Multilayered Variable Selection in QSPR: 
A Case Study of Modeling Melting Point of Bromide Ionic Liquids

Souvik Das, Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India  
Probir K Ojha, Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata India  
Kunal Roy, Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata India

ABSTRACT

Ionic liquids (ILs) are widely used in industry as green solvent alternatives because of their exceptional solvating ability and extremely low vapor pressure. For many applications of ILs, a low melting point temperature is desirable. Several simple ILs do not exhibit a clear melting point in the accessible temperature range of the DSC apparatus. Therefore, a computational approach is required to understand the relationship between the melting point of ILs and their structural characteristics. In the present study, the authors have developed predictive quantitative structure-property relationship (QSPR) models for melting point of ILs. A pool of 376 bromide ILs having quantitative melting point data were used to develop predictive models. A multilayered variable selection strategy has been adopted for development of final QSPR models. The models would provide an important guidance for the chemists to predict melting point of bromide ILs theoretically thereby saving the time and resources involved in the experimental determination.

KEYWORDS

Ionic Liquids, Melting Point, Multilayered Variable Selection, PLS, QSPR

1. INTRODUCTION

Ionic liquids are chemical salts formed by two oppositely charged ions held together by the ionic force. Such compounds are physically liquid with a melting point less than the boiling point of water (i.e., 100°C) (Stark, & Seddon, 2007). ILs that are liquid in room temperature are called room temperature ionic liquids (RTIL). Properties of ionic liquids vary with the choice of the combination of the anion and cation (Seddon, 1999). ILs are widely used in chemical and pharmaceutical industries for various applications such as extraction process, organic synthesis, electrochemistry (Buzzeo et al., 2004), nanotechnology (Ohno, 2011), drug delivery system (Jaitley et al., 2008) as pharmaceutical salts (Marrucho et al., 2014) etc. The possible cationic head groups in ionic liquids include imidazolium, pyridinium, benzimidazolium, ammonium, phosphonium, guanidinium etc. whereas the anionic groups include halides, pseudohalides, sulfates, bromides and metal complexes. ILs have some interesting properties like low inflammability, low vapor pressure, high conductivity, variable range of viscosity and density, tuneable polarity and solubility as well as their stability at high temperatures (300-400°C) and ability to retain the liquid state for a wide range of temperatures
Due to low vapor pressure, ILs have minimal discharge into the environment thus decreasing the risk of air pollution. Therefore, ILs have been considered as “Green solvents” for last few years (Jadhav et al., 2014). However, some recent studies have shown that ILs have a negative impact on living organisms due to the increased probability of their release to the aquatic environment. It has also been found that the anions have a much lower impact on the toxicity of ILs than the cations (Smiglak, et al., 2007). Thus, it is important to estimate the ability of these compounds to spread into the environment. The melting point ($T_m$) of ILs is one of the most widely used fundamental physical properties due to their direct relation with toxicity and safety (Kudlak et al., 2015). There are some evidences that ILs with higher ranges of melting point are highly toxic than those with a lower range of melting point (Kudlak et al., 2015). Thus, estimation of melting point of ILs is important for determining their environmental safety. ILs with lower melting points are also desired for specific industrial applications (Drake et al., 2003). However, it is very difficult to estimate the melting point of ILs experimentally due to the following reasons (Canongia et al., 2005; Wasserscheid, & Welton, 2007): 1) ILs do not exhibit a clear melting point in the accessible temperature range (-100°C to +100°C) of the differential scanning calorimetry (DSC) apparatus; 2) ILs are glass-transforming materials and therefore do not exhibit a clear freezing point. The knowledge of melting and glass transition temperatures of ILs is needed for their use at an industrial scale requiring to set a feasible temperature operation range. Experimental determination of solid–liquid phase transitions of ILs cannot be clearly distinguished into melting points and glass transition temperatures as many samples after glass transition begin to melt and no distinct peaks can be observed (Sashina et al., 2013). Thus, there is a need for some alternative processes to address these difficulties. In this regard, a computational approach such as quantitative structure-property relationship (QSPR) can be used as an alternative method to predict the melting point of ILs (Das, & Roy, 2013) as well as to design ILs with lower range of melting point in order to reduce the toxicity. QSPR correlates a response property with structural features or attributes of chemicals leading to statistically meaningful quantitative relationship giving precise predictions which can obviate the difficulties associated with experimental determinations of properties. Thus, QSPR reduces the requirement of resources, time and expenses involved in experiments (Dearden, 2016; Roy, Kar, & Das, 2015).

In the present study, we have developed predictive QSPR models using a dataset composed of four classes of ILs (ammonium, pyridinium, benzimidazolium and imidazolium bromides) with reported experimental melting point data. There is always a chance for overfitting while working with a large number of variables for development of predictive QSAR models. To reduce the chance of overfitting and extent of noise and/or redundant information in the set of input descriptors, variable selection prior to development of the final model is very important. In case of inappropriate selection of descriptors, the prediction quality of the developed model for the external set of compounds may be suboptimal. There have been different strategies for optimum selection of descriptors reported for final QSAR models (Mehmood et al., 2012; Frank et al., 1987; Mehmood et al., 2011; Forina et al., 1999). In the present work, we have used a novel multilayered variable selection strategy to explore the chemical structural features or molecular descriptors that are responsible for regulating melting point of ILs. The initial pool of descriptors was reduced using a multilayered variable selection strategy such as stepwise regression followed by double cross validation followed by genetic function approximation (GFA) followed by the best subset selection. The final models were developed using partial least squares (PLS) regression. We have also developed a consensus model using the developed PLS models for robust prediction of melting point of ILs. The present paper emphasizes on the variable selection strategy using a multilayered approach taking modeling of melting point of ionic liquids as a case study. The authors hope that this case study will help the readers to understand the strategies involved in the suggested workflow of the multi-layered variable selection.
The History and Development of Quantitative Structure-Activity Relationships (QSARs)
www.igi-global.com/article/the-history-and-development-of-quantitative-structure-activity-relationships-qsars/144688?camid=4v1a

(Q)SPR Models for Prediction of Hydrophobicity of Isatins
www.igi-global.com/article/qspr-models-for-prediction-of-hydrophobicity-of-isatins/191195?camid=4v1a