Quantitative structure–property relationships for pesticides in biopartitioning micellar chromatography

Weiping Ma [a], Feng Luan [a], Haixia Zhang [a], Xiaoyun Zhang [a,*],
Mancang Liu [a], Zhide Hu [a], Botao Fan [b]

[a] *Department of Chemistry, Lanzhou University, Lanzhou 730000, Gansu, PR China*
[b] *Université Paris 7-Denis Diderot, ITODYS 1, Rue Guy de la Brosse, 75005 Paris, France*

Abstract

The retention factor ($\log k$) in the biopartitioning micellar chromatography (BMC) of 79 heterogeneous pesticides was studied by quantitative structure–property relationships (QSPR) method. Heuristic method (HM) and support vector machine (SVM) method were used to build linear and nonlinear models, respectively. Compared the results of these two methods, those obtained by the SVM model are much better. For the test set, a predictive correlation coefficient (R) of 0.9755 and root-mean-square (RMS) error of 0.1403 were obtained. The proposed QSPR models, both by HM and SVM, contain the same descriptors that agree with the classical Abraham parameters of well-known linear solvation energy relationships (LSER).

© 2006 Elsevier B.V. All rights reserved.

Keywords: Heuristic methods; Quantitative structure–property relationship; Support vector machine; Biopartitioning micellar chromatography

1. Introduction

To control plague, pesticides embrace an enormous diversity of products that are used in a number of different activities. These include agriculture, amateur gardening, woodworm treatment, and public health applications. Due to the physicochemical properties of these chemical agents, such as water solubility, vapor pressure and partition coefficients between organic matter (in soil or sediment) and water, they can disperse in various environmental media. The range of damages across environmental media and different receptors is equally great, providing a particularly complex example of multidimensional environmental impacts. Loss of aquatic and terrestrial biodiversity, contamination of groundwater and agricultural produce, and poisoning of agricultural workers are among the potential consequences of pesticide use in agriculture alone. Therefore, there is a need of evaluating the toxicity of pesticides for risk assessment.

Chromatography has been established for years as the technique of choice for the analysis of pesticides [1,2]. As a mode of micellar liquid chromatography, biopartitioning micellar chromatography (BMC) is a simple and reproducible approach in emulating the partitioning of chemicals in biomembranes. Usually, it comprises a C_{18} reversed stationary phase and polioxyethylene (23) lauryl ether (Brij35) micellar mobile phases. The retention data obtained in this chromatographic system under adequate experimental conditions can be related to the biological behavior of different kinds of drugs [3–8] and the toxicity of chemicals [9–11]. This could be attributed to the fact that the characteristics of the BMC systems are similar to biological barriers and extracellular fluids. First, the stationary phase modified by hydrophobic adsorption of Brij35 surfactant monomers structurally resembles the ordered array of the membranous hydrocarbon chains. Second, the hydrophilic/hydrophobic character of the adsorbed surfactant monomers resembles the polar membrane regions. In addition, Brij35 micellar mobile phases prepared at the specific physiological conditions could also mimic the environment of partitioning.

Linear solvation energy relationships (LSER) methodology has been extensively applied in conventional reversed-phase liquid chromatography [12–18], gas chromatography [19,20], micellar liquid chromatography (MLC) [21–24] and micellar electrokinetic chromatography [21,25–28]. The general

* Corresponding author. Tel.: +86 931 891 2578; fax: +86 931 891 2582.
  E-mail address: xyzhang@lzu.edu.cn (X. Zhang).

solvation equation of LSER proposed by Abraham [29] is defined as:

$$\log SP = c + eE + sS + aA + bB + vV \tag{1}$$

In the equation, E is an excess molar refraction that is obtained from the refractive index. S is the dipolarity/polarizability that can be obtained from gas–liquid chromatographic measurements on polar stationary phases or more generally from water/solvent partitions. The parameters A and B are the overall or effective hydrogen bond acidity and basicity, respectively. V is the McGowan characteristic volume that can be calculated promptly from bond and atom contribution. These parameters represent the solute influence on various solute/solvent phase interactions. In the present paper, the descriptors were calculated from the pesticides structures by CODESSA. However, the interpretation of them is complicated by the fact that the system is commonly described using a three-phase model (mobile, stationary, and micellar phases) with three accompanying partition coefficients (mobile to stationary phase, mobile to micelle phase, and stationary to micelle phase transfers) [24]. LSER shows the fundamental chemical interactions governing the retention of MLC.

In this study, quantitative structure–property relationships (QSPR) method was performed to predict the chromatographic retention behavior of pesticides using a large number of calculated descriptors instead of Abraham parameters of LSER. To better understand the retention mechanism of BMC, the chemical meaning of the calculated descriptors was compared with that of Abraham parameters of LSER.

QSPR studies have been demonstrated to be an effective computational tool in understanding the correlation between the structure of molecules and their properties [30–32]. In a QSPR study one seeks to find a mathematical relationship between the property and one or more descriptors. Thus, this study can indicate which of the structural factors may play an important role in the determination of the property. And its advantage over other methods lies in the fact that the descriptors used to build the models can be calculated from the structure alone and are not dependent on any experimental properties. However, the main problems encountered in this kind of research are still the description of the molecular structure using appropriate molecular descriptors and selection of suitable modeling methods. At present, many types of molecular descriptors such as constitutional, topological, geometrical, electrostatic, and quantum-chemical descriptors have been proposed to describe the structural features of molecules [33–35]. The same as the diversity of molecular descriptors, different chemometrics and chemoinformatics methods, such as multiple linear regression (MLR), principal component regression (PCR), partial least squares (PLS), different types of artificial neural networks (ANN) and genetic algorithms (GA), can be employed to derive correlation models between the molecular structures and properties.

Recently, there is a growing interest in the use of support vector machine (SVM) to chemical problems due to its remarkable generalization performance in modeling non-linear problem.

SVM is a new algorithm developed from the machine learning community and has attracted attention and gained extensive application, such as pattern recognition problems [36–38], drug design [39], prediction of protein structure [40], identifying genes [41], quantitative structure–activity relationship (QSAR) [42], and QSPR analysis [43–45]. Nevertheless, to the best of our knowledge, there is no prediction of retention factor ($\log k$) of BMC by the QSPR approach based on SVM.

In the present work, HM and SVM were used for the prediction of $\log k$ of 79 pesticides in BMC using descriptors calculated and selected by the software CODESSA. The aim was to establish a QSPR model that could be used for the prediction of $\log k$, to show the flexible modeling ability of SVM, and, at the same time, to seek the important structure features related to the retention behavior of pesticides.

2. Experimental section

2.1. Data set

In our study, a set of 79 pesticides collected from ref [11] is investigated. A complete list of the pesticides' names and their corresponding experimental retention data ($\log k$) is given in Table 1. The retention factor (k) of pesticides was estimated by Eq. (2):

$$k = \left[\left(\frac{t_R}{t_{R(REF)}} \right) (1 + k_{REF}) - 1 \right] \tag{2}$$

where t_R is the experimental retention time of the pesticides assayed, $t_{R(REF)}$ the experimental retention time of a reference compound (acetanilide) injected during the working session and k_{REF} is the retention factor of acetanilide. Micellar mobile phases were prepared by dissolving Brij35 in aqueous solutions of 0.05 M phosphate buffer to get a final surfactant concentration of 0.04 M. The buffer solution was prepared with sodium dihydrogen phosphate. The pH was potentiometrically adjusted at 7.0 by addition of sodium hydroxide aqueous solution. The entire set of pesticides was divided into two subsets randomly: a training set of 63 compounds and a test set of 16 compounds. The training set was used to build the actual models and the test set was used to evaluate the models once they were built.

2.2. Descriptors calculation

All structures of the pesticides were drawn with the Hyper-Chem 4.0 program [46]. The final structural optimizations of pesticides were performed using the PM3 parameterization within the semi-empirical quantum-chemical program MOPAC 6.0 [47]. The geometry optimization was performed without symmetry restrictions. In all cases frequency calculations had been performed in order to ensure that all the calculated energy of the geometries correspond to true minima. The output files exported from MOPAC were transferred into the Microsoft Windows version of the CODESSA program developed by Katritzky et al. [48], to calculate molecular descriptors. Five types of molecular descriptors were calculated: constitutional,