Journal of Hazardous Materials 303 (2016) 137–144

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

Filling environmental data gaps with QSPR for ionic liquids: Modeling *n*-octanol/water coefficient

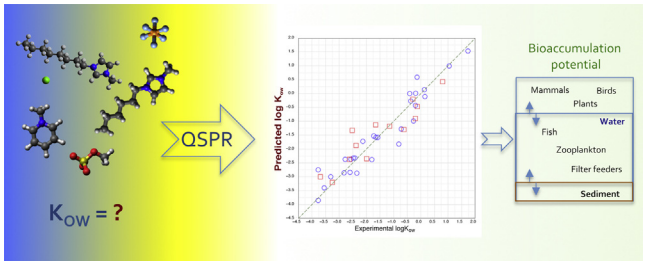
Anna Rybinska, Anita Sosnowska, Monika Grzonkowska, Maciej Barycki, Tomasz Puzyn *

Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland

HIGHLIGHTS

- We developed a QSPR model to predict the values of log K_{OW} for ionic liquids.
- Effect of the cation and anion structures on the modeled property was determinated.
- Increase in the length of alkyl chain in cation causes significant increase of the log K_{OW}.
- Majority of ILs could be transported with the water mass.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 June 2015
Received in revised form
22 September 2015
Accepted 12 October 2015
Available online 23 October 2015

Keywords:
Ionic liquids
ILs
QSPR
K_{OW}
Octanol–water partition coefficient

ABSTRACT

Ionic liquids (ILs) form a wide group of compounds characterized by specific properties that allow using ILs in different fields of science and industry. Regarding that the growing production and use of ionic liquids increase probability of their emission to the environment, it is important to estimate the ability of these compounds to spread in the environment. One of the most important parameters that allow evaluating environmental mobility of compound is *n*-octanol/water partition coefficient (K_{OW}). Experimental measuring of the K_{OW} values for a large number of compounds could be time consuming and costly. Instead, computational predictions are nowadays being used more often. The paper presents new Quantitative Structure–Property Relationship (QSPR) model that allows predicting the logarithmic values of K_{OW} for 335 ILs, for which the experimentally measured values had been unavailable. We also estimated bioaccumulation potential and point out which group of ILs could have negative impact on environment.

1. Introduction

Design of sustainable 'green' products is nowadays one of the most important challenges for chemical industry. New chemical materials should be not only useful and inexpensive, but also safe for human health and the environment. Ionic Liquids (ILs) – salts consisting of a large organic cation and a small inorganic anion, having their melting point lower than 100 °C – have being considered as 'green chemicals' for last few years. This is because ILs are characterized by low flammability, low vapor pressure, stability at high temperatures and ability to retain the liquid state for a wide range of temperatures. These properties, together with the possibility of easy modification of the structure, decide on employing ILs in many disciplines such as chemistry, biotechnology, chemical engineering and industry [1–3]. Nowadays synthesized ILs belong to the third generation of that compounds; they are designed to possess certain biological activity combined with selected physical properties. Ionic liquids, having antibacterial activity and being soluble in water, are examples of the third-generation ILs [4].

However, recent studies confirm negative impact of ILs on living organisms, when the organisms are exposed at those novel materials. Regarding that the increasing production and use of ILs increase probability of their emission to the environment, it is important to estimate the ability of these compounds to spread in the

* Corresponding author.
E-mail address: t.puzyn@qsar.eu.org (T. Puzyn).

http://dx.doi.org/10.1016/j.jhazmat.2015.10.023
0304-3894/

environment. Low vapors pressures that characterize ILs decrease the risk of air pollution by these substances and reduces their long-range atmospheric transport potential. However, since ILs have significant solubility in water, natural waters are the most likely media through which ILs would be transported in the environment. Moreover, the physical–chemical properties that make ILs useful from the industrial perspective (i.e., high chemical and thermal stability) may suggest potential problems with degradation of ILs and their high persistence in environment [5–7]. For these reasons, it is of vital importance to assess the possible transport and fate of novel ILs, before they are introduced at the market.

N-octanol/water partition coefficient (K_{OW}) is a parameter that, regarding low vapor pressure and high water solubility of ILs, is crucial for the assessment of the environmental transport and fate. It describes distribution of a substance between the organic and the aqueous phase at equilibrium state. Logarithmic values of K_{OW} are commonly used in exposure assessment to express lipophilicity of a chemical [8]. Highly lipophilic substances ($\log K_{OW} \gg 0$) could be accumulated in water organisms, whereas more hydrophilic ones ($\log K_{OW} \ll 0$) are preferably transported with water masses and, thus, the uptake by organisms is significantly lower [9,10]. The values of *n*-octanol/water partition coefficient are also used as input data for more comprehensive environmental multimedia models that allows investigating the behavior of chemicals in the total environment [11].

Unfortunately, data on physicochemical properties of ionic liquids are often incomplete or not available in the literature [12]. Therefore, regarding the large number of the existing and theoretically possible ILs, it is reasonable developing computational models that enable predicting missing data in relatively short time, without necessity of performing additional experiments. An example of such methods is Quantitative Structure–Activity/Property Relationship (QSAR/QSPR) approach, according to which the property of interest might be predicted from the variance in chemical structures of the investigated group of compounds, when experimental data are available only for a part of the group [13–19]. Since the experimentally measured values of *n*-octanol/water partition coefficient are currently available only for a small part of the ionic liquids, in this work we have developed a QSPR model that allows predicting the logarithmic values of K_{OW} of a wider set of ionic liquids from their chemical structure. Furthermore, with this model we tried to determine the effect of structural variation of cations and anions on the modeled property in the context of environmental transport and fate of ILs. We also estimated bioaccumulation potential of all ionic liquids from prediction set by comparing predicted values of $\log K_{OW}$ with criteria of the Stockholm Convention [20].

2. Materials and methods

The process of developing QSPR models consists of several basic steps, namely: (i) collecting available experimental data and splitting them into training and validation sets, (ii) calculating molecular descriptors, (iii) selecting the optimal, physically interpretable combination of the descriptors and training a QSPR model, (iv) external validation of the model, (v) providing physical interpretation of the model [21]. Only appropriately developed and validated model can be further used for making valuable predictions.

2.1. Experimental data

At first, we collected the experimental values of *n*-octanol/water partition coefficient from available literature sources [22–28]. We have found the data only for 53 ionic liquids. In addition, when evaluating the data, we were forced to discard 10 compounds, because the experimental details (i.e., the information about the applied measurement method and the temperature at measurements) were missing. In effect, we obtained a set of 43 ILs, in which the experimental values of the studied partition coefficient ranged from −3.77 to 1.73 logarithmic units. The data have been measured at 297.15 ± 2 K. We accepted this temperature range, because the temperature variation up to 10 K does not affect the measurement in a significant way [22].

In the next step, 43 ILs were sorted according to the increasing values of $\log K_{OW}$. Then, the compounds were split into a training set and a validation set. Data splitting procedure was performed using so-called "*Z*:1 algorithm", in which every Z^{th} compound in a group of the compounds sorted according to the predicted property (here $\log K_{OW}$) is assigned to the validation set, whereas the remaining ones form the training set. In this case $Z = 3$ and we obtained the training set containing 29 ionic liquids (67%) and the validation set containing 14 compounds (Table A in the electronic Supplementary material).

2.2. Molecular descriptors

In order to obtain molecular descriptors (numerical variables that characterize molecular structures of the compounds used as input variables in the QSPR model), we created molecular models of all cations and anions present in the 43 studied ILs using the ChemSketch software [29]. Subsequently molecular geometries of cations and anions (separately) were optimized with use of quantum-mechanical methods at the semi-empirical PM7 level [30] with the MOPAC 2012 software [31]. Molecular models of anionic and cationic moieties with optimized geometry were then used for calculating molecular descriptors. We obtained 1025 descriptors of the cations' and 1311 descriptors of the anions' structures (Tables B and C in the electronic Supplementary material, respectively). This includes:

- Constitutional and topological descriptors (1D–2D),
- Weighted Holistic Invariant Molecular descriptors (WHIM) (3D),
- Quantum-mechanical descriptors (3D).

Constitutional and topological descriptors that take into account one- and two-dimensional features of the molecules (1D–2D) as well as three-dimensional (3D) WHIM descriptors were separately calculated with Dragon software (version 6.0), whereas quantum-mechanical ones (e.g., frontier orbital energies, dipole moments) were extracted directly from MOPAC 2012 output files after the geometry optimization [31,32].

2.3. QSPR modeling

Next, we applied multiple linear regression method (MLR) to find the quantitative relationship between molecular descriptors (input variables) and $\log K_{OW}$ (the modeled value) [33,34]. The optimal combination of molecular descriptors was selected with genetic algorithm (GA) [35] implemented in the QSARINS software [36,37]. The following set of parameters has been applied to control the genetic algorithm: the size of a population: 100, the mutation rate: 45%. For more details, please refer to Fig. 1S in the electronic Supplementary material. It should be noted that there are more software packages for automated QSAR/QSPR modeling, which offer various modeling techniques, such as Partial-Least-Squares (PLS), Supporting Vector Machines (SVM) or neural networks (NN) [38,39].

To assure credibility of the model, the following recommendations published by Organization for Economic Cooperation and Development (OECD) were fulfilled [40]. Indeed, our model is based