Ecotoxicology and Environmental Safety **52,** 143–149 (2002)
Environmental Research, Section B
doi:10.1006/eesa.2002.2168, available online at http://www.idealibrary.com on IDEAL®

A Case Study of Logistic QSAR Modeling Methods and Robustness Tests

Wei Dongbin, Zhang Aiqian, Wei Zhongbo, Han Shuokui, and Wang Liansheng[1]

State Key Laboratory of Pollution Control and Resource Reuse Research, P.R. China; and School of Environment, Nanjing University, Nanjing 210093, P.R. China

Received December 18, 2001

Acute toxicity (15-min EC_{50}) determination of 16 substituted naphthalene compounds to *Photobacterium phosphoreum* was undertaken according to the standard procedures, while the effects of molecular structures of selected compounds on their toxicity to test microorganisms were logistically conducted using the quantitative structure–activity relationship (QSAR) technique. The relationship was developed as $-\log EC_{50} = 5.5916 (\pm 0.1189) - 7.4893(\pm 0.4900)q\mathrm{H}^+ + 0.7771(\pm 0.0619)E_{\mathrm{lumo}} + 0.0088(\pm 0.0009)\alpha$ ($N = 16$, $R^2_{\mathrm{adj}} = 0.9698$, $\mathrm{SE} = 0.0892$, $P = 0.0000$). The cluster analyses of individual structure descriptors as well as the quality control chart and Monte Carlo simulation test indicated that the prediction model was reasonable and robust even if tested with several different methods. Furthermore, the quantum chemical parameters entering into the QSAR model were used to discuss the possible toxicity pathways, and the results revealed that the selected compounds were reactive and their toxicity behaviors were complex processes containing physical partition stages as well as biochemical reaction stages. © 2002 Elsevier Science (USA)

Key Words: **acute toxicity; QSARs; *Photobacterium phosphoreum*; Monte Carlo simulation; substituted naphthalene compounds.**

INTRODUCTION

It has been widely recognized that knowledge on the acute and chronic toxicity of pollutants is a basic requirement in ecological risk assessment. However, there are an overwhelming number of chemicals being introduced into the environment, and toxicity determination is costly and time-consuming. Fortunately, quantitative structure–activity relationships (QSARs) can reveal the relationship between the toxicity of a compound and its structural descriptors (Blum and Speece, 1990). Moreover, they are of benefit in the development of property/toxicity data because they allow estimation of the toxicity to an organism based on easily measured or calculated characteristics. As a result, there are large numbers of relevant examples in QSAR studies (Baj and David, 1994; Lewis, 1989; Nevalainen and Kolehaminen, 1994; Mekenyan *et al*., 1994; Xu *et al*., 1994) depending on quantum chemical descriptors with obvious advantages (Dai *et al*., 1998). However, the collinearities among some quantum chemical descriptors limit the predict ability of QSAR models; that is, if some of the information was obtained from a weak correlation equation, it would be unreliable and would not reflect the real toxicity pathway. Therefore, to a certain extent, the robustness of a model is much more important than a high R^2 and a low SE.

One objective of this study is to determine acute toxicities of 16 substituted naphthalene compounds to *Photobacterium phosphoreum* and to develop a prediction model based on the structural descriptors with clear "physico chemical sense" through the logistic QSAR modeling method. The other objective is to test the robustness of the model using several methods and attempt to clarify the possible toxicity pathways depending on the information extracted from the prediction model.

MATERIALS AND METHODS

The 16 selected naphthalene compounds, listed in Table 1, are of analytical reagent grade (some purchased from Aldrich Chemical Co.). The test species, *P. phosphoreum* (T_3 mutation), was supplied in freeze-dried powder form by the Institute of Soil Science, Academia Sinica, Nanjing, P. R. China. Stock cultures were maintained on agar slants at 4°C. The culture broth was yeast–tryptone–salts–glycerol, pH 7.0 ± 0.5, and the medium was sterilized with high pressure (*Experimental Technology of Environmental Biology*, 1989).

The concentration for 50% inhibition of bioluminescence of *P. phosphoreum* after 15 min exposure, expressed as EC_{50} (mol/L), was measured. The Microtox test instrument (toxicity analyzer Model DXY-2), made by the Institute of Soil Science, Academia Sinica, Nanjing, P. R. China, was used. The experiment was performed at 20°C, near neutral pH, according to the procedures described by the instrument manual (*Experimental Technology of Environmental Biology*, 1989). The working solutions were colorless, and the toxicity data ($\log 1/EC_{50}$) are listed in Table 1.

0147-6513/02 $35.00

TABLE 1
Compounds, Quantum Chemical Parameters, and Toxicity Data

No.	Name	α	E_{lumo}	qH^+	log K_{ow}	log(1/EC_{50}) Exp.	Est.	Res.
1	Naphthalene	95.66	− 1.268	0.1565	3.316	4.244	4.280	− 0.036
2	2-Bromo-naphthalene	101.32	− 0.551	0.1467	4.179	4.968	4.980	− 0.012
3	1,2-Dibromo-naphthalene	111.11	− 0.831	0.1518	4.842	4.844	4.803	0.041
4	1-Nitro-naphthalene	106.17	− 1.311	0.1686	3.059	4.262	4.248	0.014
5	1-Naphthalamine	99.37	− 0.175	0.1858	2.089	4.953	4.973	− 0.020
6	2-Naphthalamine	101.13	− 0.175	0.1844	2.089	5.034	4.999	0.035
7	4-Chloro-1-naphthalamine	103.89	− 0.642	0.1581	2.901	4.786	4.828	− 0.042
8	1-Naphthalenesulfonic acid	112.01	− 1.269	0.2703	1.459	3.552	3.572	− 0.020
9	2-Naphthalenesulfonic acid	114.10	− 1.218	0.2769	1.459	3.607	3.583	0.024
10	2-Amino-1-naphthalenesulfonic acid	121.80	− 0.912	0.2739	0.232	3.935	3.919	0.016
11	5-Amino-2-naphthalenesulfonic acid	124.65	− 1.076	0.2783	0.232	3.971	3.780	0.191
12	7-Amino-4-hydroxyl-2-naphthalene-sulfonic acid	132.67	− 1.091	0.2769	− 0.435	3.648	3.849	− 0.201
13	1-Naphthalenol	94.67	− 0.368	0.2191	2.649	4.492	4.527	− 0.035
14	2-Naphthalenol	95.82	− 0.343	0.2174	2.690	4.470	4.544	− 0.074
15	5-Amino-1-naphthalenol	104.88	− 0.279	0.2198	1.770	4.634	4.656	− 0.022
16	2,2′-Binaphthalenol	224.10	− 1.453	0.2209	4.504	4.788	4.784	0.004

All molecules were built and optimized by the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Broyden 1970; Fletcher, 1970, 1980; Goldfarb 1970; Shanno, 1970), and quantum descriptors were calculated using the AM1 Hamiltonian of MOPAC program (Cambridgesoft Corp., 1997). The obtained quantum descriptors included average molecular polarizability (α), dipole moment (μ), energy of the lowest unoccupied molecular orbital (E_{lumo}), energy of the highest occupied molecular orbital (E_{homo}), the largest negative atomic charge on atoms ($\bar{q}$), and the most positive atomic charges on hydrogen atoms (qH^+). The bulkiness-related (nonspecific) parameters included final heat of formation (HOF), total energy (TE), electronic energy (EE), and core–core repulsion (CCR). Units of the heat of formation, energy, charge, dipole, and polarizability were kilo-calories, electron volts (eV), atomic charge unit (acu), and atomic unit (au), respectively.

All stepwise multiple regressions reported were performed using the STATGRAPHICS program (STSC, Inc., 1985), and the multicollinearities were measured with variance inflation factors. (VIF_i, defined as $1/(1 - R_i^2)$, where R_i^2 represents the correlation coefficient of ith subject variable regressed on all the other explanatory variables.) The correlation of relevant pairs of explanatory variables needed to be tested since three or more variables were involved in regression analyses. If the ith variable is involved in multicollinearity, then R_i^2 will be close to unity. The "ideal" situation in regression is for all R_i^2 values to be zero and all VIF values to be in unity. Generally, a VIF_i value under 5 indicates acceptable multicollearity, while the regression will be considered unstable when the value is near 10.0.

RESULTS

The logistic modeling method is very useful in QSAR studies. The current model includes several steps. First, suitable parameters that well-describe the property/activity are selected and their sensitivity to selected compounds is tested through various methods. Second, a prediction equation is developed on the basis of selected parameters with clear physiochemical sense, such as quantum chemical parameters. Third, the robustness of the established model is tested. Last, the possible toxicity pathways are discussed according to the information extracted from the established models.

In general, if more than 90% of the members of selected compounds are classified into one cluster through a parameter, the parameter would be considered invalid in a prediction model. However, the results of cluster analyses in the current study indicated that there were no ineffective data among the log(1/EC_{50}) (Fig. 1) and molecular structure descriptors. In addition, it has been suggested that the biological response is related to both the transport of chemicals from water phase to biophase and the interaction between the chemical and the biotarget molecule. Regardless of the biodegradation of chemicals in the organism, the biological response could be expressed as a linear function of three main properties: hydrophobicity, and electrostatic and steric effects (Hansch and Leo, 1995). Consequently, quantum chemical parameters with celar "sense" were the preferential selection for toxicity simulation to be of benefit in gaining insight into the intrinsic toxicity pathway. However, the results of partial correlation analyses indicated that there were remarkable multicollinearities among the

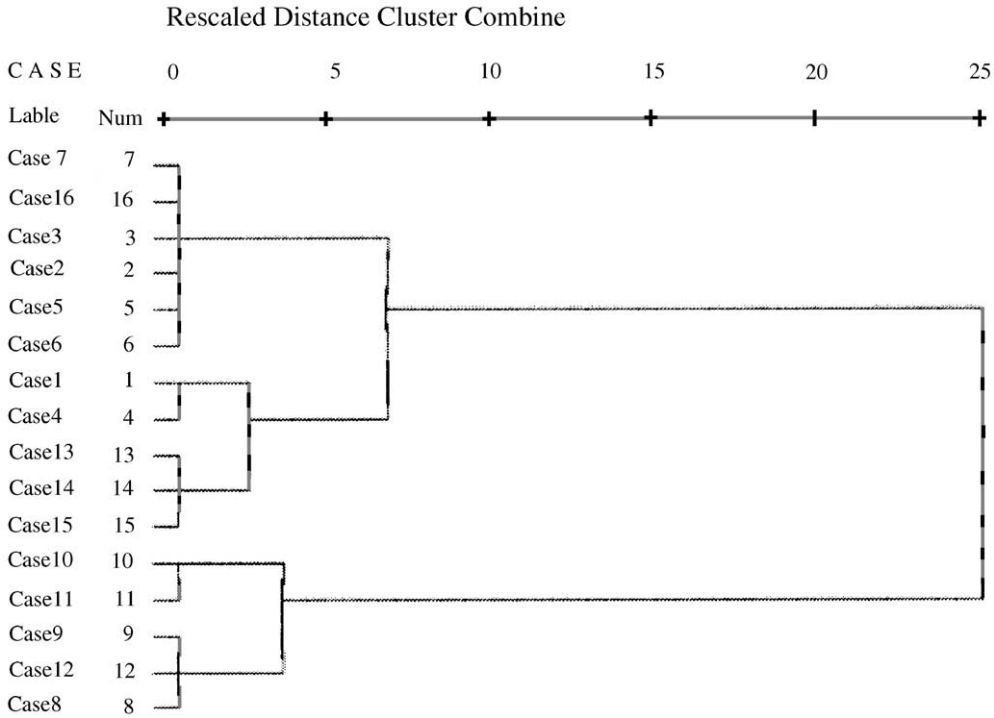

FIG. 1. Dendrogram of hierarchical cluster analysis for $\log(1/EC_{50})$ data set.

nonspecific quantum chemical parameters; therefore, descriptors with weak collinearities α, E_{lumo}, E_{homo}, $\bar{q}$, $q\mathrm{H}^+$, and HOF, along with $\log K_{ow}$ (*n*-octanol/water partition coefficient, which were estimated using CLOGP software (Manual S., 1985)) were selected as candidate parameters for $\log(1/EC_{50})$ simulation and prediction. The stepwise regression equation is

$$-\log EC_{50} = 5.5916\ (\pm 0.1189) - 7.4893\ (\pm 0.4900)q\mathrm{H}^+ + 0.7771\ (\pm 0.0619)E_{lumo} + 0.0088\ (\pm 0.0009)\alpha, \quad (1)$$

where $N = 16$, $R^2 = 0.9758$, $R^2_{adj} = 0.9698$, SE = 0.0892, $F = 161.358$, $P = 0.0000$, and N represents the number of samples, R^2 is the multiple correlation coefficient, SE is the standard error, F denotes the F test value, P is the significance level of the whole equation, and the values in parentheses are 95% confidence intervals associated with each coefficient.

Considering the R^2 and SE, Eq. (1) is adequate to simulate $\log(1/EC_{50})$ of the 16 selected compounds, and basically describes the electrostatic and steric effects on the toxicities when the chemicals interact with biotarget molecules. However, the effects of hydrophobicity on the toxicity of selected compounds cannot be extracted directly from Eq. (1); therefore, a simple regression between $\log(1/EC_{50})$ and $\log K_{ow}$ was developed,

$$-\log EC_{50} = 3.8565\ (\pm 0.1798) + 0.2300\ (\pm 0.0655)\log K_{ow}, \quad (2)$$

where $N = 16$, $R^2 = 0.4685$, $R^2_{adj} = 0.4305$, SE = 0.3869, $F = 12.3390$, and $P = 0.0034$. Equation (2) indicates that the toxicity process includes the partition stage, and the relationship between $\log EC_{50}$ and $\log K_{ow}$ is still uncertain, which probably results from the great differences in the molecular structures of the selected compounds. Most of the selected compounds have polar functional groups such as $-NH_2$, $-SO_3H$, and $-OH$, and they tend to form a hydrogen bond with water molecules, thus reducing the sensitivity of K_{ow} for distinguishing the differences in toxicity. Furthermore, the test microorganism, *P. phosphoreum*, is a monad with low fat content, and therefore the toxicities of the compounds are not dependent on lipophilicity but on the interaction between the chemical and the enzymes of the microorganism.

In order to shed more light on the toxicity process and verify the reliability of Eq. (1), equations based on molecular connectivity indices (MCIs) and the linear solvation energy relationship (LSER) were developed; 42 MCIs, including path, cluster, path/cluster, chain indices, and nondispersive force factors, were calculated (Kier and Hall, 1976, Bahnick and Doucette, 1988). Only ${}^0\chi$, ${}^1\chi$, ${}^2\chi_p$, ${}^3\chi_p$, ${}^4\chi_p$, ${}^3\chi_c$, ${}^4\chi_{pc}$, ${}^0\chi^v$, ${}^1\chi^v$, ${}^2\chi^v_p$, ${}^3\chi^v_p$, ${}^4\chi^v_p$, ${}^3\chi^v_c$, ${}^4\chi^v_{pc}$, $\Delta^0\chi^v$, $\Delta^1\chi^v$, $\Delta^2\chi^v_p$, $\Delta^3\chi^v_p$, $\Delta^4\chi^v_p$, $\Delta^3\chi^v_c$, and $\Delta^4\chi^v_{pc}$ were selected as candidate parameters. Four LSER parameters, α_m, β_m, π^*, and $V_{i/100}$, were estimated according to the "Rules of thumbs" method (Hickey and Passino-Reader, 1991). Among them, α_m and β_m measure the exoergic effects of complexation between hydrogen bond donor solvents and hydrogen bond acceptor solutes or vice

verse, π^* reflects the effects of solute–solvent, dipole–dipole, and dipole-induced dipole interactions, and $V_{i/100}$ measures the free energy or enthalpy input necessary to separate the solvent molecules and provide a suitably sized cavity for the solute:

MCI

$$-\log \mathrm{EC}_{50} = 5.8454\ (\pm 0.3577) - 0.4250$$

$$(\pm 0.0606)^2\chi_p + 1.4211\ (\pm 0.4531)^3\chi_p$$

$$N = 16,\ R^2 = 0.8049,\ R^2_{adj} = 0.7724,\ \mathrm{SE} = 0.2477,$$

$$F = 24.7521, \text{ and } P = 0.0001 \quad (3)$$

$$-\log \mathrm{EC}_{50} = 4.8002\ (\pm 0.1084) + 0.7209\ (\pm 0.1285)\Delta^2\chi^v_p$$

$$\text{where } N = 16,\ R^2 = 0.7077,\ R^2_{adj} = 0.6853,$$

$$\mathrm{SE} = 0.2912,\ F = 31.4805, \text{ and } P = 0.0001 \quad (4)$$

LSER

$$-\log \mathrm{EC}_{50} = 4.9462\ (\pm 0.1746) - 0.8189\ (\pm 0.2167)\beta_m$$

$$N = 16,\ R^2 = 0.5048,\ R^2_{adj} = 0.4695,\ \mathrm{SE} = 0.3734,$$

$$F = 14.2736, \text{ and } P = 0.0020. \quad (5)$$

In Eq. (3), the descriptor $^2\chi_p$ is positively interrelated to the molecular volume, $^3\chi_p$ is relevant to numbers of substituting groups, and both descriptors imply the volume characteristics of molecules. It is easy to understand that the larger the molecular volume is, the more difficult it is for the compound to transport through the cell membrane, and the lower the toxicity is. However, the characteristics of the substituting groups are often more important than their numbers in contributing to toxicity. Equation (4) is based on a special kind of molecule connectivity index, nondispersive force factor $\Delta\chi^v$, which reflects the polar characteristics. The higher the $\Delta^2\chi^v_p$ is, the higher toxicity is; therefore the nondispersive force (electrostatic force, hydrogen bond interaction, etc.) plays an important role in the toxicity process. Unfortunately, MCIs have no clear meaning, and are not of benefit in investigating toxicity pathways. In Eq. (5), the higher descriptor β_m is, the stronger the basicity of compound is, the easier it is for the compound to form a hydrogen bond with water, the less compound partition into the organism there is, and the lower the toxicity is.

DISCUSSION

Equations (2)–(5) agree with Eq. (1), and all of them imply that the toxicities of these selected compounds are determined by their molecular structures, and mainly rely on electrostatic and steric effect, while the hydrophobicity is not significant enough. Therefore, Eqs. (2)–(5) are forceful evidences to prove that Eq. (1) is satisfactory for toxicity prediction. Additionally, the correlativity of Eq. (1) was found to be significant through the correlation coefficient test and the variables entered into the model are reasonable according to the results of multicollinearity test listed in Table 2. Moreover, the equation is successful in evaluating $\log(1/\mathrm{EC}_{50})$ for the 16 compounds; the results are plotted in Fig. 2. In order to test the robustness of Eq. (1), a modified "jackknife test" method was applied to the data set where a random umber of observations were deleted at a time, and the regression was rerun for the remaining observations. A randomly selected number of regressions were run with different members deleted each time, and all of the regression statistics were averaged. To maintain stability, the number of deletions was kept below 20% of the total number of observations. Figure 3 vividly depicts the effects of each compound on the robustness of Eq. (1) by comparing the correlation coefficients (R^2_{adj}) in leave-one-out test. It is obvious that there are comparatively great increases in R^2_{adj} when the $\log(1/\mathrm{EC}_{50})$ values of compounds 11 and 12 are deleted separately, which indicates that some of the selected compounds with great differences in substituents have significant effects on the correlativity of the model. For example, the order of toxicity of the different substituents is $-NH_2 > -OH > -Br > -NO_2 > -SO_3H$, and therefore it

TABLE 2
Correlation Coefficient Matrix of Independent Variables and Corresponding *VIF*

	Correlation matrix			*VIF*
	α	E_{lumo}	$q\mathrm{H}^+$	Eq. (1)
α	1.0000			1.2015
E_{lumo}	0.4863	1.0000		1.1434
$q\mathrm{H}^+$	− 0.1489	0.1392	1.0000	1.0000

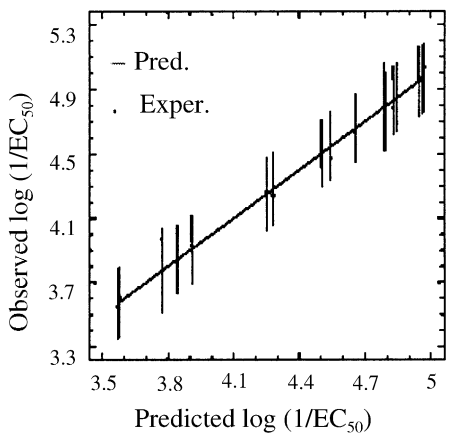

FIG. 2. Predicted and observed values with 95% intervals for predictions.

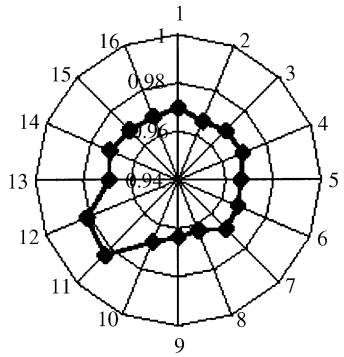

FIG. 3. Plot of R^2_{adj} for Eq. (1) by leave-one-out method.

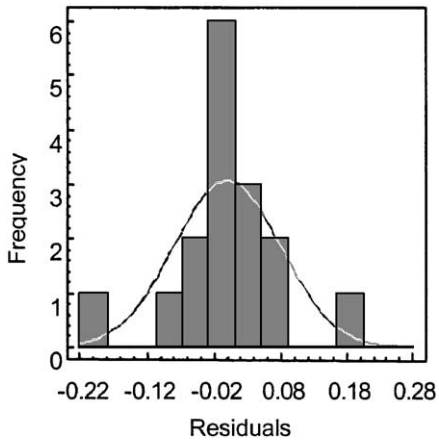

FIG. 4. Frequency distribution of residuals for Eq. (1).

can be concluded that the introduction of $-SO_3H$ would decrease toxicity, while the others groups further up the list would increase toxicity. This indicates that chemical interactions play an important role in the toxicity process. The results of leave-one/several-out tests are listed in Table 3, which indicates that Eq. (1) is robust enough. Moreover, the K–S (Kolmogorov-Smirnov) method (Andersen and Borgan, 1998) was used to test the frequency distribution of residuals from Eq. (1) (which was a small sample because there were fewer than 50 compounds tested). The results are presented graphically in Fig. 4 and listed in Table 4. These results indicate that the frequency distribution of the residuals is in accord with the normal function $X \sim N(2.7756 \times 10^{-17}, 0.0892^2)$. The coefficient of skewness was 0.0273 and the coefficient of kurtosis was 3.7874.

The Monte Carlo technique is a useful tool for testing the prediction performance of a regression equation of small sample (Lund, 1970), it permits the use of all data available in deriving the regression equation, and it does not require estimating degrees of freedom. Due to the fact that only 16 observations were available in the present work, the Monte Carlo simulation was conducted to determine the reliability of regression equation.

The null hypothesis is that the observations ($-\log EC_{50}$) of the model are independent of structural descriptors when using Monte Carlo simulation to test Eq. (1). Sixteen bogus values of the predited $-\log EC_{50}$ were generated randomly using a stochastic sampling from a normal distribution ($4.3868, 0.5132^2$) of the experimental observations. A regression equation was developed between these 16 bogus values of $-\log EC_{50}$ and the candidate parameters, and the correlation coefficient of this spurious equation was recorded as R^2*. The candidate parameters were not changed in any manner when the spurious equations were derived; only the predicted values were varied. Repeating the simulation 250 times resulted in 250 sets of 16 random numbers of $-\log EC_{50}$, as well as 250 spurious equations and their corresponding R^2* in the range 0.0082 to 0.6310.

On the basis of results of a chi-square goodness-of-fit test applied to the fitted empirical distribution of R^2*, obtained from the Monte Carlo simulation, the chi-square value is 3.2489, and the significance level 0.9985 greater than 0.05 suggests significant sufficiency of fit and no distinct difference from the beta function ($\acute{a}_1 = 1.6389$, $\acute{a}_2 = 5.4007$) (Fig. 5). 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 0.5799 to determine the significant difference between the values of R^2 and R^2*. Since R^2 (0.9698) in Eq. (1) is greater than R^2* in this study, it can be concluded that the prediction by Eq. (1) differs considerably from the random prediction by spurious equations, and, therefore, Eq. (1) is reliable. Based on Eq. (1), the $-\log EC_{50}$ value could be predicted, to some extent, from quantum chemical parameters.

TABLE 3
Results of Leave-One/Several-Out Test Eq. (1)

No. of cases deleted (< 20% of N)	No. of regression runs	$-\log EC_{50}$ av. R^2	av. R^2_{adj}	av. SE
1	16	0.9766	0.9695	0.0891
2	30	0.9753	0.9679	0.0913
3	50	0.9750	0.9666	0.0925
Average		0.9756	0.9680	0.0910

TABLE 4
Results of K–S Test

	$\log(1/EC_{50})$
Estimated KOLMOGOROV statistic DPLUS	0.1801
Estimated KOLMOGOROV statistic DMINUS	0.1742
Estimated overall statistic DN	0.1801
Approximate significance level	0.6770

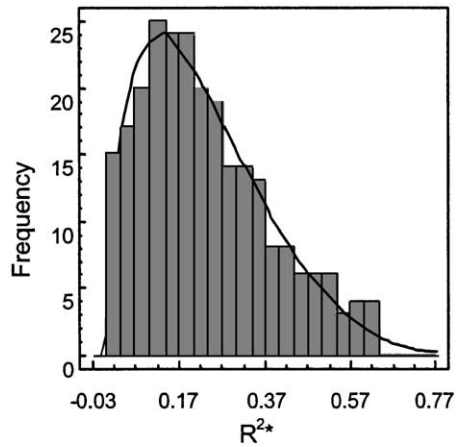

FIG. 5. Empirical distribution of R^2*.

Additionally, the selected naphthalene compounds were divided into four groups to roughly study the distribution of the robustness of the obtained model to each group with special structure characteristics. Similarly, the modified jackknife test was applied again to test the data set when one group of observations was deleted at a time, and the regression was rerun for the remaining three groups. Figure 6 graphically depicts the distribution of correlation coefficients for each group when the corresponding observations were deleted. It indicates that the robustness of Eq. (1) for predicting toxicity greatly depends on the NP subgroup (NP is a symbol denoting compounds 1, 2, 3, 4, and 7) with large differences in molecular structure. However, it needs to be mentioned that the result is only a rough one since the number of compounds within each subgroup is beyond the valid limit of the jackknife test ($\leq 20\%$). In order to gain further insight into the toxicity mechanism, the luminance process of *P. phosphoreum* is described as follows:

$$\mathrm{FMNH_2 + RCHO + O_2} \xrightarrow{\text{bacterial fluroenzyme}} \mathrm{FMN + RCOOH + H_2O + light.}$$

In the above process, FMN is flavin mononucleotide and $FMNH_2$ is its reductive form. $FMNH_2$ is an important coenzyme that transfers hydrogen. In the $FMNH_2$ molecule, the –NH-group reacts readily with other molecules by hydrogen action. Some functional groups involving high electronegativity atoms, such as O and N, form a hydrogen bond with $FMNH_2$ and the hydrogen transfer in the luminance process mentioned above is hindered; therefore the light emission of *P. phosphoreum* is inhibited.

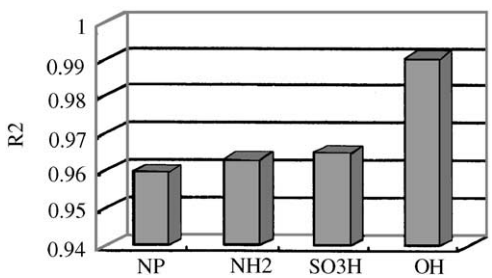

FIG. 6. Distribution plot of R^2_{adj} for each deleted group.

Furthermore, the Student t test method was used to investigate the correlation of each independent variable and dependent variable in Eq. (1), and the t values were -15.2851, 12.5589, and 9.9851 for $q\mathrm{H}^+$, E_{lumo}, and α, respectively. The most positive formal charge on hydrogen atom ($q\mathrm{H}^+$) was the most statistically significant term influencing toxicity, and the toxicity decreased as the increasing of $q\mathrm{H}^+$, which implies that the oxygen atoms in water molecules and hydrogen atoms in chemical molecules interact and form a hydrogen bond. Thus, the compounds with high $q\mathrm{H}^+$ tend to be partitioned in water rather than in cells and result in low toxicity. For example, the selected compounds with a $-SO_3H$ group have comparatively low K_{ow} values while their $q\mathrm{H}^+$ descriptors are higher than the others, as indicated in Table 1. It can therefore be concluded that there exists a strong solvation effect between the compound molecules and solvent water molecules. The energy of lowest unoccupied molecular orbital (E_{lumo}) could be used as a measure of molecular ability to accept an electron pair. An increase in E_{lumo} would lead to an increase in toxicity, which implies that the hydrogen bond exists between the chemical and the "target molecules" in the microorganism: the target molecules accept electron pairs and the chemicals donote them. This point can be explained by the fact that some atoms with high electronegativity, such as O and N, contained in the selected compounds donate lone-pair electrons and accept proton H of $FMNH_2$ and therefore inhibit the luminance ration. The compounds containing $-SO_3H$ with low toxicity are exceptions due to their high solubility in water. In addition, the greater is the polarizability (α) the greater is the toxicity. Descriptor α is involved in Eq. (1) as expected since it is in direct proportion to intrinsic molecular volume, and molecular volume is a measure of the energy needed to form a cavity in solvent. The higher α was, the larger the molecular volume was, and the larger the molecular deformation was. The result was that the solute molecule transferred much more easily and partitioned into the cell with less polarity through the cell membrane.

CONCLUSION

In summary, it can be concluded from the information presented in this logistic study that a model based on quantum chemical descriptors can be used to predict toxicity to a certain extent, and that it can provide a useful starting point for predicting potential environmental contaminants. However, for an established model, it is

essential to test robustness in order to guarantee a wide application range and accurate predictive ability. It was proved that the Monte Carlo simulation test was a reliable method for uncertainty analysis. Also, useful information extracted from a reasonable and robust model would shed light on the action mechanism, especially on pathway of toxicity. Toxicity behavior is not a simple partition process, and it is restricted by many factors, such as biochemical processes.

ACKNOWLEDGMENTS

This project was supported by National Natural Science Fund and the Research Fund for the Doctoral Program of Higher Education of P. R. China.

REFERENCES

Andersen, per K., Borgan, O., Gill, R. O., and Keiding, N. (1998). *Statistical Models Based on Counting Processes*. Springer, New York.

Bahnick, D. A., and Doucette, W. J. (1988). Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. *Chemosphere* **17**(9), 1703–1715.

Baj, S., and David, M. (1994). Correlation between the chemical structures of dialkyl peroxides and their retention in reversed-phase high-performance liquid chromatography. *J. Liquid Chromat.* **17**(18), 3933–3949.

Blum, D. J., and Speece, R. E. (1990). Determining chemical toxicity to aquatic species. *Environ. Sci. Technol.* **24,** 284.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. 1, 2. *J. Inst. Math. Appl.* **6,** 76–90, 222–231.

Dai, J. Y., *et al.* (1998). Prediction of water solubility and toxicity of substituted indoles to *Photobacterium phosphoreum* by using molecular connectivity indices and quantum chemical parameters. *Bull. Environ. Contam. Toxicol.* **61**(5), 591–599.

Dewar, M. J. S., Zoebisch, E. G., *et al.* (1985). AM1. A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**(13), 3902–3909.

Environmental Biology Teaching and Research Section of Department of Environmental Science and Engineering of Nanjing University (1989). *Experimental Technology of Environmental Biology* (in Chinese), pp. 145–151. Nanjing University Press, Nanjing.

Fletcher, R. (1970). A new approach to variable metric algorithms. *Comp. J.* **13,** 317–322.

Fletcher, R. (1980). *Practical Methods of Optimization*, Vol. 1. Wiley, New York.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Math. Comp.* **24,** 23–26.

Hansch, C., and Leo, A. (1995). *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society, Washington, DC.

Hickey, J. P., and Passino-Reader, D. R. (1991). Linear solvation energy relationships: "Rules of thumbs" for estimation of variable values. *Environ. Sci. Technol.* **25**(10), 1753–1760.

Kaliszan, R. (1993). Quantitative structure–retention relationship applied to reversed-phase high-performance liquid chromatography. *J. Chromat. A* **656,** 417–435.

Kier, L. B., and Hall, L. H. (1976). *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, New York.

Lewis, D. F. V. (1989). The calculation of molar polarizabilities by the CNDO/3 method: Correlation with the hydrophobic parameter LogP. *J. Comput. Chem.* **10**(2), 145–151.

Lund, I. A. (1970). A Monte Carlo method for testing the statistical significance of a regression equation. *J. Appl. Meteor.* **9,** 330–332.

Manual, S. (1985). *MedChem. Manual. Release 3*. Claremont, CA: Pomona College.

Mekenyan, O. G., Ankley, G. T., *et al.* (1994). QSARs for photoinduced toxicity I: Acute lethality of polycyclic aromatic hydrocarbons to *Daphnia magna*. *Chemosphere* **28**(3), 567–582.

Nevalainen, T., and Kolehaminen, E. (1994). New QSAR models for polyhalogenated aromatics. *Environ. Toxicol. Chem.* **13**(10), 1699–1706.

Shanno, D. F. (1970). Condition of quasi-Newton methods for function minimization. *Math. Comp.* **24,** 647–656.

Stewart, J. J. P. (1997). *CS MOPAC* 97. Cambridgesoft Corp.

STSC, Inc. and Statistical Graphics Corp. (1985). *STATGRAPHIC*, Version 4.0.

Tripos Associates, Inc. (1988). *A Modeling System for the IBM PC*.

Xu, L., Ball, J. W., *et al.* (11994). Quantitative structure activity relationships for toxicity of phenols using regression analysis and computational neural networks. *Environ. Toxicol. Chem.* **13**(5), 841–851.