Bull Environ Contam Toxicol (2008) 80:319–323
DOI 10.1007/s00128-008-9369-x

# Determination of Solubilities and *n*-Octanol/Water Partition Coefficients and QSPR Study for Substituted Phenols

**Y. J. Xie · H. Liu · H. X. Liu · Z. C. Zhai ·
Z. Y. Wang**

Received: 26 June 2007 / Accepted: 28 January 2008 / Published online: 14 March 2008

**Abstract** A shake-flask method was employed to determine the water solubility ($-\lg S_w$) and *n*-octanol/water partition coefficient ($\lg K_{ow}$) of 20 substituted phenols at 298.15 K. And optimized calculation was carried out at B3LYP/6-311G** level with DFT method. Afterwards the obtained parameters were taken as theoretical descriptors to establish the QSPR models for predicting $-\lg S_w$ and $\lg K_{ow}$, in which the conventional correlation coefficients ($R^2$) are 0.9800 and 0.9941, respectively. The two models were further validated by variance inflation factors (VIF) and *t*-test. Upon comparison, the stability and predictive power are more advantageous than those based on AM1 molecular orbital method and molecular connectivity method.

**Keywords** Water solubility ($-\lg S_w$) ·
*n*-Octanol/water partition coefficient ($\lg K_{ow}$) ·
Quantitative structure–property relationships (QSPR) ·
Substituted phenol

Substituted phenols are important materials in organic chemical industry. Meanwhile the environmental hazards of phenolic compounds have led to wide concern by researchers, and many works have been done for their QSPR/QSAR models in recent years (Yao et al. 2005; Lu et al. 2007). For example, Urbanczyk et al. (2002) predicted the gas chromatogrphic retention time, retention index and the correction coefficient for substituted phenols by Abraham model. On the other hand, Loader et al. (2006) established QSAR model of biotoxin for alkyl phenols based on theoretic bond length and electronic density, and Gao et al. (2002) reported QSPR model for $\lg K_{ow}$ of substituted phenols. Liu et al. (2005) also established QSPR model for predicting $K_{oc}$ of substituted phenols according to their $-\lg S_w$, molecular connectivity index and quantum parameters. This study aimed to determine the experimental values of $-\lg S_w$ and $\lg K_{ow}$ of 20 substituted phenols by the shake-flask method and to assess, model, and predict $-\lg S_w$ and $\lg K_{ow}$ of substituted phenols using the structural and thermodynamic parameters as theoretical descriptors.

## Materials and Methods

In the present study, all the samples of 20 substituted phenols and *n*-octanol are analytical pure; the conductance of deionized water is less than 2 μS/cm. To design standard curve, four unsaturated solutions with different concentrations were prepared and introduced into each flask. And then the flasks were put into a THZ-C constant temperature shaker (Jiangsu Taicang Experimental Equipment Company, China) at 298.15 ± 0.5 K to ensure complete dissolution of the specific phenol. After 24 h, the absorbencies of the solutions were determined with 752s UV-spectrophtometer (Lengguang Tech, Shanghai, China) at the maximun absorption wavelength. With plot of absorbency versus concentration, the standard curve was thus obtained. Similarly, for determination of $-\lg S_w$, excessive phenolic compound and ultrapure water were introduced

---

Y. J. Xie · H. X. Liu · Z. C. Zhai · Z. Y. Wang (✉)
School of Biological and Chemical Engineering,
Jiaxing University, Zhejiang, Jiaxing 314001,
People's Republic of China
e-mail: wangzun315cn@163.com

H. Liu
Department of Material and Chemical Engineering,
Guilin Institute of Technology, Guangxi, Guilin 541004,
People's Republic of China

into a volumetric flask of 50 mL, and the solution was then vibrated in the shaker at 298.15 ± 0.5 K. After 24 h, sample of the solution was centrifuged at 1,500 r/min and its absorbency was determined at UV-spectrophotometer against water blank. With two parallel measurements for each sample, the average value of solubility was calculated by the standard curve. While to determine $\lg K_{ow}$, 2.00 mL solution of known concentration of substituted phenol, of which the solvent is *n*-octanol saturated by water, was mixed with 8.00 mL water saturated by *n*-octanol in a volumetric flask of 50 mL. The flask with the resultant solution was vibrated in the shaker at 298.15 ± 0.5 K. After 24 h, the solution was centrifuged at 1,500 r/min and then the organic phase was removed. The absorbency of water phase was determined and the concentration was obtained in the same way as $-\lg S_w$. $\lg K_{ow}$ can be calculated with following equation:

$$K_{ow} = \frac{C_0 V_0 - C_1 V_1}{C_1 V_0} \tag{1}$$

where $C_0$ is the initial concentration, $V_0$ is the volume of octanol phase, $C_1$ is the concentration of substituted phenol in water, and $V_1$ is the volume of water phase.

Furthermore, 29 substituted phenols were calculated at B3LYP/6-311G** level in Gaussian 98 program (Frisch et al. 1998), resulting in useful structural and thermodynamic parameters of these compounds. Structural parameters used in this study include: the most negative atomic net charges of the molecular ($q^-$), the most positive atomic net charges of the molecular on the hydrogen ($qH^+$), molecular volume ($V_i$) and molecular average polarizability ($\alpha$) calculated from vibration analysis. Thermodynamic parameters include: total energy (TE).

Based on the theoretical linear solvation energy relationship (TLSER) method (Wilson et al. 1991; Famini et al. 1997), with GQSARF2.0 program (Liu et al. 2003) and SPSS for Windows program, the obtained parameters were taken as theoretical descriptors to establish the novel QSPR models for predicting $-\lg S_w$ and $\lg K_{ow}$ of substituted phenols. To determine the optimum number of components for the correlation models, the leave-one-out (LOO) cross-validation procedure was subjected to validate the derived QSPR models by using GQSARF2.0 program (Liu et al. 2003). The quality of the models was evaluated in terms of the LOO cross-validation correlation coefficient $q^2$, the squared conventional correlation coefficient $R^2$, the standard deviation SD and *t*-score.

## Results and Discussion

The experimental data of $-\lg S_w$ and $\lg K_{ow}$ for 20 substituted phenols are listed in Table 1. For comparison, some data from literature are also presented in the table (Hansch and Leo 1979). Because the experimental values for compounds No. 1–20 were determined at 298.15 K, some observed values at 293.15 K from other literatures were not shown in Table 1 due to poor comparability at different conditions.

Structural parameters and thermodynamic parameters of 20 substituted phenols (No. 1–20) referred to in this study are listed in Table 2. To obtain correlation equation of $-\lg S_w$, these parameters were taken as theoretical descriptors. By employing GQSARF2.0 program, forward stepwise regression analysis was carried out on observed values of $-\lg S_w$, and the results are presented in Table 3.

With observation of the three equations in Table 3, it indicates that there is no obvious increase in $R^2$ and $q^2$ for Eq. No. 4 with three variables compared with Eq. No. 3 with two variables. Furthermore Eq. No. 4 does not pass the *t*-test. Equation No. 3 is thus recommended and taken as model (I) in this study. By analysis of Table 3, the following proposals can be presented: (a) when one variable is used, molecular average polarizability ($\alpha$) appears first in the equation. And $-\lg S_w$ increases with increasing $\alpha$. This is because polarizability usually represents the charge deformability of a molecule, and molecules with large $\alpha$ tend to enter organic phase, resulting in small water solubility. (b) When two variables are employed, both $\alpha$ and TE are included in the equation, with $R^2$ being increased from 0.9510 to 0.9800 and SD decreased from 0.220 to 0.141. It indicates the prediction accuracy of the equation is approved with TE variable, and $-\lg S_w$ increases with the increase of absolute TE (the value is negative). This is because TE is a parameter relevant to molecular volume. With increase of absolute value of TE, the molecular volume increases. The molecule tends to be more hydrophobic and the substituted phenol is thus difficult to enter water phase, result in low $S_w$. Using GQSARF2.0 program and the same regression method as mentioned above, the correlation equations of $\lg K_{ow}$ were obtained as presented in Table 4, in which $R^2$, SD and $q^2$ are also included.

Table 4 indicates that equation with three variables is better than that with one or two variables. If four variables are included, $R^2$ and $q^2$ of the model would not increase obviously, and its SD did not decrease clearly. Therefore Eq. 7 containing three variables is relatively good and called model (II) herein. From these equations it is found that $\lg K_{ow}$ increases with increase of $V_i$ and the absolute TE (the value is negative). This is because molecular volume and the absolute value of TE are in direct proportion to the enhancement of hydrophobicity of a molecule, and substituted phenol molecule with large $V_i$ can thus easily enter *n*-octanol phase, resulting in high $\lg K_{ow}$ value.

**Table 1** Experimental and predicted data of $-\lg S_{w}$ and $\lg K_{ow}$

| No. | Compound | $-\lg S_{w}$ | | | | $\lg K_{ow}$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Experimental | Predicted | Diff | Literature[a] | Experimental | Predicted | Diff | Literature |
| 1 | 1-Nitroso-2-naphthol | 3.095 | 3.05 | −0.045 | | 2.677 | 2.694 | 0.017 | |
| 2 | 2-Nitroso-1-naphthol | 3.08 | 3.078 | −0.002 | | 2.46 | 2.496 | 0.036 | |
| 3 | 3,4-Dimethylphenol | 1.401 | 1.231 | −0.170 | | 2.374 | 2.395 | 0.021 | 2.42[b] |
| 4 | *p*-Cresol | 0.661 | 0.64 | −0.021 | | 1.955 | 2.006 | 0.051 | |
| 5 | *p*-Nitrophenol | 0.94 | 0.966 | 0.026 | | 1.958 | 1.982 | 0.024 | 1.90[c] |
| 6 | *m*-Cresol | 0.647 | 0.623 | −0.024 | | 1.962 | 2.019 | 0.057 | |
| 7 | *m*-Nitrophenol | 1.032 | 0.864 | −0.168 | | 1.985 | 2.029 | 0.044 | 1.85[b] |
| 8 | *o*-Cresol | 0.603 | 0.617 | 0.014 | | 1.943 | 1.992 | 0.049 | 1.96[c] |
| 9 | 2-Naphthol | 2.25 | 2.265 | 0.015 | | 2.684 | 2.659 | −0.025 | |
| 10 | 1-Naphthol | 2.016 | 2.202 | 0.186 | | 2.836 | 2.694 | −0.142 | |
| 11 | 2-Chlorophenol | 0.781 | 0.821 | 0.040 | 0.65 | 2.158 | 2.021 | −0.137 | 2.15[c] |
| 12 | 4-Chlorphenol | 0.673 | 0.847 | 0.174 | 1.12 | 2.338 | 2.288 | −0.050 | 2.40[c] |
| 13 | 2,4-Dichlorophenol | 1.473 | 1.699 | 0.226 | 1.55 | 2.978 | 3.018 | 0.040 | 3.06[d] |
| 14 | 2,3-Dichlorophenol | 1.644 | 1.649 | 0.005 | | 2.975 | 2.937 | −0.038 | 2.84[d] |
| 15 | 2,5-Dichlorophenol | 1.947 | 1.712 | −0.235 | | 2.972 | 2.929 | −0.043 | 2.96[b] |
| 16 | 2,6-Dichlorophenol | 1.984 | 1.674 | −0.310 | | 2.697 | 2.687 | −0.010 | 2.63[b] |
| 17 | 2,4,5-Trichlorophenol | 2.35 | 2.545 | 0.195 | | 3.586 | 3.626 | 0.040 | 3.58[b] |
| 18 | 2,4,6-Trichlorophenol | 2.599 | 2.569 | −0.030 | 2.39 | 3.559 | 3.549 | −0.010 | 3.37[b] |
| 19 | Pentachlorophenol | 4.280 | 4.243 | −0.037 | 4.28 | 5.011 | 5.035 | 0.024 | 5.04[c] |
| 20 | 3-Chlorphenol | 0.679 | 0.840 | 0.161 | 1.18 | 2.148 | 2.200 | 0.052 | |
| 21 | 3,4-Dichlorophenol | | 1.664 | | | | 3.057 | | 3.17[b] |
| 22 | 3,5-Dichlorophenol | | 1.704 | | | | 2.956 | | |
| 23 | 2,3,4-Thichlorophenol | | 2.493 | | | | 3.655 | | |
| 24 | 2,3,5-Thichlorophenol | | 2.553 | | | | 3.515 | | 3.58[b] |
| 25 | 2,3,6-Thichlorophenol | | 2.534 | | | | 3.439 | | 3.77[d] |
| 26 | 3,4,5-Thichlorophenol | | 2.502 | | | | 3.910 | | 4.01[d] |
| 27 | 2,3,4,5-Tetrachlorophenol | | 4.749 | | | | 4.461 | | |
| 28 | 2,3,4,6-Tetrachlorophenol | | 3.399 | | | | 4.393 | | 4.45[d] |
| 29 | 2,3,5,6-Tetrachlorophenol | | 3.406 | | | | 4.268 | | |

[a] Taken from Wang et al. (1992)

[b] Taken from Yao et al. (2005)

[c] Taken from Hansch et al. (1979)

[d] Taken from Dai et al. (2000)

In the present study, the correlation of each independent variable in the new models was evaluated by calculating their variation inflation factors (VIF) (Wei et al. 2001). The definition of VIF is: VIF $= 1/(1 - r^{2})$. Here $r$ indicates the correlation coefficient of multiple regression between one variable and others in the equation. VIF = 1.0 suggests no self-correlation among each varibles. If VIF ranges from 1.0 to 5.0, the correlation equation is thus acceptable. If *VIF* is larger than 10.0, that means the regression equation is unstable and re-check of the variance correlation is needed. In our study, all the VIF values of variables are presented in Table 5. As shown in Table 5, the VIF values of variables are 1.1048, 1.1048 for model (I), 2.1095, 2.1548 and 1.6444 for model (II). It is obvious the VIFs herein are all smaller than 5.0, indicating both two models have obvious statistic significance and are thus acceptable.

In order to check the reliability of the models, *t*-test was carried out. The *t*-values of variables and standard regression coefficients (SR) of variables in model (I) and model (II) are also listed in Table 5, for which the $t_{\alpha/2}$ with reliable level of 95% is 2.093. It can be seen from Table 5 that the *t*-value is 19.471, −7.436 for model (I) and 19.961, −9.877, −30.530 for model (II), respectively. Their absolute values are all larger than $t_{\alpha/2}$, suggesting a high significance level for the variables of the models. The standard regression coefficient (SR) is 0.7974, −0.3880 for model

**Table 2** Partial structural and thermodynamic descriptors by B3LYP/6-311G** for substituted phenol

| No.[a] | $V_i$ / ($cm^3$/mol) | $qH^+$/(e) | $q^-$/(e) | $\alpha$/ ($10^{-30}$esu) | TE/(Hartree) |
|---|---|---|---|---|---|
| 1 | 122.33 | 0.2540 | −0.3206 | 127.28 | −590.56145 |
| 2 | 126.28 | 0.2439 | −0.2884 | 127.92 | −590.54311 |
| 3 | 103.36 | 0.2446 | −0.3676 | 89.90 | −386.20577 |
| 4 | 92.06 | 0.2448 | −0.3657 | 77.36 | −346.87856 |
| 5 | 93.99 | 0.2555 | −0.3411 | 82.15 | −512.11027 |
| 6 | 93.21 | 0.2450 | −0.3636 | 76.99 | −346.87952 |
| 7 | 92.84 | 0.2522 | −0.3494 | 79.89 | −512.10725 |
| 8 | 90.64 | 0.2457 | −0.3686 | 76.85 | −346.88027 |
| 9 | 112.15 | 0.2477 | −0.3607 | 111.76 | −461.22876 |
| 10 | 115.68 | 0.2430 | −0.3536 | 110.37 | −461.22619 |
| 11 | 87.61 | 0.2707 | −0.3392 | 75.15 | −767.17527 |
| 12 | 90.33 | 0.2491 | −0.3583 | 75.72 | −767.17296 |
| 13 | 106.32 | 0.2787 | −0.3394 | 87.83 | −1,226.78996 |
| 14 | 104.12 | 0.2780 | −0.3378 | 86.73 | −1,226.78655 |
| 15 | 104.46 | 0.2554 | −0.3359 | 88.13 | −1,226.79081 |
| 16 | 101.64 | 0.3983 | −0.3196 | 87.28 | −1,226.79240 |
| 17 | 114.63 | 0.2783 | −0.3313 | 99.81 | −1,686.40980 |
| 18 | 116.50 | 0.4025 | −0.3177 | 100.35 | −1,686.41042 |
| 19 | 139.41 | 0.4287 | −0.3101 | 123.90 | −2,605.62975 |
| 20 | 88.41 | 0.2505 | −0.3551 | 75.57 | −767.17364 |
| 21 | 103.11 | 0.2523 | −0.3532 | 87.08 | −1,226.78910 |
| 22 | 100.96 | 0.2541 | −0.3492 | 87.97 | −1,226.79355 |
| 23 | 114.19 | 0.2560 | −0.3357 | 98.67 | −1,686.40085 |
| 24 | 110.65 | 0.2580 | −0.3320 | 100.00 | −1,686.40562 |
| 25 | 113.59 | 0.4052 | −−0.3152 | 99.58 | −1,686.40758 |
| 26 | 118.73 | 0.2554 | −0.3482 | 98.87 | −1,686.40381 |
| 27 | 127.95 | 0.2589 | −0.3315 | 141.98 | −2,146.01482 |
| 28 | 131.35 | 0.4163 | −0.3142 | 111.99 | −2,146.02063 |
| 29 | 128.13 | 0.4124 | −0.3111 | 112.13 | −2,146.02185 |

[a] The structures are the same as Table 1

**Table 5** Self-correlation coefficients ($r^2$), variance inflation factors (VIF), standard regression coefficients (SR) and *t*-values for model (I) and model (II)

| Model | Variable | $r^2$ | VIF | $t(t_{\alpha/2} = 2.093)$, $\alpha = 0.05$ | SR |
|---|---|---|---|---|---|
| I | $\alpha$ | 0.0978 | 1.1084 | 19.471 | 0.7974 |
|  | TE | 0.0978 | 1.1084 | −7.436 | −0.3880 |
| II | $V_i$ | 0.5260 | 2.1095 | 19.961 | 0.5634 |
|  | $q^-$ | 0.5360 | 2.1548 | −9.877 | −0.2810 |
|  | TE | 0.3919 | 1.6444 | −30.530 | −0.7602 |

(I) and 0.5634, −0.2810, −0.7602 for model (II), respectively. Based on the *t*-values and *SR* in Table 5, $\alpha$ has remarkable influence on $-\lg S_w$ in model (I), while TE is the determinant factor on $\lg K_{ow}$ in model (II). In addition, the stability of the models in this study was validated by cross-validation and the relevant $q^2$ is 0.9729 and 0.9864, respectively, which suggests that the models exhibit optimum stability and better predictive power. So the models can be used to predict $-\lg S_w$ and $\lg K_{ow}$ of the same kind of compounds. With models (I) and (II), the predicted values of $-\lg S_w$ and $\lg K_{ow}$ of substituted phenols (No. 1–20) and their corresponding differences against the experimental values, are presented in Table 1, in which the maximum discrepancy of $-\lg S_w$ (−0.310) is found for compound No. 16 (2,6-dichlorophenol) and the maximum discrepancy of $\lg K_{ow}$ (−0.142) for compound No. 10 (1-naphthol).

At the same time, nine compounds were used to predict the values of $-\lg S_w$ and $\lg K_{ow}$, in which only experimental values of $\lg K_{ow}$ of five compounds have been reported in literatures. As listed in Table 1, for the specific five substituted phenols, the maximum discrepancy (−0.33) between the predicted and literature values is found for compound No. 25 (2,3,6-thichlorophenol). While the results predicted for other compounds are very close to

**Table 3** The obtained equations of $-\lg S_w$

Where *n* is the number of sample, $R^2$ is the squared correlation coefficient, $q^2$ is cross-correlation coefficient, SD is the standard deviation

| No. | Equation | *n* | $R^2$ | $q^2$ | SD |
|---|---|---|---|---|---|
| 2 | $-\lg S_w = -5.555 + 6.929\alpha/100$ | 20 | 0.9510 | 0.9388 | 0.220 |
| 3 | $-\lg S_w = -3.075 + 4.501\alpha/100 -0.668\text{TE}/1000$ | 20 | 0.9800 | 0.9729 | 0.141 |
| 4 | $-\lg S_w = -3.512 + 4.498\alpha/100 -0.498\text{TE}/1000 + 2.124qH^+$ | 20 | 0.9849 | 0.9740 | 0.121 |

**Table 4** The obtained equations of $\lg K_{ow}$

| No. | Equation | *n* | $R^2$ | $q^2$ | SD |
|---|---|---|---|---|---|
| 5 | $\lg K_{ow} = 1.632 - 1.162\text{TE}/1000$ | 20 | 0.8429 | 0.7926 | 0.290 |
| 6 | $\lg K_{ow} = -0.355 - 0.865\text{TE}/1000 +2.148V_1/100$ | 20 | 0.9574 | 0.9227 | 0.151 |
| 7 | $\lg K_{ow} = -4.601 - 0.963\text{TE}/1000 +2.946V_i/100 - 9.736q^-$ | 20 | 0.9940 | 0.9861 | 0.057 |

those of literatures, for all the discrepancies are smaller than ±0.11.

Up to the present, there have been rarely data of $-\lg S_w$ reported from different literatures. Compared $-\lg S_w$ values of six compounds by Wang et al. (1992) in Table 1, the results determined in this study are close to those of literatures except compounds No. 4 and No. 20 (the differences are 0.55 and 0.50, respectively). In addition, $\lg K_{ow}$ values of 18 compounds from literature are also listed in Table 1, where the largest difference (0.19) is found for compound No. 18 and the results determined in this study are also close to those of literatures for others compounds.

Based on the quantum parameters by AM1 method, Dai et al. (2000) established QSPR model predicting $\lg K_{ow}$ for substituted phenols, PCDDs, PCDFs, PCBs and so on, the equation as follows:

$$\lg K_{\mathrm{ow}} = 13.07 - 0.50\mu - 1.05E_{\mathrm{homo}} - 5.08q\mathrm{H}^{+} - 5.84q^{-} + 1.47\mathrm{TE} \quad (8)$$

$$R^2 = 0.9254,\ \mathrm{SD} = 0.40$$

With Eq. 8, the predicted discrepancy is 0.57 for chlorinated phenol (2,3,5-thichlorophenol) and larger than that of model (I) (−0.142). The model in this study only includes three variables, of which $R^2$ (0.9941) is higher than that of literature (0.9254) and SD (0.056) is lower than the corresponding value of literature.

Hong et al. (1995) predicted $-1\mathrm{g}S_w$ by molecular connectivity index for halogenated benzenes, amines, phenols, ketones, aethers, alcohols, the equation as follows:

$$-1\mathrm{g}S_{\mathrm{w}} = -0.440X^{\mathrm{v}} - 0.36X + 3.16\Delta X^{\mathrm{v}} + 1.53 \quad (9)$$

$$R^2 = 0.945,\ \mathrm{SD} = 0.33$$

With Eq. 9, the predicted discrepancy is −0.57 for 4-allyl−2-methoxyphenol and also larger than that of this study (−0.310). Comparison with Eq. 9, the model in this study only including two variables is more proper with higher $R^2$ (0.9800) and lower SD (0.141).

In conclusion, the experimental values of $-\lg S_w$ and $\lg K_{ow}$ of 20 substituted phenols were determined by the shake-flask method. And the novel QSPR models were developed successfully. The two models were further validated by variance inflation factors (VIF) and *t*-test. The results of comparison with literatures also showed the models of this study exhibit optimum stability and better predictive power than those based on AM1 molecular orbital method and molecular connectivity method.

**Acknowledgements** We are grateful to the financially supports from the National Natural Science Foundation of China (20737001, 20477018).

## References

Dai JY, Han SK, Wang LS (2000) Prediction of *n*-octanol/water partition coefficients and soil sorption coefficients for polychlorinated organic compounds (PCOCs) using quantum chemical parameters. Acta Sci Circum 20:693–698

Famini GR, Wilson LY (1997) Using theoretical descriptors in quantitative structure activity relationships: application to partition properties of alkyl (1-phenylsulfonyl) cycloalkane-carboxylates. Chemosphere 35:2417–2447

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery Jr JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) GAUSSIAN 98, revision A.9. Gaussian, Inc., Pittsburgh, PA

Gao H, Shanmugasundaram V, Lee P (2002) Estimation of aqueous solubility of organic compounds with QSPR approach. Pharm Res 19:497–503

Hansch C, Leo A (1979) Substituent constants for correlation analysis in chemistry and biology. John Wiley & Sons, New York

Hong H, Zhang YH, Han SK, Wang LS (1995) Estimation of solubility of organic compounds by molecular connectivity index. Chin Sci Bull 40:1108–1110

Liu G.S, Yu JG. (2005) QSAR analysis of soil sorption coefficients for polar organic chemicals: Substituted anilines and phenols. Water Res 39:2048–2055

Liu SS, Liu HL, Yin CS, Wang LS (2003) VSMP: a novel variable selection and modeling method based on the prediction. J Chem Inf Comput Sci 43:964–969

Loader RJ, Singh N, O'Malley PJ, Popelier PLA (2006) The cytotoxicity of ortho alkyl substituted 4-X-phenols: a QSAR based on theoretical bond lengths and electron densities. Bioorgan Med Chem Lett 16:1249–1254

Lu GH, Wang C, Tang ZY, Guo XL (2007) Quantitative structure-activity relationships for predicting the joint toxicity of substituted anilines and phenols to algae. Bull Environ Contam Toxicol 78:66–70

Urbanczyk A, Staniewski J, Szymanowski J (2002) Abraham model in gas chromatography of phenol pollutants. Anal Chim Acta 466:51–159

Wang LS, Zhao YH, Gao H (1992) Estimation of solubility of hydrophobic organic compounds, organic acids and bases. Environ Chem 11:56–70

Wei DB, Zhang AQ, Wu CD, Han SK, Wang LS (2001) Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds (PCOCs). Chemosphere 44:1421–1428

Wilson LY, Famini GR (1991) Using theoretical descriptors in quantitative structure–activity relationships: some toxicological indices. J Med Chem 34:1668–1674

Yao XJ, Panaye A, Doucet JP, Chen HF, Zhang RS, Fan BT, Liu MC, Hu ZD (2005) Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks. Anal Chim Acta 535:259–273