



Article

# Predicting the Toxicity of Ionic Liquids toward Acetylcholinesterase Enzymes Using Novel QSAR Models

Peng Zhu <sup>1,†</sup>, Xuejing Kang <sup>2,†</sup>, Yongsheng Zhao <sup>3,\*</sup>, Ullah Latif <sup>4</sup> and Hongzhong Zhang <sup>2,\*</sup>

<sup>1</sup> School of Materials Science and Energy Engineering, Foshan University, Foshan 528000, China; yszhao@sjtu.edu.cn

<sup>2</sup> School of Material and Chemical Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China; xuejing\_kang@hotmail.com

<sup>3</sup> Department of Chemical Engineering, University of California, Santa Barbara, CA 93106-5080, USA

<sup>4</sup> Department Beijing Key Laboratory of Ionic Liquids Clean Process, Key Laboratory of Green Process and Engineering, State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China; latifucas@hotmail.com

\* Correspondence: yzhao01@ucsb.edu (Y.Z.); zhz@zzuli.edu.cn (H.Z.)

† These authors contributed equally to this work.

Received: 31 March 2019; Accepted: 26 April 2019; Published: 2 May 2019



**Abstract:** Limited information on the potential toxicity of ionic liquids (ILs) becomes the bottleneck that creates a barrier in their large-scale application. In this work, two quantitative structure-activity relationships (QSAR) models were used to evaluate the toxicity of ILs toward the acetylcholinesterase enzyme using multiple linear regression (MLR) and extreme learning machine (ELM) algorithms. The structures of 57 cations and 21 anions were optimized using quantum chemistry calculations. The electrostatic potential surface area ( $S_{EP}$ ) and the screening charge density distribution area ( $S_{\sigma}$ ) descriptors were calculated and used for prediction of IL toxicity. Performance and predictive aptitude between MLR and ELM models were analyzed. Highest squared correlation coefficient ( $R^2$ ), and also lowest average absolute relative deviation (AARD%) and root-mean-square error (RMSE) were observed for training set, test set, and total set for the ELM model. These findings validated the superior performance of ELM over the MLR toxicity prediction model.

**Keywords:** toxicity; ionic liquids; acetylcholinesterase enzyme; extreme learning machine; multiple linear regression

## 1. Introduction

With ever-increasing demand for clean energy and stricter environmental regulations, scientists have always strived for sustainable and green chemical products [1]. The apparent benefits of ionic liquids (ILs) has spurred intensive research in order to replace the conventional hazardous solvents for broad scale industrial applications, such as organic reaction [2–8], catalytic processes [9–12], gas separation [13–16], storage materials [17–20], etc. Among the unique properties of ILs, negligible vapor pressure has primarily attracted the attention of many researchers, and the field has sought improvements in positively affecting overall efficiency of many chemical processes under the principles of green chemistry [21,22].

State-of-the-art research shows that the previously accepted advantage of low toxicity for ILs has proved to be overestimated, meaning that to some extent ILs in reality pose hazard potentials to humans and environment [23]. Rogers [22] et al. challenged the notion of green ILs because of their hazardous properties, e.g., unknown toxicity and stability. Meanwhile, Jastorff [24] et al. discussed the

toxicity and ecotoxicity of ILs using quantitative structure-activity relationships (QSAR), which could be used to aid the rational design of optimal solvents from a technical and environmental perspective. Docherty [25] et al. found that ILs with hexyl- and octyl-imidazolium and pyridinium bromides had significant antimicrobial activity to pure cultures of *Saccharomyces cerevisiae*, *Escherichia coli*, *Bacillus subtilis*, and others. Later, Pang [26] et al. reviewed the environmentally relevant issues of ILs, such as environmental application and toxicity. Karunanithi [27] et al. studied the life cycle of aquatic ecotoxicity impacts for five common ILs by integrating physical properties, toxicity data, and transport parameters in the model. For the first time, they reported the freshwater ecotoxicity characterization factors for ILs. They found that an average of 83% of ecotoxicity impact was due to chemicals released during the upstream synthesis steps, while the remaining 17% ecotoxicity was related to the life-cycle energy consumption. All these findings underscore the need to develop sustainable and nontoxic ILs in future research. Unlike other basic physicochemical properties, such as density, viscosity, thermal stability, etc., to the best of our knowledge, research studies on the toxicity of ILs are still insufficient, and therefore need to be addressed further. In this regard, key developments in toxicity of ILs will highlight the designable feasibility of ILs to be environmentally benign with huge potential benefits for sustainable and green chemistry [28]. On the other hand, it is worthy to emphasize that around  $10^8$  ILs are accessible due to the combination of enormously different cations and anions. Consequently, challenges have been encountered when investigating the toxicity of ILs [28,29]. To deal with these issues, it is highly necessary to develop efficient models to predict the toxicity of ILs with high accuracy.

Recently, QSAR studies have been reported in correlation with, and prediction of, properties of ILs [30–36]. QSAR is a powerful tool for accurately predicting the physical and chemical properties when applied to a set of molecular descriptors. In this regard, QSAR is supposed to be useful for estimating and screening chemical compounds for specific applications, having the desired characteristics for elucidating the underlying relation between micro-structures and macro-properties. Since the pioneering studies by Jastorff and co-authors [37], research attempts have been devoted to improving understanding and estimating the toxicity of ILs [30,38–42]. In subsequent years, Pereira et al. presented their toxicological assessment of a group of environmentally-friendly ILs with benign cholinium cations and linear alkanoate anions, using filamentous fungi as model eukaryotic organisms [39]. They found that the toxicity of ILs was increased with elongation in the linear chains of the anion, while branching resulted in reduced toxicity due to depressed lipophilicity [39]. Similar conclusions were drawn by Lima and Coutinho [43]. In order to further these studies, Ranke and co-workers studied up to 74 ionic liquids with different cations and anions to show the influence of cation lipophilicity on the cytotoxicity in *IPC-81* leukemia cells of rats. They found that substituents in the cation increased the toxicity of ILs [41]. Torrecilla et al. developed QSAR models based on MLR and neural network for prediction of toxicological effects of 96 ILs on *IPC-81* [42]. Romero and co-workers studied  $EC_{50}$  values of each compound in an aqueous solution using the microtox standard procedure. They found that the short length of the  $R_2$  side chain on imidazolium cation was favorable to diminish the toxic effect, while anions had a minor effect on the toxicity of ILs [44]. Yan et al. investigated the toxicity of ILs in acetyl cholinesterase enzyme (AChE) by the QSAR method using topological indices based on the atom characters. Their results demonstrated that the MLR model was capable of predicting the  $\log EC_{50}$  (AChE) of ILs by using a 177 unit training set and a 44 unit testing set of topological indices generated from cations and anions [45]. Recently, Singh et al. investigated the chemical attributes of a wide variety of ILs towards their inhibitory potential of AChE using a supported vector machine (SVM) and a cascade correlation network (CCN) [46]. Their results showed that the proposed QSAR models had more statistical confidence, especially with respect to external validation, which has not been focused upon in other studies. This work addresses the success in predicting different toxicity classes and precise toxicity end-points of ILs using proposed QSAR. Like other researchers, we have previously focused on establishing the toxicity database of ILs, including over 4000 pieces of data and using QSAR models based on quantum chemistry descriptors and an artificial intelligence algorithm to study the toxicity toward the *IPC-81* [40]. The results revealed that

the nonlinear model developed by the SVM algorithm was more reliable in the prediction of toxicity of ILs. Until now, very limited information has been available on the toxicity of ILs towards AChE. Our main objective in this work was to enhance the understanding of IL's toxicity towards AChE by using novel extreme learning machine (ELM) algorithms with better accuracy in a bid to shed light on designing novel environmentally benign ILs for future applications.

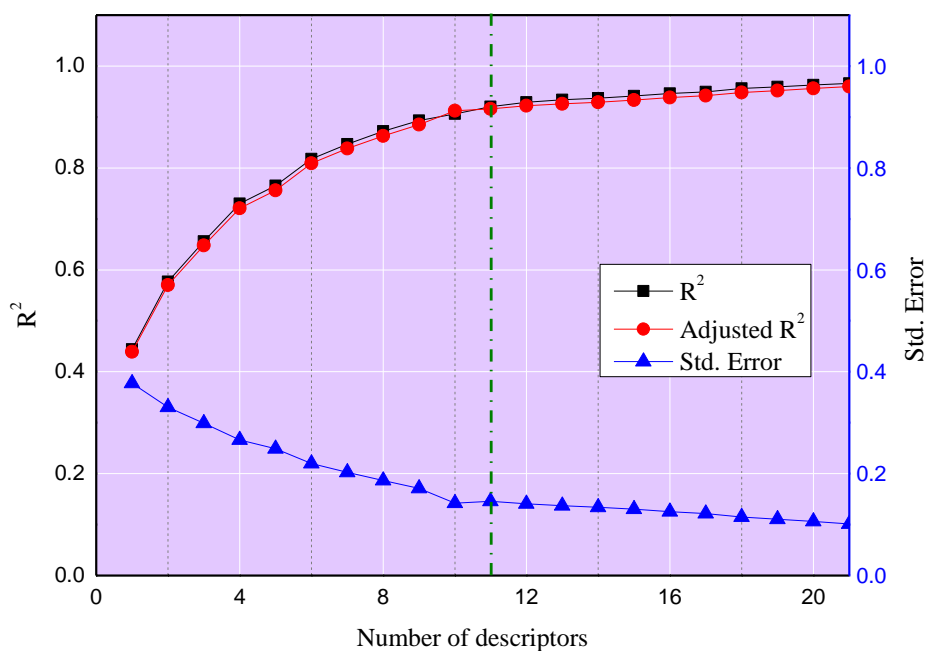
## 2. Results and Discussion

### 2.1. Quantitative Prediction of Multiple Linear Regression (MLR) Model

To obtain the predictive model with an optimal number of descriptors, the stepwise regression approach was utilized to choose the effective input descriptors. As shown in Figure 1, the number of descriptors gradually grows as the squared correlation coefficient ( $R^2$ ) and adjusted squared correlation coefficient ( $R^2_{\text{adjusted}}$ ) of the model grow, with simultaneous decrease in standard error (Std. Error). When the number of input parameters overtakes 11, the  $R^2$ ,  $R^2_{\text{adjusted}}$ , and standard error do not change evidently. Thus, the MLR model based on the optimal set of descriptors was determined. The final model is shown as Equation (1), where  $P_0$  stands for the intercept of the prediction model, while  $C_i$  and  $C_j$  mean the  $i$ th and  $j$ th coefficients respectively;  $S_{EP-h}$  is the electrostatic potential surface area for cations or anions at the electrostatic potential  $h$  while  $S_{\sigma-k}$  means the screening charge density distribution area for cations or anions at the screening charge density  $k$ .

$$\log EC_{50}(\text{AChE}) = P_0 + \sum_{i=1}^8 C_i S_{EP-h} + \sum_{j=9}^{11} C_j S_{\sigma-k} \quad (1)$$

Details of the coefficients and the parameters, including the  $t$  values of the descriptors, can be found in Table 1. In total, eight  $S_{EP}$  parameters and three  $S_{\sigma}$  parameters were chosen for the model, where  $C$  is the cation and the numbers are the corresponding electrostatic potential and surface screening charge density, respectively. Since  $t$  values indicate the significance of the parameters, it can be seen that  $S_{\sigma-C0.013}$  is the most vital parameter for the prediction of the toxicity of ILs. The  $\sigma$ -profile range can be qualitatively divided into three main parts: the hydrogen-bond (HB) donor region ( $\sigma < -0.0082 \text{ e}/\text{\AA}^2$ ) and the HB acceptor region ( $\sigma > 0.0082 \text{ e}/\text{\AA}^2$ ), as well as the non-polar region ( $-0.0082 \text{ e}/\text{\AA}^2 \leq \sigma \leq 0.0082 \text{ e}/\text{\AA}^2$ ). Therefore, it can be concluded that the larger the  $S_{\sigma-C0.013}$  value of the ILs, the lower the toxicity of the ILs. The  $S_{EP-C36.25}$  and  $S_{EP-C82.75}$  are the second and third crucial parameters, while their coefficients are 0.361 and  $-0.213$ , respectively. It should be mentioned that all the selected parameters are from cations, which illustrates that the cations of ILs have a vital influence on the toxicity of ILs, while the anions have little effect. This finding is consistent with the result that the cation makes the dominant contribution to toxicity for most commercial ILs [45].



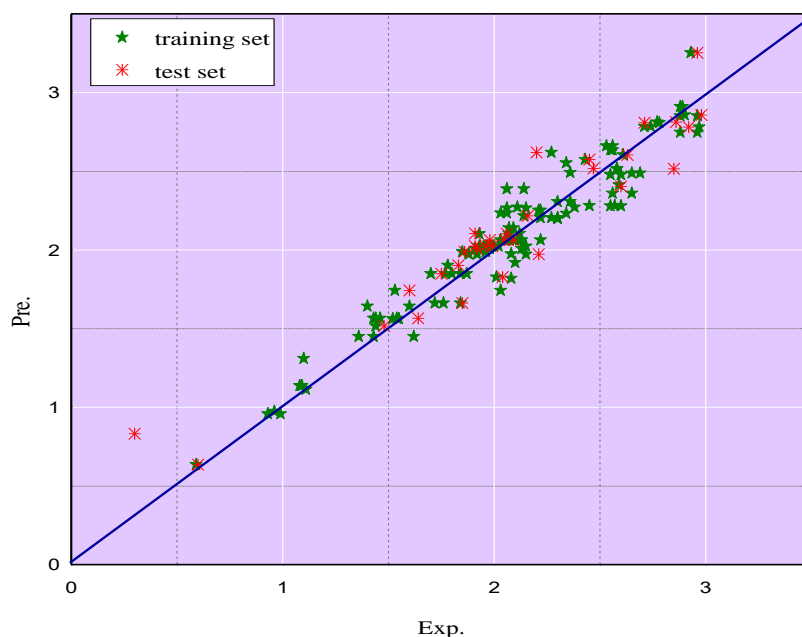
**Figure 1.** The squared correlation coefficient ( $R^2$ ) versus parameter number for the training data set.

**Table 1.** Coefficients ( $C_i$ ), parameters ( $S_{EP-h}$  and  $S_{\sigma-k}$ ) and the  $t$  values for Equation (1).

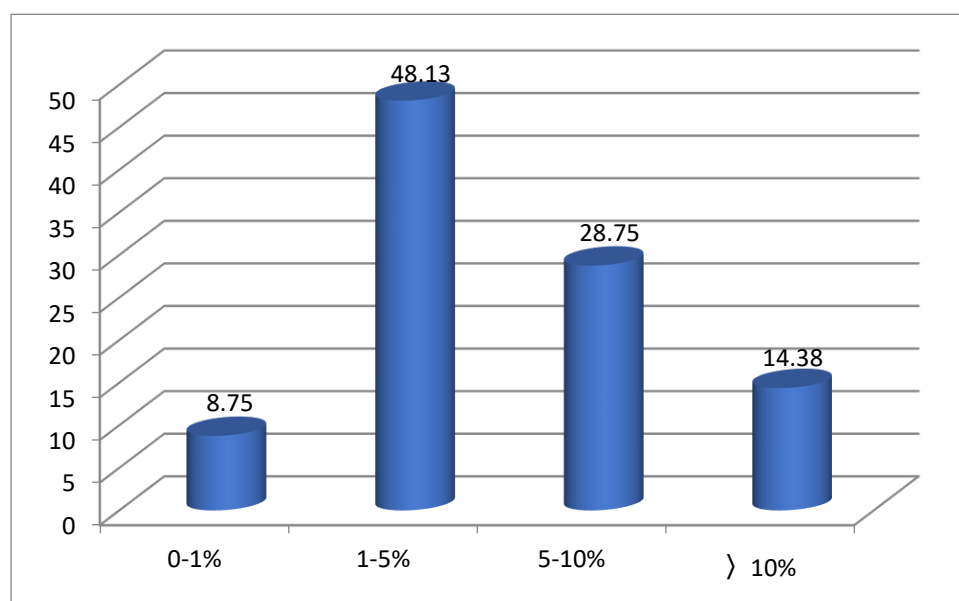
$i$	Coefficients- $C_i$	<sup>1</sup> $S_{EP-h}$	<sup>3</sup> $t$
1	-0.185	$S_{EP-C88.75}$	-6.582
2	-0.266	$S_{EP-C35.25}$	-6.731
3	0.361	$S_{EP-C36.25}$	10.565
4	-0.213	$S_{EP-C82.75}$	-10.501
5	-0.289	$S_{EP-C31.25}$	-7.917
6	0.133	$S_{EP-C89.75}$	4.936
7	0.204	$S_{EP-C64.25}$	7.241
8	-0.142	$S_{EP-C53.25}$	-5.498
$j$	Coefficients- $C_j$	<sup>2</sup> $S_{\sigma-k}$	$t$
9	0.537	$S_{\sigma-C0.013}$	18.957
10	-0.102	$S_{\sigma-C-0.012}$	-9.583
11	0.234	$S_{\sigma-C-0.016}$	8.303
$P_0$	2.712		

<sup>1</sup> The electrostatic potential surface area ( $S_{EP}$ ) at the electrostatic potential  $h$ ; <sup>2</sup> The screening charge density distribution area ( $S_{\sigma}$ ) at the screening charge density  $k$ ; <sup>3</sup>  $t$  value stands for the importance of the parameter.

Experimental and predicted the  $\log EC_{50}$  values of ILs towards acetyl cholinesterase enzyme (AChE) ( $\log EC_{50}$  (AChE)) values of ILs (as shown in Supporting Information Table S1) via the MLR model for both the training set as well as the test set are plotted in Figure 2, where it can be seen that the predicted values are close to the experimental results. Figure 3 compares the absolute relative deviation (ARD%) within the range of  $\log EC_{50}$  (AChE) values from the MLR model. It is evident that the largest proportion of the predicted values falls in the range of 1–5%, which accounts for 48.13%. Similarly, the percentages of predicted data fall in the range of 5–10% and over 10%, accounting for 28.75% and 14.38%, respectively. Only 8.75% of the predicted ARD% values are in the range of 0–1%. These conclusions showed the good performance of the MLR model to predict the toxicity of ILs.



**Figure 2.** Comparisons of predicted the  $\log EC_{50}$  values of ILs towards acetyl cholinesterase enzyme (AChE) ( $\log EC_{50}$  (AChE)) by multiple linear regression (MLR) in external validation with the experimental  $\log EC_{50}$  (AChE).

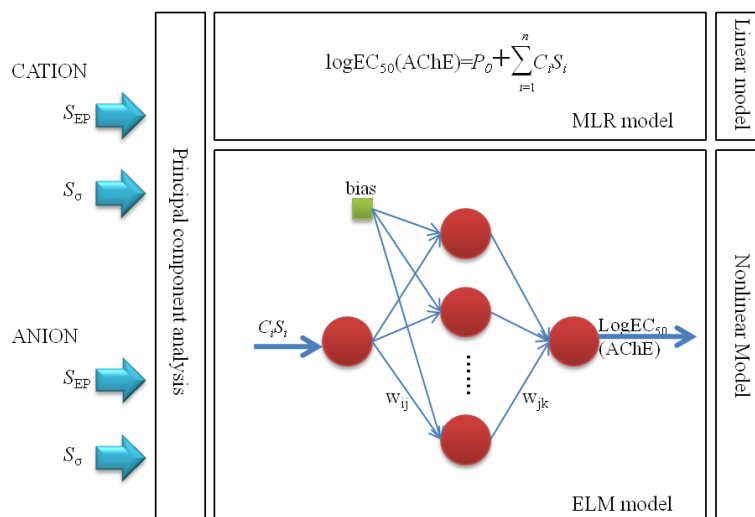


**Figure 3.** Percent of  $\log EC_{50}$  (AChE) values in different absolute relative deviation (ARD%) ranges of the MLR model.

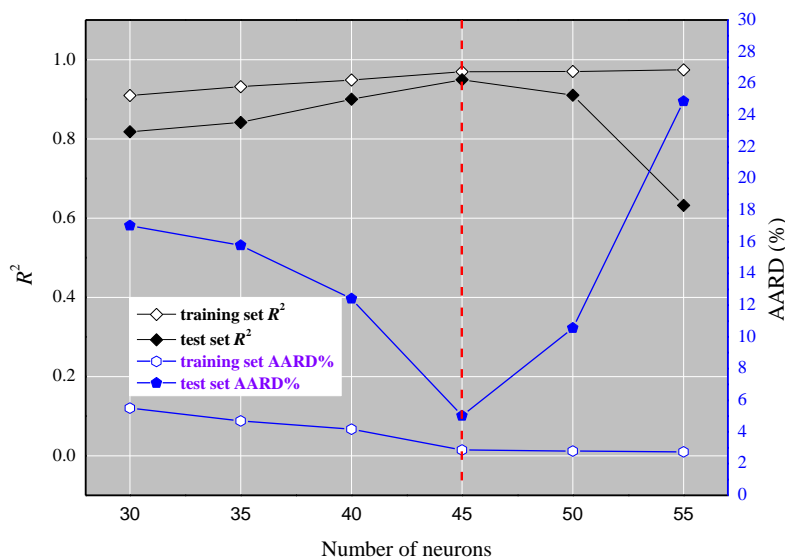
## 2.2. Quantitative Prediction of Extreme Learning Machine (ELM) Model

Based upon the foundation of 11 descriptors selected via stepwise regression method and the data sets from the former model, a more effective nonlinear ELM model was established. In the ELM model, there are three layers, as shown in Figure 4. The sine function in the ELM model acted as an activation function between the input and intermediate layers, while the linear function was from the intermediate neurons to the output variables. The bias and  $\omega_{ij}$  produced randomly were the parameters between the first and intermediate layers. The  $\omega_{jk}$  was the coefficient needed to calculate between the intermediate and final layer. After the determination of the functions, the next step was to acquire the ideal number of neurons between the input variables and intermediate neurons. As shown

in Figure 5, in terms of the training set, the  $R^2$  value slightly increases with the growth in the number of neurons, whereas the average absolute relative deviation (AARD%) considerably drops when the number of neurons is less than 45, beyond which it dramatically rises. The variation tendency of  $R^2$  and AARD% for the test set are almost opposite to the growing number of neurons. Therefore, the best number of neurons (45) is acquired for the prediction the toxicity of ILs by the ELM model.

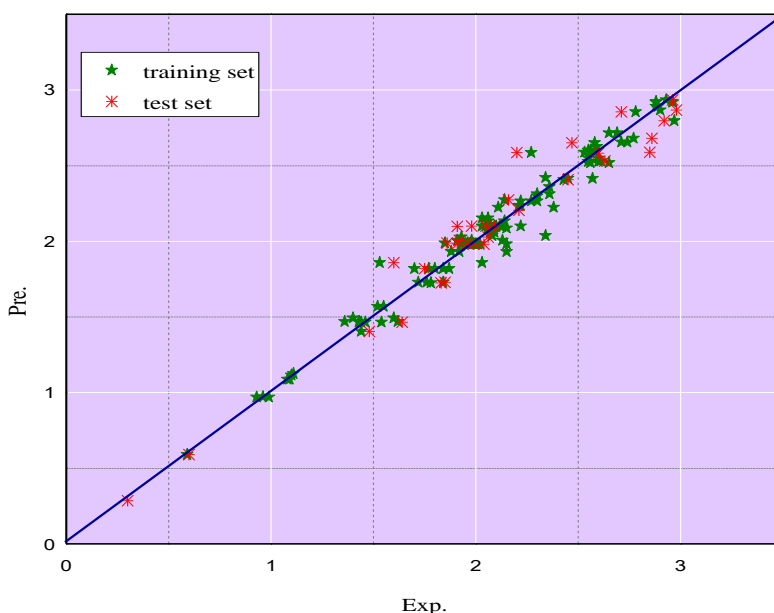


**Figure 4.** Scheme of linear and non-linear models programming methodology. The screening charge density distribution area ( $S_{\sigma}$ ) and the electrostatic potential surface area ( $S_{EP}$ ) of cations and anions were calculated employed to develop models optimal number of descriptors. The effective input descriptors were chosen by the stepwise regression approach and then the multiple linear regression (MLR) model was built with the obtained coefficients ( $C_i$ ) and selected descriptor ( $S_i$ ) to predict the the toxicity values ( $\log EC_{50}$ ) of ionic liquids (ILs) towards acetyl cholinesterase enzyme (AChE). The extreme learning machine (ELM) comprising three parts—the input layer, hidden layer, and output layer—was established based on the same input descriptors. The parameters  $w_{ij}$  and bias from the input layer to hidden layer were generated randomly and the parameters  $w_{jk}$  were coefficients from the hidden layer to output layer.

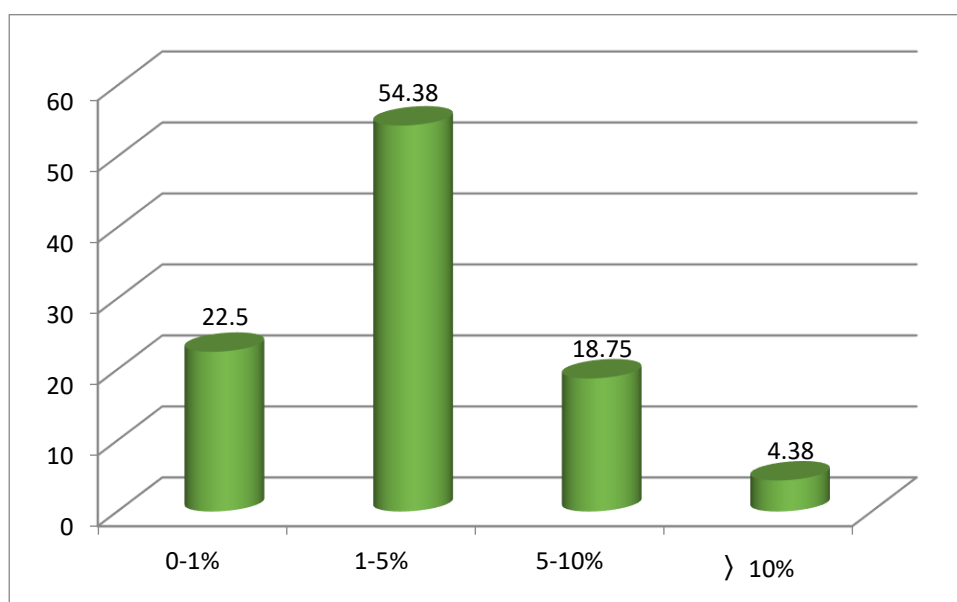


**Figure 5.** The  $R^2$  and average absolute relative deviation (AARD%) versus the parameter number for the ELM) model.

The comparisons of the predicted  $\log EC_{50}$  (AChE) values by ELM (as shown in Supporting Information Table S1) in both the training and test sets with the experimental  $\log EC_{50}$  (AChE) values are shown in the Figure 6, which reveals that the predicted values match well with the experimental data. Figure 7 illustrates the proportion of  $\log EC_{50}$  (AChE) values in the different ARD% ranges of the ELM model. It is clear that the largest percentage of ARD% from the  $\log EC_{50}$  (AChE) values is 54.38% within the range of 1–5%, followed by the 22.5% in the range of 0–1%. The percentage of ARD% of the  $\log EC_{50}$  (AChE) values in the range of 5–10% and over 10% account for 18.75% and 4.38%, respectively. These results showcase the validity and reliability of the ELM model to predict the toxicity of ILs.



**Figure 6.** Comparisons of the predicted  $\log EC_{50}$  (AChE) by ELM in external validation with the experimental  $\log EC_{50}$  (AChE).



**Figure 7.** Percent of  $\log EC_{50}$  (AChE) values in different ARD% ranges of the ELM model.



### 2.3. Comparison between Two Quantitative Structure-activity Relationships (QSAR) Models

As discussed above, satisfactory results are obtained for the two models having high  $R^2$  values. Detailed comparison of the statistical parameters of the training set, test set, and total data of different QSAR models are given in Table 2.

**Table 2.** Comparison of the statistical parameters by different quantitative structure-activity relationships (QSAR) models.

Model	Dataset	No.	<sup>1</sup> $R^2$	<sup>2</sup> AARD%	<sup>3</sup> RMSE
MLR	Training	128	0.920	5.18	0.136
	Test	32	0.914	11.05	0.180
	Total	160	0.917	6.35	0.145
ELM	Training	128	0.969	2.86	0.084
	Test	32	0.950	5.02	0.134
	Total	160	0.964	3.29	0.096

<sup>1</sup> Squared correlation coefficient ( $R^2$ ); <sup>2</sup> Average absolute relative deviation (AARD%); <sup>3</sup> Root-mean-square error (RMSE).

It is evident from Table 2 that the  $R^2$  values of the training set, test set, and total data using the ELM model are 0.969, 0.950, and 0.964, respectively, which are comparatively higher than of the  $R^2$  values of the MLR model, i.e., 0.920 for the training set, 0.914 for the test set, and 0.917 for the total set. This trend is valid for both AARD% and RMSE, meaning that the non-linear ELM shows better predictive performance than MLR. For example, AARD% values for the training set by the MLR and ELM models are 5.18 and 2.86, while the RMSE values are 1.534 and 0.950, respectively. A similar situation can easily be observed for the test set and all data sets using two models. Undoubtedly, the ELM algorithm is a powerful prediction tool with better accuracy.

In addition, according to the effective criteria used in the literature [47–50], a QSAR model with acceptable predictive ability should satisfy the conditions in Equations (2)–(5). As can be seen from Table 3, all the coefficients for both the MLR and ELM models satisfy the above-mentioned criteria, illustrating the good predictive power of the models in this study. To further verify the reliability of the models established in this study, we also predicted the toxicity values of five new ILs. The specific prediction results can be found in the Supporting Information (Table S2). It can be seen that the predicted values agree well with the experimental values, indicating that the models we built are reliable, and the ELM model has better prediction performance than the MLR model.

$$R^2 > 0.7 \quad (2)$$

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \text{ and } 0.85 < k < 1.15 \quad (3)$$

$$\frac{R^2 - R_0'^2}{R^2} < 0.1 \text{ and } 0.85 < k' < 1.15 \quad (4)$$

$$|R_0^2 - R_0'^2| < 0.3 \quad (5)$$



**Table 3.** The results of external validation of the test set in the multiple linear regression (MLR) and extreme learning machine (ELM) models.

Model	$R^2$	$^1 k$	$^2 k'$	$^3 R_0^2$	$^4 R_0'^2$	$(R^2 - R_0^2)/R^2$
MLR	0.914	0.9876	1.0058	0.9975	0.9996	-0.094
ELM	0.950	0.9944	1.0018	0.9996	1.0000	-0.051

<sup>1</sup>  $k$  is slope of the calculated versus experimental values; <sup>2</sup>  $k'$  is slope of the experimental versus calculated values; <sup>3</sup> the coefficients of calculated values versus experimental values ( $R_0^2$ ); <sup>4</sup> the coefficients of experimental values versus calculated values ( $R_0'^2$ ).

We compared our estimated results with the reported values, as summarized in Table 4. Our two QSAR models, MLR ( $R^2 = 0.917$ ) and ELM ( $R^2 = 0.964$ ), have higher  $R^2$  values compared with those of other models, except the Multiple linear regression (MLP) model ( $R^2 = 0.973$ ). Although MLP exhibits the highest value of  $R^2$ , our models used less input parameters. Our models show AARD% values of 6.35 for MLR and 3.29 for ELM, which fall in the range of the reported values, i.e., 2.8 to 9.35 (Table 4). In addition, as can be seen from Table 4, the RMSE values of our models are also relatively low compared to the literature [45]. Taking these factors into consideration, our models are reliable and meaningful for accurate prediction of the toxicity of ILs towards AChE.

**Table 4.** Comparisons of our work with the reported works.

Method	No. Parameter	$R^2$	AARD%	RMSE	Ref.
<sup>1</sup> MLR	12	0.814	7.7	/	[42]
<sup>2</sup> MLP	12	0.973	2.8	/	[42]
<sup>3</sup> RB	12	0.842	7.1	/	[42]
MLR	17	0.877	9.35	0.212	[45]
MLR	11	0.917	6.35	0.145	This work
<sup>4</sup> ELM	11	0.964	3.29	0.096	This work

<sup>1</sup> Multiple linear regression (MLR); <sup>2</sup> Multilayer perceptron (MLP); <sup>3</sup> Radial-basis function (RB); <sup>4</sup> Extreme learning machine (ELM); / indicates that the RMSE value cannot be obtained from literature.

### 3. Framework

As we know, quantitative structure-activity relationships (QSAR) can be implemented in the form of linear and non-linear algorithms. The former are relatively easy and intuitively present the impact of each parameter upon the properties, while the latter are considered to be suitable for accurate prediction in real-world scenarios. In this work, the toxicity of ionic liquids (ILs) was predicted according to the scheme shown in Figure 4.

### 4. Dataset and Structural Descriptors

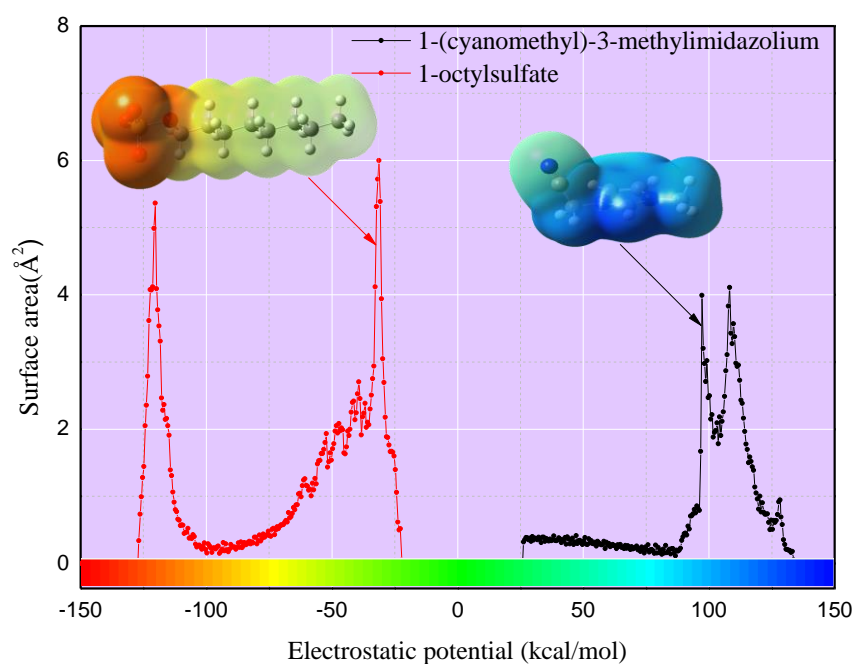
#### 4.1. Dataset

Herein, toxicity data in acetyl cholinesterase enzyme (AChE) of 160 ILs was collected from the widely acknowledged ionic liquid (IL) database [40,51,52]. Concentration for 50% of maximal effect ( $EC_{50}$ ) values ( $\mu\text{M}$ ) for AChE, written as  $\log EC_{50}$ , were converted into the form of a logarithm of half maximal effective concentration. The whole data set was divided into two parts: a training set of 80% ILs to build the model and a set of 20% ionic liquids (ILs) to evaluate the model's predictability. Detailed information on the ILs studied in this paper, including the names, SMILES strings,  $\log_{10}(EC_{50})$  values, etc., can be found in Supporting Information.

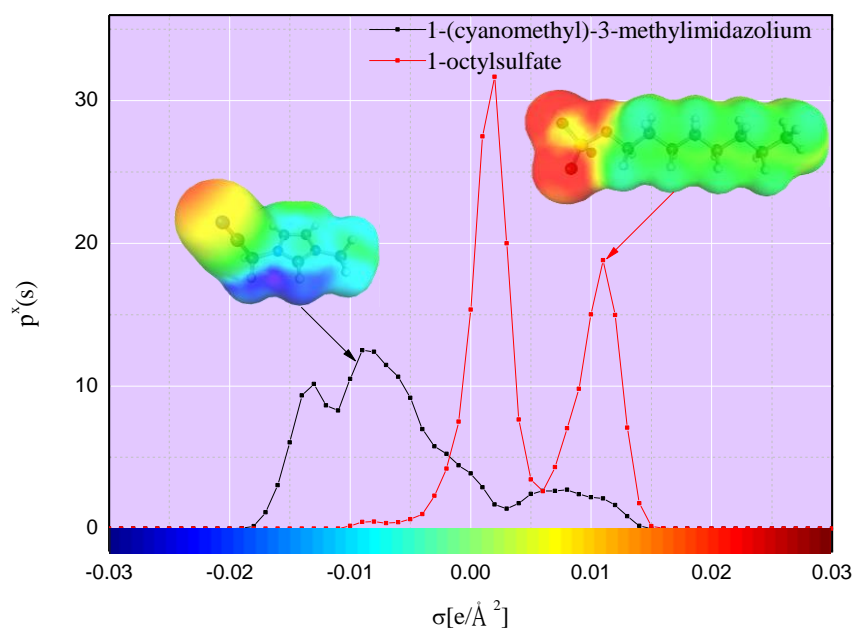
#### 4.2. Calculation of Descriptors

As established, the related compounds are represented by theoretical molecular descriptors, which are of key significance for the predictive performance of the models [40]. The screening charge

density distribution area ( $S_\sigma$ ), obtained from the histogram function of the  $\sigma$ -profiles given by the conductor-like screening model for real solvent (COSMO-RS) computation, is an a priori quantum chemistry descriptor that could quantitatively represent the molecule's polar surface screen charge on the polarity scale [36]. The electrostatic potential surface area ( $S_{EP}$ ) of molecules refers to the surface areas of the molecules within the interval of the different electrostatic potential, and it can be used for the prediction of the material's properties due to its rich information at the electron level. In this work, two descriptors  $S_{EP}$  and  $S_{\sigma\text{-profiles}}$  were employed to develop models and were calculated by different programs based on the optimal structures of cations and anions. First of all, the  $S_{EP}$  files of the corresponding cations were calculated using Multiwfn software with an electrostatic potential range of 0~150 kcal/mol, and was kept the same for the anions in the range of -150~0 kcal/mol, with a step size of 0.5 kcal/mol [53]. For example, the  $S_{EP-C88.75}$  means the molecular surface areas in the electrostatic potential scale of 88.5–89 kcal/mol. Then, the conductor-like screening model (COSMO) files of cations and anions were obtained using Gaussian 03 software [54]. Finally,  $S_{\sigma\text{-profiles}}$  files of corresponding cations and anions were calculated with a  $\sigma$ -profile ranging from -0.03 to 0.03 e/Å<sup>2</sup> and with a step size of 0.001 e/Å<sup>2</sup>, respectively. We chose cation (1-(cyanomethyl)-3-methylimidazolium) and anion (1-octylsulfate) to present the representative  $S_{EP}$  and  $S_{\sigma\text{-profiles}}$ , and the results were depicted in Figure 8; Figure 9, respectively. As shown in Figure 8, the darker shade of red for 1-(cyanomethyl)-3-methylimidazolium or blue for 1-octylsulfate indicates stronger polarity. Figure 9 presents a similar situation of electrostatic potential surfaces.



**Figure 8.** The electrostatic potential surface area ( $S_{EP}$ ) of representative cations and anions of ionic liquids (ILs).



**Figure 9.** The screening charge density distribution area ( $S_{\sigma}$ ) of representative cations and anions of ILs.

#### 4.3. Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) is widely harnessed in quantitative structure-activity relationships (QSAR) to determinate the relationship between a dependent variable and various independent variables. The linear combination relationship of the independent variables is described by a set of coefficients in this model. The structural characteristics to the toxicity can be linked as follows:

$$\log EC_{50} = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (6)$$

where  $b_0$  is the intercept and  $b_1$ ,  $b_2$ , and  $b_n$  are regression coefficients of the corresponding descriptors;  $n$  is the number of descriptors used in the equation in order to figure out the optimal regression model;  $x$  denotes the descriptor used to describe the chemical structure of the compound. MLR consists of stepwise selection of descriptors, which combines the forward and backward procedures.

#### 4.4. Extreme Learning Machine (ELM)

Extreme learning machine (ELM) architecture was proposed for the first time by Huang et al., as a single hidden layer feed-forward neural network that has a very strong learning ability for application in basic classification and regression [55]. The most key characteristic of ELM is that it has faster learning speed compared with the traditional learning algorithms. In addition, ELM can be easily used and is a fast-learning and effective algorithm for the feed-forward neural network with single hidden layer interpolation capability and universal approximation capability, which are the two important characteristics of the feed-forward neural network. To some extent, ELM is similar to an artificial neural network (ANN), and its weights and biases in the first layer are randomly initialized and kept constant, while the weights of the second layer are selected by diminishing the least squared error. In practical applications, the training data set is mainly concerned with the specific issues [56]. The data sets include actual results and the related factors. Through iterations to finish the learning process, the impact factors and corresponding results were put into the ELM for training during the training process. Afterward, using the trained ELM, we only needed to input the training data set, similar to the influencing factors. It has been reported that ELM is feasible in many fields, such as protein sequence classification, regression problems, etc., to easily achieve good performance with fast speed [57]. Evidently, increasingly more research in this regard shows that excellent predictive

performance can be achieved using the non-linear model developed by the ELM algorithm. More information on the theory and application of ELM can be found elsewhere [55–57].

#### 4.5. Evaluation of Quantitative Structure-activity Relationships (QSAR)

In this work, several parameters were defined to evaluate the model performance, as measured by different metrics: adjusted squared correlation coefficient ( $R^2_{\text{adjusted}}$ ), squared correlation coefficient ( $R^2$ ), average absolute relative deviation (AARD%), root-mean-square error (RMSE), absolute relative deviation (ARD%), the coefficients of calculated values versus experimental values ( $R_0^2$ ), experimental values versus calculated values ( $R_0'^2$ ), slope  $k$  of the calculated versus experimental values, and slope  $k'$  of the experimental versus calculated values. The corresponding definition could be expressed in Equations (7)–(15).

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i^{\text{exp}} - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{\text{cal}} - y_i^{\text{exp}})^2}{\sum_{i=1}^{N_p} (y_i^{\text{exp}} - \bar{y}_m)^2} \quad (7)$$

$$R^2_{\text{adjusted}} = 1 - \frac{(1 - R^2)(N_p - 1)}{N_p - p - 1} \quad (8)$$

$$\text{AARD}(\%) = 100 \times \sum_{i=1}^{N_p} \left( \left| \frac{y_i^{\text{cal}} - y_i^{\text{exp}}}{y_i^{\text{exp}}} \right| \right) / N_p \quad (9)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N_p} (y_i^{\text{cal}} - y_i^{\text{exp}})^2 / N_p} \quad (10)$$

$$\text{ARD}(\%) = |100 \times (y_i^{\text{cal}} / y_i^{\text{exp}} - 1.0)| \quad (11)$$

$$R_0^2 = 1 - \frac{\sum_{i=1}^{N_p} (y_i^{\text{cal}} - k y_i^{\text{cal}})^2}{\sum_{i=1}^{N_p} (y_i^{\text{cal}} - \bar{y}^{\text{cal}})^2} \quad (12)$$

$$R_0'^2 = 1 - \frac{\sum_{i=1}^{N_p} (y_i^{\text{exp}} - k y_i^{\text{exp}})^2}{\sum_{i=1}^{N_p} (y_i^{\text{exp}} - k' \bar{y}^{\text{exp}})^2} \quad (13)$$

$$k = \frac{\sum_{i=1}^{N_p} y_i^{\text{exp}} y_i^{\text{cal}}}{\sum_{i=1}^{N_p} (y_i^{\text{cal}})^2} \quad (14)$$

$$k' = \frac{\sum_{i=1}^{N_p} y_i^{\text{exp}} y_i^{\text{cal}}}{\sum_{i=1}^{N_p} (y_i^{\text{exp}})^2} \quad (15)$$

where  $y$  means  $\log EC_{50}$  (AChE) value of ILs while the “cal” and “exp” superscripts stand for calculated and experimental values, respectively.  $N_p$  denotes the number of data points for the corresponding dataset;  $\bar{y}_m$  represents the average  $\log EC_{50}$  (AChE) value of ILs for all of the selected data.

## 5. Conclusions

Safe operation and benign utilization are crucial for ILs, therefore the scope of accurately predicting the toxicity of ILs is of vital interest for scientific and technological advancement. In this work, detailed data analyses of the toxicity of ILs were made based on the established database and novel QSAR models were developed for toxicity prediction of 160 ILs in AChE. Good correlation between the estimated and original data was observed for the training set, test set, and the total set. It was found that the newly used ELM model gave the highest  $R^2$  values and the lowest values for AARD% and RMSE. The results validated the superior performance of ELM for estimation of toxicity of ILs. This study will be helpful in design and screening of environmentally friendly commercial ILs.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/9/2186/s1>.

**Author Contributions:** Writing—original draft, P.Z.; Writing—review & editing, X.K., Y.Z., U.L. and H.Z.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ranke, J.; Stolte, S.; Störmann, R.; Arning, J.; Jastorff, B. Design of sustainable chemical products the example of ionic liquids. *Chem. Rev.* **2007**, *107*, 2183–2206. [[CrossRef](#)]
2. Niedermaier, I.; Kolbeck, C.; Taccardi, N.; Schulz, P.S.; Li, J.; Drewello, T.; Wasserscheid, P.; Steinrück, H.P.; Maier, F. Organic reactions in ionic liquids studied by in situ XPS. *Chem. Phys.* **2012**, *13*, 1725–1735. [[CrossRef](#)] [[PubMed](#)]
3. Yao, W.; Wang, H.; Cui, G.; Li, Z.; Zhu, A.; Zhang, S.; Wang, J. Tuning the hydrophilicity and hydrophobicity of the respective cation and anion: Reversible phase transfer of ionic liquids. *Angew. Chem. Int. Ed.* **2016**, *55*, 7934–7938. [[CrossRef](#)]
4. Chen, K.; Shi, G.; Zhang, W.; Li, H.; Wang, C. Computer-Assisted Design of Ionic Liquids for Efficient Synthesis of 3 (2 H)-Furanones: A Domino Reaction Triggered by CO<sub>2</sub>. *J. Am. Chem. Soc.* **2016**, *138*, 14198–14201. [[CrossRef](#)] [[PubMed](#)]
5. Biswas, K.; Zhang, Q.; Chung, I.; Song, J.-H.; Androulakis, J.; Freeman, A.J.; Kanatzidis, M.G. Synthesis in ionic liquids: [Bi<sub>2</sub>Te<sub>2</sub>Br](AlCl<sub>4</sub>), a direct gap semiconductor with a cationic framework. *J. Am. Chem. Soc.* **2010**, *132*, 14760–14762. [[CrossRef](#)]
6. Sawant, A.; Raut, D.; Darvatkar, N.; Salunkhe, M. Recent developments of task-specific ionic liquids in organic synthesis. *Green Chem. Lett. Rev.* **2011**, *4*, 41–54. [[CrossRef](#)]
7. Cooper, E.R.; Andrews, C.D.; Wheatley, P.S.; Webb, P.B.; Wormald, P.; Morris, R.E. Ionic liquids and eutectic mixtures as solvent and template in synthesis of zeolite analogues. *Nature* **2004**, *430*, 1012. [[CrossRef](#)]
8. Plechkova, N.V.; Seddon, K.R. Applications of ionic liquids in the chemical industry. *Chem. Soc. Rev.* **2008**, *37*, 123–150. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, G.R.; Munoz, M.; Etzold, B.J. Accelerating Oxygen-Reduction Catalysts through Preventing Poisoning with Non-Reactive Species by Using Hydrophobic Ionic Liquids. *Angew. Chem. Int. Ed.* **2016**, *55*, 2257–2261. [[CrossRef](#)]
10. Asadi, M.; Kim, K.; Liu, C.; Addepalli, A.V.; Abbasi, P.; Yasaei, P.; Phillips, P.; Behranginia, A.; Cerrato, J.M.; Haasch, R. Nanostructured transition metal dichalcogenide electrocatalysts for CO<sub>2</sub> reduction in ionic liquid. *Science* **2016**, *353*, 467–470. [[CrossRef](#)]
11. Olivier-Bourbigou, H.; Magna, L.; Morvan, D. Ionic liquids and catalysis: Recent progress from knowledge to applications. *Appl. Catal. A: Gen.* **2010**, *373*, 1–56. [[CrossRef](#)]
12. Wasserscheid, P.; Keim, W. Ionic liquids—new “solutions” for transition metal catalysis. *Angew. Chem. Int. Ed.* **2000**, *39*, 3772–3789. [[CrossRef](#)]
13. Gurkan, B.E.; de la Fuente, J.C.; Mindrup, E.M.; Ficke, L.E.; Goodrich, B.F.; Price, E.A.; Schneider, W.F.; Brennecke, J.F. Equimolar CO<sub>2</sub> absorption by anion-functionalized ionic liquids. *J. Am. Chem. Soc.* **2010**, *132*, 2116–2117. [[CrossRef](#)] [[PubMed](#)]
14. Cao, L.; Gao, J.; Zeng, S.; Dong, H.; Gao, H.; Zhang, X.; Huang, J. Feasible ionic liquid-amine hybrid solvents for carbon dioxide capture. *Int. J. Greenh. Gas. Control.* **2017**, *66*, 120–128. [[CrossRef](#)]

15. Wang, C.; Luo, H.; Jiang, D.E.; Li, H.; Dai, S. Carbon dioxide capture by superbase-derived protic ionic liquids. *Angew. Chem. Int. Ed.* **2010**, *49*, 5978–5981. [[CrossRef](#)]
16. Cao, L.; Huang, J.; Zhang, X.; Zhang, S.; Gao, J.; Zeng, S. Imidazole tailored deep eutectic solvents for CO<sub>2</sub> capture enhanced by hydrogen bonds. *Phys. Chem. Chem. Phys.* **2015**, *17*, 27306–27316. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, S.; Sun, J.; Zhang, X.; Xin, J.; Miao, Q.; Wang, J. Ionic liquid-based green processes for energy production. *Chem. Soc. Rev.* **2014**, *43*, 7838–7869. [[CrossRef](#)] [[PubMed](#)]
18. Watanabe, M.; Thomas, M.L.; Zhang, S.; Ueno, K.; Yasuda, T.; Dokko, K. Application of ionic liquids to energy storage and conversion materials and devices. *Chem. Rev.* **2017**, *117*, 7190–7239. [[CrossRef](#)] [[PubMed](#)]
19. MacFarlane, D.R.; Forsyth, M.; Howlett, P.C.; Kar, M.; Passerini, S.; Pringle, J.M.; Ohno, H.; Watanabe, M.; Yan, F.; Zheng, W. Ionic liquids and their solid-state analogues as materials for energy generation and storage. *Nat. Rev. Mater.* **2016**, *1*, 15005. [[CrossRef](#)]
20. Guo, F.; Zhang, S.; Wang, J.; Teng, B.; Zhang, T.; Fan, M. Synthesis and applications of ionic liquids in clean energy and environment: A review. *Curr. Org. Chem.* **2015**, *19*, 455–468. [[CrossRef](#)]
21. Anastas, P.T.; Zimmerman, J.B. Peer reviewed: Design through the 12 principles of green engineering. *Environ. Sci. Technol.* **2003**, *37*, 94A–101A. [[CrossRef](#)]
22. Swatloski, R.P.; Holbrey, J.D.; Rogers, R.D. Ionic liquids are not always green: Hydrolysis of 1-butyl-3-methylimidazolium hexafluorophosphate. *Green Chem.* **2003**, *5*, 361–363. [[CrossRef](#)]
23. Zhao, D.; Liao, Y.; Zhang, Z. Toxicity of ionic liquids. *Clean-Soilairwater* **2007**, *35*, 42–48. [[CrossRef](#)]
24. Jastorff, B.; Störmann, R.; Ranke, J.; Mölter, K.; Stock, F.; Oberheitmann, B.; Hoffmann, W.; Hoffmann, J.; Nüchter, M.; Ondruschka, B. How hazardous are ionic liquids? Structure–activity relationships and biological testing as important elements for sustainability evaluation. *Green Chem.* **2003**, *5*, 136–142. [[CrossRef](#)]
25. Docherty, K.M.; Kulpa, C.F., Jr. Toxicity and antimicrobial activity of imidazolium and pyridinium ionic liquids. *Green Chem.* **2005**, *7*, 185–189. [[CrossRef](#)]
26. Amde, M.; Liu, J.-F.; Pang, L. Environmental application, fate, effects, and concerns of ionic liquids: A review. *Environ. Sci. Technol.* **2015**, *49*, 12611–12627. [[CrossRef](#)] [[PubMed](#)]
27. Mehrkesh, A.; Karunanithi, A.T. Life-cycle perspectives on aquatic ecotoxicity of common ionic liquids. *Environ. Sci. Technol.* **2015**, *50*, 6814–6821. [[CrossRef](#)] [[PubMed](#)]
28. Rogers, R.D.; Seddon, K.R. Ionic liquids—solvents of the future? *Science* **2003**, *302*, 792–793. [[CrossRef](#)]
29. Bubalo, M.C.; Radošević, K.; Redovniković, I.R.; Halambek, J.; Srček, V.G. A brief overview of the potential environmental hazards of ionic liquids. *Ecotoxicol. Environ. Saf.* **2014**, *99*, 1–12. [[CrossRef](#)]
30. Scire, S.; Musumarra, G. A QSPR approach to the ecotoxicity of ionic liquids (*Vibrio fischeri*) using VolSurf principal properties. *Toxicol. Res.* **2016**, *5*, 1090–1096.
31. Yousefinejad, S.; Hemmateenejad, B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 177–204. [[CrossRef](#)]
32. Gupta, S.; Basant, N.; Singh, K.P. Nonlinear QSAR modeling for predicting cytotoxicity of ionic liquids in leukemia rat cell line: An aid to green chemicals designing. *Environ. Sci. Pollut. Res.* **2015**, *22*, 1–12. [[CrossRef](#)] [[PubMed](#)]
33. Ghaedi, A. Predicting the cytotoxicity of ionic liquids using QSAR model based on SMILES optimal descriptors. *J. Mol. Liq.* **2015**, *208*, 269–279. [[CrossRef](#)]
34. Zhao, Y.; Gani, R.; Afzal, R.M.; Zhang, X.; Zhang, S. Ionic liquids for absorption and separation of gases: An extensive database and a systematic screening method. *Aiche J.* **2017**, *63*, 1353–1367. [[CrossRef](#)]
35. Rybinska, A.; Sosnowska, A.; Grzonkowska, M.; Barycki, M.; Puzyn, T. Filling environmental data gaps with QSPR for ionic liquids: Modeling n-octanol/water coefficient. *J. Hazard. Mater.* **2016**, *303*, 137. [[CrossRef](#)] [[PubMed](#)]
36. Zhao, Y.; Huang, Y.; Zhang, X.; Zhang, S. A quantitative prediction of the viscosity of ionic liquids using  $\Sigma\sigma$ -profile molecular descriptors. *Phys. Chem. Chem. Phys.* **2015**, *17*, 3761–3767. [[CrossRef](#)]
37. Ranke, J.; Jastorff, B. Multidimensional risk analysis of antifouling biocides. *Environ. Sci. Pollut. Res.* **2000**, *7*, 105–114. [[CrossRef](#)]
38. Bailey, M.M.; Townsend, M.B.; Jernigan, P.L.; Sturdivant, J.; Hough-Troutman, W.L.; Rasco, J.F.; Swatloski, R.P.; Rogers, R.D.; Hood, R.D. Developmental toxicity assessment of the ionic liquid 1-butyl-3-methylimidazolium chloride in CD-1 mice. *Green Chem.* **2008**, *10*, 1213–1217. [[CrossRef](#)]



39. Petkovic, M.; Ferguson, J.L.; Gunaratne, H.N.; Ferreira, R.; Leitão, M.C.; Seddon, K.R.; Rebelo, L.P.N.; Pereira, C.S. Novel biocompatible cholinium-based ionic liquids-toxicity and biodegradability. *Green Chem.* **2010**, *12*, 643–649. [CrossRef]
40. Zhao, Y.; Zhao, J.; Huang, Y.; Zhou, Q.; Zhang, X.; Zhang, S. Toxicity of ionic liquids: Database and prediction via quantitative structure-activity relationship method. *J. Hazard. Mater.* **2014**, *278*, 320–329. [CrossRef]
41. Ranke, J.; Bottin-Weber, M.U.; Stock, F.; Stolte, S.; Arning, J.; Stoermann, R.; Jastorff, B. Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity. *Ecotoxicol. Environ. Saf.* **2007**, *67*, 430–438. [CrossRef]
42. Torrecilla, J.S.; García, J.; Rojo, E.; Rodríguez, F. Estimation of toxicity of ionic liquids in Leukemia Rat Cell Line and Acetylcholinesterase enzyme by principal component analysis, neural networks and multiple lineal regressions. *J. Hazard. Mater.* **2009**, *164*, 182. [CrossRef]
43. Oliveira, M.V.; Vidal, B.T.; Melo, C.M.; de Miranda, R.d.C.; Soares, C.M.; Coutinho, J.A.; Ventura, S.P.; Mattedi, S.; Lima, Á.S. (Eco) toxicity and biodegradability of protic ionic liquids. *Chemosphere* **2016**, *147*, 460–466. [CrossRef]
44. Romero, A.; Santos, A.; Tojo, J.; Rodriguez, A. Toxicity and biodegradability of imidazolium ionic liquids. *J. Hazard. Mater.* **2008**, *151*, 268–273. [CrossRef]
45. Yan, F.; Xia, S.; Wang, Q.; Ma, P. Predicting Toxicity of Ionic Liquids in Acetylcholinesterase Enzyme by the Quantitative Structure–Activity Relationship Method Using Topological Indexes. *J. Chem. Eng. Data* **2012**, *57*, 2252–2257. [CrossRef]
46. Basant, N.; Gupta, S.; Singh, K.P. Predicting acetyl cholinesterase enzyme inhibition potential of ionic liquids using machine learning approaches: An aid to green chemicals designing. *J. Mol. Liq.* **2015**, *209*, 404–412. [CrossRef]
47. Golbraikh, A.; Tropsha, A. Beware of q<sup>2</sup>! *J. Mol. Graph. Model.* **2002**, *20*, 269–276. [CrossRef]
48. Afantitis, A.; Melagraki, G.; Tsoumanis, A.; Valsami-Hones, E.; Lynch, I. A nanoinformatics decision support tool for the virtual screening of gold nanoparticle cellular association using protein corona fingerprints. *Nanotoxicology* **2018**, *12*, 1148–1165. [CrossRef] [PubMed]
49. Zhang, S.; Golbraikh, A.; Oloff, S.; Kohn, H. A novel automated lazy learning QSAR (ALL-QSAR) approach: Method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Modeling* **2006**, *46*, 1984–1995. [CrossRef]
50. Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P.A.; Markopoulos, J. Development and evaluation of a QSPR model for the prediction of diamagnetic susceptibility. *QSAR & Comb. Sci.* **2008**, *27*, 432–436.
51. Zhang, S.; Sun, N.; He, X.; Lu, X.; Zhang, X. Physical Properties of Ionic Liquids: Database and Evaluation. *J. Phys. Chem. Ref. Data* **2006**, *35*, 1475–1517. [CrossRef]
52. The UFT/Merck Ionic Liquids Biological Effects Database. Available online: <http://www.il-eco.uft.uni-bremen.de/> (accessed on 22 April 2014).
53. Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580. [CrossRef] [PubMed]
54. Frisch, M.; Schlegel, G.T.H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J., Jr.; Vreven, T.; Kudin, K.; Burant, J. *Gaussian 03, revision B. 03*; revision B. 03; Gaussian Inc.: Pittsburgh, PA, USA, 2003.
55. Chorowski, J.; Wang, J.; Zurada, J.M. Review and performance comparison of SVM- and ELM-based classifiers. *Neurocomputing* **2014**, *128*, 507–516. [CrossRef]
56. Ding, S.; Zhao, H.; Zhang, Y.; Xu, X.; Nie, R. Extreme learning machine: Algorithm, theory and applications. *Artif. Intell. Rev.* **2015**, *44*, 103–115. [CrossRef]
57. Huang, G.B.; Siew, C.K. Extreme learning machine: RBF network case. In Proceedings of the ICARCV 2004 8th Control, Automation, Robotics and Vision Conference, Kunming, China, 6–9 December 2004; pp. 1029–1036.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



© 2019. This work is licensed under  
<https://creativecommons.org/licenses/by/4.0/> (the “License”).  
Notwithstanding the ProQuest Terms and Conditions, you may use this  
content in accordance with the terms of the License.