



QSPR analysis of infinite dilution activity coefficients of chlorinated organic compounds in water

Jie Xu*, Hui Zhang, Luoxin Wang, Wenxiang Ye, Weilin Xu, Zengchang Li

Key Lab of Green Processing & Functional Textiles of New Textile Materials, Ministry of Education, Wuhan University of Science & Engineering, No. 1, Fangzhi Road, Hongshan District, 430073, Wuhan, Hubei, PR China

ARTICLE INFO

Article history:

Received 9 October 2009

Received in revised form

22 December 2009

Accepted 24 December 2009

Available online 11 January 2010

Keywords:

QSPR

Infinite dilution activity coefficients

Chlorinated organic compounds

Stepwise multilinear regression analysis

Molecular descriptors

ABSTRACT

A quantitative structure–property relationship (QSPR) study was performed for the prediction of the natural logarithm of infinite dilution activity coefficients, $\ln(\gamma^\infty)$, of 45 chlorinated organic compounds in water. A five-descriptor correlation equation, with a squared correlation coefficient (R^2) of 0.9655 and a standard error of estimation (s) of 0.613, was produced by using the stepwise multilinear regression analysis on a training set of 35 compounds. The reliability of the proposed model was further illustrated using various evaluation techniques: leave-one-out cross-validation procedure, randomization tests, and validation through an external test set of 10 compounds. All descriptors involved in the model were derived solely from the chemical structure of the compounds, which makes the model very useful in predicting $\ln(\gamma^\infty)$ of other chlorinated organic compounds not present in the dataset used for the development of the present model. In addition, the descriptors were discussed in detail to understand the interactions that affect infinite dilution activity coefficients.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The infinite dilution activity coefficient, γ^∞ , is an important parameter related to the thermodynamic behavior of dilute solutions. This coefficient is extremely meaningful for both theoretical and practical purposes. Theoretically, it provides valuable information regarding the solute–solvent interactions in the absence of solute–solute interactions. The magnitude of these infinite dilution activity coefficients gives insight into the type and strength of interactions between the solvent and the solute molecules [1]. From a practical point of view, knowledge of γ^∞ is crucial for the design of many separation processes such as removing dilute impurities as encountered in many environmental applications. Because the most common industrial solvent is water, the study of γ^∞ of aqueous solutions is of great interest.

Although numerous experimental techniques have been used to measure γ^∞ , experimental values are available only for relatively low molecular weight compounds [2]. For reasons of cost, time, safety, and availability of chemical samples, it is useful to be able to predict γ^∞ rather than to measure it [3]. Many predictive methods have been reported, including group contribution methods (e.g. ASOG [4,5], UNIFAC [2,6] and diverse modified UNIFAC versions [7–9]), methods derived from the regular solution

theory [10], linear solvation energy relationships (LSER) [11], modified excess Gibbs energy model [12], and free energy perturbation simulations [13]. However, predictions from group contribution methods are highly dependent on the quality of the structural parameters that are developed from experimental data, and do not work well for aqueous systems where the variation of γ^∞ is quite large. Methods derived from the regular solution theory, LSER and Gibbs energy model depend on the availability of some experimental values. The free energy perturbation method with Monte Carlo simulations is very computationally expensive and time-consuming.

Alternatively, the quantitative structure–property relationship (QSPR) provides a promising method for the estimation of γ^∞ based on descriptors derived solely from the molecular structure to fit experimental data [14–18]. The QSPR is based on the assumption that the variation of the behavior of the compounds, as expressed by any measured physicochemical properties, can be correlated with numerical changes in structural features of all compounds, termed “molecular descriptors” [19–23]. The advantage of this method lies in the fact that it requires only the knowledge of the chemical structure and is not dependent on any experimental properties. Once a correlation is established, it can be applicable for the prediction of the property of new compounds that have not been synthesized or found. Thus the QSPR method can expedite the process of development of new molecules and materials with desired properties. The QSPR method has been successfully applied to predict the chemical, physical, biochemical, and pharmacological properties of

* Corresponding author. Tel.: +86 27 87426559; fax: +86 27 87426559.

E-mail address: xujie0@ustc.edu (J. Xu).

Table 1
Experimental and calculated values of $\ln(\gamma^\infty)$ for chlorinated organic compounds.

No.	Compound	Expt.	Calc.	Residual
1	1,1,1,2-Tetrachloroethane	9.10	8.62	0.48
2	1,1,1-Trichloroethane	8.68	7.37	1.31
3	1,1,2,2-Tetrachloroethane	8.15	8.58	-0.43
4	1,1,2-Trichloroethane ^a	7.31	7.50	-0.19
5	1,1,2-Trichlorotrifluoroethane	11.00	11.15	-0.15
6	1,1-Dichloroethane	6.99	6.74	0.25
7	1,2,3-Trichloropropane	8.37	8.29	0.08
8	1,2-Dichlorobenzene ^a	11.10	11.44	-0.34
9	1,2-Dichloroethane	6.46	6.97	-0.51
10	1,2-Dichloropropane	7.75	7.61	0.14
11	1,2-Dichlorotetrafluoroethane	11.20	11.20	0.00
12	1,3-Dichlorobenzene ^a	11.10	11.45	-0.35
13	1,3-Dichloropropane	7.74	7.63	0.11
14	1,3-Dichloropropene	7.24	7.82	-0.58
15	1,4-Dichlorobenzene	11.50	11.36	0.14
16	1-Chloro-2-methylpropane ^a	8.62	8.16	0.46
17	1-Chlorobutane	8.94	8.34	0.60
18	1-Chlorohexane	11.90	11.65	0.25
19	1-Chloropentane	10.40	9.74	0.66
20	1-Chloropropane ^a	7.47	7.36	0.11
21	2-Chloro-2-methylbutane	7.48	9.26	-1.78
22	2-Chlorobutane	8.55	8.26	0.29
23	2-Chloropropane	7.30	7.30	0.00
24	3-Chloro-1-propene	6.97	7.19	-0.22
25	Benzylchloride	10.40	10.93	-0.53
26	Chlorobenzene ^a	9.55	9.76	-0.21
27	Chlorodifluoromethane	7.46	6.38	1.08
28	Chloroethane	5.98	6.66	-0.68
29	Chloroethene ^a	7.16	6.50	0.66
30	Chloromethane	6.17	6.19	-0.02
31	Chloropentafluoroethane	11.90	11.95	-0.05
32	Chlorotrifluoromethane	11.10	10.93	0.17
33	cis-1,2-Dichloroethene ^a	6.77	6.95	-0.18
34	Dichlorodifluoromethane	10.00	10.02	-0.01
35	Dichlorofluoromethane	5.72	6.12	-0.40
36	Dichloromethane	5.53	6.24	-0.71
37	Hexachlorobenzene	21.90	22.22	-0.32
38	Hexachloroethane	14.30	13.37	0.93
39	Pentachloroethane ^a	10.10	10.53	-0.43
40	Tetrachloroethene	10.50	10.31	0.19
41	Tetrachloromethane	9.64	9.53	0.11
42	trans-1,2-Dichloroethene	7.14	6.95	0.19
43	Trichloroethene ^a	9.08	8.21	0.87
44	Trichlorofluoromethane	8.86	9.34	-0.48
45	Trichloromethane	6.81	6.91	-0.10

^a Data used for the test set.

compounds. Delgado and Alderete [15] correlated the $\ln(\gamma^\infty)$ values of 45 chlorinated organic compounds in water with their quantum-chemical descriptors and got a relatively good correlation with R^2 of 0.949. He and Zhong [16] have applied the molecular connectivity indices to develop QSPR models for $\ln(\gamma^\infty)$ of hydrocarbons, oxygen containing organic compounds and halogenated hydrocarbons in water, respectively. With the same dataset, Delgado and co-workers [18] recently obtained several QSPR equations for the prediction of $\ln(\gamma^\infty)$ by using quantum-chemical descriptors. However, the abovementioned QSPR models have not been evaluated with the external validation set. In fact, validation is a crucial aspect of any QSPR/QSAR (quantitative structure–activity relationship) modeling [24]. Furthermore, all authors tried to develop their QSPR models based on a special group of molecular descriptors (connectivity indices in the work of He and Zhong [16], quantum-chemical descriptors in the work of Delgado et al. [15,18]). This phenomenon caused the obtained results are not the best results which can be obtained by QSPR analysis.

The goal of this study was to produce a robust QSPR model that could predict the $\ln(\gamma^\infty)$ values from the training set of 35 chlorinated organic compounds by using various types of descriptors and then the model was validated with 10 chlorinated organic compounds in the test set.

2. Materials and methods

2.1. Dataset

The experimental data of γ^∞ of 45 structurally diverse chlorinated organic compounds in water (Table 1) were taken from the article by Delgado and Alderete [15]. The reported $\ln(\gamma^\infty)$ values ranged from about 5 to 22 ln units. Among them, 35 compounds were randomly chosen as the training set, and the other 10 compounds were used as the test set.

2.2. Descriptor generation

The structures of all molecules were preoptimized using MM+ molecular mechanics force field (Polak–Ribiere algorithm) in the HYPERCHEM program [25]. The final geometries of the minimum energy conformation were obtained by the semi-empirical AM1 method at a restricted Hartree–Fock level with no configuration interaction, applying a gradient norm limit of $0.01 \text{ kcal } \text{Å}^{-1} \text{ mol}^{-1}$ as a stopping criterion. Then totally 1664 molecular descriptors for each molecule were calculated on the resulting geometry with Dragon software [26]. These descriptors include (a) 0D-constitutional (atom and group counts); (b) 1D-functional groups

Download English Version:

<https://daneshyari.com/en/article/202792>

Download Persian Version:

<https://daneshyari.com/article/202792>

[Daneshyari.com](https://daneshyari.com)