Chemosphere 170 (2017) 242–250

Contents lists available at ScienceDirect

Chemosphere

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

Development of QSAR model to predict the ecotoxicity of *Vibrio fischeri* using COSMO-RS descriptors

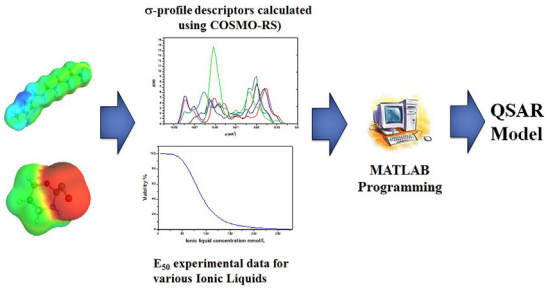
Ouahid Ben Ghanem [a, *], M.I. Abdul Mutalib [a], Jean-Marc Lévêque [b], Mohanad El-Harbawi [c]

[a] Faculty of Chemical Engineering, Universiti Teknologi Petronas, Bandar Seri Iskandar, 31750, Tronoh, Perak, Malaysia
[b] Fundamental & Applied Sciences Department, Universiti Teknologi Petronas, 31750, Bandar Seri Iskandar, Tronoh, Perak, Malaysia
[c] Chemical Engineering Department, King Saud University, Riyadh, 11421, Saudi Arabia

HIGHLIGHTS

- The Log EC_{50} of 110 ILs towards *Vibrio fischeri* were obtained from literatures.
- Linear and non-linear QSAR models were developed to predict the ILs ecotoxicity.
- σ-profile descriptors were used for molecules representation.
- The selected descriptors demonstrated dominant behaviour of cations on the toxicity of ILs.
- Highly accurate models can be used for estimation of the ILs toxicity.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 August 2016
Received in revised form
16 November 2016
Accepted 1 December 2016
Available online 5 December 2016

Handling Editor: Andreas Gies

Keywords:
Ionic liquids
Eco-toxicity
Vibrio fischeri
QSAR
Sigma profile
Molecular descriptors

ABSTRACT

Ionic liquids (ILs) are class of solvent whose properties can be modified and tuned to meet industrial requirements. However, a high number of potentially available cations and anions leads to an even increasing members of newly-synthesized ionic liquids, adding to the complexity of understanding on their impact on aquatic organisms. Quantitative structure activity\property relationship (QSAR\QSPR) technique has been proven to be a useful method for toxicity prediction. In this work,σ-profile descriptors were used to build linear and non-linear QSAR models to predict the ecotoxicities of a wide variety of ILs towards bioluminescent bacterium *Vibrio fischeri*. Linear model was constructed using five descriptors resulting in high accuracy prediction of 0.906. The model performance and stability were ascertained using k-fold cross validation method. The selected descriptors set from the linear model was then used in multilayer perceptron (MLP) technique to develop the non-linear model, the accuracy of the model was further enhanced achieving high correlation coefficient with the lowest value being 0.961 with the highest mean square error of 0.157.

1. Introduction

Ionic liquids (ILs) have probably been one of the most explored class of chemicals in the academic and industrial fields for the last two decades. They have gained widespread attention due to their

* Corresponding author.
E-mail address: wahidghanem@gmail.com (O.B. Ghanem).

http://dx.doi.org/10.1016/j.chemosphere.2016.12.003
0045-6535/

unique physicochemical properties such as hardly measurable vapour pressure, non-flammability, non-volatility, high thermal and electrochemical stability, high ionic conductivity and ability to solvate compounds of very different polarity (Swatloski et al., 2002; Fukaya et al., 2008; Greaves and Drummond, 2015; Pham et al., 2010). Their remarkable characteristics have made them a subject of exploration in various fields including biomass conversion (van Putten et al., 2013), medical research (Kumar and Malhotra, 2009; Malhotra and Kumar, 2010), solar and thermal energy conversion (Cosar et al., 2013; Lin et al., 2012; Lee et al., 2012), extraction (Nasir Shah et al., 2014; Shah et al., 2016), lubricants and fuel additives (Qu et al., 2009; Bermúdez et al., 2009), and so forth.

The extraordinary low vapour pressure of ILs has been one of the main reason for their classification as 'environmentally-friendly' compounds (Plechkova and Seddon, 2007). Meanwhile, numerous studies have reported the toxicity of diverse ILs sub-families against different organisms including *green algae* (Kulacki and Lamberti, 2008; Latała et al., 2009; Pretti et al., 2009), *Vibrio fischeri* (Viboud et al., 2012; Stolte et al., 2007; Costa et al., 2014), *gram-positive* and *gram-negative bacteria* (Yu et al., 2016; Coleman et al., 2012; Garcia et al., 2013), and *fish* (Radošević et al., 2013; Dong et al., 2013). All of these studies concluded that ILs could not be considered as a fully green class of chemicals but instead as a safer one to the atmosphere with lesser risk for heavy air pollution caused mainly by volatile organic components (VOCs). On the contrary, their toxicity is actually associated with their good water solubility, leading to possible untoward effects to the aquatic living organisms by effluent or accidental spillage (Viboud et al., 2012; Egorova et al., 2015; Ventura et al., 2013; Alvarez-Guerra and Irabien, 2011). Nevertheless, the non-water soluble ILs are normally found to be more toxic than their water soluble counterparts, and hence emphasis should be given to study the ILs impact of all classes, i.e. hydrophobic/hydrophilic ones (Salminen et al., 2007; Papaiconomou et al., 2010; Łuczak et al., 2010).

Currently, there is an exponential growth in the number of newly synthesized ILs most notably due to the availability of a multitude of cations and anions causing the toxicity assessment for all of them close to be experimentally impossible as it is not only time consuming but also very costly. Therefore, developing predictive quantitative structure activity relationship (QSAR) models to establish a prior estimation of their toxicity is highly desired.

The technique of QSAR\QSPR has been implemented to predict several physical, chemical and biological properties. The molecular descriptors play an important role in the model development and interpretation, but do not always have a straightforward physical meaning (Járvás et al., 2011). In other words, the link between the selected molecular descriptors and the studied properties of the ILs is generally not obvious. There are several software tools capable of providing a large number of molecular descriptors such as DRAGON™ 6 and CODESSA. For example, DRAGON™ 6 can provide up to 4885 molecular descriptors for a single molecule (Dragon software website). As such, it could lead to a complex and false methodology for optimizing and selecting the proper molecular descriptors for molecule representation. Hence, it is important to highlight the need of identifying the correct molecular descriptors which can be applied for representing the molecules in QSAR toxicity models as well as having a straightforward physical meaning related to the studied property.

Therefore, this work aims at developing highly predictive linear and non-linear QSAR models constructed with a few but reasonable descriptors to predict the toxicity of a relatively large dataset containing the ecotoxicity of 110 ILs towards *Vibrio fischeri*. For this purpose, σ-profile descriptors which derived from COSMO-RS were used to generate the input dataset. COSMO-RS theory is a continuum solvation model that combines quantum chemical theory, dielectric continuum models, surface interactions, and statistical thermodynamics. σ-profile calculates the molecular interactions from the screening (polarization) charge densities on the molecular surface segments (Klamt et al., 2001; Klamt and Schüürmann, 1993). It was successfully implemented in several QSAR\QSPR studies including density (Palomar et al., 2008), cytotoxicity applied on leukemia rat cell line (Torrecilla et al., 2010), enzymes performance (Mai and Koo, 2014), carbon dioxide absorption/desorption capacities (Gonfa et al., 2016a) and antimicrobial activity employed on four human pathogens bacteria (Ghanem et al., 2015a). σ-profile showed superiority in molecule representation, but have yet to be applied for ecotoxicity prediction.

Selecting the appropriate descriptor for the QSAR/QSPR model development is a challenging task, and to find the link between the selected descriptors and the molecules character is even more complicated. Therefore, the second goal of this work is to establish the link between the selected descriptors and the toxicity character of the molecules such as length of the alkyl chain, presence of the oxygen atom, and the water nature of the anions.

2. Ionic liquids data set and EC_{50} values

A diverse set of 110 ILs comprising of a combination of 49cations and 29 anions and their ecotoxicity data corresponding to the logarithmic value of EC_{50} (EC_{50} in mmol.l^{-1}) towards bioluminescent bacterium *Vibrio fischeri* were collected from literature (Viboud et al., 2012; Stolte et al., 2007; Ventura et al., 2013; Alvarez-Guerra and Irabien, 2011; Luis et al., 2010; Montalbán et al., 2016; Romero et al., 2008; Ghanem et al., 2015b; Ranke et al., 2004; Hernández-Fernández et al., 2015; Garcia et al., 2005; Matzke et al., 2007; Docherty and Kulpa, 2005; Peric et al., 2013). The database includes the logarithm of 5- 15- and 30-min EC_{50} (log EC_{50}) values for *Vibrio fischeri*, considering that normally the log EC_{50} values for a particular chemical rarely vary much with such exposure times (Kaiser and Palabrica, 1991; Schultz et al., 2003). The full IL names along with their abbreviation and log EC_{50} values is provided as Supplementary information S1.

3. σ-profile descriptors

The first step for QSAR models development is the structure optimization of all the ILs molecules made by combinations of the 49 cations and 29 anions. Geometry optimization was performed using TURBOMOLE package (Version 6.2) through the implementation of the density functional theory level, utilizing the BP functional (Vosko et al., 1980; Becke, 1988; Steffen et al., 2010) with resolution of identity approximation and a triple-ξ valence polarized basis set (Eichkorn et al., 1997; Schäfer et al., 1994). The COSMO-RS was then used to generate the molecular descriptors for each pair of ions. The optimized geometry file for each ion was then converted into a surface composition function (σ -Profile) by COSMO-RS.

To translate the structural information of ILs into numerical variables, molecular descriptors were developed using σ-profiles of the cations and anions involved in this study. According to COSMO-RS theory, the (σ-profile) of molecules spans the range of (−3.0 to 3.0) e/nm^2. The σ-profile of a molecule shows the probability of the relative amount of surface segments with polarity_σ on the surface of the molecule. The σ-profile of a molecule is considered as the characteristic properties of the molecule. In other words, σ-profile includes the main chemical information necessary to predict the possible electrostatic, hydrogen-bonding, and dispersion interactions of the compound (Palomar et al., 2008; Diedenhofen and Klamt, 2010; Gonfa et al., 2016b). The position of the charge distribution, width and the height of the peaks vary with the nature of