Quantitative Structure-Toxicity Relationship for Predicting Acute Toxicity of Phenols

Zhi-Xiang ZHOU[1, a, *], Meng-Nan QIN[1], Yang-Hua LIU[1], Xiao-Long ZHANG[1], Han-Dong LI[2]

[1]College of Life science and Bioengineering, Beijing University of Technology China

[2]Chinese Research Academy of Environmental Sciences, China

[a]email: zhouzhixiang@bjut.edu.cn

*Corresponding author

Keywords: Phenols, QSTR, Acute toxicity.

Abstract. Phenols represent one of the most important classes of environmental chemicals. Most of them may cause serious public health and environmental problems. The present work is to develop an effective QSTR model for acute toxicity, a toxicological endpoint of Phenols. We calculated various descriptors and used linear regression way to select relevant parameters, and built a QSTR model. The model showed a good forecasting ability. Based on the descriptors, a further discussion was presented for the toxic mechanism.

Introduction

Phenols are important materials or intermediates of explosives, pesticides, organic synthesis, and dyestuffs etc. With the development of industry, thousands of these compounds have being introduced into the environment every year. Most of these chemicals exhibit toxicity, which may cause serious public health problems [1]. With the development of computer technology and quantum chemistry, methods based on quantitative structure-toxicity relationship (QSTR) have been an increasing role in environmental hazard assessment [2]. With this method, we can use parameters named descriptors in QSTR software package. Descriptors can be classified into different types such as topochemical, geometric, constitution, and electron descriptors etc [3]. The values of these descriptors can reflect the structure information of chemicals and help us to have a better insight into the action of mechanism [4]. If the experimental data and the value of descriptors are linearly related to toxicity for a set of molecules, it means the model is successfully established. Once the descriptors selected in QSTR model are related to acute toxicity of the chemical, we can analysis which descriptors influenced the toxicity. If the toxicity of a chemical is unknown, we can calculate its parameters and give it a predicted value for the hazard assessment.

LD_{50} (50% lethal dose concentration) is the toxicological endpoint which has great effect on human health. Therefore, the goal of this study is to use the multiple linear regressions (MLR) technique to develop QSTR models to predict the LD_{50} of Phenols, based on the most comprehensive data collection available from databases and literature.

Material and Data

Descriptors Calculation

The 2D structures of the compounds obtained from the EPI (Estimation Programs Interface) Suite™ were optimized based on the AM1 semi empirical method. The descriptors of phenols were calculated by Projectleader of Scigress 7.7 and Dragon software to obtain their LD_{50}.

Model Building

After the calculation of the molecular descriptors, multiple linear regression (MLR) was carried out to select the most relevant descriptors from the pool of calculated descriptors, and the SPSS program package was used to analyze data and select descriptors, and establish the linear relationship between structure and toxicity in the MLR way at the confidence level of 95%. When the regression was completed, SPSS showed a form filling with the regression coefficient (R^2), standard error (s), and Fisher statistic value (F). The best model can be selected with consideration of these values.

Model Validation

External validation has been considered more reliable for judging the prediction potential of QSAR models than internal validation techniques. For extreme cases, appropriate external datasets are not available for prediction purposes. The (external) predictive capacity of a given model was judged by its application for prediction of the test set toxicity values, and the model's stability was established by a cross-validated regression coefficient (Q^2). The closer to 1 of the Q^2 value, the more stable the model was.

Dataset the experimental values of mouse LD_{50} were collected from the ChemIDplus database (http://chem.sis.nlm.nih.gov/chemidplus/chemidheavy.jsp), and are shown in Table 1.

Table 1 Experimental $LogLD_{50}$ value and predictive $LogLD_{50}$ value of phenols

No.	CAS. No.	Experimental	Predictive	Reside
1	100-02-7	2.45	2.79	-0.34
2	106-41-2	2.72	2.89	-0.17
3	106-44-5	2.54	2.72	-0.18
4	106-48-9	2.56	2.84	-0.28
5	108-39-4	2.92	2.66	0.26
6	108-43-0	2.72	2.86	-0.14
7	108-73-6	3.66	3.29	0.37
8	108-95-2	2.43	1.95	0.48
9	119-34-6	3.17	2.95	0.22
10	119-36-8	3.05	3.43	-0.38
11	120-80-9	2.41	2.12	0.29
12	120-83-2	3.11	3.10	0.01
13	121-33-5	3.59	3.62	-0.03
14	123-31-9	2.39	2.35	0.04
15	14008-60-7	2.88	2.94	-0.06
16	150-19-6	2.49	2.97	-0.48
17	1570-64-5	3.12	2.98	0.14
18	17788-00-0	2.9	2.86	0.04
19	2144-08-3	3.34	3.04	0.30
20	2295-58-1	3.44	3.20	0.24
21	487-70-7	3.51	3.49	0.02
22	490-79-9	2.6	2.78	-0.18
23	4901-51-3	3.65	3.39	0.26
24	498-02-2	3.95	3.52	0.43
25	50-85-1	3.26	3.36	-0.10
26	51-28-5	1.65	1.65	0.00
27	527-60-6	4	3.30	0.70
28	534-52-1	1.32	1.41	-0.09

No.	CAS. No.	Experimental	Predictive	Reside	No.	CAS. No.	Experimental	Predictive	Reside
29	552-41-0	2.69	3.05	-0.36	60	95-55-6	2.9	2.60	0.30
30	554-84-7	3.03	2.86	0.17	61	95-56-7	2.81	2.69	0.12
31	567-61-3	2.4	2.64	-0.24	62	95-57-8	2.54	2.72	-0.18
32	576-26-1	2.65	2.77	-0.12	63	95-65-8	2.6	2.82	-0.22
33	583-78-8	2.98	2.56	0.42	64	95-77-2	3.23	3.24	-0.01
34	59-50-7	2.78	3.13	-0.35	65	95-85-2	3.01	3.12	-0.11
35	591-27-5	2.6	2.81	-0.21	66	95-95-4	2.78	2.68	0.10
36	591-35-5	3.38	3.29	0.09	67	97-51-8	2.83	2.74	0.09
37	609-99-4	2.43	2.60	-0.17	68	99-06-9	3.3	3.10	0.20
38	618-45-1	3.21	2.72	0.49	69	99-24-1	3.23	3.25	-0.02
39	626-02-8	3.46	3.32	0.14	70	99-57-0	2.93	3.12	-0.19
40	65-45-2	2.48	2.86	-0.38	71	99-76-3	3.9	3.82	0.08
41	65-49-6	3.6	3.56	0.04	72	99-89-8	2.94	3.30	-0.36
42	654-42-2	2.27	2.50	-0.23	73	99-93-4	3.18	3.42	-0.24
43	69-72-7	2.68	3.07	-0.39	74	99-96-7	3.34	3.43	-0.09
44	7693-52-9	3.4	3.25	0.15	75	105-67-9*	2.91	2.83	0.08
45	831-61-8	3.76	3.62	0.14	76	108-46-3*	2.3	2.64	-0.34
46	87-64-9	2.85	2.78	0.07	77	108-68-9*	2.68	3.01	-0.33
47	87-65-0	3.33	3.21	0.12	78	118-95-6*	1.65	2.75	-1.10
48	87-66-1	2.48	2.76	-0.28	79	1198-55-6*	2.5	0.60	1.90
49	87-86-5	1.56	1.85	-0.29	80	504-15-4*	2.89	2.91	-0.02
50	88-04-0	3	3.17	-0.17	81	576-24-9*	3.38	3.16	0.22
51	88-75-5	3.11	2.73	0.38	82	58-90-2*	2.12	2.86	-0.74
52	89-56-5	3	3.15	-0.15	83	615-58-7*	2.45	3.10	-0.65
53	89-57-6	3.53	3.33	0.20	84	618-83-7*	3.84	3.04	0.80
54	89-83-8	2.81	2.91	-0.10	85	619-19-2*	2.81	3.26	-0.45
55	90-00-6	2.78	2.45	0.33	86	767-00-0*	2.65	2.44	0.21
56	90-05-1	2.79	2.70	0.09	87	83-40-9*	3	3.06	-0.06
57	91-10-1	3.4	3.61	-0.21	88	95-48-7*	2.54	2.43	0.11
58	935-95-5	2.04	2.28	-0.24	89	95-71-6*	2.6	2.40	0.20
59	95-01-2	3.14	2.89	0.25	90	95-87-4*	2.58	2.46	0.12

“*” is test set.

Results and Discussion

Establishment of QSTR. Different descriptors were calculated for the model development with SPSS package. The inter correlation of descriptors were taken into consideration and used for dependent variables. The Log LD_{50} value was served as independent value from Table1.This model was established by MLR method and shown as follows:

$$\begin{aligned}\mathrm{LogLD_{50}} &= 0.199\mathrm{Mor24s}+0.012\mathrm{RDF040s}+0.281\mathrm{SM06\_AEA(dm)}-0.495\mathrm{F04[O\text{-}Cl]}\\ &\quad -0.528\mathrm{MATS5p}-0.974\mathrm{VE3\_RG}-0.126\mathrm{CATS2D\_03\_DL}\\ &\quad +0.165\mathrm{RDF065e}+0.264\mathrm{Mor16s}+1.759\mathrm{IVDE}+0.86\\ R^2&=0.763 \quad Q^2=0.59 \quad s=0.271 \quad F=20.289 \quad N_{train}=74 \quad N_{test}=16\end{aligned} \tag{1}$$

Where R^2 is the square of correlation coefficient, Q^2 is a cross-validated regression coefficient, s is the standard error, F is the mean square radio, and N is the number of compounds.

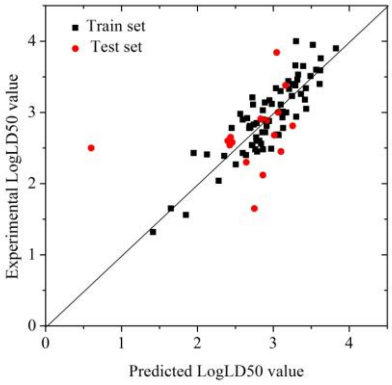

Figure 1. Linear analysis for LD_{50} QSTR model of Phenols

The MLR prediction figure (Fig. 1) showed the linear relationship between the predicted and experimental values were good. The predicted value of LD_{50} and residue can be calculated by this model, the result shows that the residues of the training and test sets are low (Table 1). The correlation coefficient of R^2 is 0.763 and the cross-validated regression coefficient of Q^2 is 0.59. This signifies that the model has a strong predicting ability and high stability.

Table 2. Definition of descriptors in LD_{50} model of Phenols

Descriptors	Definition
Mor24s	signal 24 / weighted by I-state
RDF040s	Radial Distribution Function - 040 / weighted by I-state
SM06_AEA(dm)	spectral moment of order 6 from augmented edge adjacency mat. weighted by dipole moment
F04[O-Cl}	Frequency of O - Cl at topological distance 4
MATS5p	Moran autocorrelation of lag 5 weighted by polarizability
VE3_RG	logarithmic coefficient sum of the last eigenvector from reciprocal squared geometrical matrix
CATS2D_03_DL	CATS2D Donor-Lipophilic at lag 03
RDF065e	Radial Distribution Function - 065 / weighted by Sanderson electronegativity
Mor16s	signal 16 / weighted by I-state
L3s	3rd component size directional WHIM index / weighted by I-state

The definitions of the 10 descriptors are shown in Table 2. From Eq. (1), we can see that the *F04[O-Cl}*, *MATS5p*, *VE3_RG* and *CATS2D_03_DLC-025* were positively correlated with the acute toxicity. As these descriptor parameter values increase, the LogLD50 value decreases, and the acute toxicity of chemicals increases. The other descriptors are negatively correlated with the acute toxicity, wherein, *Mor24s* was signal 24 / weighted by I-state, *Mor16s* associated signal 16 / weighted by I-state. This signifies that the reducing polarizability of the molecules tend to decrease the toxicity of the chemical.

Summary

With MLR analysis, QSTR model for LD50 of phenols was successfully built with the descriptors from AM1 and E-dragon software. The QSTR model showed high stability and excellent predicting properties.

Acknowledgements

This study was supported by the Ministry of environmental protection of China (Grant No. 2011467037).

References

[1] Oberg, T.G. Prediction of vapor pressures for halogenated diphenylether congeners from molecular descriptors. Environ. Sci. Pollut. Res. 2002, 9, 405–411.

[2] Leong MK, Lin SW, Chen HB, Tsai FY. Predicting acute toxicity of aromatic amines by various machine learning approaches. Toxicol Sci.Vol. 116 (2010),p:498-513.

[3] Simon-Hettich, B.; Rothfuss, A.; Steger-Hartmann, T. Use of computer-assisted prediction of toxic effects of chemical substances. Toxicology 2006, 224(1-2), 156-162.

[4] Patlewicz, G.; Rodford, R.; Walker, J.D. Quantitative structure-activity relationships for predicting carcinogenicity and carcinogenicity. Environ. Toxicol. Chem. 2003, 22(8), 1885-1893