Chemometrics and Intelligent Laboratory Systems 144 (2015) 138–147

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Computational study of the effects of cations and anions to the cytotoxicity of diverse ionic liquids by supervised machine learning

Shuying Ma, Min Lv, Xiaoyun Zhang *, Honglin Zhai, Wenjuan Lv

Department of Chemistry, Lanzhou University, Lanzhou 730000, Gansu, People's Republic of China

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 25 March 2015
Accepted 27 March 2015
Available online 18 April 2015

Keywords:
Ionic liquids
Genetic function approximation
Least squares support vector machine
Cytotoxicity
Leukemia Rat Cell Line

ABSTRACT

Ionic liquids (ILs) have been widely used in many fields due to their unique physicochemical properties, and even they were considered as green solvents. However, the recent researches showed that ILs might bring potential risk to environment and humans. In this work, genetic function approximation (GFA) and least squares support vector machine (LSSVM) models were developed for predicting the cytotoxicity of a great variety of ILs, including 9 types of cations and 44 types of anions, to Leukemia Rat Cell Line (IPC-81) based on the structural descriptors calculated from the combination of cations and anions. Seven descriptors were selected by GFA to develop the linear QSAR model. According to the discussion of descriptors, the cation structure was the main factor of the toxicity which mainly depends on the hydrophobicity and space structure of cations. The LSSVM model was built to predict accurately the cytotoxicity of ILs to capture the nonlinear nature. The rigorous internal and external validation and applicability domain (AD) were performed to verify the reliability and predictability for GFA and LSSVM models. The results indicated that both models could be used for estimating the cytotoxicity of new ILs to IPC-81, and the discovered key structural characteristics could provide reference information for designing and synthesizing safer ILs.

1. Introduction

Ionic liquids (ILs) can be defined as ionic molten salts, whose melting points are below 100 °C and are completely formed by organic/inorganic anions and organic cations. They have attracted extensive attention in the past decades due to their unique physicochemical properties, such as low vapor pressure, non-volatility, non-flammability, high thermal and electrochemical stability and strong dissolution ability for a wide range of chemicals. Therefore, they have been regarded as "green or sustainable solvents" that replace the conventional volatile organic solvents [1–6]. The most interesting part of ILs is that their structure characteristic, which provides a very flexible design as long as the combination of cations and anions is changed. Thus, the physical and chemical properties of ILs can be customized by reasonably designing and modifying the different combinations of cations and anions [7]. Such advantage brings up a wide variety of ILs with more excellent performance. So, a variety of ILs are presented in diverse fields, such as organic synthesis [8–10], catalysis and biocatalysis [11,12], electrochemistry [13,14], solar cells [15,16], biomass processing [17,18], drug delivery [19,20] and even in the reprocessing of nuclear waste [21,22].

Although ILs have been widely used as green solvents, the potential consequent risks should not be neglected in the process of industrial production and application. With the increasing industrial manufacture and application, they would be more and more likely released into the environment. Because of the non-volatility of ILs, and most of them are hardly decomposed by microorganisms, they would be bioaccumulative [23–27]. What is worse, the high chemical and thermal stability would intensify the bioaccumulation. They may induce the serious pollution and bring potential risk for human beings and animals [28,29]. Therefore, it is necessary to conduct the toxicity assay before application. At present, Leukemia Rat Cell Line (IPC-81) is frequently used for the cytotoxicity assays of ILs [30]. However, these common toxicity assays are time-consuming, costly and resource-wasted because of the enormous number of ILs. So, it is necessary to establish a convenient and available approach to estimate the toxicity of ILs without experimental operation.

Quantitative structure–activity relationship (QSAR) is based on the assumption that the change in observed activity of compounds is reflected by the change in microscopic molecular structures [31]. The major objective of QSAR modeling is to obtain a reliable and accurate mathematical model that correlates the biological activity with several descriptive parameters, which derive from the molecular structure of compounds, and can be used for predicting the activity of compounds that are newly synthesized and yet have not been tested. On the other hand, QSAR model can reveal the intrinsic relationship between the information of molecular characteristic and biological activity and provide some insights into major factors in order to better understand the mechanism of action [32,33]. Now, QSAR models have been widely used for predicting the activity of compounds, including the toxicity of ILs.

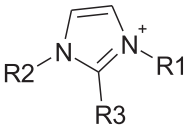
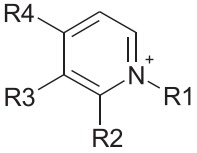
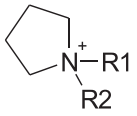
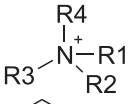
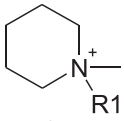
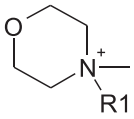
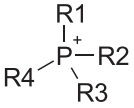
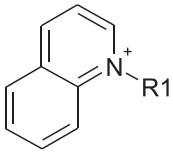
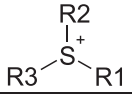
* Corresponding author. Tel.: +86 931 8912578; fax: +86 931 8912582.
E-mail address: xyzhang@lzu.edu.cn (X. Zhang).

http://dx.doi.org/10.1016/j.chemolab.2015.03.014
0169-7439/

Torrecilla et al. [34,35] reported the QSAR models that estimated the cytotoxicity of 153 ILs to IPC-81 using empirical formulas (element composition) and molecular weights as descriptors for performing the principal component analysis (PCA). 12 principal components were used to build the multiple linear regression (MLR), radial basis (RB) network, and multilayer perception (MLP) neural networks (NN) models with the square of correlation coefficient R^2 of 0.867, 0.861 and 0.982, respectively. Then, they also developed QSAR models that included MLR and NN to predict the toxicity of 96 ILs by COSMO-RS descriptor. Fatemi et al. [36,37] constructed the QSTR models based on MLR and MLP NN for correlating the cytotoxicity to IPC-81 of 50 ILs using 6 descriptors selected by genetic algorithm (GA). Cytotoxicity estimation of ILs to IPC-81 was explored again by their group using 2D and 3D structural descriptors, GA-MLR and MLP NN models were developed for the prediction of the toxicity of ILs. Yan [38] obtained a linear QSAR model for predicting the toxicity of 173 ILs to IPC-81 using topological indexes proposed by themselves. Cruz-Monteagudo et al. [39,40] introduced a Classification and Regression Trees (CART) classifier to enable the prioritization of the cytotoxicity to IPC-81 of ILs, with 81% of accuracy and 75% of sensitivity. Besides, they applied the qualitative method of network like similarity graph (NSG) to mine the relevant structure–cytotoxicity relationship as a complement of previous findings. Das and Roy [41] reviewed the researches of QSPR/QSTR models to different microorganisms including IPC-81 for designing greener and safer ILs. The researches above presented good statistical performance; however, few data for estimation of the cytotoxicity of ILs were described and the NN models including RB NN and MLP NN were relative complex.

The present work was aimed to develop a novel QSAR model for predicting the cytotoxicity of a wide variety of ILs to IPC-81 based on the descriptors calculated by the combination of cation and anion structures, and find out the key structural features that affect the cytotoxicity of ILs. In this study, self-organizing map (SOM) network was employed to divide the dataset into the training set and the test set, genetic function approximation (GFA) was used for selecting the best subset of descriptors and developing the linear model. Besides, the least squares support vector machine (LSSVM) model was built to capture the nonlinear nature existing between molecular structures of ILs and the cytotoxicity to IPC-81 by the descriptors selected by GFA. Moreover, models could contribute to gain a profound insight into mode of cytotoxicity action and provide theoretical guidance for designing and synthesizing safer and greener ILs.

Table 1
The diverse types of cations in the present work.

Type	Structure	Numbers
Imidazolium		100
Pyridinium		68
Pyrrolidinium		23
Ammonium		26
Piperidinium		15
Morpholinium		25
Phosphonium		6
Quinolinium		5
Sulfonium		2

2. Materials and methods

2.1. Datasets

The structures of 270 ILs and their corresponding cytotoxicity data to IPC-81 were taken from the UFT/Merck Ionic Liquids Biological Effects Database—center for environmental research and sustainable technology [42] and literatures [43,44]. The cytotoxicity values, denoting as the half maximal effective concentration (EC_{50}, in μmol/L) were converted into the form of logarithm ($logEC_{50}$). The dataset of ILs and their experimental cytotoxicity values were listed in Table 1S (Appendix A, Supplementary data). In the present work, the types of cations including imidazolium (IM), pyridinium (Py), pyrrolidinium (Pyr), ammonium (N), piperidinium (Pip), morpholinium (Mor), phosphonium (P), quinolinium (Quin) and sulfonium (S) were listed in Table 1. Also, the structures of 44 types of anions including organic and inorganic were described in Table 2.

2.2. Molecular descriptors

In order to relate the cytotoxicity to IPC-81 to the structures of their ionic constituents by cation and anion-based descriptors, the cation and anion descriptors for each ionic liquid should be calculated separately. The structures of all cations and anions of 270 ILs were drawn separately by ACD/ChemSketch software [45] and were initially optimized by the molecular mechanics (MM+) force field. A more precise optimization was performed with a semi-empirical PM3 method in the HyperChem Program 8.0 [46]. In the process, the corresponding total ionic charge was labeled for each cation and anion. The structures of ILs were optimized by the Polak–Ribiere conjugate gradient algorithm with a root mean square gradient of 0.1 and 0.01 kcal/(mol · Å) as stop criterion for MM+ and PM3, respectively. Finally, the output files of HyperChem geometry optimization were used as input information of Dragon software 6.0 [47] to calculate all 29 types of molecular descriptors for cations and anions. As a result, 2182 and 2114 descriptors were obtained for cations and anions, respectively. So, it is necessary to prescreen those descriptors to reduce the redundant and non-useful information. The prescreen criteria are: (1) removing the constant or near constant descriptors; (2) eliminating descriptors with relative standard deviation less than 0.05; and (3) for the absolute pairwise correlation coefficients among these descriptors larger than or equal to 0.95, one of them was deleted randomly due to multicollinearity. Finally, the descriptors were decreased to 745 and 436 for cations and anions, respectively.