Validation and updating of QSAR models for partitioning coefficients of ionic liquids in octanol-water and development of a new LFER model

Chul-Woong Cho [a], Stefan Stolte [b,c], Yeoung-Sang Yun [a,*]

[a] School of Chemical Engineering, Chonbuk National University, 567 Beakje-dearo, Deokjin-gu, Jeonju, Jeonbuk 561-756, South Korea
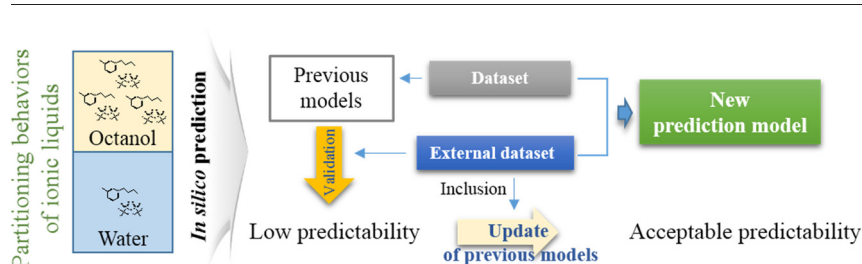[b] Department of Hydrosciences, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany
[c] Institute of Water Chemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany

HIGHLIGHTS

- Log P values of unstudied ionic liquids (ILs) were experimentally measured.
- Previous QSAR models for log P of ILs were validated using the external validation set.
- Limitations of previous QSAR models for log P of ILs and their updated versions were introduced.
- An accurate and simple linear free energy relationship model for predicting log P of ILs was newly suggested.

GRAPHICAL ABSTRACT

ABSTRACT

Since estimating the octanol-water partitioning coefficients (log P) of numerous ionic liquids (ILs) is tedious, time & material consuming and labor intensive, predicting by quantitative structure-activity relationship (QSAR) approach is necessary. Although several researchers presented the QSAR models for the property, validation assessment of the models were not sufficiently performed due to lack of log P dataset. In this study, the log P values of external ILs were measured by a shaking-flask method or collected from literatures. The newly obtained external log P values were applied for the validation study of previous models. In results, it was found that previous models showed rather low predictabilities and/or non-ignorable prediction limits to some IL structures whose anions were not involved in the previous studies. Accordingly, to achieve better predictability, the parameters used for previous modeling were re-selected and also their coefficients were re-calculated by multiple linear regression analysis with an inclusion of the external validation set to previous training set. Moreover, for reasonable understanding of chemical meanings in octanol-water partitioning behavior of ILs, we developed a new prediction model with a few number of descriptors, which has a good accuracy of $R^2 = 0.862$ and standard error $= 0.564$ log units.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Ionic liquids (ILs), as molten salts under 100 °C, have been used in various industrial applications because they have attractive properties, e.g., thermal stability, ignorable volatility, and material dissolution ability (Petkovic et al., 2011). Moreover, since they have structural tenability – approximately 10^8 types of ILs theoretically predicted (Holbrey and Seddon, 1999) – their physicochemical properties can be appropriately controlled. This implies that applicability of ILs may be unlimited. However, in last two decades it has also been reported that several ILs can play roles as hazardous materials in aquatic environments (Plechkova and Seddon, 2008; Ranke et al., 2007; Amde et al., 2015; Cho et al., 2016). Therefore, prior to the synthesis or use of ILs, it is important to understand their environmental properties and fate.

* Corresponding author.
  E-mail address: ysyun@jbnu.ac.kr (Y.-S. Yun).

As an indicator to help forecast the environmental fate e.g., toxicity, adsorption, and distribution etc., the octanol-water partitioning behavior (log P) of a substance, which is defined as the ratio of the concentration of a chemical in the binary phases, has been frequently used because it is related to the chemical and environmental properties of the substance. For example, log P values have been correlated with toxicities (Saarikoski and Viluksela, 1982; Konemann and Musch, 1981; Raevsky et al., 2008), adsorption (Sabljic et al., 1995), solubility (Valvani et al., 1981), distribution (Chiou, 1985), and bioconcentration factors (Mackay and Fraser, 2000) of non-charged molecules.

In order to determine the log P values of the ILs of interest in this study, several methods (e.g., shaking flask, slow stirring methods, and HPLC measurement) can be applied. However, the experimental processes based on shaking flask and slow stirring methods are very tedious, time & material consuming, and dependent on the solubility of the substance. In the case of poorly soluble chemicals in water or octanol, it can be biased to highly soluble part, thus the amount chemicals in the opposite part is below detection limits. In the case of the HPLC measurement method, it may have limitation in determination of log P of ILs because the salt may be dissociated in the mobile phase of HPLC; thus only single ion's effect of an IL may be considered while ion-paring of a cation and anion and the opposite ion's effect may be neglected. To solve these problems and also to develop faster, safer, and less expensive log P estimation methods, a quantitative structure activity relationship (QSAR) approach (Roy et al., 2015; Dearden, 2016) - finding some relationship between chemical structures and a certain property - is effective.

Although there are several prediction methods for the log P values of chemicals (i.e., atom-based, fragment-based, and knowledge-based methods), QSAR models for ILs have been rarely used. Kamath et al. (2012) predicted log P values of six imidazolium-based ILs with adaptive bias force-molecular dynamics simulations. Similarly, Lee and Lin (2014) developed an in silico method using the Pitzer-Debye-Hueckel model and the COSMO-SAC model, and achieved predictability of $R^2 = 0.75$. Although the reported methods achieved rather reasonable predictabilities, they could not support understanding of partitioning behaviors of ILs on a molecular basis, because they did not use fitting parameters which visibly explain the chemical meanings of the property. Additionally, Das and Roy (2016) and Roy and Popelier (2014) predicted chromatographic lipophilic parameter (log k_o) using extended topochemical atom indices and/or quantum topological molecular similarity. However, the log k_o reflects only cationic lipophilicity of ILs, thus it cannot explain log P of ILs, determined by the effect of both cation and anion.

In 2011, members of our group presented a linear free energy relationship (LFER) model, whose concept was originated by Abraham et al. (2004) using in silico calculated descriptors (Cho et al., 2015). The calculation of these descriptors was based on a conductor-like screening model (COSMO) calculation using the structures of 470 neutral compounds (Abraham et al., 2004) as shown in Eq. (S1) to (S5). The calculated values are given in supporting information. For the prediction in silico method, first anionic hydrophobicity (Ha) at the equilibrium status of ILs in octanol-water needed to be calculated using Eq. (1), as shown below (Cho et al., 2011):

$$H_a[\text{dimensionless}] = 0.298 - 0.841E_a + 0.263S_a + 1.691A_a - 0.571B_a + 1.318V_a \tag{1}$$

where the subscript 'a' stands for the anion of ILs. The meanings of the solute-dependent descriptors (capital letters) are as follows: E - Excess molar refraction [$cm^3\ mol^{-1}$/100]; S - dipolarity/polarizability [dimensionless]; A and B - Hydrogen bonding acidity [dimensionless] and hydrogen bonding basicity [dimensionless], respectively; V - McGowan volume [$cm^3\ mol^{-1}$/10]; system coefficient were determined for predicting H_a.

Next, using the H_a values calculated by Eq. (1), the log P values of the ILs can be predicted using Eq. (2) (Cho et al., 2011):

$$\text{Log P of ILs} = -6.239 - 0.603E_c + 0.794S_c - 0.901\ A_c - 6.765B_c + 2.976V_c + 1.007\ H_a \tag{2}$$

where the subscript 'c' stands for the cation of ILs. From the calculations using Eqs. (1) and (2), the log P values of the ILs could be predicted to a coefficient of determination (R^2) of 0.977 and a standard error (SE) of 0.217 log unit.

In 2016, Rybinska et al. (2016) presented a QSAR model by using three constitutional and topological descriptors, as Eq. (3) below:

$$\text{Log}P = -1.48 + 0.47\ \text{L3m}^A + 0.93\ \text{ON0V}^C + 0.55\ \text{X5Av} \tag{3}$$

where the employed parameters were all calculated values and their meanings can be defined as follows:

L3m^A – the 3rd component size directional WHIM index, weighted by atomic masses (calculated for the anion);

ON0V^C – the overall modified Zagreb index of order 0, by valence vertex degrees (calculated for the cation); and.

X5Av^C – the average valence connectivity index of order 5 (calculated for the cation).

The system parameters in Eq. (3) could not be directly used by the parameters presented by Rybinska et al. (2016) without rescaling of the solute parameters (Rybinska et al., 2017). Therefore, for convenient application of the solute parameters from Rybinska et al. (2017), we recalculated the system parameters as Eq. (4) (Cho et al., 2017):

$$\text{Log}P = -6.22 + 1.80\ \text{L3m}^A + 0.54\ \text{ON0V}^C + 25.15\ \text{X5Av}^C \tag{4}$$

The log P value calculated by Eq. (4) could be correlated with the measured values of ILs, studied by Rybinska et al. (2016), in R^2 of 0.91 and SE of 0.42 log units.

However, previous models i.e., Eqs. (2) and (4) should need further validation study to be considered for checking appropriate application domain (Gadaleta et al., 2016) because the models were built with an insufficient dataset. Therefore, in this study, the validation studies for the two models were performed using external data set. That is, since the log P values of the ILs were not sufficient, we experimentally measured and collected log P values of additional ILs as the external validation set. The log P values of the external validation set were then compared with those calculated by the previous models. For more validation study, the estimation of MAE-based criteria for models were applied (Roy et al., 2016). Finally, to better understand the partitioning behaviors of the ILs in octanol-water, we developed a simple prediction model using the re-calculated LFER descriptors. These descriptors were based on models established using both neutral and ionic compounds (Cho et al., 2015), which differs from previous case (Cho et al., 2011) that used only neutral compounds (Zissimos et al., 2002).

2. Material and methods

2.1. Abbreviations of IL ions

[IM12] – 1 ethyl 1 methylimidazolium; [IM13] – 1 methyl 1 propylimidazolium; [IM14] – 1 butyl 1 methylimidazolium; [IM15] – 1 methyl 3 pentylimidazolium; [IM16] – 1 hexyl 1 methylimidazolium; [IM17] – 1 heptyl 1 methylimidazolium; [IM18] – 1 methyl 3 octylimidazolium; [IM19] – 1 methyl 3 nonylimidazolium; [IM1–10] – 1 decyl 3 methylimidazolium; [IM1 1Ph] – 1 benzyl 3 methylimidazolium; [IM1 (1Ph 4Me)] – 1 methyl 3 (4 methylbenzyl)imidazolium; [IM1 2Ph] – 1 methyl 3 (2 phenylethyl) imidazolium; [IM1–2 = 1] – 1 methyl 3 (2 propenyl) imidazolium; [IM11O2] - 1 (ethoxymethyl) 3 methylimidazolium; [IM12O1] – 1 (2 methoxyethyl) 3 methylimidazolium; [IM13OH] – 1 (3 hydroxypropyl) 3 methylimidazolium;