mackay, D.; Paterson, S. Fugacity Revisited. *Environ. Sci. Technol.* **1982**, *16*, 654A–660A.
- (3) Mackay, D.; Shiu, W. Y. A Critical Review of Henry's Law Constants for Chemicals of Environmental Interest. *J. Phys. Chem. Ref. Data* **1981**, *10*, 1175–1199.
- (4) (a) Thomas, R. G. In *Handbook of Chemical Property Estimation Methods*; Lyman, W. J., Reehl, W. F., Rosenblatt, D. H., Eds.; McGraw-Hill: New York, 1982; Chapter 15. (b) Site, A. D. The Vapor Pressure of Environmentally Significant Organic Chemicals: A Review of Methods and Data at Ambient Temperature. *J. Phys. Chem. Ref. Data* **1997**, *26*, 157–193.
- (5) (a) Macknick, A. B.; Prausnitz, J. M. Vapor Pressures of Heavy Liquid Hydrocarbons by a Group-Contribution Methodol. *Ind. Eng. Chem. Fundam.* **1979**, *18*, 348–353. (b) Edwards, D. R.; Prausnitz, J. M. Estimation of Vapor Pressures of Heavy Liquid Hydrocarbons Containing Nitrogen or Sulfur by a Group-Contribution Methodol. *Ind. Eng. Chem. Fundam.* **1981**, *20*, 280–283.
- (6) Burkhard, L. P.; Andren, A. W.; Armstrong, D. E. Structure Activity Relationships Using Molecular Connectivity Indices with Principle Component Analysis. *Chemosphere* **1983**, *12*, 935–943.
- (7) Burkhard, L. P.; Andren, A. W.; Armstrong, D. E. Estimation of Vapor Pressures for Polychlorinated Biphenyls: A Comparison of Eleven Predictive Methods. *Environ. Sci. Technol.* **1985**, *19*, 500–507.
- (8) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Structure–Activity Analysis*; Wiley: New York, 1986.
- (9) Banerjee, S.; Howard, P. H.; Lande, S. S. General Structure–Vapor Pressure Relationships for Organics. *Chemosphere* **1990**, *21*, 1173–1180.
- (10) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA Version 2.0 Reference Manual*; 1994.
- (11) (a) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279–287. (b) Katritzky, A. R.; Karelson, M.; Lobanov, V. S. QSPR as a Means of Predicting and Understanding Chemical and Physical Properties in Terms of Structure. *Pure Appl. Chem.* **1997**, *69*, 245–248.
- (12) Yalkowsky, S. H.; Banerjee, S. In *Aqueous Solubility. Methods of Estimation for Organic Compounds*; Marcel Dekker: New York, 1992.
- (13) Bodor, N.; Huang, M. A New Method for the Estimation of the Aqueous Solubility of Organic Compounds. *J. Pharm. Sci.* **1992**, *81*, 954–960.
- (14) Suzuki, T.; Ohtaguchi, K.; Koide, K. Computer-Aided Prediction of Solubilities of Organic Compounds in Water. *J. Chem. Eng. Jpn.* **1992**, *25*, 729–734.
- (15) Nelson, T. M.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 601–609.
- (16) Sutter, J. M.; Jurs, P. C. Prediction of Aqueous Solubility for a Diverse Set of Heteroatom-Containing Organic Compounds Using a Quantitative Structure–Property Relationship. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 100–107.
- (17) Huibers, P. D. T.; Katritzky, A. R. Correlation of Aqueous Solubility of Hydrocarbons and Halogenated Hydrocarbons with Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 283–292.
- (18) (a) Hansch, C.; Quinn, J. E.; Lawrence, G. L. The Linear Free-Energy Relationship between Partition Coefficients and the Aqueous Solubility of Organic Liquids. *J. Org. Chem.* **1968**, *33*, 347–350. (b) Banerjee, S.; Yalkowsky, S. H.; Valvani, S. C. Water Solubility and Octanol/Water Partition Coefficients of Organics. Limitations of the Solubility–Partition Coefficient Correlation. *Environ. Sci. Technol.* **1980**, *14*, 1227–1229. (c) Mackay, D.; Bobra, A.; Shiu, W. Y.; Yalkowsky, S. H. Relationships between Aqueous Solubility and Octanol–Water Partition Coefficients. *Chemosphere* **1980**, *9*, 701–712.
- (19) (a) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors that Influence the Solubility of Gases and Vapours in Water at 298 K, and a New Method for its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791. (b) Staudinger, J.; Roberts, P. V. A Critical Review of Henry's Law Constants for Environmental Applications. *Critical Rev. Environ. Sci. Technol.* **1996**, *26*, 205–297.
- (20) (a) Miller, M. M.; Wasik, S. P.; Huang, G. L.; Shiu, W. Y.; Mackay, D. Relationships between Octanol–Water Partition Coefficient and Aqueous Solubility. *Environ. Sci. Technol.* **1985**, *19*, 522–529. (b) Yalkowsky, S. H.; Valvani, S. C. Solubilities and Partitioning 2. Relationships between Aqueous Solubilities, Partition Coefficients, and Molecular Surface Areas of Rigid Aromatic Hydrocarbons. *J. Chem. Eng. Data* **1979**, *24*, 127–129. (c) Isnard, P.; Lambert, S. Aqueous Solubility and *n*-Octanol/Water Partition Coefficient Correlations. *Chemosphere* **1989**, *18*, 1837–1853. (d) Krop, H. B.; van Velzen, M. J. M.; Parsons, J. R.; Govers, H. A. J. *n*-Octanol–Water Partition Coefficients, Aqueous Solubilities, and Henry's Law Constants of Fatty Acid Esters. *Chemosphere* **1997**, *34*, 107–119.
- (21) Meylan, W. M.; Howard, P. H. Bond Contribution Method for Estimating Henry's Law Constants. *Environ. Toxicol. Chem.* **1991**, *10*, 1283–1293.
- (22) Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure–Property Relationship Approach. 1. Nonionic Surfactants. *Langmuir* **1996**, *12*, 1462–1470.
- (23) Serena Software, *PCMODEL User Manual* 1992.
- (24) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (25) Stewart, J. J. P. *MOPAC 6.0*; QCPE No 455, **1989**.
- (26) Katritzky, A. R.; Mu, L.; Karelson, M. A QSPR Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162–1168.
- (27) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Point with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, *100*, 10400–10407.
- (28) Liang, C. K.; Gallagher, D. A. QSPR Prediction of Vapor Pressure from Solely Theoretically-Derived Descriptors. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 321–324.

CI980022T