Chemosphere 91 (2013) 229–232

Contents lists available at SciVerse ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

QSPR modeling of *n*-octanol/air partition coefficients and liquid vapor pressures of polychlorinated dibenzo-*p*-dioxins

Xiao-Lan Zeng *, Xiu-Lan Zhang, Yan Wang

College of Chemistry and Chemical Engineering, Xinyang Normal University, Henan, Xinyang 464000, People's Republic of China

HIGHLIGHTS

- ▶ Molecular geometries of 75 PCDDs were optimized using DFT method.
- ▶ Two QSPR models for PCDDs were established.
- ▶ Two established QSPR models exhibit optimum stability and good predictive power.
- ▶ Two derived QSPRs models can be used to predict log K_{OA} and log P_L for PCDDs.

ARTICLE INFO

Article history:
Received 26 July 2012
Received in revised form 9 December 2012
Accepted 15 December 2012
Available online 26 January 2013

Keywords:
Polychlorinated dibenzo-*p*-dioxins (PCDDs)
n-Octanol/air partition coefficient (log K_{OA})
Subcooled liquid vapor pressure (log P_L)
Quantitative structure–property relationship (QSPR)
DFT method

ABSTRACT

The molecular geometries of 75 polychlorinated dibenzo-*p*-dioxins (PCDDs) were optimized using B3LYP/6-31G* method. The calculated structural parameters were taken as theoretical descriptors to establish two new novel QSPR models for *n*-octanol/air partition coefficients (log K_{OA}) and subcooled liquid vapor pressure (log P_L) of PCDDs. The R^2 values of the two models are 0.983 and 0.985 respectively. Their standard deviations of prediction in modeling (*SD*) are 0.174 and 0.230 respectively. The results of leave-one-out (LOO) cross-validation for training set show that the two models exhibited optimum stability and good predictive power.

1. Introduction

Polychlorinated dibenzo-*p*-dioxins (PCDDs) are typical persistent organic pollutants that are formed during some natural processes like volcanic eruptions and forest fires (Gribble, 1994; Freeman and de Tejada, 2002) and various industrial processes (Zheng et al., 1999; Hylander et al., 2003; Chang et al., 2004; Chen, 2004; Coutinho et al., 2006). As generally considered, PCDDs are comprised of 75 individual congeners with various degrees of chlorination and substitution patterns at available sites. They are ubiquitous and have been detected throughout the world in various environmental compartments. Due to their persistence and high lipophilicity, PCDDs are bioaccumulated and biomagnified through food chain into different trophic levels of organisms, and subsequently into humans (Smith and Gangolli, 2002; Naito et al., 2003).

The environmental fate of persistent organic pollutants such as PCDDs depends on strongly on their distributions between different environment compartments. The liquid vapor pressure (P_L) and *n*-octanol/air partition coefficients (K_{OA}) play a major role in their transport and mobility in the global environment. The liquid vapor pressure (P_L) describes air–surface exchange property of POPs, which is able to describe volatility of non-polar organic pollutants successfully (Xiao and Wania, 2003). Another significant partition coefficient, $K_{OA,}$ is the ratio of the solute concentration in the octanol to its concentration in air when the octanol–air system is at equilibrium. It is a useful descriptor of a chemical's mobility in the atmospheric environment and thought to be the key descriptor of the absorptive partitioning of semivolatile compounds such as PCDDs between the atmosphere and organic phases found in soil, in vegetation, and on aerosols (Harner and Shoeib, 2002). Recently, some models employed to screen and rank chemicals for persistence and long-range transport potential have incorporated K_{OA} (Chen et al., 2002; Meylan and Howard, 2005; Zhao et al., 2010).

The physicochemical properties of organic compounds are important in determining their distribution and fate in the environment. However, the significant amount of time, expense, and safety precautions are required to measure their distributions between different environment compartments directly, thus accurate esti-

---

* Corresponding author. Tel.: +86 376 6390702.
E-mail address: zxl6688@163.com (X.-L. Zeng).

0045-6535/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chemosphere.2012.12.060

mation or prediction of environmental distribution coefficients are essential for any reliable ecological model. On the other hand, a statistically significant quantitative structure–property relationship (QSPR) model that requires only chemical structure data could serve as a tool to predict reliable P_L and K_{OA} data in a fraction of the time and expense. Many studies showed the method could be well used to develop the predictive models for *n*-octanol/air partition coefficients and liquid vapor pressure (Puzyn and Falandysz, 2005; Zhao et al., 2005; Puzyn et al., 2006; Xu et al., 2007; Zeng et al., 2007; Mostrag et al., 2010). The structural descriptors derived from quantum-chemical calculation have been successfully applied in QSPR researches. In this paper, the molecular properties computed using density functional theory (DFT) method were taken as theoretical descriptors being incorporated into QSPRs for log K_{OA} and log P_L of all PCDDs. The models obtained could be applied to extend the current database of experimental determinations.

2. Materials and methods

All possible 75 PCDD congeners together with dibenzo-*p*-dixoin are included in this study. There are experimental physical–chemical data only available for some of the PCDD congeners. Data on *n*-octanol/air partition coefficients (log K_{OA}) for 10 PCDDs at 298 K were taken from the literature (Harner et al., 2000) and the values range from 7.86 (1-MCDD) to 11.42 (1,2,3,4,6,7,8-Hepta-CDD). Data on subcooled liquid vapor pressures (log P_L) for 15 PCDDs were taken from another literature (Mackay et al.,1992) and their values range from −0.291 (DD) to −6.751 (1,2,3,4,6,7,8-Hepta-CDD). These compounds (10 with log K_{OA} data and 15 with log P_L data) were selected as training sets and the remaining compounds without experimental data were selected as predictive sets respectively. These data, together with the chemical names and the predicted properties from QSPR models for all 75 PCDD congeners, were collected in Table A1 (see Appendix A) to aid discussion.

Density functional theory (DFT) is a quantum mechanical modeling method based on the Hohenberg–Kohn theorem (Hohenberg and Kohn, 1964) and the Kohn–Sham method (Kohn and Sham, 1965) and is used to calculate the ground-state electronic energy E_0 and other ground-state molecular properties from the ground-state electronic density ρ_0 instead of the electronic wave function. From the mid-1990s the B3LYP hybrid functional (Lee et al., 1988; Becke, 1993) has been the most widely used for molecular calculations due to the accuracy of the B3LYP results obtained for a large range of compounds, particularly organic molecules. Recent studies (Puzyn and Falandysz, 2007; Wei et al., 2007; Xie et al., 2007; Zeng et al., 2007; Puzyn et al., 2009; Piliszek et al., 2011, 2012a,b; Wilczyńska-Piliszek et al., 2012a,b,c,d,e) indicated that B3LYP functional is also adequate to calculate molecular descriptors. In this article, the geometries of 75 PCDD congeners were optimized with density functional theory (DFT) method at the B3LYP/6-31G* level (Ditchfield et al., 1971; Hehre et al., 1972; Hariharan and Pople, 1973, 1974; Gordon, 1980) and frequency calculations were performed at the same level to ensure they were minimal on the potential energy surface. All calculations were carried out with Gaussian 98 program (Frisch et al., 1998).

The structural descriptors employed in this work, such as polarizability (α), dipole moment (μ), energy of the highest occupied molecular orbital (E_{HOMO}), energy of the lowest unoccupied molecular orbital (E_{LUMO}), negative atomic partial charge on oxygen atom (q_O^-), and most positive partial charge on a hydrogen atom (q_H^+), are all obtained directly from the Gaussian output files.

In the present work, the QSPR equations were obtained by forward stepwise multiple linear regression technique. To determine the optimum number of components for the correlation models, the leave-one-out (LOO) cross-validation technique has been applied to validate the stability and predictive ability of constructed models. Based on the residuals, it is possible to calculate the cross-validation coefficient R^2_{CV}, which can assess internal predictive ability of the model. In this study, the conventional correlation coefficient (R^2) and the standard deviations of prediction (*SD*) are used to characterize the fitness of the QSPR models, while the LOO cross-validation correlation coefficient R^2_{CV} and the root mean square error of cross-validated (*RMSEP*) were used to measure the robustness and predictive power. The statistic significance of the model was validated by *t*-tests.

Moreover, all correlations reported in this paper adhere to two conditions. First, this cross correlation of descriptors is minimized, as determined by the variance inflation factor (*VIF*). Second, all independent variables in the regressions have a confidence degree at least 95% (based on *F*-test). If *F*-values are evidently higher than critical values $F_{0.05}$ and then suggests that model is highly significant.

The applicability domains of the models were explored by calculating both standardized residuals and leverages (h) for chemicals as suggested in recent minireview (Gramatica, 2007). If a compound has a standardized residual value larger than three or has a value of h greater than the critical leverage ($h^* = 3p'/n$, where p' is the number of model variables plus one, and n is the number of the objects used to calculate the model), it can be considered outside the applicability domain of the model. However, in the assessment of a chemical for which there is no experimental value, the conclusion can only be based on the leverage values.

3. Results and discussion

3.1. QSPR model of log K_{OA}

In order to work out QSPR model for predicting log K_{OA} of PCDDs, the stepwise multiple linear regression analysis was performed for the training set on the experimental data of log K_{OA} and all structural parameters using SPSS 13.0 for windows. One-parameter optimal equation was obtained as the following equation:

$$\begin{aligned}&\log K_{OA} = (1.007 \pm 0.414) + (0.049 \pm 0.002)\alpha\\&n = 10, R^2 = 0.983\ (R = 0.991), SD = 0.174, \quad F = 452.368, P = 0.000\ \text{(Modeling)}\\&R^2_{CV} = 0.959, RMSEP = 0.25 \quad \text{(LOO prediction)}\end{aligned} \tag{1}$$

As it can be observed from Eq. (1), the non-cross-validated R^2 = 0.983 and F = 452.368 indicted that the log K_{OA} model is highly satisfactory. Next, the model created was validated internally, the cross-validated $R^2_{CV}\left(R^2_{CV} = 0.959\right)$ is highly significant and calculated root mean square error of cross-validation *RMSEP* (*RMSEP* = 0.25) is low which confirmed goodness-it-fit, robustness and the good stability of the model. The plot of observed vs. predicted values of log K_{OA} (see Fig. 1) also confirmed the linear character of the model and its very good fitting.

Based on the calibrated model, the values of log K_{OA} for 75 PCDD congeners were predicted and the results are presented in Table A1. Their corresponding residual values are also listed in Table A1, in which the maximum discrepancy of 0.27 log units is found for 1,2,3,4,7-Penta-CDD for the training set, the absolute average residual values is 0.13. When confidence degree is 95%, *F*-values (452.368) are evidently higher than critical values $F_{0.05}$ (5.318), which shows that the model is highly significant. As Eq. (1) has high correlation coefficients (R^2 and R^2_{CV} values) and small standard deviation, it can be used to predict log K_{OA} of all PCDDs excellently.

Observation of Eq. (1) may lead to the following significant interpretations: Increasing α value of PCDD molecule leads to increasing the log K_{OA} values. This is because intermolecular dispersive forces are directly proportional to the product of α for two interacting molecules, and molecules with great α values