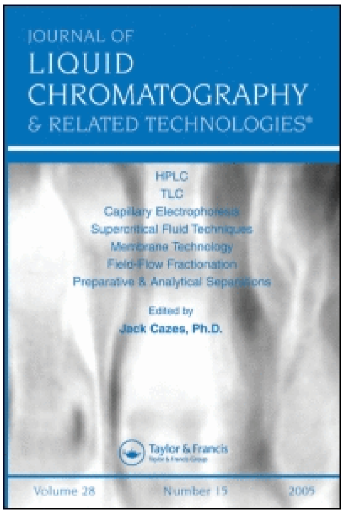


Journal of Liquid Chromatography & Related Technologies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

QSPR Prediction of Three Partition Properties for Phenylacrylates

D. B. Wei[a][b]; L. S. Wang[b]; Z. F. Lin[b]; H. Y. Hu[a]
[a] Department of Environmental Science and Engineering, ESPC, Tsinghua University, Beijing, P.R. China [b] School of Environment, Nanjing University, Nanjing, P.R. China

To cite this Article Wei, D. B. , Wang, L. S. , Lin, Z. F. and Hu, H. Y.(2003) 'QSPR Prediction of Three Partition Properties for Phenylacrylates', Journal of Liquid Chromatography & Related Technologies, 26: 13, 2065 — 2082
To link to this Article: DOI: 10.1081/JLC-120022394
URL: http://dx.doi.org/10.1081/JLC-120022394

JOURNAL OF LIQUID CHROMATOGRAPHY & RELATED TECHNOLOGIES®
Vol. 26, No. 13, pp. 2065–2082, 2003

QSPR Prediction of Three Partition Properties for Phenylacrylates

D. B. Wei,[1,2,*] L. S. Wang,[2] Z. F. Lin,[2] and H. Y. Hu[1]

[1]Department of Environmental Science and Engineering, ESPC,
Tsinghua University, Beijing, P.R. China
[2]School of Environment, Nanjing University, Nanjing, P.R. China

ABSTRACT

The *n*-octanol/water partition coefficients (K_{ow}) and water solubilities (S_w) of 11 phenylacrylates were determined by the shake-flask method. Their capacity factors (k') were determined by reversed-phase high performance liquid chromatography (RP-HPLC) with C_{18} column and methanol water eluent. It was found that both $\log K_{ow}$ and $\log S_w$ had good correlations with $\log k'$, respectively. Linear solvation energy parameters, describing the interaction of molecules accurately, were calculated for these compounds and used to predict their partition properties. Results demonstrated that three partition properties could be predicted precisely

---

[*] Correspondence: D. B. Wei, Department of Environmental Science and Engineering, ESPC, Tsinghua University, Beijing 100084, P.R. China/School of Environment, Nanjing University, Nanjing 210093, P.R. China; E-mail: weidongbin@ tsinghua.org.cn.

DOI: 10.1081/JLC-120022394
1082-6076 (Print); 1520-572X (Online)
www.dekker.com

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

with similar models based on linear solvation energy relationships, which indicated the similarity of their partition mechanisms.

Key Words: QSPR prediction; Partition properties; Phenylacrylates; LSER; Octanol/water partition coefficient; Water solubility; Capacity factor.

INTRODUCTION

Partitioning is one of the fundamental processes of pollutants in the environment. Many physicochemical properties, such as water solubility (S_w), octanol/water partition coefficient (K_{ow}), adsorption coefficient for soils and sediments (K_{oc}), and the bioconcentration factor (BCF), are relevant to the partition process.[1–5] Moreover, partition behaviors are also involved in some biological activity processes.

Derivates of phenylacrylates used widely as intermediary products to synthesize pesticides, drugs, and spices, are being introduced into the environment. However, these studied compounds are relatively new, the researches on their environmental behaviors are few, and selecting appropriate descriptors to describe their physicochemical and biological properties tends to be difficult. The linear solvation energy relationship (LSER) concept, developed by Kamlet et al.,[6] is a general approach to describe solvation and partition or related properties in diverse media. Within this approach there are three contributions to the total solute–solvent interaction:

$$\text{Total} = \text{Cavity term} + \text{Dipolarity/Polarizability term} \\ + \text{Hydrogen bonding term}$$

The cavity term measures the free energy or enthalpy input necessary to separate the solvent molecules, i.e., to overcome solvent–solvent cohesive interactions to provide a suitably sized cavity for the solute. The dipolarity/polarizability term measures the effects of solute–solvent dipole–dipole, and dipole–induced dipole interactions, and usually is exoergic (release energy). The hydrogen bonding term measures the exoergic effects of complexation between hydrogen bond donor solvents and hydrogen bond acceptor solutes or vice verse. A number of studies have demonstrated that many physicochemical and biological properties, such as water solubility, K_{ow}, BCF, K_{oc}, solute retention in HPLC, and biological activity, depend on the interaction of molecules, which can be predicted through LSERs.[7–11]

In this study, solubility (S_w) and *n*-octanol/water partition coefficient (K_{ow}) of a series of ethyl 2-formamido-phenylacrylates were predicted from chromatographic capacity factor (k'), which is a fast, accurate, and easily available method. Furthermore, all of three partition properties (S_w, K_{ow}, and k') were described through molecular structure parameters based on LSER models. The robustness of the obtained models was tested through a modified Jacknife method and Monte Carlo simulation test.

EXPERIMENTAL

Instruments and Reagents

The HPLC system (Shimadzu, Japan) consisted of a SCL-8A system monitor, a LC-8A pump, a C-R4A integrator, and a SPD-6AV ultraviolet spectrophotometer as detector.

Methanol (bought from Merck, German); doubly distilled water; sodium nitrate (analytical reagent); octanol (analytical reagent).

A total of 11 ethyl 2-formamido-phenylacrylates, kindly provided by the Chemistry Department of Nanjing University, were proven to have no interference peaks by HPLC determination. Their structures are shown in Table 1.

Calculation of Linear Solvation Energy Relationship Parameters

Linear solvation energy parameters, called "solvatochromic parameters," were previously measured by spectroscopic methods and were afflicted by bad availability and experimental errors. Wilson[12] introduced quantum chemical parameters into the calculation of LSER parameters, and developed "theoretical LSER" (TLSER). Subsequently, a handy estimation method "Rule of Thumb" was put forward by Hickey and Passino-Reader.[13] The LSER parameters included $V_i/100$, π^*, α_m, and β_m, which corresponded to cavity, dipolarity/polarizability, and hydrogen bonding term, respectively. The calculated LSER parameters of 11 compounds are also listed in Table 1.

Determination of Capacity Factors (*k*′)

A Nucleusil7 C_{18} column, 15 cm × 4.6 mm i.d. (made by Shanghai Institute of Organic Chemistry, Academic Sinica) was used. Mobile phases were made by mixing methanol with water in the proportions 100 : 0, 95 : 5, 90 : 10, 85 : 15, 80 : 20, 75 : 25, 70 : 30 (v/v). The flow rate was 1.0 mL min^{-1}. An aqueous solution of sodium nitrate was used for the measurement of dead time. All measurements were, at least, duplicated. The average reproducibility

Table 1. The molecular structures and LSER parameters of 11 tested compounds.

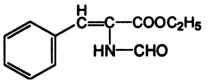
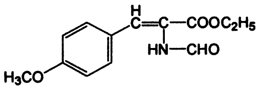
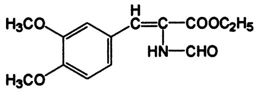
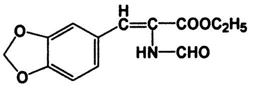
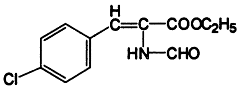
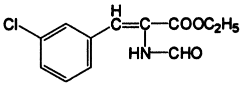
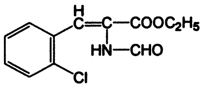
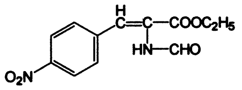
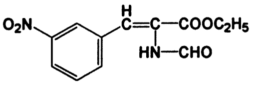
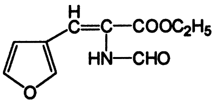
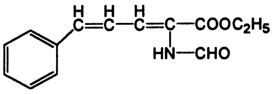
No.	Structures	α_m	β_m	V_i	π^*
1		0.50	1.43	1.314	2.25
2		0.50	1.65	1.457	2.38
3		0.50	1.87	1.600	2.51
4		0.50	2.45	1.512	2.33
5		0.50	1.39	1.404	2.37
6		0.50	1.39	1.404	2.37
7		0.50	1.39	1.404	2.37
8		0.66	1.63	1.454	2.67
9		0.66	1.63	1.454	2.67
10		0.50	1.64	1.193	2.06
11		0.63	1.60	1.617	2.45

Table 2. Experimental and predicted values of $\log K_{ow}$, $\log S_w$, and $\log k'_W$ from corresponding equations.

	$\log K_{ow}$			$\log S_w$				$\log k'_W$	
No.	Experimental	Predicted[a]	Predicted[b]	Experimental	Predicted[c]	Predicted[d]	Predicted[e]	Experimental	Predicted[f]
1	2.475	2.335	2.431	−2.889	−2.823	−2.956	−2.946	1.372	1.474
2	2.224	2.177	2.248	−2.736	−2.661	−2.712	−2.758	1.222	1.302
3	1.970	2.033	2.064	−2.414	−2.512	−2.464	−2.570	1.085	1.130
4	1.661	1.648	1.581	−2.144	−2.117	−2.163	−2.075	0.720	0.677
5	2.648	2.582	2.464	−3.012	−3.078	−3.125	−2.980	1.607	1.505
6	2.510	2.549	2.464	−3.110	−3.044	−2.990	−2.980	1.576	1.505
7	2.406	2.504	2.464	−3.004	−2.997	−2.889	−2.980	1.532	1.505
8	2.049	2.020	2.044	−2.455	−2.499	−2.541	−2.455	1.073	1.062
9	2.047	1.999	2.044	−2.466	−2.478	−2.539	−2.455	1.053	1.062
10	2.081	2.255	2.256	−2.748	−2.741	−2.753	−2.766	1.296	1.310
11	2.909	2.878	2.918	−3.354	−3.382	−3.379	−3.369	1.887	1.891

[a–f] corresponding to the Eqs. (1)–(6), respectively.

of each determination was better than 1.0% relative. Table 2 shows the extrapolated capacity factors at pure water.

Determination of Water Solubility (S_w)

S_w values were measured with the shake-flask method at $25 \pm 0.5°C$, as described by the Organization for Economic Cooperation and Development's (OECD) guidelines for testing of chemicals.[14] Each measure was done in triplicate bottles. The exact concentration of the sample was determined with UV 2201 UV–VIS spectrophotometer.

Determination of *n*-Octanol/Water Partition Coefficients (K_{ow})

K_{ow} values were determined with the shake-flask method according to the standard procedure of OECD[14] at 25°C. After equilibrium was achieved, the mixed solution was centrifuged, and the concentration of compounds in aqueous phase was determined with a spectrophotometer. K_{ow} values were calculated from the quotient of the volume-based *n*-octanol and water concentrations. The average reproducibility of each determination was better than 1% relative.

All of the determined values of K_{ow}, S_w, and extrapolated k'_W at pure water eluent, are listed in Table 2. The statistical analyses were done using the STATGRAPHICS program.[15] The QSPRs were obtained with the stepwise procedure using forward selection.

RESULTS AND DISCUSSION

Prediction of $\log K_{ow}$ and $\log S_w$ from HPLC Capacity Factor ($\log k'$)

The relationship between $\log k'$ and methanol concentration in mobile phase given by Snyder et al.,[16] is: $\log k' = \log k'_W - s \times \varphi_{CH_3OH}$. Where k'_W represents the k' value for a compound if pure water is used as eluent, s is the slope of regression curve, and φ_{CH_3OH} is the volume percentage of methanol in eluent. The regression coefficients for the studied compounds are all above 0.99.

A lot of studies have shown that capacity factors obtained by RP-HPLC correlate well with K_{ow} and S_w,[5,10] and HPLC, used for determining K_{ow} and S_w, has become a formal method.[17] In this study, it was found that there were good linear relationships between $\log K_{ow}$ and $\log k'_W$ [Eq. (1) in Table 3], $\log S_w$ and $\log k'_W$ (Eq. (2) in Table 3), respectively.

Table 3. The summary of obtained predictive equations.

Eq.	Dependent variable	Indep. variable	Coefficient	t-value	Constant	n	r/r^2_{adj}	s	F	p (<)
(1)	$\log K_{\text{ow}}$	$\log k'_{\text{W}}$	1.053	11.90	0.890	11	0.960	0.092	141.53	0.0001
(2)	$\log S_{\text{w}}$	$\log k'_{\text{W}}$	−1.085	−18.85	−1.336	11	−0.988	0.060	355.32	0.0001
(3)	$\log S_{\text{w}}$	$\log K_{\text{ow}}$	−0.974	−10.81	−0.545	11	−0.964	0.101	116.89	0.0001
(4)	$\log K_{\text{ow}}$	α_m	−1.376	−2.78	4.310	11	0.913	0.105	53.72	0.0001
		β_m	−0.833	−10.35						
(5)	$\log S_{\text{w}}$	α_m	2.002	5.25	−5.168	11	0.950	0.081	95.45	0.0001
		β_m	0.854	13.78						
(6)	$\log k'_{\text{W}}$	α_m	−1.601	−4.97	3.393	11	0.957	0.068	111.62	0.0001
		β_m	−0.782	−14.94						

Note: r, correlation coefficient; r^2_{adj} multiple correlation coefficient adjusted with free degree; s, standard error; F, F statistic; and n, number of observations. α_m, β_m are solvatochromic parameters, which represent the abilities of donating/accepting proton when the solute molecules combine and form hydrogen bond with solvent molecules.

The statistical significance of the equation is indicated by the fact that the calculated F value exceeds the tabulated value at the 99% confidence level ($F_{(1,9,0.01)} = 10.6$). From Eqs. (1) and (2), it appears that K_{ow} and S_w can be predicted by the HPLC capacity factor. The determined, calculated $\log K_{ow}$ and $\log S_w$ values are shown in Table 2. From Eq. (3), it can be seen that there is a strong correlation between $\log K_{ow}$ and $\log S_w$, as of previous reports.[18]

Prediction of $\log K_{ow}$, $\log S_w$, and $\log k'_W$ from Linear Solvation Energy Relationships

The linear solvation energy parameters are powerful tools in predicting both physicochemical properties and biological activities of pollutants, and have especial priority in describing the interaction between solutes and solvents. Equations (4)–(6), presented in Table 3, were developed by using multiple stepwise regression analyses. They describe the relationships of $\log K_{ow}$, $\log S_w$, and $\log k'_W$ based on LSER parameters, respectively. It can be seen that both of descriptors α_m and β_m are positively correlated with $\log S_w$, while negatively to $\log K_{ow}$ and $\log k'_W$. However, most interesting, is that the same two variables α_m and β_m enter in all of three equations. Considering the molecular structures of these studied compounds, it is not strange that they have similar parent structure, which include some proton accepting atoms, such as the O atom in $\text{—CH}\underline{\text{O}}$, $\text{—C}\underline{\text{O}}\text{OH}$, and N atom in $\text{—}\underline{\text{N}}\text{H—}$. This is in accordance with the higher t-value of β_m in Eqs. (4)–(6). In addition, the whole molecule is a large conjugated system.

The introduction of electron-donating groups (such as —OCH_3, $\text{—OCH}_2\text{—}$) into the benzene ring, would increase the electronic cloud density on the ring, activate the benzene ring, and increase the aqueous solubility. On the contrary, the electron-withdrawing group, such as —Cl, would decrease electron cloud density on the phenyl ring, and lead to low solubility. However, the introduction of a —NO_2, electron-withdrawing group, increased the solubility. It might be considered that —NO_2 decreased the cloud density on phenyl ring, but would form a hydrogen bond with the solvent water molecule and increase solubility. Therefore, the compound is much easier to partition into aqueous phase than into organic phase (octanol, methanol) due to the strong interaction between solute molecule and solvent water molecule. Equations (4)–(6) are highly significant at the level better than 0.01 ($F_{(2,8,0.01)} = 8.65$). The student t-values for each coefficient also are very large compared to the significance level of 0.05 ($t_{(11,0.05)} = 1.80$). Besides, there was no serious multicolinearity between the two entering variables (α_m and β_m). The plot of fitted values against observed values based on Eqs. (4)–(6) are shown in Fig. 1.

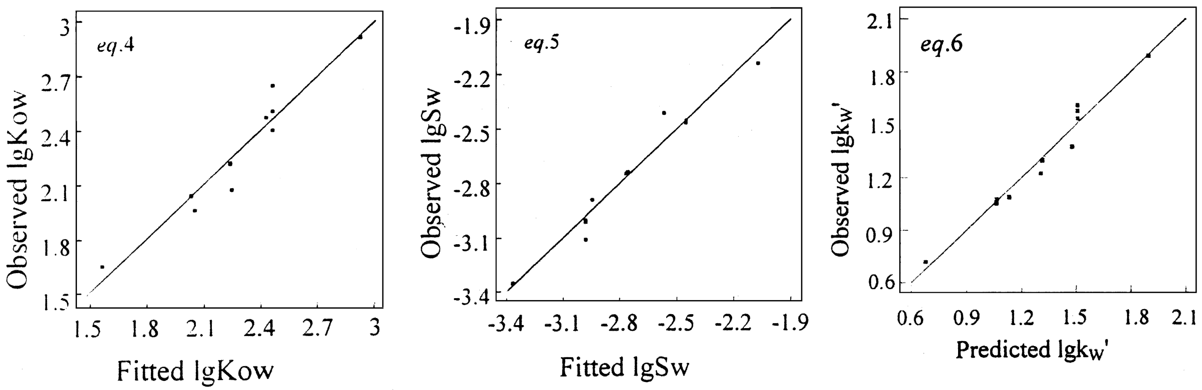


Figure 1. The plots of fitted values vs. observed values based on Eqs. (4)–(6), respectively.

Robustness Test

In order to test the robustness of the obtained equations, a modified *Jacknife* test was used. In this study the test sample is a small one ($n = 11$), so leave-one-out method was selected, one observation was randomly deleted each time, and regression was rerun for the rest of the 10 observations. The procedure was repeated 11 times, and all of 11 regression statistics were averaged. The overall results of the deletion study are summarized in Table 4. It is seen that the Eqs. (4)–(6) are quite robust. None of the regression parameters for the diminished data sets are significantly different from those for the full data sets. Furthermore, when an observation was deleted, the predicted value of the deleted one could be computed, and the residuals could also be obtained. The average residuals for the diminished data sets are 0.065, 0.055, 0.050; close to the original residuals 0.062, 0.049, 0.046 for full data sets based on Eqs. (4)–(6), respectively. This aspect of the study tests the predictive quality of the model, and the results demonstrate that the obtained equations have high predictive ability. The frequency distribution of multiple correlation coefficients of regressions is shown in Fig. 2.

Considering the structures of tested compounds and the frequency distribution of multiple correlation coefficients, it can be concluded that the distribution of structure fragment would affect the robustness of the model to a great extent. In addition, some useful information can be extracted from the frequency distribution of residuals that came from the difference of observed values and the predicted values of the model (Fig. 3). It is necessary to test the distribution of residuals for evaluating the model. The *K-S* method was selected, and the results showed that the frequency distribution of $\log K_{\mathrm{ow}}$, $\log S_{\mathrm{w}}$, and $\log k'_{\mathrm{W}}$ conformed to the normal function $X \sim N(-5.450 \times 10^{-16},\ 0.011^2)$, $X \sim N(8.478 \times 10^{-16},\ 0.007^2)$, and $X \sim N(-3.734 \times 10^{-16},\ 0.005^2)$, respectively. Based upon the above analyses, the obtained models are so robust that they are suitable to predict the $\log K_{\mathrm{ow}}$, $\log S_{\mathrm{w}}$, and $\log k'_{\mathrm{W}}$ values of new chemicals.

Monte Carlo Simulation Test

The Monte Carlo technique for testing the prediction performance of a regression equation permits using all data available for deriving the regression equation and does not require estimating degrees of freedom. These are especially important when the sample is small.[19] Due to only 11 observations being available in our present work, the Monte Carlo simulation was conducted to determine whether the prediction of the regression equation was reliable.

Table 4. Comparison of regression variables for diminished and full log K_{ow}, log S_w, and log k'_W data sets.

	Coefficients				
Data set	α_m	β_m	Constant	r	s
Eq. (4)					
Full	−1.376 (0.495)	−0.833 (0.081)	4.310 (0.331)	0.913	0.105
Diminished	−1.376 (0.531)	−0.837 (0.087)	4.315 (0.354)	0.910	0.105
Eq. (5)					
Full	2.002 (0.381)	0.854 (0.062)	−5.168 (0.255)	0.950	0.081
Diminished	2.003 (0.408)	0.858 (0.067)	−5.173 (0.272)	0.949	0.081
Eq. (6)					
Full	−1.601 (0.322)	−0.782 (0.052)	3.393 (0.215)	0.957	0.068
Diminished	−1.598 (0.346)	−0.783 (0.057)	3.393 (0.231)	0.955	0.069

Note: From Eqs. (4)–(6), respectively. See text for definition. Values are averages of 11 separate runs, and the standard error in the coefficient estimates are given in parentheses.

Table 2. Experimental and predicted values of $\log K_{ow}$, $\log S_w$, and $\log k'_W$ from corresponding equations.

	$\log K_{ow}$			$\log S_w$				$\log k'_W$	
No.	Experimental	Predicted[a]	Predicted[b]	Experimental	Predicted[c]	Predicted[d]	Predicted[e]	Experimental	Predicted[f]
1	2.475	2.335	2.431	−2.889	−2.823	−2.956	−2.946	1.372	1.474
2	2.224	2.177	2.248	−2.736	−2.661	−2.712	−2.758	1.222	1.302
3	1.970	2.033	2.064	−2.414	−2.512	−2.464	−2.570	1.085	1.130
4	1.661	1.648	1.581	−2.144	−2.117	−2.163	−2.075	0.720	0.677
5	2.648	2.582	2.464	−3.012	−3.078	−3.125	−2.980	1.607	1.505
6	2.510	2.549	2.464	−3.110	−3.044	−2.990	−2.980	1.576	1.505
7	2.406	2.504	2.464	−3.004	−2.997	−2.889	−2.980	1.532	1.505
8	2.049	2.020	2.044	−2.455	−2.499	−2.541	−2.455	1.073	1.062
9	2.047	1.999	2.044	−2.466	−2.478	−2.539	−2.455	1.053	1.062
10	2.081	2.255	2.256	−2.748	−2.741	−2.753	−2.766	1.296	1.310
11	2.909	2.878	2.918	−3.354	−3.382	−3.379	−3.369	1.887	1.891

[a–f] corresponding to the Eqs. (1)–(6), respectively.

Table 3. The summary of obtained predictive equations.

Eq.	Dependent variable	Indep. variable	Coefficient	*t*-value	Constant	*n*	r/r^2_{adj}	*s*	*F*	*p* (<)
(1)	$\log K_{\mathrm{ow}}$	$\log k'_{\mathrm{W}}$	1.053	11.90	0.890	11	0.960	0.092	141.53	0.0001
(2)	$\log S_{\mathrm{w}}$	$\log k'_{\mathrm{W}}$	−1.085	−18.85	−1.336	11	−0.988	0.060	355.32	0.0001
(3)	$\log S_{\mathrm{w}}$	$\log K_{\mathrm{ow}}$	−0.974	−10.81	−0.545	11	−0.964	0.101	116.89	0.0001
(4)	$\log K_{\mathrm{ow}}$	α_m	−1.376	−2.78	4.310	11	0.913	0.105	53.72	0.0001
		β_m	−0.833	−10.35						
(5)	$\log S_{\mathrm{w}}$	α_m	2.002	5.25	−5.168	11	0.950	0.081	95.45	0.0001
		β_m	0.854	13.78						
(6)	$\log k'_{\mathrm{W}}$	α_m	−1.601	−4.97	3.393	11	0.957	0.068	111.62	0.0001
		β_m	−0.782	−14.94						

Note: *r*, correlation coefficient; r^2_{adj} multiple correlation coefficient adjusted with free degree; *s*, standard error; *F*, *F* statistic; and *n*, number of observations. α_m, β_m are solvatochromic parameters, which represent the abilities of donating/accepting proton when the solute molecules combine and form hydrogen bond with solvent molecules.

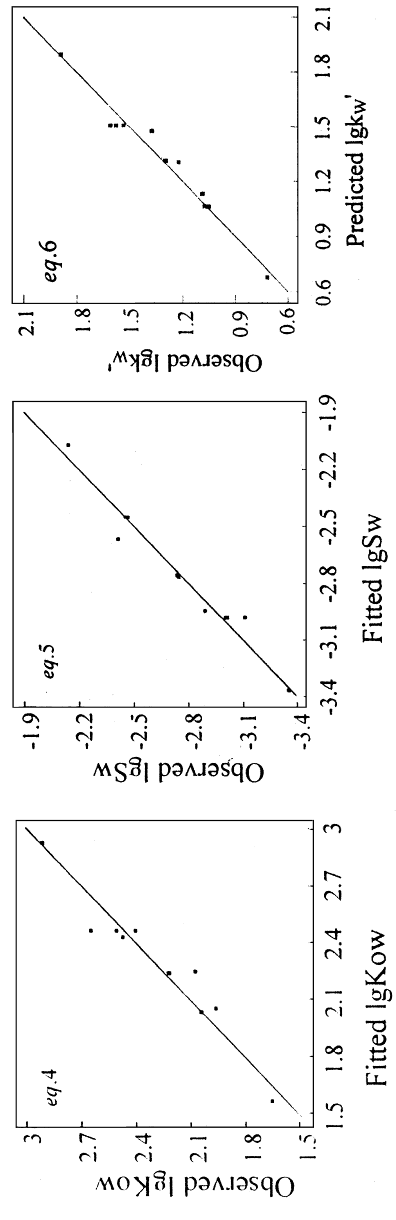


Figure 1. The plots of fitted values vs. observed values based on Eqs. (4)–(6), respectively.

Table 4. Comparison of regression variables for diminished and full log K_{ow}, log S_{w}, and log k'_{W} data sets.

	Coefficients				
Data set	α_m	β_m	Constant	r	s
Eq. (4)					
Full	−1.376 (0.495)	−0.833 (0.081)	4.310 (0.331)	0.913	0.105
Diminished	−1.376 (0.531)	−0.837 (0.087)	4.315 (0.354)	0.910	0.105
Eq. (5)					
Full	2.002 (0.381)	0.854 (0.062)	−5.168 (0.255)	0.950	0.081
Diminished	2.003 (0.408)	0.858 (0.067)	−5.173 (0.272)	0.949	0.081
Eq. (6)					
Full	−1.601 (0.322)	−0.782 (0.052)	3.393 (0.215)	0.957	0.068
Diminished	−1.598 (0.346)	−0.783 (0.057)	3.393 (0.231)	0.955	0.069

Note: From Eqs. (4)–(6), respectively. See text for definition. Values are averages of 11 separate runs, and the standard error in the coefficient estimates are given in parentheses.

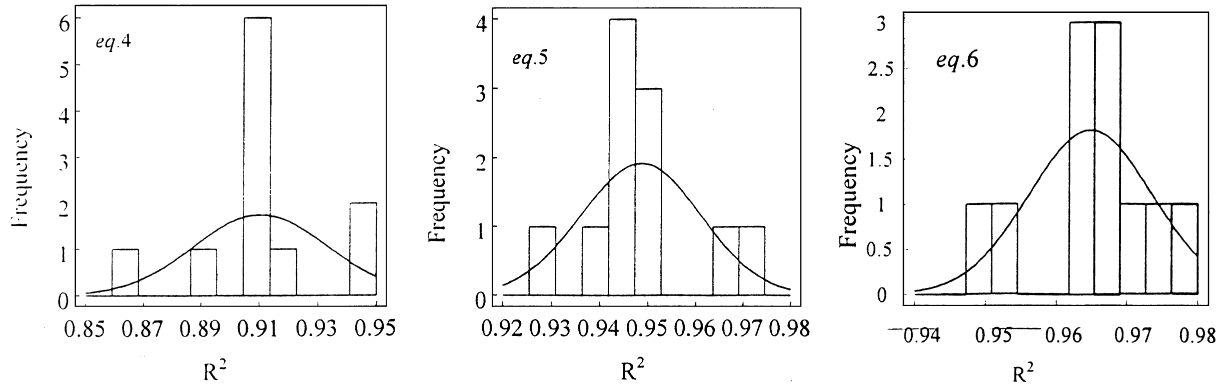


Figure 2. The frequency distribution of coefficients of Eqs. (4)–(6) using leave-one-out method, respectively.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

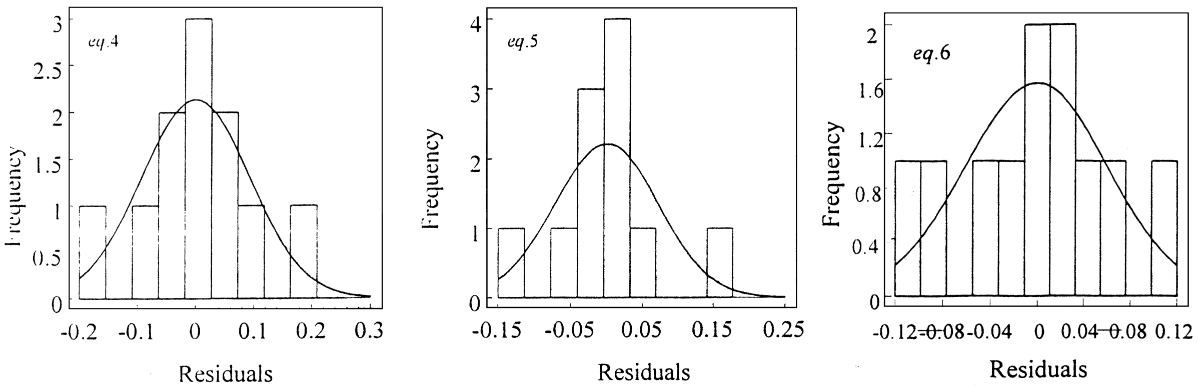


Figure 3. The frequency distribution of residuals from Eqs. (4)–(6), respectively.

Table 4. Comparison of regression variables for diminished and full $\log K_{\mathrm{ow}}$, $\log S_{\mathrm{w}}$, and $\log k'_{\mathrm{W}}$ data sets.

	Coefficients				
Data set	α_m	β_m	Constant	r	s
Eq. (4)					
Full	−1.376 (0.495)	−0.833 (0.081)	4.310 (0.331)	0.913	0.105
Diminished	−1.376 (0.531)	−0.837 (0.087)	4.315 (0.354)	0.910	0.105
Eq. (5)					
Full	2.002 (0.381)	0.854 (0.062)	−5.168 (0.255)	0.950	0.081
Diminished	2.003 (0.408)	0.858 (0.067)	−5.173 (0.272)	0.949	0.081
Eq. (6)					
Full	−1.601 (0.322)	−0.782 (0.052)	3.393 (0.215)	0.957	0.068
Diminished	−1.598 (0.346)	−0.783 (0.057)	3.393 (0.231)	0.955	0.069

Note: From Eqs. (4)–(6), respectively. See text for definition. Values are averages of 11 separate runs, and the standard error in the coefficient estimates are given in parentheses.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 2. Experimental and predicted values of $\log K_{ow}$, $\log S_w$, and $\log k'_W$ from corresponding equations.

	$\log K_{ow}$			$\log S_w$				$\log k'_W$	
No.	Experimental	Predicted[a]	Predicted[b]	Experimental	Predicted[c]	Predicted[d]	Predicted[e]	Experimental	Predicted[f]
1	2.475	2.335	2.431	−2.889	−2.823	−2.956	−2.946	1.372	1.474
2	2.224	2.177	2.248	−2.736	−2.661	−2.712	−2.758	1.222	1.302
3	1.970	2.033	2.064	−2.414	−2.512	−2.464	−2.570	1.085	1.130
4	1.661	1.648	1.581	−2.144	−2.117	−2.163	−2.075	0.720	0.677
5	2.648	2.582	2.464	−3.012	−3.078	−3.125	−2.980	1.607	1.505
6	2.510	2.549	2.464	−3.110	−3.044	−2.990	−2.980	1.576	1.505
7	2.406	2.504	2.464	−3.004	−2.997	−2.889	−2.980	1.532	1.505
8	2.049	2.020	2.044	−2.455	−2.499	−2.541	−2.455	1.073	1.062
9	2.047	1.999	2.044	−2.466	−2.478	−2.539	−2.455	1.053	1.062
10	2.081	2.255	2.256	−2.748	−2.741	−2.753	−2.766	1.296	1.310
11	2.909	2.878	2.918	−3.354	−3.382	−3.379	−3.369	1.887	1.891

[a–f] corresponding to the Eqs. (1)–(6), respectively.

Table 3. The summary of obtained predictive equations.

Eq.	Dependent variable	Indep. variable	Coefficient	*t*-value	Constant	n	r/r^2_{adj}	s	F	p (<)
(1)	$\log K_{\text{ow}}$	$\log k'_{\text{W}}$	1.053	11.90	0.890	11	0.960	0.092	141.53	0.0001
(2)	$\log S_{\text{w}}$	$\log k'_{\text{W}}$	−1.085	−18.85	−1.336	11	−0.988	0.060	355.32	0.0001
(3)	$\log S_{\text{w}}$	$\log K_{\text{ow}}$	−0.974	−10.81	−0.545	11	−0.964	0.101	116.89	0.0001
(4)	$\log K_{\text{ow}}$	α_m	−1.376	−2.78	4.310	11	0.913	0.105	53.72	0.0001
		β_m	−0.833	−10.35						
(5)	$\log S_{\text{w}}$	α_m	2.002	5.25	−5.168	11	0.950	0.081	95.45	0.0001
		β_m	0.854	13.78						
(6)	$\log k'_{\text{W}}$	α_m	−1.601	−4.97	3.393	11	0.957	0.068	111.62	0.0001
		β_m	−0.782	−14.94						

Note: r, correlation coefficient; r^2_{adj} multiple correlation coefficient adjusted with free degree; s, standard error; F, F statistic; and n, number of observations. α_m, β_m are solvatochromic parameters, which represent the abilities of donating/accepting proton when the solute molecules combine and form hydrogen bond with solvent molecules.

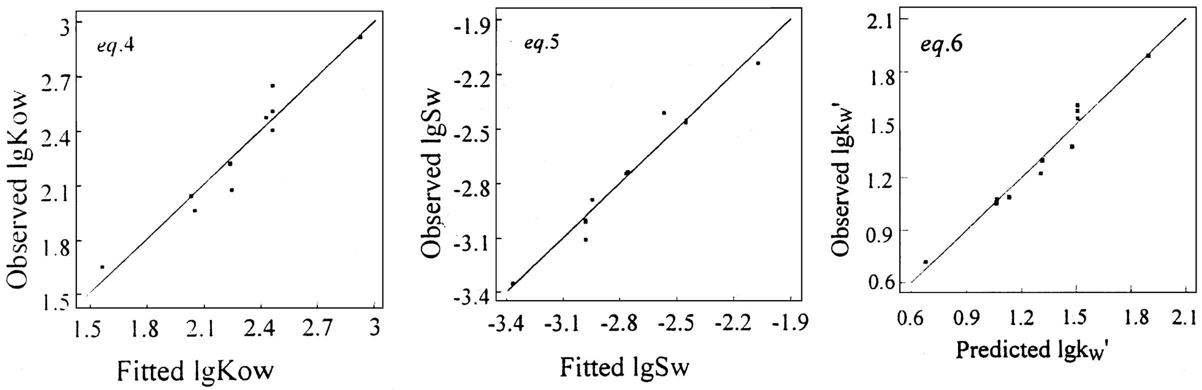


Figure 1. The plots of fitted values vs. observed values based on Eqs. (4)–(6), respectively.

Table 4. Comparison of regression variables for diminished and full $\log K_{\mathrm{ow}}$, $\log S_{\mathrm{w}}$, and $\log k'_{\mathrm{W}}$ data sets.

	Coefficients				
Data set	α_m	β_m	Constant	r	s
Eq. (4)					
Full	−1.376 (0.495)	−0.833 (0.081)	4.310 (0.331)	0.913	0.105
Diminished	−1.376 (0.531)	−0.837 (0.087)	4.315 (0.354)	0.910	0.105
Eq. (5)					
Full	2.002 (0.381)	0.854 (0.062)	−5.168 (0.255)	0.950	0.081
Diminished	2.003 (0.408)	0.858 (0.067)	−5.173 (0.272)	0.949	0.081
Eq. (6)					
Full	−1.601 (0.322)	−0.782 (0.052)	3.393 (0.215)	0.957	0.068
Diminished	−1.598 (0.346)	−0.783 (0.057)	3.393 (0.231)	0.955	0.069

Note: From Eqs. (4)–(6), respectively. See text for definition. Values are averages of 11 separate runs, and the standard error in the coefficient estimates are given in parentheses.

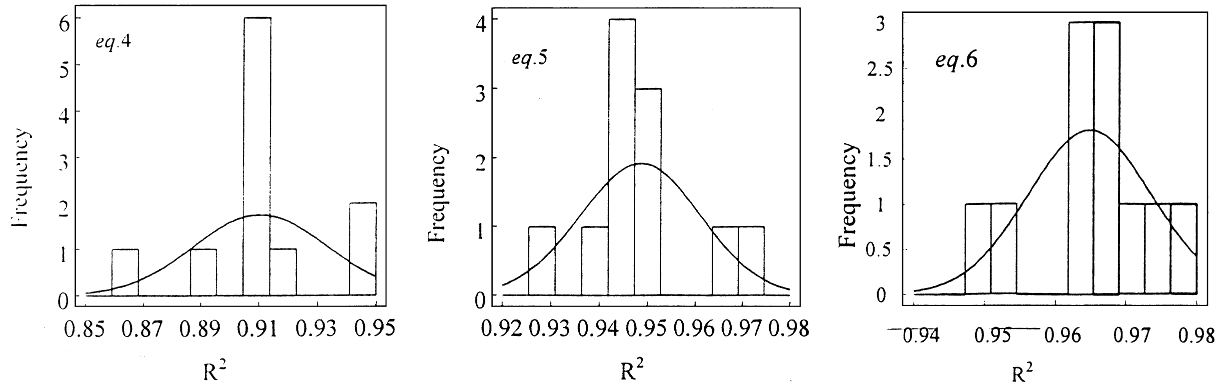


Figure 2. The frequency distribution of coefficients of Eqs. (4)–(6) using leave-one-out method, respectively.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

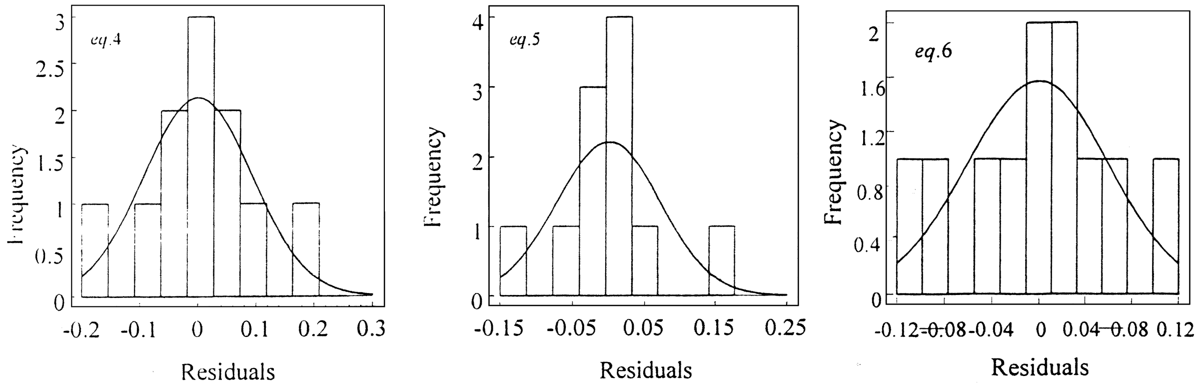
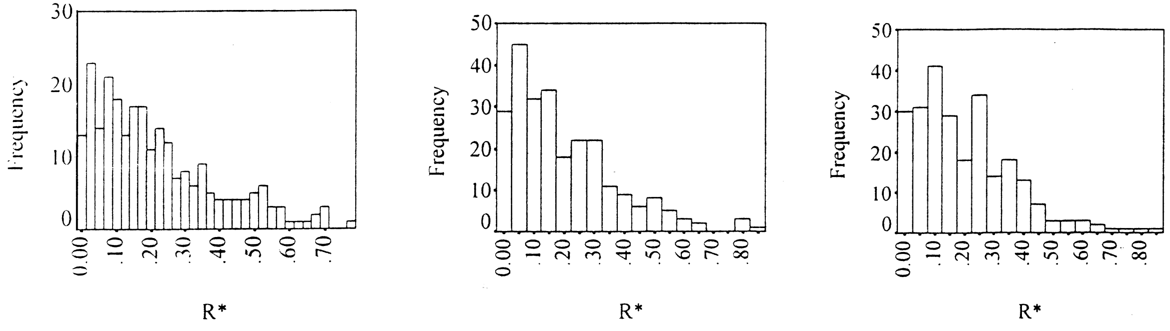


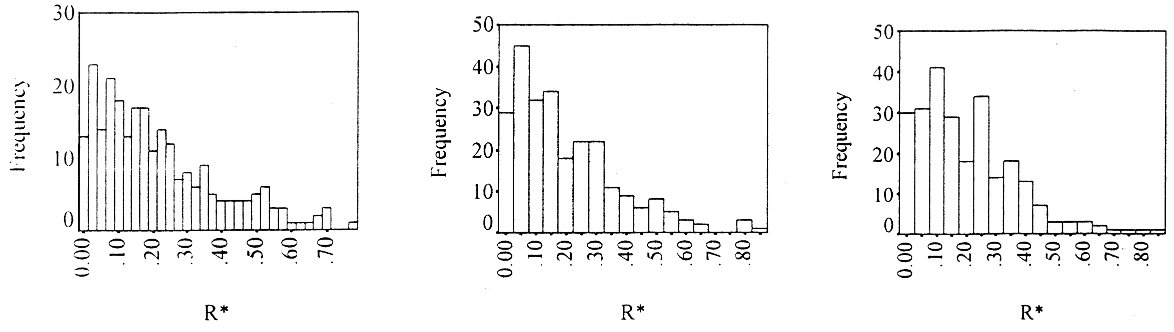
Figure 3. The frequency distribution of residuals from Eqs. (4)–(6), respectively.



(a) lgKow (b) lgSw (c) lgk'w

Figure 4. Empirical distribution of R^2* in Monte Carlo simulation test.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016



(a) lgKow

(b) lgSw

(c) lgk'w

Figure 4. Empirical distribution of R^2* in Monte Carlo simulation test.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 5. The results of Monte Carlo simulation test.

	Original distribution function	Range of R^2	Significance level (*K-S* test)	Frequency distribution function
Eq. (4)	N(2.271, 0.356^2)	0.002~0.772	0.641	Beta(0.968, 3.548)
Eq. (5)	N(−2.758, 0.360^2)	0.0001~0.822	0.630	Beta(0.844, 3.439)
Eq. (6)	N(1.311, 0.328^2)	0.002~0.828	0.607	Beta(0.991, 3.932)

Take, $\log K_{ow}$, for example. The null hypothesis is that the observed $\log K_{ow}$ of Eq. (4) is independent of structural descriptors (α_m, β_m) when using the Monte Carlo simulation to test the regression in Eq. (4). Eleven bogus values of predicted $\log K_{ow}$ were generated with a random number generator, stochastically sampling from a normal distribution (2.271, 0.356^2) of experimental observations. A regression equation was developed between these 11 bogus values of $\log K_{ow}$ and the structural descriptors, and the correlation coefficient of this spurious equation was recorded as R^{2*}. The structural descriptors were not changed in any manner when the spurious equations were

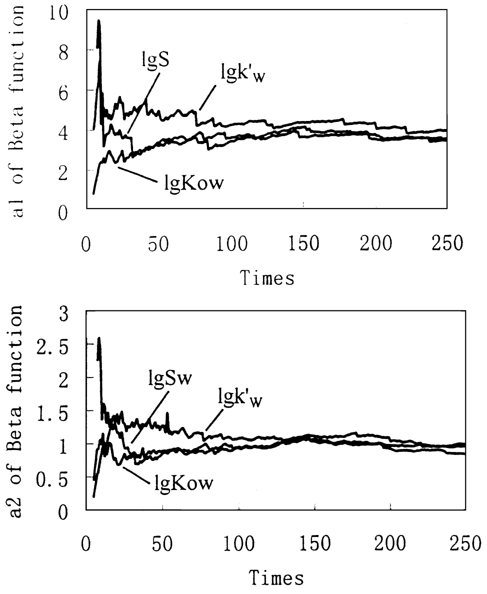


Figure 5. The changes of variables α_1, α_2 as the adding of R^{2*} one by one.

derived; only the predicted values were varied. Repeating such a Monte Carlo simulation 250 times resulted in 250 sets of 11 random numbers of $\log K_{ow}$, as well as 250 spurious equations and their corresponding R^2*, with the range from 0.002 to 0.772.

On the basis of results of *K*-*S* goodness-of-fit test, applied to the fitted empirical distribution of R^2*, obtained from the Monte Carlo simulation, the significance level is 0.988; greater than 0.05 suggests significant sufficiency of fit and no distinct difference from a Beta distribution ($\alpha_1 = 0.968$, $\alpha_2 = 3.548$) (Fig. 4). From the Beta probability distribution of R^2*, it is known, that for a probability of 0.975 (two sided), it is necessary to exceed a critical value of 0.641 to determine the significant difference between the values of R^2 and R^2*. Since R^2 0.913 in Eq. (4) is greater than R^2* in our study, it can be regarded that the prediction by Eq. (4) differs considerably from the random prediction by spurious equations; therefore, Eq. (4) is reliable. Based on Eq. (4), the $\log K_{ow}$ values could be predicted, to some extent, from quantum chemical parameters. The results of $\log S_w$ and $\log k'_W$ are similar to $\log K_{ow}$'s, which are listed in Table 5.

As for the Monte Carlo simulation test, the more times the random numbers are generated, the higher the reliability of the test. Theoretically, the natural distribution function would be obtained if there were enough generation times, even if it was irrelevant to the original distribution of the simulation sample. In this paper, the random numbers were generated from normal distribution functions, all of 250 R^2* of three samples conform to Beta distribution functions. It was obvious, that the variables α_1, α_2 tended to a relatively stable value as the 250 R^2* were added in one by one (shown in Fig. 5).

ACKNOWLEDGMENTS

This project was supported by National Natural Science Fund of P.R. China and China Postdoctoral Science Foundation.

REFERENCES

1. Thomsen, M.; Rasmussen, A.G.; Carlsen, L. SAR/QSAR approaches to solubility, partitioning and sorption of phthalates. Chemosphere **1999**, *38*, 2613–2624.
2. Dai, J.Y.; Jin, L.J.; Yao, S.C.; Wang, L.S. Prediction of partition coefficient and toxicity for benzaldehyde compounds by their capacity factors and various molecular descriptors. Chemosphere **2001**, *42*, 899–907.

3. Ohlenbusch, G.; Frimmel, F.H. Investigations on the sorption of phenols to dissolved organic matter by a QSAR study. Chemosphere **2001**, *45*, 323–327.
4. Wei, D.B.; Tan, Y.R.; Liu, X.H.; Wang, L.S.; Hu, H.Y. Partition properties of 18 polychlorinated organic compounds (PCOCs): correlation with molecular structural descriptors. J. Liq. Chromatogr. & Relat. Technol. **2002**, *25*, 627–637.
5. Hong, H.; Han, S.K.; Wang, X.R.; Wang, L.S. Prediction of partition coefficient and toxicity for phenylthio, phenylsulfinyl and phenylsulfonyl acetates. Environ. Sci. Technol. **1995**, *29*, 3044–3048.
6. Kamlet, M.J.; Doherty, R.M.; Carr, P.W. Linear solvation energy relationships: 44. Parameter estimation rules that allow accurate prediction of octanol/water partition coefficients and other solubility and toxicity properties of polychlorinated biphenyls and polycyclic aromatic hydrocarbons. Environ. Sci. Technol. **1988**, *22*, 503–509.
7. Famini, G.R.; Wilson, L.Y. Using theoretical descriptors in quantitative structure activity relationships: application to partition properties of alkyl(1-phenylsulfonyl) cycloalkane-carboxylates. Chemosphere **1997**, *35*, 2417–2447.
8. Carlsen, L.; Sørensen, P.B.; Thomsen, M. Partial order ranking-based QSAR's: estimation of solubilities and octanol-water partitioning. Chemosphere **2001**, *43*, 295–302.
9. Kamlet, M.J.; Doherty, R.M.; Taft, R.W. Solubility properties in polymers biological media, 8. An analysis of the factors that influence toxicities of organic nonelectrolytes to the golden orfe fish (*Leuciscus idus melanotus*). Environ. Sci. Technol. **1987**, *21*, 149–155.
10. He, Y.B.; Wang, L.S.; Han, S.K.; Zhao, Y.H.; Zhang, Z.; Zou, G.W. Determination and estimation of physicochemical properties for phenylsulfonyl acetates. Chemosphere **1995**, *30*, 117–125.
11. Klotz, W.L.; Schure, M.R.; Foley, J.P. Rapid estimation of octanol-water partition coefficients using synthesized vesicles in electrokinetic chromatography. J. Chromatography A **2002**, *962*, 207–219.
12. Wilson, L.Y. Using theoretical descriptors in quantitative structure activity relationships: some toxicological indices. J. Med. Chem. **1991**, *34*, 1668–1674.
13. Hickey, J.P.; Passino-Reader, D.R. Linear solvation energy relationships: "Rules of Thumb" for estimation of variable values. Environ. Sci. Technol. **1991**, *25*, 1753–1760.
14. *OECD Guideline for Testing of Chemicals*. Organization for Economic Cooperation and Development: Paris, France, 1981.
15. *STATGRAPHICS*, Version 4.0 (Serial Number: 3017435); STSC Inc. and Statistical Graphics Cooperation: USA, 1985.

16. Snyder, L.R.; Dolan, J.W.; Gant, J.R. Gradient elution in high performance liquid chromatography theoretic basis for reversed system. J. Chromatogr. **1979**, *165*, 3–30.
17. Klein, W.; Kordel, W.; Weib, M.; Poremski, H.J. Updating of the OECD test guideline 107 "partition coefficient n-octanol/water": OECD laboratory intercomparison test on the HPLC method. Chemosphere **1988**, *17*, 361–386.
18. Dai, J.Y.; Jin, L.J.; Han, S.K.; Wang, L.S. Determination and estimation of the water solubilities and the octanol/water partition coefficients for derivates of benzanilides. Chemosphere **1998**, *37*, 1419–1426.
19. Lund, I.A. A Monte Carlo method for testing the statistical significance of a regression equation. J. Appl. Meteror. **1970**, *9*, 330–332.

Received November 30, 2002
Accepted February 12, 2003
Manuscript 6025

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016