ЖУРНАЛ СТРУКТУРНОЙ ХИМИИ

2011. Том 52, № 6 *Ноябрь – декабрь* *С. 1148 – 1151*

КРАТКИЕ СООБЩЕНИЯ

UDC 548.737

STUDY ON OCTANOL-WATER PARTITION COEFFICIENTS OF BENZENE HALIDES BY MOLECULAR VERTEX DISTANCE INDEX

© 2011 **L.M. Liao[1,2], J.F. Li[1,2], G.D. Lei[1], B. Wang[2]***

[1]*College of Resource and Environment Science, Neijiang Normal University, Neijiang, Sichuan 641112, P. R. China*
[2]*College of Chemistry and Life Science, Neijiang Normal University, Neijiang, Sichuan 641112, P. R. China*

Received December, 14, 2009

A new method of molecular structural characterization (MSC) called the molecular vertex distance index (MVDI) is constructed and used to describe the structures of benzene halides. Two quantitative structure-property relationship (QSPR) models of octanol-water partition coefficients (lg*K*ow) are obtained through multiple linear regression (MLR) and partial least squares regression (PLSR). The estimation stability and generalization ability of the models are analyzed by both internal and external validations. The results show that the models constructed in this work can provide satisfactory estimation stability and favorable predictive ability.

Keywords: structural descriptor, benzene halides, molecular vertex distance index (MVDI), QSPR.

The hydrophobic constant of organic compounds plays very important role in the environmental risk assessment and it is directly related to the migration and distribution of pollutants in the atmosphere, water, sediments, soil, and other environmental media as well as the organisms. The octanol-water partition coefficient (lg*K*ow) is the most common form of the hydrophobic constant of organic compounds. The construction of a quantitative structure-property relationship (QSPR) requires the structural descriptors of related chemical compounds. Much meaningful work on QSPR has been done for a long time [1—3]. We took 38 benzene halides as the samples of this study and their lg*K*ow listed in Table 1 were taken from the literature [4]. As for the QSPR research, our laboratory has developed a molecular structural characterization method called the molecular electronegativity-distance vector (MEDV) [5—8]. Based on the MEDV, the main aim of this study was to develop another molecular structural characterization method called the molecular vertex distance index (MVDI). A useful QSPR model of lg*K*ow of benzene halides was also established in this work.

Principles and methods. The property of an organic compound is directly related to the molecular structure. The octanol-water partition coefficient of organic matter is usually affected by the molecular size, flexibility and polarity, the hydrogen bonding between the molecules and so forth. For the molecular matrix skeleton, each non-hydrogen atom is a molecular vertex. The vertex is the basic unit of a molecule, so the external property of a compound can be reflected from the level of its vertices. Hence, the characteristic value of the vertex should be defined firstly for the correlative index. This definition is not a definite restriction and its purpose is to correlate the defined characteristic value with the practical problems of the research. The characteristic value of the vertex should exhibit the atomic structure and its surrounding environment. Without regard to non-framework hydrogen atoms, these vertex atoms were classified as four atomic types according to the atom number of each atom connecting to through some chemical bond/bonds. If an atom is linked to k ($k = 1, 2, 3, 4$) non-

* E-mail: wangbi05@126.com

Table 1

Compounds and their lg*K*ow

No	Compounds	lg*K*ow /*Exp.*	*Cal.*(1)	*Cal.*(2)	No	Compounds (*Test set)	lg*K*ow /*Exp.*	*Cal.*(1)	*Cal.*(2)
1	Chlorobenzene	2.81	2.6882	2.7025	20	1,3-Difluorobenzene	2.58	2.6453	2.6687
2	1,3-Dichlorobenzene	3.38	3.4575	3.4593	21	1,2-Difluorobenzene	2.59	2.7044	2.7183
3	1,2-Dichlorobenzene	3.55	3.4746	3.4821	22	Iodobenzene	3.36	3.2785	3.2556
4	1,3,5-Trichlorobenzene	4.27	4.2398	4.2057	23	1,4-Diiodobenzene	4.64	4.6506	4.5870
5	1,2,4-Trichlorobenzene	4.27	4.2093	4.2154	24	1,3-Diiodobenzene	4.64	4.5978	4.5684
6	1,2,3-Trichlorobenzene	4.27	4.2597	4.2305	25	1,2-Diiodobenzene	4.65	4.5836	4.5764
7	1,2,3,5-Tetrachlorobenzene	5.05	4.9544	4.9543	26	1,2,3-Triiodobenzene	5.86	5.8279	5.8798
8	1,2,4,5-Tetrachlorobenzene	5.05	4.9909	4.9663	27	1,2,4-Triiodobenzene	5.85	5.8444	5.8753
9	1,2,3,4-Tetrachlorobenzene	5.03	4.9715	4.9771	28	1,3,5-Triiodobenzene	5.85	5.8898	5.8740
10	1,2,3,4,5-Pentachlorobenzene	5.69	5.6820	5.6925	29	1-Bromo-2-chlorobenzene	3.83	3.7423	3.7512
11	Perchlorobenzene	6.42	6.4897	6.4405	30	1-Bromo-3-chlorobenzene	3.83	3.7373	3.7352
12	Bromobenzene	2.99	2.9897	2.9874	31	1-Bromo-4-chlorobenzene	3.83	3.7800	3.7505
13	1,2-Dibromobenzene	4.07	4.0222	4.0303	32	1-Chloro-2-iodobenzene	3.54	4.0042	4.0085
14	1,3-Dibromobenzene	3.75	4.0270	4.0194	33	1-Chloro-4-iodobenzene	4.12	4.0543	4.0142
15	1,2,3-Tribromobenzene	4.98	5.0004	5.0433	34	1-Bromo-2-iodobenzene	4.36	4.2964	4.2980
16	1,2,4-Tribromobenzene	4.98	5.0023	5.0346	35	*1,4-Dichlorobenzene	3.59	3.4940	3.4713
17	1,3,5-Tribromobenzene	4.98	5.0436	5.0314	36	*1,4-Dibromobenzene	4.07	4.0745	4.0367
18	1,2,4,5-Tetrabromobenzene	6.01	5.9738	6.0349	37	*1,4-Difluorobenzene	2.58	2.6586	2.6664
19	Fluorobenzene	2.27	2.2366	2.2817	38	*1-Bromo-4-iodobenzene	4.36	4.3577	4.3079

hydrogen atoms through chemical bonds, the atomic type belongs to the *k*th one. The characteristic value of the vertex was defined as Eq. (1)

$$Z_i = \sqrt{(n_i - 1)[(q_c - 1)/(q_i - 1)]m_i - h_i}, \qquad (1)$$

where n_i represents the number of atomic electronic shells of atoms i; q_c is the electronegativity (Pauling) of carbon; q_i is the electronegativity (Pauling) of atom i; m_i represents the number of valence shell electrons of atoms i; h_i is the number of hydrogen atoms linked to the atom i.

The properties of a molecule depend mainly on various correlations between the vertex atoms in the molecule. The main ides of MVDI is focused on an indirect reflection of the overall molecular structures by expressing correlations between different characteristic values of the vertices. These correlations should reflect the two changing trends in a specific interaction manner: one varies in a negative manner with an increase or decrease in interatomic distances and the other varies in a positive manner with a change in characteristic values. The correlation between four types of vertex atoms was defined as Eq. (2)

$$M_{kl} = \sum_{i \in k, j \in l} \frac{Z_i \times Z_j}{d_{i,j}^2}, \qquad (k = 1,\ 2,\ 3,\ 4,\quad k \le l \le 4), \qquad (2)$$

where k or l represent the atomic types of atoms i and j; i or j is the code of an atom in the molecular skeleton graph; in the MEDV [5—8], Z_i and Z_j are the relative electronegativities of atoms i and j to atom C. But in this paper, Z_i and Z_j are the characteristic values of atoms i and j; d_{ij} represents the relative distance between the ith and jth atoms (*viz*. sum of the experienced shortest path with respect to the relative C—C single bond length).

According to Eq. (2), the correlation between non-hydrogen atoms could be assembled as follows: M_{11}, M_{12}, M_{13}, M_{14}, M_{22}, M_{23}, M_{24}, M_{33}, M_{34}, and M_{44}, shortened as x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, and x_{10}; they are all called the molecular vertex distance indices (MVDIs) in this paper. According to the principles showed above, all MVDI values of the samples can be obtained. Since the samples stu-

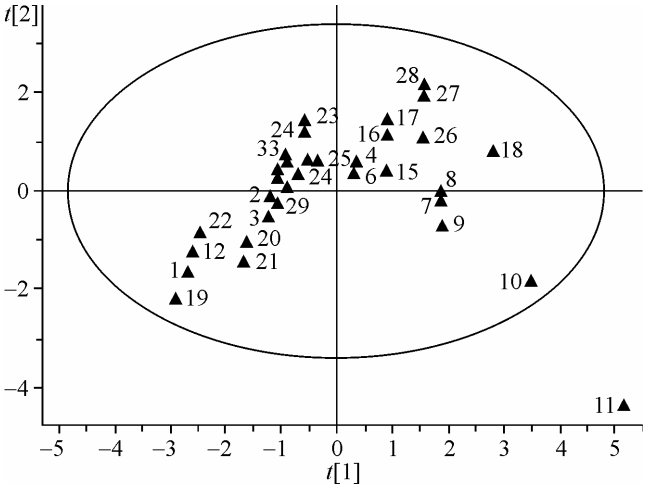

Fig. 1. Front two principal component score distribution plots by the PLS model

died in this paper contain no vertex atoms whose atomic type belongs to the 4th one, the four related items (x_4, x_7, x_9 and x_{10}) obtained are zero.

Results and analysis. The data set was randomly divided into two groups: a training set consisting 34 molecules was used for the model construction and an external data set of 4 compounds was chosen to test the prediction quality of the QSPR model. A MLR analysis was used to construct the QSRR model (M1) using six variables. The relationship between MVDI and the lg*K*ow of 34 components was modeled by Eq. (3)

$$\lg Kow = 4.7794 + 0.0597x_1 + 0.1150x_2 - 0.0067x_3 - 0.0836x_5 - 0.0909x_6 + 0.0013x_8. \quad (3)$$

Model fitting: $N = 34$, $R = 0.9945$, $SD = 0.1239$, $F = 406.3598$; Cross-Validation: $N = 34$, $R_{CV} = 0.9913$, $SD_{CV} = 0.1560$, $F_{CV} = 254.4619$.

N is the number of samples of the regression; SD is the standard deviation; R is the correlation coefficient; F is the Fischer test value. The model established above obeyed the experiential rule of ″samples/number of variables ≧5″, and the multiple correlation coefficient (R) reached 0.99. The multiple correlation coefficient (R_{CV}) of cross-validation reached 0.99, too. The results show that the model constructed above can provide estimation stability and favorable predictive ability.

To further explore the relationship between MVDI and lg*K*ow, PLSR was used to construct another model (M2), and the number of the principal components was 3. R was 0.9940 and R_{CV} was 0.9925. The correlation coefficients (R and R_{CV}) exceeded 0.99. The results show that the fitting effect, predictive ability, and stability of the model are perfect.

The PLSR score (Fig. 1) showed that most sample dots were in confident intervals of the Hotelling T^2 confidence ellipse with 95 %. Only one sample (No 11) was not in the ring. Its molecular structure contains six chlorine atoms (all six positions of benzene been replaced); the molecular structure is relatively special, and this may be the reason why it was out of the ring. All the results show that MVDI can successfully characterize the structural features of organic compounds.

The lg*K*ow of the training set was estimated and the lg*K*ow of the test set was predicted by the two models, and the results are listed in Table 1 (*Cal.* (1) and *Cal.* (2)). For the test set, the calculated values are near to the observed values and the errors are very small, and the satisfactory predictive ability of the two models has been tested. Fig. 2 presents a plot of the observed lg*K*ow (Exp.) values versus the calculated (*Cal*.) values. It can be seen that all samples are near to the diagonal line of the square. All the results show that the models can provide estimation stability and favorable predictive ability.

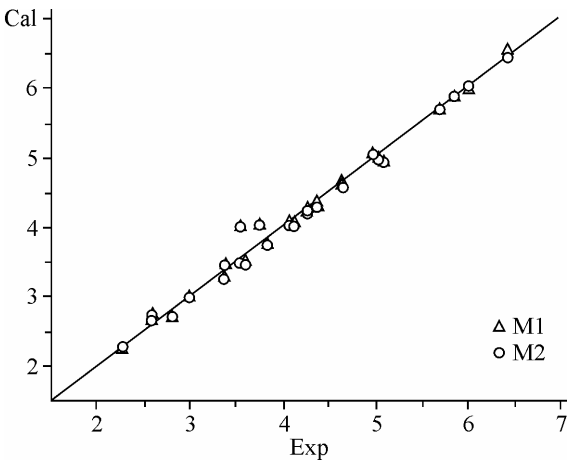

Fig. 2. Plot of the calculated values vs. the experimental values

Conclusions. The molecular vertex distance index (MVDI) was constructed and successfully used to describe the molecular structure of a set of benzene halides. MVDI can be obtained directly from a molecular structure computation with fewer components of experience and without the introduction of other structural parameters, so it is simple

and convenient. As compared to CoMFA [9, 10] or other methods [11, 12] that have been proved useful in many cases, it does not need to suffer from some limitations such as the requirement of molecular superposition. By using MVDI descriptors, we are likely to develop another simple and rapid method to estimate and predict the properties of organic compounds.

This work was supported by the Youth Foundation of Education Bureau, Sichuan Province (09ZB036).

REFERENCES

1. *Liu P.X., Long W.* // Int. J. Mol. Sci. – 2009. – **10**. – P. 1978 – 1998.
2. *Lauria A., Ippolito M., Almerico A.M.* // QSAR Comb. Sci. – 2009. – **28**. – P. 387 – 395.
3. *Yu X.L., Yu W.H., Wang X.Y.* // J. Struct. Chem. – 2009. – **50**. – P. 821 – 826.
4. *Wang L.S., Han S.K.* Advances in Chemistry of Organic Pollutants. – Beijing: Chemical Industry Press, 1998. – P. 12 – 15.
5. *Liao L.M., Mei H., Li J.F.* // J. Mol. St-TH. – 2008. – **850**. – P. 1 – 8.
6. *Zhou Y., Sun L.L., Mei H.* // Chromatography. – 2006. – **64**. – P. 565 – 570.
7. *Liu S.S., Liu Y., Yin D.W.* // J. Sep. Sci. – 2006. – **29**. – P. 296 – 301.
8. *Sun L.L., Zhou L.P., Yu Y.* // Chemosphere. – 2007. – **66**. – P. 1039 – 1051.
9. *Ju X.L., Hao Y.L., Pei J.F. et al.* // Chemosphere. – 2007. – **69**. – P. 864 – 871.
10. *Sadler B.R., Cho S.J., Ishaq K.S. et al.* // J. Med. Chem. – 1998. – **41**. – P. 2261 – 2267.
11. *Robinson D.D., Winn P.J., Lyne P.D. et al.* // J. Med. Chem. – 1999. – **42**. – P. 573 – 583.
12. *Wu B., Li M.Y., Xia L.* // Comput. Appl. Chem. (China). – 2004. – **24**. – P. 841 – 844.