

Quantitative structure retention relationship studies for predicting relative retention times of chlorinated phenols on gas chromatography

LI Shi-yin, SUN Cheng^{*}, WANG Yu, XU Shi-fen, YAO Shu-chun, WANG Lian-sheng

(State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing 210093, China. E-mail: chengsun@public1.ptt.js.cn)

Abstract: A new method of quantitative structure-retention relationship(QSRR) studies was reported for predicting gas chromatography(GC) relative retention times(RRTs) of chlorinated phenols (CPs) using a DB-5 column. Chemical descriptors were calculated from the molecular structure of CPs and related to their gas chromatographic RRTs by using multiple linear regression analysis. The proposed model had a multiple square correlation coefficient $R^2 = 0.970$, standard error $SE = 0.0472$, and significant level $P = 0.0000$. The QSRR model also reveals that the gas chromatographic relative retention times of CPs are associated with physicochemical property interactions with the stationary phase, and influenced by the number of chlorine and oxygen in the CP molecules.

Keywords: QSRR; relative retention times; molecular descriptors; chlorinated phenols

Introduction

Chlorinated phenols(CPs) from mono-to penta-chlorine-substituted are widely used as either synthesis intermediates in dyestuffs, or insecticides, fungicides, antiseptic and disinfectants. In our recent study on investigating the contamination status of CPs in Yangtze River sediment, it is found that CPs concentration in the river sediment commonly ranged from 0.30 – 1.90 ng/g of monochlorophenols to 0.49 – 4.57 ng/g of pentachlorophenol(Xu, 2000). It is well known that CPs released from various sources usually present potential hazards to human and the environment. Therefore, it is of importance in predicting the transformation process and toxicity of CPs in the environment. CPs have been analyzed by high performance liquid chromatography(HPLC; Butte, 1981), however, isomer-specific analysis requires purified standards of individual isomers. Although gas chromatography-mass spectrometry(GC-MS) is the best analytical technique to identify these compounds at congener level, resolving complex mixtures might require additional tools involving the knowledge of molecular interactions of CPs with a stationary phase. This may be achieved using quantitative structure-retention relationship(QSRR) which provides statistical equations that relate molecular structure with the retention phenomena(Payares, 1997; Kang, 1998; Thomas, 1992). The objective of this paper is to provide a statistical model to predict the chromatographic retention of individual CPs in a DB-5 column(5% phenyl, methypolysiloxane).

1 Experimental

1.1 Data set

A Fisons8000 mass spectrometer system equipped with a GC8000, an autosampler AS8000(operating in the splitless mode) and a quadrupoles-MSMD were used to perform the GC-MS analysis. Relative retention times were determined for 16 CPs in a DB-5 column (30m × 0.33 mm I. D., 0.25 μm film thickness, J&W Scientific Folsom, CA, USA). Chromatographic conditions were: the oven temperature from setpoint at 60°C and held for 2 min, followed by ramping to 260°C at a temperature gradient of 10°C/min, again held for 5 minutes. The injector and the detector temperature was set at 220°C and 260°C, respectively. Helium was used as a carrier gas and injection was made splitless. Electron impact mass

spectra were obtained at 70 eV ionizing energy. RRTs were the gas chromatographic retention time of chlorinated phenols relative to 2-chlorophenol (Table 1).

1.2 Descriptor generation

To determine the quantum chemical descriptors, the molecular modeling package ALCHEMY II (Tripos Associates, Inc, 1988) was used to construct and view the molecular structures. Molecular descriptors used in this study were derived using the semi-empirical orbital MOPAC 6.0 program (Stewart, Co80840). Molecules were entered and fully optimized using the semiempirical quantum chemical method AM1 (Dewar, 1985). Calculated descriptors of 16 chlorinated phenols include heat of formation (HF), energy of the highest occupied molecular orbital (HOMO), energy of the lowest

Table 1 Observed and calculated RRTs for CPs in a DB-5 column and descriptors included in the QSRR model

Molecule	RRTs		NCIO	log K_{ow}
	Observed	Calculated		
2-chlorophenol	1.000	0.996	2	2.15
3-chlorophenol	1.045	1.044	2	2.50
4-chlorophenol	1.057	1.029	2	2.39
2,3-dichlorophenol	1.174	1.236	3	2.84
2,4-dichlorophenol	1.250	1.266	3	3.06
2,5-dichlorophenol	1.271	1.266	3	3.06
3,4-dichlorophenol	1.344	1.303	3	3.33
3,5-dichlorophenol	1.301	1.342	3	3.62
2,3,4-trichlorophenol	1.567	1.513	4	3.80
2,4,6-trichlorophenol	1.407	1.498	4	3.69
2,3,6-trichlorophenol	1.478	1.509	4	3.77
2,3,5-trichlorophenol	1.491	1.518	4	3.84
3,4,5-trichlorophenol	1.590	1.541	4	4.01
2,3,5,6-tetrachlorophenol	1.687	1.670	5	3.88
2,3,4,5-tetrachlorophenol	1.779	1.715	5	4.21
Pentachlorophenol	1.948	1.984	6	5.12

occupied molecular orbital (LUMO), energy of the core-core repulsion (CCR), dipolar moment (DIP), electronic energy (EE), energy of the ion potential (IP), most negative charge (MNC), most positive charge (MPC), number of chlorine and oxygen (NCIO). Octanol-water partition coefficient (log K_{ow}) was obtained by the computer program CLOGP (Ghose, 1988) which was widely demonstrated to be the most convenient, advanced and accurate computer program for the calculation of *n*-octanol-water partition coefficients.

1.3 Statistical analysis

In order to reduce the number of descriptors prior to submitting them to regression routines, the concept of nonredundant descriptors (NRD) was applied (Katritzky, 1993). When two descriptors are correlated by a linear correlation coefficient value high than 0.9, both are correlated with the dependent variable. The best correlation is used for further analysis, the another which shows lower correlation is left out. The QSRR equation was obtained by forward stepwise multiple regression techniques following the multilinear form: $RRT = b_1 D_1 + b_2 D_2 + \dots + b_n D_n$, where D_1 , D_2 and D_n are the descriptors and b_1 , b_2 and b_n are the respective regression coefficients (Neter, 1985). All the stepwise regression analyses were performed using the STATISTICAL for windows release 5.0 program (Statsoft Inc., USA, 1984 - 1995). The best model was selected on the basis of the multiple correlation coefficient (R^2), standard error of estimation (SE), and significant level (P).

1.4 Model validation

Validation is a means to determine the stability and robustness of a model. Two methods for validation are available, internal and external validation. In this study, for data were initially small, internal validation was chosen to test the quality of a model. Here an observation was held out of data set, and the model was recalculated. The resulting model was then used to predict the observation withheld. This process continued for each observation in the entire data set. Then predictions were compared with the

actual observations (Thomas, 1992).

2 Results and discussion

Regression equation obtained with selected molecular descriptors, which include NCIO and $\log K_{ow}$ are presented in Table 1. It reveals that chromatographic retention of CPs for the DB-5 column is linked to the extent of chlorination and oxygen in the aromatic rings (NCIO) as expected for a homologous series. The stepwise regression analysis of variance shows a statistical significance at $P = 0.0000$ for the calculated model, where the R^2 and SE values for the regression equation are 0.970 and 0.0472, respectively. The high R^2 value also implies that the developed model is suitable to predict the RRTs of CPs.

Fig. 1 illustrates a plot of observed vs. calculated RRTs values for all the 16 congeners. The result shows that the developed model is suitable to predict the RRTs of CPs. The best equation developed by the procedure described above is presented in Table 2.

Table 2 The best two descriptors QSRR model^a for gas chromatography retention of CPs in a DB-5 column

Descriptor	Regression coefficient \pm standard error	<i>t</i> -value	Significant level
NCIO	0.146 \pm 0.033	4.3512	0.0001
Log K_{ow}	0.136 \pm 0.051	2.6830	0.0008
Intercept	0.412 \pm 0.073	5.6328	0.0188
$R^2 = 0.970$, $F = 241.80$, $SE = 0.047$, $P = 0.0000$			

^aRRT = $b_1 D_1 + b_2 D_2 + \dots + b_n D_n$; RRT: retention times relative to 2-chlorophenol; b_1 , b_2 , b_n : regression coefficients; D_1 , D_2 , D_n : molecular descriptors

polarizability (Katritzky, 1993). Not surprisingly it accounts for most of the variation in the RRTs of CPs on a DB-5 column ($R^2 = 0.956$, $P = 0.0000$), similar to that observed for other homologous series of compounds such as polychlorinated dibenzofurans (Robbat, 1990). NCIO was highly correlated with EE ($R^2 = 0.997$, $P = 0.0000$) and CCR ($R^2 = 0.995$, $P = 0.0000$) without considering the $\log K_{ow}$. It is quite reasonable to suggest that chlorine and oxygen may play an important role in the ability of these compounds to interact electronically with the stationary phase. This quantity of the energy of core-core repulsion describes a summary of the repulsion of electrons of atoms constituting a molecule and is probably correlated with the inclination of the thermally cracked products to undergo chemi-ionization (Katritzky, 1994).

A model containing $\log K_{ow}$ without considering the degree of chlorination and hydroxylation can predict RRTs of CPs with $R^2 = 0.931$, $P = 0.0000$. $\log K_{ow}$ is correlated with LUMO ($R^2 = 0.867$, $P = 0.0000$). This suggests that the retention of CPs in the DB-5 column involves physicochemical property interactions with the stationary phase. $\log K_{ow}$ is a physicochemical property descriptor. The value of $\log K_{ow}$ embodies the feature of molecular structure, the interaction between the internal and the external of molecular. Partial correlation controlling for NCIO reveals that without considering the degree of chlorination and hydroxylation, chromatographic relative retention time is highly correlated with LUMO ($R^2 = 0.919$, $P = 0.0001$) and HF ($R^2 = 0.873$, $P = 0.0000$). LUMO is the energy of lowest unoccupied molecular orbital. From the molecular orbital theory of chemical reactivity, transition states are formed

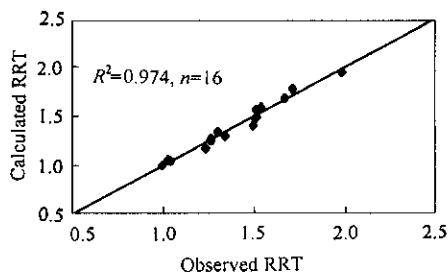


Fig. 1 Plot of calculated RRT vs. observed RRT values for the two-descriptor equation

QSRR models developed using quantum-derived, physicochemical and geometrical descriptors are found elsewhere (Payares, 1997; Gergakopoulos, 1991; Stanton, 1989). As expected, the descriptor NCIO is the most significant in the model. NCIO is a topological descriptor which encodes information about molecular size, electron number and

during interaction between LUMO (electron acceptor) and HOMO (electron donor) of reacting species (Fukui, 1975). The interaction between donor and acceptor can lead to the formation of charge transfer complexes (Franke, 1984) between CPs and the stationary phase. The high correlation suggests that HOMO and LUMO play important parts in chemical reactivity and the formation of charge transfer complexes (Osmidowski, 1985; Zhou, 1990) between CPs and the stationary phase. HF is a quantum chemical index obtained by subtracting atomic heats of formation from the binding energy. The high correlation between RRTs and HF suggests that the retention of CPs in the DB-5 column is related to the formation of a hydrogen bond between the CPs and the stationary phase (Katritzky, 1994).

The results suggest that the retention of CPs in the DB-5 column not only involves molecular size, but also physicochemical property interactions with the stationary phase. It is possible to employ the developed model for both further structural identification and more accurate quantitative measurement of these substances.

The robustness of the regression equation obtained in this study was evaluated by calculating the cross-validated correlation coefficient R^2_{crossval} (Walpole, 1990). Internal jack-knifing technique was used to test the quality of a model. This methodology was conducted leaving out one compound from the data set and regenerating the model coefficients. The correlation of these data with those obtained from the principal model yields the R_{crossval} .

In this study, the internal jack-knifing validation for the DB-5 model gave a R_{crossval} value of 0.981, meaning that the model was able to predict relative retention times of CPs. The radar graph of R^2 (adjusted) from 16 times leave-one-out is given in Fig.2. There are 16 dots between 0.960 and 0.980, which account for 94 percent of the whole data set. Results showed that the established model had a good quality.

3 Conclusions

Results of this study demonstrated that the QSRR can generate a very good model for the relative retention times of CPs in a DB-5 column. The model could be applicable for predict the RRT of CPs. Octanol-water partition coefficients ($\log K_{ow}$) and the number of chlorine and oxygen (NCIO) were found to be important descriptors. The descriptors give information related to the different molecular properties, which can participate in the physicochemical process in gas chromatographic retention of chlorinated phenols.

References:

- Ballschmiter K, Zell M, 1980. Analysis of polychlorinated biphenyls(PCB) by glass capillary gas chromatography[J]. Fresenius Z Analytic Chemistry, 302: 20—31.
- Butte W, Fookan C, Klussmann R *et al.*, 1981. Evaluation of lipophilic properties for a series of phenols using reversed-phase high performance liquid chromatography and high-performance thin-layer chromatography[J]. Journal of Chromatography, 214: 59—67.
- Charles V E, 1986. Application of reverse-phase h.p.l.c. for the determining of partition coefficients[J]. Pesticide Science, 17: 311—325.
- Chen J, Liao Y, Zhao Y *et al.*, 1996. Quantitative structure-activity relationships and mixture toxicity studies of heterocyclic nitrogen compounds[J]. Bulletin of Environmental Contamination and Toxicology, 57: 77—83.
- Dai J Y, Yao S C, Ding Y B *et al.*, 1999. Retention of substituted indole compounds on RP-HPLC: Correlation with molecular connectivity indices and quantum chemical descriptors[J]. Journal of Liquid Chromatography and Related Technologies, 22(15): 2271—2282.
- Dewar M F S, Zoebisch E G, Healy E *et al.*, 1985. AM1: a new general purpose quantum mechanical molecular model[J]. Journal of the American Chemical Society, 107F: 3902—3909.
- Eadsforth C V, Moser P, 1983. Assessment of reverse-phase chromatographic methods for determining partition coefficients[J]. Chemosphere,

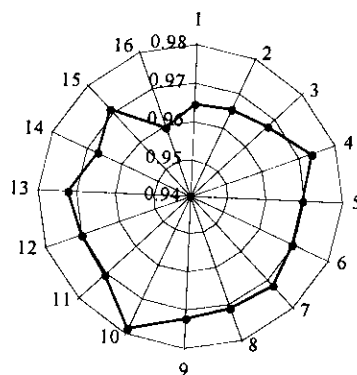


Fig.2 Radar graph of R^2 (adj.) from 16 times leave one out

- 12(11–12): 1459–1475.
- Frame G M, 1997. A collaborative study of 209 PCB congeners and 6 aroclors on 20 different HRGC columns[J]. *Fresenius Z Analytic Chemistry*, 357, 701–713.
- Franke R, 1984. *Theoretical drug design methods*[M]. New York: Elsevier. 115–123.
- Gerhard R, 1990. Gas chromatographic studies of chlorinated phenols, chlorinated anisoles, and chlorinated phenylacetates[J]. *Toxicological and Environmental Chemistry*, 27: 217–224.
- Gergakopoulos C, J Kiburis, P Jurs, 1991. Prediction of gas chromatographic relative retention times of stimulants and narcotics[J]. *Analytical Chemistry*, 63: 2021–2024.
- Ghose A, A Pritchett, G Crippen, 1988. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions[J]. *Journal of Computational Chemistry*, 9: 80–90.
- Hotaka S, Masato S, Tadayoshi S *et al.*, 1991. In vitro cytotoxicity of chlorophenols to goldfish gf-scale (GFS) cells and quantitative structure-activity relationships[J]. *Environmental Toxicology and Chemistry*, 10: 235–241.
- Hong H, Zhou D, Han S K *et al.*, 1996. Retention of aromatic sulfur containing compounds on RP-HPLC correlation with partition coefficient and molecular connectivity indices[J]. *Journal of Liquid Chromatography and Related Technologies*, 19(10): 1617.
- Hong H, Wang L S, 1996. Prediction adsorption coefficients (K_{oc}) for aromatic compounds by HPLC retention factors (K') [J]. *Chemosphere*, 32(2): 343–351.
- Järnberg U, Asplund L, Jakobson E, 1994. Gas chromatographic retention behavior of polychlorinated naphthalenes on non-polar, polarizable, polar and smectic capillary columns[J]. *Journal of Chromatography A*, 683: 385–396.
- Kang J, C Cao, Z Li, 1998. Quantitative structure-retention relationship studies for predicting the gas chromatography retention indices of polycyclic aromatic hydrocarbons[J]. *Journal of Chromatography A*, 799: 361–367.
- Katritzky A, E Gordeeva, 1993. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research[J]. *Journal of Chemical Information & Computer Science*, 33: 835–857.
- Katritzky A, E Ignatchenko, R Barcock *et al.*, 1994. Prediction of gas chromatographic retention times and response factors using a general quantitative structure-property relationship treatment[J]. *Analytical Chemistry*, 66: 1799–1807.
- Katritzky A, Y Wang, S Sild *et al.*, 1998. QSPR studies on vapor pressure, aqueous solubility, and the prediction of water-air partition coefficients[J]. *Journal of Chemical Information & Computer Science*, 38: 720–725.
- Kurunthachalam K, Takashi I, Alan L B *et al.*, 1998. Isomer-specific analysis and toxic evaluation of polychlorinated naphthalenes in soil, sediment and biota collected near the site of a former chlor-alkali plant[J]. *Environmental Science & Technology*, 32: 2507–2514.
- Marius D' A, Toshihiko H, 1982. Hydrophobicity and retention in reversed phase liquid chromatography [J]. *Journal of Liquid Chromatography and Related Technologies*, 5(2): 229–244.
- Neter J, W Wasserman, M Kutner, 1985. *Applied linear statistical models* [M]. 2nd ed Homewood, IL: Richard D. Irwin.
- Osmidowski K, J Halkiewicz, A Radecki *et al.*, 1985. Quantum chemical parameters in correlation analysis of gas-liquid chromatographic retention indices of amines[J]. *Journal of Chromatography*, 346: 53–60.
- Payares P, D Diaz, J Olivero *et al.*, 1997. Prediction of the gas chromatographic relative retention times of flavonoids from molecular structure [J]. *Journal of Chromatography A*, 771: 213–219.
- Robbat A, C Kalogeropoulos, 1990. Prediction of gas chromatographic retention indexes for polychlorinated dibenzofurans [J]. *Analytical Chemistry*, 62: 2684–2688.
- Robert M C, Robert E. C, Herbert L K, 1975. Determination of partition coefficients by liquid chromatography [J]. *Journal of Chromatography*, 107: 219–223.
- Stanton D, P Jurs, 1989. Computer-assisted prediction of gas chromatographic retention indices of pyrazines[J]. *Analytical Chemistry*, 61: 1328–1332.
- Stewart J J P, 1989. Mopac (Ver.6.0) manual [E]. Frank J Seiler. Research Laboratory, U.S. Air Force Academy, Co80840.
- Thomsen M, Rasmussen A G, Carlsen L, 1999. SAR/QASR approaches to solubility, partitioning and sorption of phthalates [J]. *Chemosphere*, 38: 2613–2624.
- Wei L, Yu H, Cao J *et al.*, 1999. Determination and prediction of partition coefficient and toxicity for sulfonylurea herbicides and their degradation products[J]. *Chemosphere*, 38: 1713–1719.
- Wolozyn T, P Jurs, 1992. Quantitative structure-retention relationship studies of sulfur vesicants[J]. *Analytical Chemistry*, 64: 3059–3063.
- Xu S, Jiang X, Dong Y Y *et al.* 2000. Polychlorinated organic compounds in Yangtze River sediments[J]. *Chemosphere*, 41: 1897–1903.
- Zhou Z, R G Parr, 1990. Activation hardness: new index for describing the orientation of electrophilic aromatic substitution[J]. *Journal of the American Chemical Society*, 112: 5720–5724.