Contents lists available at ScienceDirect

# Fluid Phase Equilibria



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

## Quantitative structure-property relationship study of liquid vapor pressures for polychlorinated diphenyl ethers



Ye Yuan<sup>a</sup>, Yongyuan Sun<sup>a</sup>, Daxiang Wang<sup>a</sup>, Ruina Liu<sup>a</sup>, Shaojin Gu<sup>a</sup>, Guijie Liang<sup>b</sup>, Jie Xu<sup>a,\*</sup>

<sup>a</sup> College of Materials Science & Engineering, State Key Laboratory Cultivation Base for New Textile Materials & Advanced Processing Technology, Wuhan Textile University, 430200 Wuhan, China

<sup>b</sup> Hubei Key Laboratory of Low Dimensional Optoelectronic Materials and Devices, Hubei University of Arts and Science, 441053 Xiangyang, China

#### ARTICLE INFO

Article history: Received 14 July 2014 Received in revised form 13 January 2015 Accepted 20 January 2015 Available online 23 January 2015

Keywords: Quantitative structure-property relationship (QSPR) Polychlorinated diphenyl ethers (PCDEs) Liquid vapor pressure

## ABSTRACT

A quantitative structure-property relationship (QSPR) study was performed to correlate the logarithmic values of liquid vapor pressure of polychlorinated diphenyl ethers (PCDEs) with their molecular structures. The entire set of 106 PCDEs and diphenyl ether was divided into a training set of 72 samples and a test set of 35 samples using the DUPLEX algorithm. An extended set of molecular descriptors was calculated to represent the molecular structures by DRAGON software. Multiple linear regression (MLR) was used to select descriptors and develop models. A three-descriptor equation was obtained for the training set, with a squared correlation coefficient ( $R^2$ ) of 0.997 and a standard error(s) of 0.069. The robustness and predictive performance of the proposed model were assessed by different approaches, including leave-many-out cross-validation, Y-randomization test, and external validation through test set. Satisfactory results of  $R^2$  = 0.995 and a mean absolute error of 0.066 for the test set confirmed the model being very useful to predict the vapor pressure of PCDEs. Furthermore, the applicability domain of the models was analyzed based on the Williams plot.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Polychlorinated diphenyl ethers (PCDEs) are a group of halogenated aromatic compounds (Fig. 1), which are structurally located between polychlorinated biphenyls (PCBs) and polychlorinated dibenzofurans (PCDFs). They are mainly originated from by-products of technical chlorophenols and chlorinated phenoxyacetic acids, fly ashes, transformer fluids, wood preservatives, incomplete combustion intermedia, and flues of municipal waste incinerators [1]. However, their ubiquitous environmental occurrence is basically the result of their presence as impurities in chlorophenol preparations, where they have been identified at levels of 100-1000 mg/kg [2]. They have been detected in a wide range of environmental and biological samples including sediments, fish, birds, marine mammals, edible marine organisms, and human adipose tissues [3-6]. Previous studies have shown that PCDEs have similar toxic properties to PCBs [7]. They can induce cytochrome P-450-dependent monoxygenase activity and have a broad spectrum of toxicity [2]. Therefore, they

http://dx.doi.org/10.1016/j.fluid.2015.01.021 0378-3812/© 2015 Elsevier B.V. All rights reserved. are also regarded as a type of persistent indicator molecules for a global pollution of the environment by organochlorine compounds [8]. PCDEs have received more and more concerns surrounding their potential to persist and bioaccumulate in the environment.

Liquid vapor pressure  $(P_{I})$  is an important factor for assessing the transport, distribution, and fate of organic pollutants in the environment. For example, the P<sub>L</sub> of organic pollutants determines their distribution between the soil and the atmosphere. Pollutants with  $P_{\rm L}$  of less than  $10^{-5}$  Pa exist almost entirely as molecules adsorbed on the solid airborne particles, while compounds with  $P_{\rm L}$ between  $10^{-2}$  and  $10^{-5}$  Pa tend to concentrate more in a gas phase rather than airborne solid particles, soil or water [9]. The P<sub>L</sub> data are also used for the estimation of air-water partition coefficient, enthalpy of vaporization, flash point, and some other important physicochemical properties of compounds [10]. In addition, the P<sub>L</sub> plays a significant role in gas separation, fire and explosion prevention, process engineering and control [11]. However, experimental  $P_1$  measurement of the ever-growing number of actual and potential chemicals is both time-consuming and expensive.

Alternatively, the quantitative structure-property relationship (QSPR) approach is highly promising to estimate the  $P_L$  based on

<sup>\*</sup> Corresponding author. E-mail address: xujie0@ustc.edu (J. Xu).



Fig. 1. The generic structure of PCDEs.

descriptors derived solely from the molecular structure to fit experimental data. The OSPR approach 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 [12–20]. The advantage of this approach 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 and validated, it can be used for the prediction of the property of new compounds that have not been synthesized or found. Thus, the QSPR approach can expedite the process of development of new molecules and materials with desired properties. To support this development, Organization for Economic Co-operation and Development (OECD) drawn up the following principles for the validation of QSPR models [21]: (i) a defined endpoint, (ii) an unambiguous algorithm, (iii) a defined domain of applicability, (iv) appropriate measures of goodness-offit, robustness and predictive power and (v) a mechanistic interpretation, if possible.

Many QSPR models have been already developed for the  $P_{\rm L}$ [1,9,22–32] and, some of them, for PCDEs. Liang and Gallagher [22] proposed a OSPR approach to predict logarithmic units of  $P_{\rm I}$  $(\log P_{\rm I})$  from computationally derived molecular descriptors for a set of 479 compounds including acids, alcohols, aldehydes, alkanes, alkenes, alkynes, amines, aromatic compounds and polycyclic aromatic hydrocarbons (PAHs), dibenzofurans, dioxins, ethers, esters, ketones, nitriles and other nitrogen-containing compounds, polychlorinated biphenyls (PCBs), and sulfur-containing compounds. The authors developed a seven-descriptor model with  $R^2 = 0.960$ ,  $Q^2 = 0.957$  and s = 0.534 using multiple linear regression (MLR) method. Chen et al. [23] reported a QSPR study on the prediction of  $\log P_{\rm L}$  for the polychlorinated dibenzo-pdioxins and dibenzofurans (PCDD/Fs) using partial least-squares (PLS) regression. The obtained model with 12 descriptors and one latent vector was characterized by satisfactory goodness-of-fit  $(R^2 = 0.986 \text{ and } Q^2 = 0.963)$ . Katritzky et al. [24] used the MLR method to predict  $\log P_{\rm L}$  for hydrocarbons, halogenated hydrocarbons, O- and N-containing compounds. A five-descriptor equation with  $R^2 = 0.937$ ,  $Q^2 = 0.936$  and s = 0.366 was obtained. Öberg [28] also published a study on predicting  $\log P_{\rm L}$  of halogenated diphenyl ether congeners (PCDEs and PBDEs) using molecular descriptors. Zeng et al. [25] established a QSPR model with  $R^2 = 0.991$  and s = 0.112 to predict  $\log P_L$  of PCDEs using the position of Cl substitution (PCS) methods. All these models were characterized by good  $R^2$  and s; but few of them fit all OECD standards and can be used.

This paper aimed to develop fully validated QSPR models to predict the  $P_{\rm L}$  of 106 PCEDs respecting all OECD principles including the determination of their applicability domains. Molecular structures of PCDEs were modeled by molecular mechanics and molecular descriptors were derived from the optimized structures to correlate with the log  $P_{\rm L}$ . A QSPR model was obtained using the MLR method. The robustness and

predictive ability of the obtained model were fully assessed by different approaches, including leave-many-out cross-validation, Y-randomization test, and external validation through test set. Additionally, the applicability domain (AD) of the model was analyzed based on the Williams plot.

## 2. Materials and method

## 2.1. Dataset

The experimental values of  $\log P_L$  (Pa, 25 °C) of the 106 PCDEs and diphenyl ether were taken from Ref. [8], which ranged from 0.38 to -5.80 logarithmic units (Table 1).

## 2.2. Descriptor generation

The chemical structure of each compound was sketched using the Hyperchem program [33] and preoptimized using MM+ molecular mechanics method (Polak-Ribiere algorithm). 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 0.03 kcal  $Å^{-1}$  mol<sup>-1</sup> as a stopping criterion. Then totally 1664 molecular descriptors for each compound were calculated from the optimized geometries using the DRAGON software [34]. These descriptors include (1) 0D-constitutional (atom and group counts); (2) 1D-functional groups and atom centered fragments; (3) 2D-topological, BCUTs, walk and path counts, autocorrelations, connectivity indices, information indices, topological charge indices, and eigenvalue-based indices; and (4) 3D-Randic molecular profiles from the geometry matrix, geometrical, WHIM, and GETAWAY descriptors.

To reduce redundant and useless information, constant or near constant values and descriptors found to be highly correlated pairwise (one of any two descriptors with a correlation greater than 0.99 [35]) were removed. Furthermore, the rug-like descriptors (including GVWAI-80, Neoplastic-80, and Infective-80) were excluded. Finally, 575 descriptors were remained to undergo subsequent descriptor selection.

## 2.3. Dataset splitting

The dataset was divided into training and test sets by the DUPLEX algorithm [36] combined with principal component analysis (PCA) due to its easy implementation and its good performance in the selection of representative training set samples. This algorithm proceeds as follows: first the two points which are furthest away from each other are selected for the training set; from the remaining points, the two objects which are furthest away from each other are included in the test set; then the remaining point which is furthest away from the two previously selected for the training set is included in the training set. The procedure is repeated selecting a single point for the test set which is furthest from the existing points in that set. Following the same procedure, points are added alternately to each set. Finally, points representing both training and test set were distributed uniformly within the whole space which is occupied by the entire dataset. Therefore, it guarantees that the composition of the training set and the test set is representative, at the same time avoids the unbalance of the two datasets. The PCA was first performed based on the 575 descriptors of the complete dataset, and the obtained principal components were put into the DUPLEX algorithm to select a training set of 72 compounds and a test set of 35 compounds.

Download English Version:

https://daneshyari.com/en/article/201825

Download Persian Version:

https://daneshyari.com/article/201825

Daneshyari.com