The molecular topological research on acute toxicities of substituted arenes to aquatic organisms

To cite this article: Hui Feng et al 2018 IOP Conf. Ser.: Earth Environ. Sci. 153 022035

View the article online for updates and enhancements.

Related content

- Introduction to Pharmaceutical Biotechnology. Volume 1: Introduction to genetic engineering
  S Bhatia and D Goli
- Ecophotonics: assessment of temperature gradient in aquatic organisms using up-conversion luminescent particles
  E.K. Volkova, I.Yu. Yanina, A.P. Popov et al.
- Study on acute toxicity of amoxicillin wastewater to Zebrafish
  Weifang Xie and Hongyan Shen
The molecular topological research on acute toxicities of substituted arenes to aquatic organisms

Hui Feng, Yan Chen, Wei Yue, Changjun Feng

School of Chemistry and Chemical Engineering, Xuzhou University of Technology, Xuzhou, 221111
corresponding author’s e-mail address: fengcj@xzit.edu.cn

ABSTRACT. A novel topological index (\(1D\)) based on the adjacency matrix and atomic characteristic values (\(d_i\), which is equal to \((n_i-1)m_i-h_i\)) was derived in this paper. The index \(1D\) was defined as \(1D=\Sigma(d_i\cdot d_j)^{0.5}\). The linear regression equations of \(1D\), \(\lg K_{ow}\) and acute toxicities for substituted arenes to aquatic organisms were proposed as follows: \(-\lg EC_{50}=2.391+0.025D+0.238\lg K_{ow}\); \(-\lg LC_{50}=2.192+0.021D+0.414\lg K_{ow}\). The correlation coefficients \((R^2)\) and the leave-one-out (LOO) cross validation \(R_{cv}^2\) for the \(EC_{50}\) and \(LC_{50}\) models were 0.881 and 0.852, 0.939 and 0.916, respectively. The QSAR models have favorable correlation, as well as robustness and good prediction capability by \(R_{cv}^2\), \(V_{IF}\) tests. These results were better than those reported in the literatures.

1. Introduction

Halogenated arenes and their derivants such as halogenated benzenes, halogenated phenols and halogenated amines, are very dangerous pollutant. A simple, economic and dependable method is in urgent need to estimated the toxicity of the substituted arenes. In recent years quantitative structure-activity relationship (QSAR)\(^{[1-6]}\) has been proved to be a very important way to estimate and predict the bioactivity of the organic compounds\(^{[7]}\). In this article, we constructed a new connectivity index \((\delta D)\) based on Randic's branching degree index \((\delta')\)\(^{[8,9]}\). The first-order index \((\delta D)\) and hydrophobicity parameters (\(\lg K_{ow}\)) are used to estimate the acute toxicity of the substituted arenes to aquatic organisms, and the results are excellent.

2. Constructing \(\delta D\)

2.1 Molecular descriptors

Randic\(^{[8,9]}\) defined the branching degree \((\delta_i)\) of the carbon atom in alkane as: \(\delta_i=4-h_i\), Where \(h_i\) is the number of hydrogen atoms bonding with the carbon atom \(i\). The first-order index \((\delta'\chi)\) is calculated by the following formula:

\[\chi=\sum(\delta_i\cdot \delta_j)^{0.5}\]

where \(i\) and \(j\) represent adjacent carbon atoms. But this method is only appropriate for hydrocarbon. The properties of the atom in molecule are correlative with the number of valence electrons and the number electron shells strongly, hence, in order to keep consistent with \(\delta_i\), we defined the characteristic value of the atom \((d_i)\) as:

\[d_i=(n_i-1)^2-m_i-h_i\]

where \(m_i\) is the number of valence electrons, and \(n_i\) is the number of electron shells. For halogenated...
hydrocarbons, the values of \(d_i\) of F, Cl, Br, I are 7, 28, 63, 112, respectively. The value of \(d_i\) of the hydrogen atom in organic compounds is zero, and the values which is used to construct topological index are constant with the structure diagram without hydrogen atom. From Wiener\cite{4}, most of topological index were calculated according to structure diagram directly without interpretation.

Based on \(d_i\), the connective index \(m(D)\) of the atom characteristic is defined as:

\[ m(D) = \sum(d_i)_{1/2} \]  

(3)

According to formula (3), the zero-order \(0(D)\) and the first-order \(1(D)\) are defined as:

\[ 0(D) = \sum(d_i)^{0.5} \]  

(4)

\[ 1(D) = \sum(d_i^2)^{0.5} \]  

(5)

For example, the structure diagram without hydrogen atoms of 2,5-dichlorotoluene, \(0(D)\) and \(1(D)\) are as following:

\[ 0(D) = (1\times4)^{0.5} + 2\times(4\times28)^{0.5} + (4\times4)^{0.5} + 4\times(4\times3)^{0.5} + (3\times3)^{0.5} = 22.790 \]

\[ 1(D) = (1\times4)^2 + 2\times(4\times28)^2 + (4\times4)^2 + 4\times(4\times3)^2 + (3\times3)^2 = 44.022 \]

Because the structure selectivity of \(0(D)\) is poor, we only use \(1(D)\) in this article.

2.2 Statistical regression analysis

For QSAR derivation, IC_{50} values act as the dependent variables, and electronegativity distance vector descriptors act as the independent variables. The regression analyses are carried out by using multiple linear regression(MLR), partial least squares(PLS), leaps-and-bounds regression (LBR) program. The correlation between variables in model was estimated by the variance inflation factor \((V_{if})\)\cite{10} is defined as follows:

\[ V_{if} = 1/(1 - \beta^2) \]  

(6)

in which \(\beta\) is the correlation coefficient of multiple regressions between one variable and the others in the equation. \(V_{if}\) = 1.0 suggests no self-correlation among each variable; if \(V_{if}\) ranges from 1.0 to 5.0, indicating that there is no obvious autocorrelation between variables, the model is stable; when \(V_{if}\) is larger than 5.0, the regression equation is unstable and recheck of variables’ correlation coefficient is necessary. The “leave-one-out”(LOO) cross-validation coefficient \(R_{cv}^2\) was considered as an indicator of the predictive performance and stability of a QSAR model. As a rule of thumb, the equations with regression coefficients \(R_{cv}^2 > 0.50\)\cite{11} are considered reasonable. Where parameter \(R_{cv}^2 > 0.50\) is used as a criterion of both robustness and predictive ability of the model.

3. Results and Discussion

3.1 QSAR equation

The action mechanism of the toxicant to the aquatic organism has not been known clearly, and the “target theory” has been accepted generally. Hansch et al\cite{5}, thought that the bioeffect of the organic compound had something to do with the process, during which the organic compound transferred from aqueous phase to biophase and acted on the target cell of the organism. Usually, the action include transfer of electrons among molecules, hydrophobic effect and dispersion force, so the linear equation about the bioeffect of organisms is defined as:

\[ \log B_l = a\log K + bE_{\text{exc}} + cH + d \]  

(7)

where \(B_l\) measures bioeffect, \(K\) is partition coefficient, \(E_{\text{exc}}\) is space-effect index, \(H\) is electric effect index, and \(a, b, c\) are all regression coefficients.

Octanol/water partition coefficients \((K_{ow})\) represent hydrophobic/hydrophilic properties of organic compounds, and they are suitable for interpreting the distribution of organic compounds in target molecules and the ability of organic compounds bonding with target molecules, so we substitute \(\log K_{ow}\) for \(\log K\). It is well known that \(\chi\) defined by Randic is correlated with the boiling point of the alkanes, which is attracted by dispersion force greatly, so we think that \(\chi\) measures dispersion force between molecules. Base on \(\delta_{i}, d_i\) is defined as: \(d_i = (n_i - 1)^2 \cdot m_i - h_i\), where \((n_i - 1)^2 \cdot m_i\) measures total energy of the valence electrons of atoms i, and \(n_i\) measures the electron level, so \((n_i - 1)^2 \cdot m_i\) represents electric properties. \(1(D)\) constructed by \(d_i\) discloses dispersion force effect and electric effect of the molecule.
After the substitution of \(1^D\) for \(E_x\) and \(H\), equation (6) turned into:
\[
\log B_f = A_1 \log K_{ow} + B_1 D + C
\]
(8)

Using equation (7), we related toxicities (data in Table 1) to \(1^D\) and \(\log K_{ow}\), and got two regression equations. To photoluminescent bacteria\[^7\]:
\[
-\log EC_{50} = 2.391 + 0.025 D + 0.238 \log K_{ow}
\]
\(f = 30, R^2=0.881 (0.774), R_{cv}^2 = 0.852, F=98.02, S=0.225\)
(9)

where \(f\) is the number of compounds included in the model. To pimephales promelas\[^7\]:
\[
-\log LC_{50} = 2.192 + 0.021 D + 0.414 \log K_{ow}
\]
\(f=20, R=0.939 (0.903), R_{cv}^2 = 0.916, F=130.12, S=0.215\)
(10)

In equation (9), \(EC_{50}\) is the concentration of the organic compound in water, which can restrain half of the photoluminescent bacterium from luminescing. In equation (10), \(LC_{50}\) is the concentration of the organic compound in water, which can kill half of fathead minnows. In Table 1, the structures of thirty organic compounds are very different, of which substituted groups include methyl, halogen, amino group and hydroxyl group. Although the errors from testing bioactivity of organisms are great, the correlations of equation (9) and (10) are better than that of literature report (1). In literature (1), \(\log K_{ow}\) and energy of the Highest Occupied Molecular Orbit (\(E_{homo}\)) were correlated with the toxicity. The results from equation (9) and (10) show that most of the errors for halogenated arenes are less than the experimental errors in Table 1. In order to compare clearly, the results of literature [7] are also listed in Table 1.

3.2 Analysis of the QSAR equation

Molecular connectivity index which represents the characteristic structure effect of molecular is a nonexpirical objective parameter of topology. Furthermore, this index represents not only the ability of the organic compound to act on the organism, but also the real toxicity of the organic compound to the organism. Hence, \(1^D\) is correlative with bioactivity excellently. The following equations:
\[
-\log EC_{50} = 2.506 + 0.0401^D, \quad r = 0.893
\]
(11)
\[
-\log LC_{50} = 2.345 + 0.0481^D, \quad r = 0.898
\]
(12)
are both better than the equations from \(E_{homo}\):
\[
-\log EC_{50} = 3.544 + 0.772 E_{homo}, \quad r = 0.207
\]
(13)
\[
-\log LC_{50} = 3.168 + 1.286 E_{homo}, \quad r = 0.226
\]
(14)

Our work shows that the \(E_{homo}\) can not represent the reactivity power of the organic compound listed in this article.

\(K_{ow}\) can represent the ability of the organism to concentrate organic compounds, so it measures the external factor of the toxicity to the organism. Because the structures of different organisms are different, the quantities of the same organic compound concentrated in different organisms are different. Therefore, the toxicities of the same organic compound to different organisms are different, namely, \(K_{ow}\)'s affects different organisms in different degree. The coefficients of \(\log K_{ow}\) in equation (9) and (10) are 0.414 and 0.238 respectively, and this means that \(K_{ow}\) has greater effect on the toxicity to fish than to bacteria. The reason is that the bacteria is a single-cell organism, which has less lipid than fish, and the concentration of different organic compounds in fish is greatly different, but the concentration difference in bacteria is not so great.

| Halogenated arenes          | \(1^D\) | \(\log K_{ow}\) | \(-\log EC_{50}\) | \(-\log LC_{50}\) |
|-----------------------------|---------|-----------------|-------------------|-------------------|
| 1,2,4,5-Tetrachlorobenzene  | 64.168  | 5.05            | 5.51              | 5.20              |
| 1,2,4-Trichlorobenzene      | 52.605  | 4.29            | 4.50              | 4.73              |
| 1,2,3-Trichlorobenzene      | 52.677  | 4.27            | 4.53              | 4.72              |
| 1,4-Dichlorobenzene         | 41.022  | 3.59            | 4.39              | 4.27              |
| 1,3-Dichlorobenzene         | 41.022  | 3.58            | 4.24              | 4.22              |
| 1,2-Dichlorobenzene         | 41.094  | 3.55            | 4.38              | 4.26              |
| Chlorobenzene               | 29.511  | 2.81            | 3.86              | 3.80              |
| 1,3-Dibromobenzene          | 51.605  | 3.75            | 4.99              | 4.57              |

Table 1 Toxities to fathead minow and photoluminescent bacteria, \(1^D\) and \(\log K_{ow}\) of halogenated arenes.
3.3 Validation of the QSAR equation
The predictive capability of a QSAR model should be tested through model validation. Cross-validation is one of the most often used model validation methods. The $R_{cv}^2$ values of models (9) and (10) are 0.852 and 0.916, respectively, which are well above 0.5, indicating that the model has good robustness and prediction ability. The $V_{if}$ of the two variables in model (9) and (10) is 2.351, which indicates that all models are statistically significant and have good stability.

4. Conclusion
The molecular connectivity index can characterize the abstract molecular structure to achieve numerical representation of 30 halogenated arenes, and show good structural selectivity. The QSAR models have showed good correlation, as well as robustness and prediction ability by statistical indicators: $R^2$, $R_{cv}^2$, and $V_{if}$ tests.

In summary, the action mechanism of the organic compounds in the organism includes two steps. The first step is the concentrating process of the organic compound in the organism, which is measured in this article by $\lg K_{ow}$. The second step is the binding process of the molecule of the organic compound with target molecule which is measured in this article by $1D$. Our work shows that topological index has many advantages, such as the computing of this index is simple, it can represent molecular structure information accurately, and it is well correlated with bioactivity. The topological index is a suitable technique for QSAR, so it will find extensive applications in the field.

Acknowledgments
This work was financially supported by the the National Natural Science Foundation of China (21272095, 21075138), special fund of State Key Laboratory of Structure Chemistry (20160028) and Scientific Research Fund of Xuzhou University of Technology (XKY2012307).

References
(1) C J Feng, W H Yang. Linear QSAR regression models for the prediction of bioconcentration factors of chloroanilines in fish by density functional theory. Chin. J. Struct. Chem., 2014, 33(6):830-834.
(2) C J Feng. Theoretical studies on quantitative structure-activity relationship and structural modification for 3-substituted sulfur-(2-hydroxyphenyl)-4H-1,2,4-triazole compounds. Acta Chim. Sin., 2012, 70, 512–518.
(3) B G Liu, J W Liu, J Q Li, et al. 3D-QSAR and interaction mechanism of flavonoids as P-glycoprotein inhibitors. Chem.J.Chinese Universities, 2017, 38, 41-46.
(4) M A Abdelrahman, I Salama, M S Gomaa, et al. Design, synthesis and 2D QSAR study of novel pyridine and quinolone hydrazone derivatives as potential antimicrobial and antitubercular agents. European Journal of Medicinal Chemistry, 2017, 138, 698-714.

(5) C Wang, C J Feng. QSAR study of the action strength of DOM of phenyl-isopropyl-amine dopes using MLR and BP-ANN. Chin. J. Struct. Chem., 2017, 36, 1720-1728.

(6) C J Feng. Theoretical Study on the Antifungal Acticity of pyridinecarboxamide cyclohexyl sulfonamides to *Rhizoctonia cerealis*. Journal of Xuzhou Institute of Technology (Natural Sciences Edition), 2017, 32, 23-27.

(7) L S Wang, S K Han. Progress of Organic pollution chemistry. Beijing: Chemical Industry Press, 1998, 138-148.

(8) M Randic. On charmancterization of molecular branching. J. Amer. Chem. Soc., 1975, 97: 6609-6615.

(9) L B Kier, L H Hall. Molecular Connectivity in Structure-Activity Analysis. England: Research Studies Press, 1986: 82.

(10) D B Wei, A Q Zhang, C D Wu, et al. Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds. Chemosphere, 2001, 44, 1421-1428.

(11) S Xia, Y Feng, J G Cheng, et al. QAAR exploration on pesticides with high solubility: An investigation on sulfonylurea herbicide dimers formed through p–p stacking interactions[J]. Chin. Chem. Let., 2014, 25: 973–977.