Journal of Environmental Sciences Vol.15, No.1, pp.88—91, 2003

ISSN 1001—0742
CN11—2629/X

Article ID: 1001-0742(2003)01-0088-04 **CLC number**: X131.2 **Document code**: A

QSBR study of substituted phenols and benzoic acids

LU Guang-hua[1], WANG Chao[1], YUAN Xing[2], ZHAO Yuan-hui[2]

(1. College of Environmental Science and Engineering, Hohai University, Nanjing 210098, China. E-mail: ghlucc@sina.com; 2. Department of Environmental Science, Northeast Normal University, Changchun 130024, China)

Abstract: The biodegradability of 30 substituted phenols and benzoic acids was determined by BOD technique. The molecular weight (M_W), heat of formation (H_f) and the energy of the highest occupied molecular orbital (E_{HOMO}) of the studied compounds were calculated by the quantum chemical method MOPAC6.0-AM1. The quantitative structure-biodegradability relationships (QSBRs) were developed by the linear regression method and neural network approach, respectively. It has been shown that the neural network method is able to provide a superior fit to the training set data and test set data and produce a lower prediction error than the linear regression method.

Keywords: BOD; regression; artificial neural network; prediction

Introduction

Most phenols and benzoic acids constitute a class of important environmental pollutants that are present in the Songhuajiang River of Jilin Province, China. The presence of many of these chemicals in the natural waters is a serious public health problem. Biodegradation is an important mechanism for removing them from the ecosystem. Biodegradation can eliminate hazardous chemicals by biotransforming them into innocuous forms, we can presume, *via* a series of enzymatic processes in microorganisms, degrading them by mineralization to carbon dioxide and water (Alexander, 1980).

Biochemical oxygen demand (BOD) is the relatively simple method of determination and the information obtained is very useful for most cases. So the parameter is used to determine the ultimate degradability, which is demanded in most of the standard tests (Pagga, 1997). However, gathering this information is labor intensive, time consuming, and expensive due to the large number of these chemicals. Therefore, it is necessary to develop correlation and predictive techniques in order to estimate biodegradability. The quantitative structure-biodegradability relationships (QSBRs) have been used to predict the fate of organic chemicals and analyze biodegradation mechanism (Dearden, 1996).

At present, the reports on QSBR studies of organic chemicals are not too much for it is not easy to gain biodegradation data at the same experimental conditions. In this paper, the biodegradability of 30 organic pollutants was determined by taking the bacteria in the Songhuajiang River as inoculums, and QSBR models were developed using the structure parameters of chemicals as descriptors by the linear regression method and artificial neural network approach.

1 Materials and methods

Water samples were gathered from Jilin section in the Songhuajiang River. Temperature of water sample: 15—20℃; pH 6.8—7.0; dissolved oxygen(DO): about 8.0 mg/L. The bacteria counts were determined by standard plate count techniques (Wang, 1988), and are about 800—3000/ml. Biodegradation of substituted phenols and benzoic acids were determined using the BOD technique(Du, 1994). The initial test concentration of chemicals was planned at 2 mg/L on the basis of their toxicity to the river bacteria (Yuan, 1997), theoretical oxygen demand (ThOD), and residual DO of at least 1 mg/L at the final day (Dinesh, 1987).

The test chemical was added to 250 ml BOD bottles. The bottles were then filled to capacity with the

Foundation item: The National Natural Science Foundation of China (No. 29877004) and the Science Innovation Foundation of Hohai University (No. 1043/401100)

river water sample, sealed, and incubated for 5d at 20 ± 1℃. There were two replicates for each chemical and each control(only inoculums), respectively. The DO concentrations were determined by the iodometric titration method(Du, 1994). The test result was expressed as BOD% by comparing the measured BOD_5 with ThOD, which is calculated from the molecular formula of the test compound.

2 Calculation methods

The molecular weight (M_w), the heat of formation (H_f) and the energy of the highest occupied molecular orbital (E_{HOMO}) of 30 phenols and benzoic acids were calculated by the quantum chemical method MOPAC6.0-AM1 on energy-minimized structures. This method can automatically optimize the bond length, the bond angle and the twist angle, and yield a lot of structural information. The parameter values of studied compounds are listed in Table 1.

3 Results and discussion

The experimental biodegradability data of 30 compounds are listed in Table 1. The results show that phenol, diphenols, benzoic acids, phthalic acid and isophthalic acid belong to readily biodegradable chemical due to their higher BOD%, these compounds under low concentration produce less effect to water environment; partial methyl, methoxy substituted phenols and benzoic acids belong to biodegradable chemical; while chloro-, amino- and nitro-substituents belong to non-readily biodegradable chemical in natural river water, so they are prior monitored and controlled pollutants because of their more harm to environment.

Guo and Xu(Guo, 1998) developed the correlation between narcosis toxicity and molecular structure using n-octanol/water partition coefficient $\log P$, $\log^2 P$ and Van Der Waal volume V, V^2, and topogical index Am_3 by the multivariance regression analysis and artificial neural network(ANN) method. The root-mean-square (RMS) of the regression analysis method is 0.159; of ANN is 0.0885 for the training set; while their RMS are 0.207 and 0.131 for the test set.

Zhang *et al*. (Zhang, 1998) developed QSBR models on the first order biodegradation rates of 26 aromatic compounds by the regression analysis and ANN too. It has shown that the RMS predicted by ANN (0.00102)is much lower than that by the linear regression method(0.01591).

On the basis of above work, QSBR models were developed for the biodegradability of 30 organic chemicals using the structure parameters of chemicals as descriptors by the linear regression method and ANN approach. 30 compounds are randomly divided into two sets. Of these, 25 compounds are included in the training set and 5 compounds in the test set.

3.1 Linear regression method

The linear regression analysis was performed using the SPSS statistical package. By the multiple linear regression analysis, one equation is developed using the BOD% values and parameters listed in Table1.

$$BOD\% = 39.50 - 0.67\,M_w - 0.08H_f - 8.88E_{HOMO},$$
$$R^2 = 0.752,\ S.E. = 9.70,\ F = 21.22,\ P = 0.000,\ n = 25. \quad (1)$$

Where, R^2 is the square of the correlation coefficient; $S.E.$ is the standard error; F is the test value; P is the significant level; and n is the number of compounds.

The obtained QSBR model shows that the biodegradation of studied compounds is related mainly to steric parameter M_w and electronic parameter E_{HOMO} and H_f. M_w is the molecular weight, and can reflect the size of a molecular. The negative correlation between BOD% and M_w suggests that the smaller the M_w, the more easily the compound is biodegraded. E_{HOMO} is the energy of the highest occupied molecular orbital, and is related to ionization potential (Liu, 1991). The heat of formation (H_f) of a compound is a measure of its stability. The result in Eq.(1) shows that the biodegradability of studied compounds also

appears to be negative correlation with E_{HOMO} and H_f.

Table 1 The experimental and predicted BOD(%) and structure parameters of 30 compounds

Compound	$-E_{HOMO}$, eV	$-H_f$, kJ/mol	M_w	BOD%				
				Exp.	Pre.1[a]	Er%.1[b]	Pre.2[c]	Er%.2[b]
Training set								
Phenol	9.12	93.89	94.11	69.8	64.8	7.16	77.8	11.6
Catechol	8.88	277.48	110.11	65.1	66.8	2.61	71.5	9.22
3-chlorophenol	9.34	120.50	128.56	31.7	45.9	44.8	42.1	32.8
4-chlorophenol	9.13	122.59	128.56	44.0	44.1	0.23	37.2	13.4
3-methylphenol	9.02	124.85	108.14	63.4	50.0	21.1	70.2	10.7
4-methylphenol	8.88	124.64	108.14	69.2	55.8	19.4	68.4	1.16
2-aminophenol	8.36	97.45	109.13	30.6	48.3	57.8	33.6	9.80
2-nitrophenol	9.93	66.07	139.11	36.0	39.8	9.55	34.6	3.89
4-nitrophenol	10.77	81.80	139.11	50.3	48.3	3.98	55.3	9.94
4-methoxyphenol	8.64	240.32	124.14	45.9	52.8	15.0	55.4	20.7
2,4-dichlorophenol	9.23	138.66	163.00	19.1	23.2	21.5	13.4	29.8
2,4,6-trichlorophenol	9.39	160.29	197.45	0.00	3.20	–	0.05	–
Benzoic acid	10.08	284.34	122.12	67.3	69.9	3.86	64.5	4.16
2-hydroxylbenzoic acid	9.47	470.11	138.12	64.9	68.8	6.01	67.3	3.70
4-hydroxylbenzoic acid	9.61	472.21	138.12	63.0	70.2	11.4	65.5	3.97
2-aminobenzoic acid	8.78	296.02	137.14	62.7	49.2	21.5	59.0	5.90
4-aminobenzoic acid	8.91	297.27	137.14	61.6	50.5	18.0	59.0	4.22
2-nitrobenzoic acid	10.91	240.66	167.12	41.9	43.6	4.06	43.0	2.63
4-nitrobenzoic acid	10.90	259.41	167.12	50.2	45.0	10.4	45.1	10.2
Phthalic acid	10.47	631.24	166.13	75.4	71.8	4.77	66.4	11.9
Isophthalic acid	10.52	656.09	166.13	60.7	74.3	22.4	66.8	10.0
3-chlorobenzoic acid	9.94	310.66	156.60	45.3	47.6	5.08	50.8	12.1
4-chlorobenzoic acid	10.02	312.17	156.60	64.6	48.5	24.9	50.6	21.7
4-methoxybenzoic acid	9.48	446.14	152.15	71.5	57.5	19.6	67.1	6.15
3-methylbenzoic acid	9.75	316.94	136.15	52.9	60.2	13.2	54.3	2.65
Test set								
Resorcinol	9.05	279.49	110.11	68.6	68.5	0.15	71.0	3.50
2-methylphenol	9.00	123.18	108.14	77.7	56.7	37.4	70.7	9.00
3-aminobenzoic acid	8.88	289.07	137.14	56.2	49.6	11.7	57.7	2.67
2-chlorobenzoic acid	9.91	288.86	156.60	55.3	45.6	17.5	48.4	12.5
2-methoxybenzoic acid	9.68	422.08	152.15	55.8	57.3	2.67	62.7	12.4

Notes: a. Pre.1 are calculated from (1); b. percentage error is defined as the absolute difference between the experimental and predicted values for BOD divided by the experimental values of BOD; c. pre.2 are calculated from ANN method

3.2 Artificial neural network method (ANN)

There is extensive literature on mathematical models of artificial neural networks. Neural networks have found application in image and speech recognition, process control, and optimization of complex functions (Bhagat, 1990). At present, the artificial neural networks have been used in QSBR studies, and showed a good prediction function (Tabak, 1993).

In our study, a three-layer neural network was used with three input nodes(M_w, H_f and E_{HOMO}), six hidden layer nodes and one output node (BOD%). The neural network was trained by using a backpropagation algorithm, which used a gradient search technique to minimize a cost function equal to the mean square difference between the desired and actual net outputs.

The predicted values and percentage errors of the regression analysis method and neural network approach are presented in Table 1. The results show that the percentage errors between the experimental and predicted values of BOD% are generally lower by ANN method(the average is 10.5% for the training

set and 8.0% for the test set), when compared with the linear regression method (15.3% and 13.9%, respectively). The errors by the linear method are significant for compounds such as 3-chlorophenol (44.8%), 2-aminophenol(57.8%), and 2-methylphenol(37.4%). However, the percentage errors for ANN predictions remain below 35% for the training set and 13% for the test set.

In order to compare further the prediction effect of linear and nonlinear models, the plot of the predicted values from Eq. (1) *versus* the experimental values is shown in Fig. 1 and the plot of the predicted values from ANN method *versus* the experimental values is shown in Fig.2. It can be seen from Fig.1 and 2 that the predicted values by ANN method fit better than by the linear analysis method.

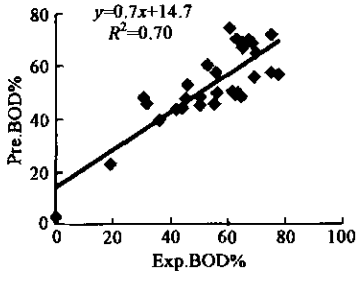

Fig.1 Plot of pre. BOD% by Eq.(1) *vs* exp. BOD%

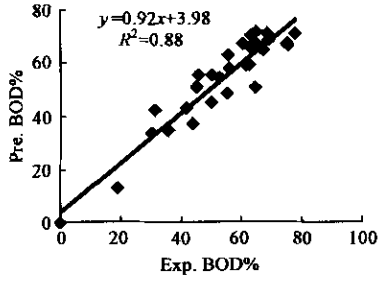

Fig.2 Plot of pre. BOD% by ANN *vs* exp. BOD%

4 Conclusion

The biodegradability of 30 substituted phenols and benzoic acids in the Songhuajiang River water was determined. Of which, phenol, diphenols, benzoic acids, phthalic acid and isophthalic acid belong to readily biodegradable chemical due to their higher BOD%; partial methyl, methoxy substituted phenols and benzoic acids belong to biodegradable chemical, while most chloro-, amino- and notro-substituents belong to non-readily biodegradable chemical in natural river water. QSBR models were developed by the linear regression method and the neural network method and used for predicting the biodegradability of substituted benzenes. It was shown that the neural networks method is able to predict the biodegradability more accurately than the linear regression approach whether for the training set or for the test set.

References:

Alexander M, 1980. Biodegradation of chemicals of environmental concern[J]. Science, 211: 132—138.

Bhagat P, 1990. An introduction to neural nets[J]. Chem Eng Prog, 86: 55—60.

Dearden J C, Cronin M T D, 1996. Biodegradation prediction[M] (Peijnenburg W. J. G. M. and Damborsky J. eds.). Kluwer: Academic Press. 93—105.

Dinesh D V, Robert S B, Leo Babeu, 1987. Quantitative structure-biodegradability relationships for alcohols, ketones and alicyclic compounds [J]. Chemosphere, 16(4): 695—703.

Du C J, Chu D S, 1994. Water quality analysis basis in environmental monitoring[M]. Harbin: Heilongjiang Science and Technology Press. 303—308.

Guo M, Xu L, 1998. Quantitative structure-activity relationship for toxicity of phenols using regression analysis and neural network[J]. Acta Scientiae Circumstantiae, 18(21): 122—126.

Liu C Q, 1991. Quantum biology and application[M]. Beijing: Higher Education Press. 16.

Pagga U, 1997. Testing biodegradability with standardized methods[J]. Chemosphere, 35 (12): 2953—2972.

Tabak H H, Govind R, 1993. Prediction of biodegradation kinetics using a nonlinear group contribution method[J]. Environ Toxicol Chem, 12: 251—260.

Wang J L, 1988. Environmental microbiology[M]. Beijing: Higher Education Press. 12.

Yuan X, Lu G, Lang P, 1997. QSAR study of the toxicity of nitrobenzenes to river bacteria and photobacterium phosphoreum[J]. Bull Environ Contam Toxicol, 58: 123—127.

Zhang A Q, Han S K, Shen Z, 1998. Prediction of aerobic biodegradation rates of aromatic compounds contained sulfur by regression analysis and artificial neural network method[J]. Environmental Science, 19(1): 37—40.

(Received for review September 12, 2001. Accepted February 18, 2002)