Available online at www.sciencedirect.com
ScienceDirect
Chemosphere 69 (2007) 469–478
CHEMOSPHERE
www.elsevier.com/locate/chemosphere

QSPR prediction of *n*-octanol/water partition coefficient for polychlorinated biphenyls

Wenjuan Lü, Yonglei Chen, Mancang Liu *, Xingguo Chen, Zhide Hu

Department of Chemistry, Lanzhou University, Lanzhou 730000, Gansu Province, China

Received 2 February 2007; received in revised form 11 April 2007; accepted 12 April 2007
Available online 12 June 2007

Abstract

The logarithmic *n*-octanol/water partition coefficient ($\log K_{ow}$) is a very important property which concerns water-solubility, bioconcentration factor, toxicity and soil absorption coefficient of organic compounds. Quantitative structure–property relationship (QSPR) model for $\log K_{ow}$ of 133 polychlorinated biphenyls (PCBs) is analyzed using heuristic method (HM) implemented in CODESSA. In order to indicate the influence of different molecular descriptors on $\log K_{ow}$ values and well understand the important structural factors affecting the experimental values, three multivariable linear models derived from three groups of different molecular descriptors were built. Moreover, each molecular descriptor in these models was discussed to well understand the relationship between molecular structures and their $\log K_{ow}$ values. The proposed models gave the following results: the square of correlation coefficient, R^2, for the models with one, two and three molecular descriptors was 0.8854, 0.9239 and 0.9285, respectively.

Keywords: Polychlorinated biphenyls; Quantitative structure–property relationship; *n*-octanol/water partition coefficient; Heuristic method

1. Introduction

Polychlorinated biphenyls (PCBs) are persistent organic contaminants and widespread environmental pollutants; they were used in plastifiers formulation of coatings, inks, adhesives, flame retardants, pesticide extenders and in the micro-encapsulation of dyes for carbonless duplicating papers. Human exposure to PCBs occurs mainly from eating food that contains these chemicals (Schecter et al., 2001; Juan et al., 2002; Kiviranta et al., 2002, 2004; Erdogrul et al., 2005; Coelhan et al., 2006). It has been reported that meat, dairy products, and fish, makes up more than 90% of the intake of PCBs for the general population (Schecter et al., 1997; Llobet et al., 2003a,b; Bocio and Domingo, 2005; Huwe and Larsen, 2005). PCBs are rapidly absorbed from the gastrointestinal tract, and are distributed and accumulated in the liver and adipose tissues. The compounds have serious ecologically harmful effects and are implicated as potent carcinogens. Thus, there is a need for prediction tools to study PCBs' properties including retention behavior, properties, and activity/toxicity for which analytical standards are currently difficult to obtain, but yet for which environmental data are needed.

The logarithmic *n*-octanol/water partition coefficient ($\log K_{ow}$) is an important property for pharmacology, toxicology and medicinal chemistry. There have been many reports on the prediction of $\log K_{ow}$ for PCBs, most of the reported prediction methods are based on thermodynamically oriented theories (Kamlet et al., 1988; Banerjee and Howard, 1988), connectivity indexes (Randic, 1975; Murray et al., 1975; Hawker and Connell, 1988; Patil, 1991; Sabljic et al., 1993) and characteristic root index (CRI) (Melek and Inel, 1995). Three-dimensional structure–property correlations for prediction of thermodynamic properties of PCBs have been recently made to predict the enthalpy of vapourization and enthalpy of sublimation (Puri et al., 2002). The octanol/water partition coefficient expressed as $\log K_{ow}$ is an important property for various applications in pharmacology, toxicology and

* Corresponding author. Tel.: +86 931 8911486; fax: +86 931 8912582.
E-mail address: chyl@lzu.edu.cn (M. Liu).

0045-6535/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2007.04.044

medicinal chemistry(Leo, 1993). $\log K_{ow}$ is used to model partitioning of chemicals between the lipophilic membrane and the relative hydrophobic cellular cytoplasmic material. $\log K_{ow}$ quantities hydrophobicity of chemicals and is important both for predicting pharmacokinetics and pharmacodynamics of drugs and toxicants (Klopman et al., 1994). Lipophilicity is traditionally measured in the octanol/water system. $\log K_{ow}$ values have been shown to be generally satisfactory for modelling protein binding and lipophilic interactions with biological membranes consisting largely of protein (Platts et al., 1999). Several methods have been described in the literature for the estimation of the octanol/water partition coefficient, which is $\log K_{ow}$ (Platts et al., 2000; Khadikar et al., 2002).

Here, based on the octanol/water partition coefficient of 133 polychlorinated biphenyl (PCBs) congeners, we report a QSAR model by the Heuristic method of CODESSA (comprehensive descriptors for structural and statistical analysis) technique. The purpose of the present study was to investigate the relationship between the octanol/water partition coefficient of 133 polychlorinated biphenyls (PCBs) and their molecular parameters. Moreover, molecular descriptors were discussed to explore the influence of structural features on the values of $\log K_{ow}$. This paper provided a simple and straightforward way to predict the $\log K_{ow}$ values of PCBs from their structures and gave some insight into structural features related to the $\log K_{ow}$ values of the compounds. The prediction results are satisfactory in all the three groups.

The software CODESSA, developed by Katritzky group (Katritzky et al., 1995; Katritzky et al., 1995–1997), has been applied successfully in a variety of QSAR analyses (Oblak et al., 2000; Katritzky and Tatham, 2001). It can calculate a comprehensive set of descriptors: constitutional descriptors, topological descriptors, geometrical descriptors, electrostatic descriptors, and semi-empirical quantum chemical descriptors (in this work, 367 descriptors were calculated).

2. Materials and methods

2.1. Data set

All data of the present investigation were obtained from the reference (Padmanabhan et al., 2006). The data set for this investigation consisted of 133 PCBs. The geometry of biphenyl template is depicted along with atom numbering (Fig. 1).

2.2. Molecular descriptor generation

To obtain a QSRR model, compounds are often represented by the molecular descriptors. The calculation process of the molecular descriptors was described as below:

The two-dimensional molecular structures of 133 PCBs were drawn by Chem draw 6.0 and were geometrical optimized by the semi-empirical AM1 method in Hyperchem 5.0 software. Then calculated in the software MOPAC (Stewart, 1989), then the output MOPAC files were transferred into software CODESSA to calculate all kinds of descriptors. The software CODESSA can calculate constitutional, topological, geometrical, electrostatic, and quantum chemical descriptors and has been successfully used in various QSPR researches (Hiob and Karelson, 2000; Katritzky et al., 2001; Delgado et al., 2003; Bosque and Sales, 2003; Hu et al., 2005).

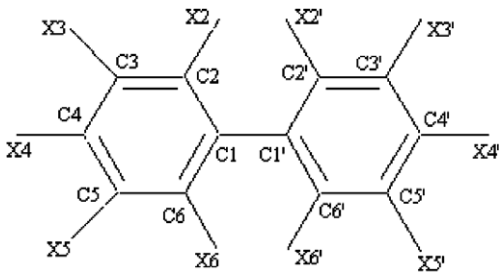

Fig. 1. Geometry of biphenyl template with atom numbering.

Constitutional descriptors are related to the number of atoms and bonds in each molecule. Topological descriptors include valence and non-valence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition, and the degree of branching of a molecule. The topological descriptors describe the atomic connectivity in the molecule. The geometrical descriptors describe the size of the molecule and require 3D-coordinates of the atoms in the given molecule. The electrostatic descriptors reflect characteristics of the charge distribution of the molecule. The quantum chemical descriptors offer information about binding and formation energies, partial atom charge, dipole moment, and molecular orbital energy levels. In the present work, five classes of structural descriptors were obtained and about 367 descriptors were provided for every compound.

2.3. Theory of CODESSA

CODESSA includes two advanced procedures for systematic development of multi-linear QSAR/QSPR equations: (i) the Heuristic method, and (ii) the best multi-linear regression method. The heuristic method for descriptor selection proceeds with a preselection of descriptors by sequentially eliminating descriptors which do not match any of the following criteria: (i) the F-test greater than one unit; (ii) R^2 value less than a value defined at the start (default 0.01); (iii) the student's t-test less than that defined (default 0.1); and (iv) duplicate descriptors having a higher squared inter-correlation coefficient than a predetermined level (usually 0.8). The next step involves correlation of the given property with (i) the top descriptor in the above list with each of the remaining descriptors, and (ii) the next one with each of the remaining descriptors, etc. The best pairs, as evidenced by the highest F-values in the