Asian Journal of Chemistry; Vol. 26, No. 9 (2014), 2707-2710

ASIAN JOURNAL OF CHEMISTRY

http://dx.doi.org/10.14233/ajchem.2014.16017

QSTR Studies on Acute Toxicity and Mutagenicity of Halogenated Benzenes

XIAO-LONG ZHANG[1], ZHI-XIANG ZHOU[1,*], XUE-LAN FAN[2], YANG-HUA LIU[1], HAN-DONG LI[3] and JIAN-TAO WANG[2]

[1]College of Life science and Bioengineering, Beijing University of Technology, Beijing 100124, P.R. China
[2]Institute of Physics, Chinese Academy of Sciences, Beijing 100190, P.R. China
[3]Chinese Research Academy of Environmental Sciences, Beijing 100012, P.R. China

*Corresponding author: Tel/Fax: +86 10 67392780; E-mail: zhouzhixiang@bjut.edu.cn

Received: 8 July 2013; *Accepted*: 19 December 2013; *Published online*: 28 April 2014; AJC-15102

In this paper, a quantitative structure toxicity relationship study was performed for the prediction of the acute toxicity and mutagenicity of halogenated benzenes. The molecular descriptors of halogenated benzenes have been calculated with semi-empirical AM1 and E-dragon methods and both quantitative structure toxicity relationship models for mice *via* the oral LD_{50} model and the *S. typhimurium* (TA98 + S9) mutagenic model of halogenated benzenes were developed using multiple linear regression (MLR) analysis. The validation results obtained from the test set indicate that the proposed models are robust and satisfactory.

Keywords: Halogenated benzenes, Quantitative structure toxicity relationship, Multiple linear regression, Mutagenicity.

INTRODUCTION

Halogenated benzenes are widely used as solvents, herbicides, antiseptics and pesticides. Due to their ubiquitous roles, persistence and bio-accumulative potential, most of these chemicals exhibit toxicity, which may cause serious public health and environmental problems[1]. With increasing concern toward human health and environmental pollution, governments and regulatory agencies from around the world currently seek to assess the toxicological risks posed by existing chemicals and those which will be released in the future[2]. Although the toxic potential of chemicals is often required to be assessed by using standard animal models, such testing is demanding of both time and resources and thus is not deemed suitable for screening of large numbers of potential toxicants[3]. It is necessary to predict the toxicity of the compounds accurately and speedily, as increasing numbers of compounds are found. In recent years, more and more researchers have focused on the prediction of toxicity with a tool known as a quantitative structure toxicity relationship (QSTR). A QSTR is a model which describes the mathematical relationship between a property of the chemical, in this case toxicity and one or more descriptors of the chemical[4]. The descriptors are chemical and physical characteristics which are obtained from the structure of the chemical and relate the chemical structure to its toxigenicity mechanism. Therefore QSTR provides a better investigation to toxigenicity through efficient computational, rather than biological activity using *in vivo* approaches[5].

Recently, various investigations of QSTR of the aquatic toxicity for the series of halogenated benzenes have been published[6-9]. However, QSTR of halogenated benzenes in mutagenicity and mice *via* oral LD_{50} has rarely been reported. Moreover, LD_{50} (50 % lethal dose concentration) and mutagenicity are toxicological endpoints which have great effect on cancers and tumors. Therefore, the goal of this study is to use the multiple linear regression (MLR) technique to develop QSTR models to predict the LD_{50} and mutagenicity of halogenated benzenes, based on the most comprehensive data collection available from databases and literature.

EXPERIMENTAL

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. The mutagenic activity is described as log R, where R is the number of revertants/nmol. The set of mutagenicity experiment data containing 24 halogenated benzenes used in this study was taken from the work of Basak *et al.*[10]. and shown in Table-3. The datasets were randomly divided into a training set and a test set for MLR analysis.

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 halogenated benzenes were calculated by Projectleader of Scigress 7.7 and Dragon software to obtain their LD_{50} and mutagenicity model constructions, respectively.

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[11,12]. 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.

RESULTS AND DISCUSSION

Mouse LD_{50} model: LD_{50} is a common factor for evaluating compound toxicity. The experimental data for the mouse oral LD_{50} model were randomly divided into two subsets, namely a training set of 48 compounds and a test set including the remaining[12]. Using the LD_{50} value as a dependent variable and the chemical descriptors as independent variables, the QSTR model was established with the SPSS method, as follows:

$$\log LD_{50} = -0.496DISPe + 0.31L2s - 0.622H\text{-}3 + 1.548M04p - 0.301H3m + 0.422F04[N\text{-}Cl] - 0.557B01[C\text{-}N] - 4.35 \quad (1)$$

or

$R^2 = 0.779$ $Q^2 = 0.65$ s = 0.224 F = 20.185 $N_{train} = 48$ $N_{test} = 12$

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. 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 0.779 and the cross-validated regression coefficient of Q^2 is 0.65. This signifies that the model has a strong predicting ability and high stability.

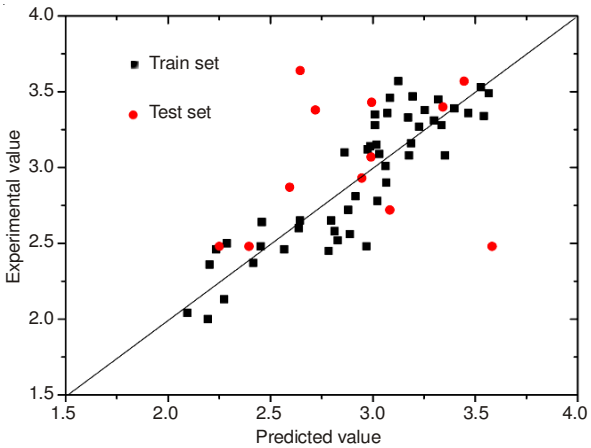

Fig. 1. Linear diagram of experimental and predictive value of log LD_{50} for halogenated benzenes model

TABLE-1
EXPERIMENTAL AND PREDICTED log LD_{50} VALUE OF HALOGENATED BENZENES

CAS number	Experimental value	Predicted value	Reside
100-00-5	2.64	2.46	0.18
104-88-1	3.15	3.02	0.13
104-92-7	3.34	3.54	-0.20
10403-47-1	2.46	2.57	-0.11
106-37-6	3.49	3.57	-0.08
106-40-1	2.46	2.24	0.22
106-41-2	2.72	2.88	-0.16
106-43-4	3.28	3.01	0.27
106-46-7	3.47	3.19	0.28
106-47-8	2	2.19	-0.19
106-48-9	2.56	2.89	-0.33
108-36-1	3.35	3.01	0.34
108-42-9	2.52	2.83	-0.31
108-70-3	3.53	3.53	0.00
108-90-7	3.36	3.07	0.29
1122-91-4	3.09	3.03	0.06
1194-65-6	3.31	3.30	0.01
1198-55-6	2.5	2.29	0.21
120-82-1	2.48	2.97	-0.49
121-73-3	2.58	2.81	-0.23
121-87-9	3.1	2.86	0.24
133-91-5	2.65	2.80	-0.15
137-19-9	3.57	3.12	0.45
147-82-0	2.65	2.64	0.01
1570-64-5	3.12	2.98	0.14
1689-83-4	2.36	2.20	0.16
1689-84-5	2.04	2.09	-0.05
1777-82-8	3.36	3.47	-0.11
17788-00-0	2.9	3.07	-0.17
1918-00-9	3.08	3.18	-0.10
554-00-7	2.6	2.64	-0.04
578-57-4	3.39	3.40	-0.01
591-18-4	3.16	3.19	-0.03
615-58-7	2.45	2.78	-0.33
623-00-7	2.37	2.42	-0.05
623-03-0	2.48	2.45	0.03
626-02-8	3.46	3.08	0.38
87-65-0	3.33	3.17	0.16
88-73-3	2.13	2.27	-0.14
89-61-2	3.45	3.32	0.13
89-98-5	3.28	3.34	-0.06
95-46-5	3.27	3.23	0.04
95-56-7	2.81	2.92	-0.11
95-73-8	3.38	3.25	0.13
95-94-3	3.01	3.06	-0.05
95-95-4	2.78	3.02	-0.24
99-54-7	3.14	2.99	0.15
99-91-2	3.08	3.35	-0.27
108-43-0*	2.72	3.08	-0.36
108-86-1*	3.43	3.00	0.43
1194-02-1*	2.48	2.40	0.08
133-90-4*	3.57	3.45	0.12
2042-37-7*	2.48	2.25	0.23
2362-12-1*	2.93	2.95	-0.02
576-24-9*	3.38	2.72	0.66
634-93-5*	3.07	2.99	0.08
874-42-0*	2.48	3.58	-1.10
95-49-8*	3.4	3.34	0.06
95-50-1*	3.64	2.65	0.99
95-76-1*	2.87	2.59	0.28

"*" is test set

TABLE-2
DEFINITION OF DESCRIPTORS IN LD_{50} MODEL OF HALOGENATED BENZENES

Descriptors	Definition	Block
DISPe	Displacement value/weighted by Sanderson electronegativity	Geometrical descriptor
L2s	2nd component size directional WHIM index/weighted by I-state	WHIM descriptor
H-3	the 3rd highest orbital energy	AMI descriptor
Mor04p	signal 04/weighted by polarizability	3D-MoRSE descriptor
H3m	H autocorrelation of lag 3/weighted by mass	GETAWAY descriptor
F04[N-Cl]	Frequency of N-Cl at topological distance 4	2D Atom Pairs
B01[C-N]	Presence/absence of C-N at topological distance 1	2D Atom Pairs

The definitions of the seven descriptors are shown in Table-2. From eqn. (1), we can see that the DISPe, H-3, H3m and B01[C-N] were positively correlated with the acute toxicity. Among these descriptors, DISPe is the replacement price and the electrical and molecular negative connection; H-3 is the third-highest molecular orbital energy; H3m represents the first three autocorrelation functions H; and B01 [C-N] is the topology of the CN bond distance 1 number. As these descriptor parameter values increase, the log LD_{50} value decreases and the acute toxicity of chemicals increases. The other three descriptors, *i.e*. L2s, Mor04p and F04[N-Cl], are negatively correlated with the acute toxicity, wherein, L2s was the second package size WHIM index, Mor04p associated polarizability of the molecule and F04 [N-Cl] was the Topology distance 4 the number of N-Cl bond. This signifies that the reducing stability of the molecule's structure tends to increase the toxicity of the chemical.

Mutagenic model of *S. typhimurium*: Mutagenicity is a toxicological endpoint which plays a great role in the appearance and development of cancers and tumors[13]. The Ames test is a simple and cost efficient method for determining the mutagenic potential of the compounds[14]. Previous studies have found that *S. typhimurium* TA98 undergoes mutations in the presence of halogenated benzenes[10]. The toxicity data from the Ames test of mutagenicity for the QSAR model building are presented in Table-3. Finally, QSAR was established by multiple linear regression analysis and shown as follows:

$$\log R = 0.137H2s - 15.476GATS2m + 16.048Sp\ Min1\_Bh(m) + 23.472SpMAD\_EA(ri) - 57.944 \quad (2)$$

$R^2 = 0.914\ Q^2 = 0.78\ s = 0.679\ F = 40.013\ N_{train} = 20\ N_{test} = 4$

In this model, log R is related to four Dragon descriptors, as shown in eqn. (2). The MLR model has a cross-validated correlation coefficient Q^2 value of 0.78, which illustrates the stability of the model by focusing on the sensitivity of the model to the elimination of any single data point. In general, the larger the magnitude of the F ratio is, the more accurately the model predicts the property values in the train set. The large F ratio of 40.013 indicates that the MLR model provides an excellent prediction of the mutagenicity of the chemicals. Fig. 2 shows that the data of the test set are close to or inside the fitline. This suggests that the experimental value has a linear relationship with the predicted value.

The definitions of the four descriptors are shown in Table-4. According to eqn. (2), the descriptors entered into the model were mainly the topology parameters (Table-4). The descriptor of GATS2m, a Geary autocorrelation function, is positively correlated with mutagenicity, *i.e*. an increased GATS2m leads to a decrease of the mutagenicity of the chemical. The other

TABLE-3
EXPERIMENTAL AND PREDICTED log R VALUE OF HALOGENATED BENZENES

CAS number	Experimental value	Predicted value	Reside
70-34-8	1.2	0.47	0.73
88-73-3	-1.72	-1.06	-0.66
89-61-2	-1.54	-2.14	0.60
89-63-4	-2	-1.68	-0.32
91-94-1	0.81	0.43	0.38
95-51-2	-3	-3.37	0.37
95-85-2	-3	-2.99	-0.01
97-00-7	0.3	0.63	-0.33
350-30-1	-1.21	-0.26	-0.95
364-74-9	-0.79	-0.82	0.02
446-35-5	-1.66	-1.57	-0.09
609-20-1	-0.69	-1.03	0.34
1806-25-3	2.68	1.85	0.83
6638-60-4	2.62	2.36	0.26
6638-61-5	3.06	3.37	-0.31
13824-23-2	0.23	-0.03	0.26
17700-09-3	-2.94	-3.13	0.19
71721-78-3	1.73	1.55	0.18
101126-67-4	-1.4	0.34	-1.74
6939-05-5	3.11	2.85	0.26
367-25-9*	-2.7	-1.90	-0.80
371-40-4*	-3.32	-4.18	0.86
1817-73-8*	-0.54	0.49	-1.03
3209-22-1*	-1.51	-1.46	-0.05

"*"is test set

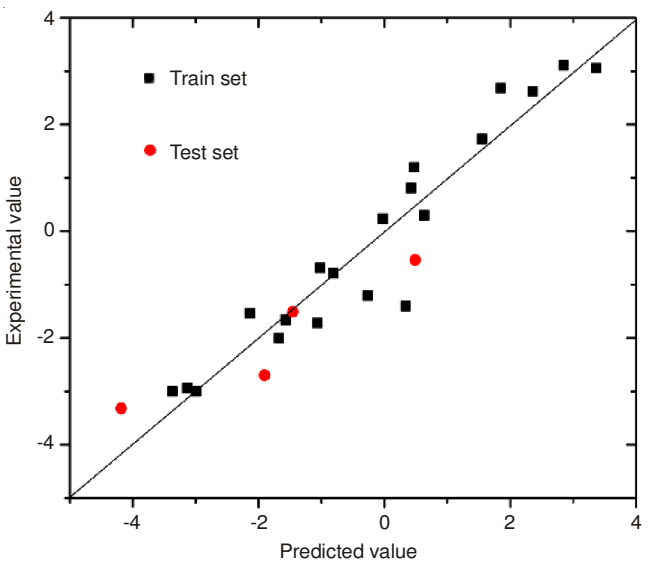

Fig. 2. Linear diagram of experimental and predictive value of mutagenicity for halogenated benzenes model

TABLE-4
DEFINITION OF DESCRIPTORS IN MUTAGENIC MODEL OF HALOGENATED BENZENES

Descriptors	Definition	Block
H2s	H autocorrelation of lag 2/weighted by I-state	GETAWAY descriptor
GATS2m	Geary autocorrelation of lag 2 weighted by mass	2D autocorrelation
SpMin1_Bh(m)	Smallest eigenvalue n. 1 of Burden matrix weighted by mass	Burden eigenvalues
SpMAD_EA(ri)	Spectral mean absolute deviation from edge adjacency mat. weighted by resonance integral	Edge adjacency indices

three descriptors, namely H2s, SpMin1_Bh(m) and SpMAD_EA(ri), were negatively correlated with mutagenicity, indicating that the increasing of the structural stability of the chemical tends to reduce the toxicity of the molecules.

The descriptors in the QSTR models for both the LD_{50} and mutagenicity of halogenated benzenes are mainly the quantum chemical and topological parameters. Although the descriptors in these models are different, they represent similar physical meanings. This signifies that the benzene compounds have similar toxicological mechanisms. Meanwhile, it was also found that there are many topological parameters associated with quantum chemical descriptors, such as molecular polarizability, electronegativity, orbital energy, *etc*. This illustrates that the toxicological mechanisms of benzene compounds are complicated.

Conclusion

With MLR analysis, QSTR models for LD_{50} and mutagencities of halogenated benzenes were successfully built with the descriptors from AM1 and E-dragon software. Both of the QSTR models 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. T.G. Oberg, *Environ. Sci. Pollut. Res*., **9**, 405 (2002).
2. J.C. Dearden, *J. Braz. Chem. Soc*., **13**, 754 (2002).
3. B. Simon-Hettich, A. Rothfuss and T. Steger-Hartmann, *Toxicology*, **224**, 156 (2006).
4. G. Patlewicz, R. Rodford and J.D. Walker, *Environ. Toxicol. Chem*., **22**, 1885 (2003).
5. G.Y. Patlewicz and D.R. Lander, *Front. Biosci*., **5**, 418 (2013).
6. M. Zeng, Z. Lin, D. Yin, Y. Zhang and D. Kong, *Bull. Environ. Contam. Toxicol*., **86**, 565 (2011).
7. M. Zeng, Z. Lin, D. Yin and K. Yin, *Bull. Environ. Contam. Toxicol*., **81**, 525 (2008).
8. G.H. Lu, C. Wang and Y.M. Li, *Biomed. Environ. Sci*., **19**, 457 (2006).
9. Q. Zhang and L.S. Wang, *Chinese J. Struct. Chem*., **26**, 933 (2007).
10. S.C. Basak, D.R. Mills, A.T. Balaban and B.D. Gute, *J. Chem. Inf. Comput. Sci*., **41**, 671 (2001).
11. M.K. Leong, S.W. Lin, H.B. Chen and F.Y. Tsai, *Toxicol. Sci*., **116**, 498 (2010).
12. B. Hemmateenejad, *J. Chemometr*., **18**, 475 (2004).
13. M. Jalali-Heravi and M. Asadollahi-Baboli, *QSAR Comb. Sci*., **27**, 750 (2008).
14. L. Tomatis and J. Huff, *Health Perspect*., **109**, a458 (2001).