Chemical Communications, Volume 47, Number 4 (pp. 1102 – 1106) 2015

QSAR study of aromatic compounds toxicity to *Chlorella vulgaris*

N. Li[1, 2], F. Chen[1, 2], D. Yang[1, 2], Y. B. Zhou[1, 2*]

[1] *Key Laboratory of Marine Bio-resources Restoration and Habitat Reparation in Liaoning Province, Dalian Ocean University, 116023 Dalian, China;*
[2]*Key Laboratory of North Mariculture, Ministry of Agriculture, Dalian Ocean University, 16023Dalian,China*

ReceivedApril 4, 2015

As demands for energy have increased worldwide, oil has become the main pollutant of the ocean. The impact of aromatic compounds, which are primary pollutants in oil, on various marine ecosystems has become a growing concern. Establishing a quantitative structure–activity relationship (QSAR) model to predict the toxicity of unknown aromatic compounds may thus serve as an important pollution-preventive measure. In this study, 21 aromatic compounds, 15 of which served as a training set and 6 as a test set, were selected. The structural parameters of the compounds were obtained by multiple linear regression, and a 2-descriptor prediction model was established. The test set was used to determine the predictive ability of the model. The model built using the proposed method showed satisfactory statistical results ($R^2 = 0.974$ vs. the test set $R^2 = 0.804$). These data show that the model provides good predictability and stability and can thus be used to predict the inhibitory effect of aromatic compounds on *Chlorella vulgaris*.

Key words:Aromatic compounds,*Chlorella vulgaris*, QSAR.

INTRODUCTION

Asdemands for energy have increased worldwide, exploration and production of offshore oil and gas have steadily expanded [1]. Oil transport via marine vessels and oil pipelines in the ocean frequently resultsin oil spills. Thus, oil has become the main pollutantcausing serious damage to the marine ecosystem [2]. While aromatic compounds are widely used in the industry, most of them are toxic;in fact, aromatic compounds are the main pollutantsin oil.Bioaccumulation of aromatic compounds can destroy human hematopoietic functions and even cause cancer. The effect of aromatic compounds in marine ecosystems has elicited great concern from environmental scientists[3].

Algae are important primary producersand the bases of marine food chain. Theseorganisms maintain thefunctions of various marine ecosystems. Previous findings show that algae are more sensitive to pollution than fish or crustaceans. *Chlorella*is an algal species with a wide ecological distribution and short growthperiod.It is easy toisolateandculture. Symptoms of *Chlorella* poisoning can be directly observed at the cellular level; thus, the species is an ideal test organism in toxicity experiments [4]. As such, using algae to evaluate the effect of toxicaromatic compounds on marine ecosystems is ecologically significant.

Quantitative structure–activity relationships (QSARs) establish links between the activity of organic matter and its structure. QSAR is based on theoretical calculations of parameters and does not rely on experimental parameters; thus, the method is both convenient and timesaving. It can predict, filter, and preliminarily evaluate the activity of organic matterby establishing a mathematical model that can predict its biological activity[5,6]. QSARis widely used in predictions of biological toxicity[7]and has become a rapid, economical, and effective method to generate basic toxicity data for risk evaluation and management of compounds. QSAR employs a variety of methods, such as multiplelinearregression(MLR)[8],partialleast-squaresregression (PLS) [9], and nonlinear regression,among others, to establish a mathematical model.

The effects of aromatic compounds on algae have recently been studied using QSAR.Sacan et al.[10] built a QSAR model describing the toxicity of substituted benzene to oblique *Scenedesmus* and found that the toxicity of the aromatic hydrocarbons is associated with their molecular size, hydrophobicity, and highest occupied molecular orbital energy(EHOMO). Modelliet al. [11} usedrelevant data on ionization energy, electron affinity, and relationship between EHOMO and lowest occupied molecular orbital energy(ELUMO) to predict the ionization energy, electron affinity, and bio-toxicity of polycyclic aromatic hydrocarbons (PAH).

* To whom all correspondence should be sent:
E-mail: Ybzhou@dlou.edu.cn

Table 1. Aromatic compounds descriptions and toxicities to *Chlorella vulgaris*

NO.	Name	SMILES*	TpiPC	MATS3v	pEC50		
					Exp.	Pre.	Dif.
1	Phenol	Oc1ccccc1	5.119	−0.233	−2.339	−2.083	−0.256
2	*o*-chloroaniline	c1(c(cccc1)Cl)N	5.384	−0.192	−2.242	−1.858	−0.384
3	*m*-chloroaniline	c1c(cccc1Cl)N	5.376	−0.316	−2.202	−2.165	−0.037
4	*p*-nitroaniline	c1([N+](=O)[O−])ccc(N)cc1	5.889	−0.300	−1.894	−1.880	−0.014
5	nitroaniline	c1(c(cccc1)N)[N+](=O)[O−]	5.924	−0.219	−1.815	−1.672	−0.143
6	*m*-nitroaniline	c1(cc(ccc1)N)[N+](=O)[O−]	5.905	−0.300	−1.782	−1.877	0.095
7	nitrobenzene	c1(ccccc1)[N+](=O)[O−]	5.690	−0.115	−1.665	−1.529	−0.136
8	nitrophenol	Oc1cc(ccc1)[N+]([O−])=O	5.905	−0.246	−1.627	−1.745	0.118
9	2,4-dinitrotoluene	Cc1ccc(cc1[N+]([O-])=O)[N+]([O−])=O	6.534	−0.289	−1.605	−1.557	−0.048
10	*p*-nitrophenol	c1(cc(ccc1c)[N+](=O)[O−])[N+](=O)[O−]	6.534	−0.289	−1.570	−1.557	−0.013
11	2-chlorotoluene	c1(c(cccc1)Cl)c	5.384	−0.188	−1.478	−1.849	0.371
12	3-nitrochlorobenzene	c1(cc(ccc1)Cl)[N+](=O)[O−]	5.905	−0.204	−1.202	−1.644	0.442
13	2,4-nitrochlorobenzene	c1(cc(ccc1Cl)[N+](=O)[O−])[N+](=O)[O−]	6.534	−0.182	−1.503	−1.297	0.244
14	*o*-dinitrobenzene	c1(c(cccc1)[N+](=O)[O-])[N+](=O)(O-)	6.416	−0.071	−0.940	−1.082	0.142
15	2,4-dichloronitrobenzene	c1(c(cc(Cl)cc1)Cl)[N+](=O)[O−]	6.109	−0.239	−0.681	−1.633	0.952
16	naphthalene	c12ccccc1cccc2	7.454	−0.076	−0.663	−0.608	−0.055
17	acenaphthene	c12c3CCc1cccc2ccc3	8.252	−0.063	−0.233	−0.202	−0.031
18	fluorine	c12c3c(ccc3)Cc1cccc2	7.694	−0.197	−0.193	−0.791	0.598
19	benzo(a)anthracene	c12c3c(ccc1cc1ccccc1c2)cccc3	9.918	−0.039	0.658	0.635	0.023
20	pyrene	c12c3c4ccc1cccc2ccc3ccc4	9.965	0.005	0.783	0.766	0.017
21	benzo(a)pyrene	c1ccc2c(c1)cc3ccc4cccc5c4c3c2cc5	10.604	0.021	1.046	1.102	−0.056

*SMILES: Simplified molecular input line entry specification.

In the present study, PaDEL-Descriptor [12] software was used to calculate the 2D descriptors of 21 aromatic hydrocarbons, andSPSS 19.0 software was employed to perform MLR and select two optimal descriptors. Regression equations considering the *Chlorella* toxicity data and these optimal descriptors were obtained by the stepwise regression method.

DATA COLLECTION AND METHODS

Data

Data of the 21 aromatic compounds, including phenol and benzo(a)pyrene, and their toxicity to *Chlorella* within 96 h of exposure were obtained from the literature [13]. These data were used to establish a QSAR model through the linear regression method with SPSS 19.0. During model building, 15 molecules (5/7) were randomly assigned to the training set, and the remaining 6 molecules were assigned to the test set, as shown in Table 1 (Note: molecules in boldface constitute the test set). The pEC_{50}(−LogEC_{50}) values of these compounds represent their inhibitory activities toward *Chlorella*.

PaDEL-Descriptor software was used to calculate the molecular structural descriptors of each of the 21 aromatic compounds; parameters with zero values were then deleted to obtain 36 groups of structure parameters. Finally, QSAR analysis was performed using MLR.

Calculation Methods

This study used the MLR method to build a QSAR model. MLR analysis is based on calculation of regression equations through analysis of the relationships between variables; this method is often used for QSAR modeling [14–15]. The mathematical model of the multivariate linear regression equation is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

where Y is the bio-activity of the aromatic compound, namely,pEC_{50}; x_1, x_2, x_3, …, x_n are the molecular structural descriptors of the compounds; β_1, β_2, β_3,…, β_n are the regression coefficients; βiis an average change of dependent variable Y caused by one unit of change of independent variable xi,when the other independent variables remain unchanged, i = 1, 2, …, k; and β_0 is a constant.

RESULTS

The simplified group of 36 molecular structural parameters wastreated as a group of independent variables, and the values of theinhibitory activity were considered as dependent variables. The stepwise selection method was used to analyze, predict, and build an ideal mathematical model containing two descriptors. The model equation is as follows:

$$pEC_{50} = -8.911 + 0.468\,TpiPC + 2.484\,MATS3v$$

(1)

The correlation coefficients obtained are as follows: R^2=0.974, F=226.452, SE=0.188,and n=15; these values indicate statistically sound results. Equation (1) reveals thatTpiPC and MATS3vinfluence the activity of the aromatic compounds. Equation (1) was then defined as Model I, andthe test value of each independent variableis shown in Table 2.

Table 2. Descriptions of the parameters and their test values

Variable	Descriptions	B	T	Sig.	VIF
TpiPC	autocorrelation	0.468	10.464	0	2.541
MATS3v	path count	20434	3.477	0.005	2.541

Table 2 shows thatall values of theindependent variable (VIF) <10, which illustratesthelack of multiple linear relationships between the independent variables; Sig<0.01 indicates the significant influence of the two structural parameters considered on the activity of the compounds andtherelative stabilityof the model. When Model I was applied to the test set, the correlation coefficients R^2 = 0.809 and SE = 0.448 were obtained. These results confirm the reliability and predictive ability of Model I.

The inhibitory activity of the 21 aromatic compounds toward *Chlorella* was predicted by using Model I, and the results are shown in Table 1. Correlations between the experimental and the predicted values are shown in Fig. 1, and residual errors between the experimental and predicted values are shown in Fig. 2. Fig. 1 demonstrates that the experimental and predicted values of the training and test sets are consistent; any variations observed were similar and good correlations were found.

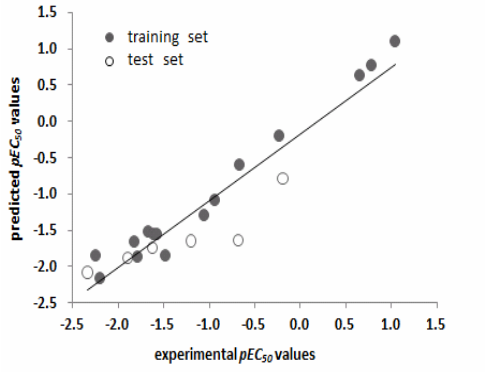

Fig. 1. Comparison between actual values and predicted values.

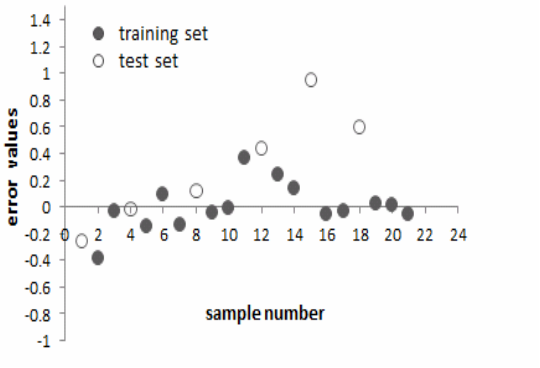

Fig. 2. Error of samples.

The values that are indicated in Fig. 2 fluctuate around the zero point, which indicates that the model achieves reliable prediction and can thus be used to predict the inhibitory toxicity pEC_{50} of aromatic compounds to *Chlorella.*

DISCUSSION

This study used PaDEL-Descriptor software to calculate the descriptors of 21 aromatic compounds; a 2DQSAR model considering the inhibitory activities of the compounds and theircompound descriptors was then built. The keyof QSAR studies is to determine how to produce structural descriptors that express the structural characteristics of various molecules [16]. Existing parameters can be divided into four categories: (1) the substituent parameter, or the linear free-energy relationship parameter, is the earliest and most commonly used structural description. It accumulates a large amount of available data and features a considerable number of successful examples. (2) With the rapid development of theoretical chemistry and computer technologies, scientistshave found that quantum chemistry, orbit energy, and geometrical features are well associated with bioactivity[17]. (3) Interaction parameters and three-dimensional molecular structure descriptors (3D descriptors). (4) Various topological parameters deduced from the 2D topology [18]. The 3DQSAR method generally well predicts relationships between the structure and properties of compounds. In case a 3D descriptor cannot be obtained for a compound, this compound should be excluded from the sample. 2D descriptors can easily beobtained as long as the molecular structure is known, and theycan serve as excellentmodels for 3D descriptors. This research used the SMILES format of the 21 aromatic compounds as input in PaDEL-Descriptor todetermine their structures. The parameters of the model in this study are ideal. We thus conclude that the 2Dmodel is stable and exhibits good ability for predicting the inhibitory effectof toxic aromatic

compounds on*Chlorella*.

The 21 bioactive aromatic compounds used in this study are toxic moleculeswith different substituents and various numbers of benzene rings. The two descriptors in the model are TpiPC, and MATS3v, otherwise known as autocorrelation and path count, respectively. The topology autocorrelation descriptor reveals the distribution of atomic properties in the molecular topology, whereas the path count reflects the number of carbon atoms and branching conditions in the molecule. Bothdescriptors are molecular topology parameters. H. Wiener[19] is the first chemist to introducethe topological index. Studies have shown that the Wiener index demonstrates good correlationsbetweenthe properties of hydrocarbon molecules, such as critical constant, viscosity, surface tension, and chromatographic retention time[20]. Our model demonstrates that molecular autocorrelation and path count are also significantly correlated with the toxicity of aromatic compounds.

Several studies onQSAR between aromatic compounds and algae have been conducted.Netzeva et al.[21] built a QSAR model that showed the inhibitory effect of 65 toxic aromatic compounds to *Chlorella*; these researchers usedMLR to build a QSAR model with two descriptors and achieve the equation Log (1/EC50) = 0.73LogKow −0.59Elumo−1.91; R^2=0.84 and SE=0.43. While the statistical results of this model are satisfactory,the regression coefficients are lower compared with that (R^2=0.974) of the model presented in this paper. Thus,TpiPC and MATS3v may be more suitable for predictthe inhibitory effect of toxicaromatic compounds on*Chlorella* than other descriptors.

CONCLUSIONS

In this paper, PaDEL-Descriptor software was used to determine the molecular structure of 21 aromatic compounds. Calculations were based on 15 randomly selected molecules constituting a training set and 6 molecules constituting a test set. An ideal model with two descriptors was built by using the stepwise regression method. The model showed good correlation and strong stability after determination of its R^2, Sig, VIF, and $R^2$by using the test set. This model can predict the inhibitory effect of toxic aromatic compounds on *C. vulgaris.*

Acknowledgements: *This work was funded by the National Marine Public Welfare Research Project (No.201305002, 201305043), the Natural Science Foundation of Dalian (No. 2012J21DW014), and the project of Marine Ecological Restoration Technology Research on the Penglai 19-3 oil spill accident(No. 201422).*

REFERENCES

1. V. Parikshit, S. R. Wate, S. Devotta, *Environ Monit Assess,***146**, 191 (2008).
2. J. Papadimitrakis, M. Psaltaki, M. Christolis, N. C. Markatos, *Environ Modell Softw*, **21**, 170 (2006).
3. E. Kriek, Aromatic amines and related compounds as carcinogenic hazards to man. In: Environmental carcinogenesis: Occurrence, risk, evalulation, mechanisms, Elsevier/North Holland Biomedical Press, Amsterdam New York Oxford, 1979.
4. J. L. Liu, P. Z. Lang, *Environmental Science*, **16,** 7 (1995).
5. P. R. Duchowicz, A. G. Mercader, F. M. Fernández, E. A. Castro, *ChemometrIntell Lab*, **90**, 97 (2008).
6. B. B. Xia, K. P. Liu, Z. G. Gong, et al., *EcotoxEnviton Safe*, **72**, 787 (2009).
7. O. Isayev, B. Rasulev, *Mol Divers*, 10, 233 (2006).
8. K. Mansouri, V. Consonni, M. K. Durjava, B. Kolar, T. Öberg, R. Todeschini, *Chemosphere*, **89**, 433 (2012).
9. E. B. de Melo*, Ecotox Environ Safe*, **75**, 213 (2012).
10. M. T. Saçan, M. Özkul, S. S. Erdem, *Chemosphere,***68**, 695 (2007).
11. A. Modelli, L. Mussoni, D. Fabbri, *J Phys Chem A,***110**, 6428 (2006).
12. C.W. Yap, ComputChem, **32**, 1466 (2011).
13. C. H. Wang, S. L. Yang, Y. Wu, *Computers and Applied Chemistry*, **29**, 297 (2012).
14. G. Moreau, P. Broto, *Nouv J Chim.,* **4**, 359 (1980).
15. Y. L. Gu, Y. M. Guo, B. Z. Li, *J Mol Sci-Int.,***19**, 155 (2003).
16. Y. J. Tseng, A. J. Hopfinger, E. X. Esposito, *J Comput Aid Mol Des*., **26**, 39 (2012).
17. A. R. Katritzky, D. C. Fara, R. O. Petrukhin, *Curr Top Med Chem*., **2**, 1333 (2002).
18. M. Karelson, V. S. Lobanov, A. R. Katritzky, *Chem Rev.*, **96**, 1027 (1996).
19. H. Wiener, *J Am Chem Soc.,***69**, 17 (1947).
20. Z. Mihalić, N. Trinajstić, *J ChemEduc,***69**, 701 (1992).
21. T. I. Netzeva, J. C. Dearden, R. Edwards, A. D. Worgan, M. T. Cronin, *J. Chem. Information Computer Sci.*, **44**, 258 (2004).

QSAR-ИЗСЛЕДВАНЕ НА ТОКСИЧНОСТТА НА АРОМАТНИ СЪЕДИНЕНИЯ СПРЯМО *Chlorella vulgaris*

Н. Ли[1, 2], Ф. Чен[1, 2], Д. Янг [1, 2], Й. Жоу[1, 2*]

[1] *Ключова лаборатория по възстановяване на морските биоресурси и хабитати в провинция Ляонинг,Университет в Далиан, Китай*

[2]*Ключова лаборатория по северни морски култури, Министерство на земеделието, Университет по океанология в Далиан, Китай*

Постъпила на 4 април, 2015 г.

(Резюме)

С увеличаването на търсенето на енергия по светапетролът е станал главния замърсител на океаните. Ароматните съединения, които са компонент на петрола са станали основна нарастваща заплаха за морските екосистеми. Съставянето на модел на количествена връзка структура-активност (QSAR) за предсказване на токсичността на неизвестни ароматни съединения може да послужи за предпазване от замърсяване. В настоящата работа са подбрани 21 ароматни съединения, от които петнадесет са за упражнение, а шест - за тестуване.Структурните параметри на съединенията са получени чрез множествена линейна регресия и е съставен 2-дескрипторен предсказващ модел. Тестовите съединения са използвани за определяне предсказващата способност на модела. Този модел, използващ предложения метод показва задоволителни статистически резултати ($R^2 = 0.974$ срещу тестовата стойност $R^2 = 0.804$).Тези данни показват, че моделът има добра предсказваща способност и затова може да се използва за предсказването на ефекта на инхибиране на ароматни съединения спрямо *Chlorella vulgaris*.