Estimation of the Heat Capacity of Ionic Liquids: A Quantitative Structure−Property Relationship Approach

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · SEPTEMBER 2013

Impact Factor: 2.59 · DOI: 10.1021/ie401782n

5 AUTHORS, INCLUDING:

Mehdi Sattari
University of KwaZulu-Natal
40 PUBLICATIONS **322** CITATIONS
SEE PROFILE

Poorandokht Ilani-kashkouli
47 PUBLICATIONS **273** CITATIONS
SEE PROFILE

Amir H. Mohammadi
549 PUBLICATIONS **4,714** CITATIONS
SEE PROFILE

Deresh Ramjugernath
University of KwaZulu-Natal
332 PUBLICATIONS **1,965** CITATIONS
SEE PROFILE

I&EC research
Industrial & Engineering Chemistry Research

Article

pubs.acs.org/IECR

Estimation of the Heat Capacity of Ionic Liquids: A Quantitative Structure−Property Relationship Approach

Mehdi Sattari,[†] Farhad Gharagheizi,[†] Poorandokht Ilani-Kashkouli,[†] Amir H. Mohammadi,*[,†,‡] and Deresh Ramjugernath*[,†]

[†]Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa

[‡]Institut de Recherche en Génie Chimique et Pétrolier, Paris Cedex, France

Supporting Information

ABSTRACT: In this paper, a quantitative structure−property relationship model is developed using genetic function approximation (GFA) to predict the liquid heat capacity at constant pressure (C_{pL}) for ionic liquids at atmospheric pressure. The NIST Standard Reference Database was used to prepare a data set of C_{pL} data consisting of 3726 experimental data points comprised of 82 ionic liquids. The data set was split into two subsets, with 80% of the data used as a training set and 20% as a test set. Instead of using nonlinear modeling like artificial neural networks and a support vector machine, the GFA method was used to determine a model by a binary combination of descriptors rather than using single ones. Statistical analysis of the model shows that it has an overall AARD % of 1.70%, a coefficient of determination (R^2) of 0.993, and a root mean square of error of 15.11 J mol^{-1} K^{-1}.

1. INTRODUCTION

Ionic liquids (ILs) are molten salts that are generally liquid at temperatures below 100 °C.[1] They have interesting physical and chemical properties, such as negligible volatility, high thermal conductivity, chemical and physical stability, and nonflammability. As a result, they have received great interest from researchers, especially for applications in reaction chemistry and separation technology. Because ILs consist of ions and any variation in the combination of the ions affects the thermophysical properties drastically, they can in theory be designed for a particular application by selecting a proper combination of anion and cation pairs.[1,2] This task can, however, not be done by random selection of the cation and anion pairs without some knowledge about the influence of the cation and anion pairs on the target property. Therefore, having a predictive model that relates the ion chemical structures to the desired property is useful for estimation before attempting to synthesize the IL.

The heat capacity at constant pressure (C_p) is equal to the change in the enthalpy divided by the change in the temperature. C_p is related to several other thermodynamic properties, such as entropy, enthalpy, and Gibbs free energy. Knowledge of the heat capacity and its relationship to thermodynamic properties enables the study of other thermophysical properties. Heat capacity data are also used routinely for heat-transfer calculations for chemical engineering unit operations such as heat exchangers and reactors.

There are currently models available in the literature for prediction of the C_{pL} values of the ILs, with each having their own advantages and disadvantages. Some models[2−5] proposed in the literature use the molar volume (V_m) to predict the C_{pL} values of the ILs. One of the problems in the application of these models is the need for experimental values of density, in order to calculate V_m. Instead of using V_m, some models have been developed using group contribution methods (GCMs). Gardas and Coutinho[3] used a second-order group additivity method for the development of a 12-parameter model. Their database consisted of 19 ILs with 2396 data points over a wide temperature range of (196.36−663.10 K). They used only three cations and six anions for the model development, which limits the applicability of their model. The number of model parameters was also relatively high in contrast with the number of ILs used (12 parameters for 19 ILs). Similar approaches were applied by Soriano et al.[6] using cations and anions instead of functional groups for 32 ILs and 3149 data points with a temperature range of 188.06−663.10 K. Because they used certain types of ions, the model could not predict the heat capacity of the ILs comprised of cations or anions that were not in their data set.

Valderrama and Rojas[7] used a mass connectivity index, based on the molecular connectivity concept introduced by Randic.[8] They developed a 40-parameter GCM with two specific constants for each IL based on a data set consisting of 15 ILs with 541 data points. In another study, Valderrama et al.[9] used an approach for 31 ILs and 477 data points and developed an artificial neural network (ANN) model for the prediction of C_p. The disadvantage of their models is the use of a limited number of ILs, which reduced the applicability of the models for new ILs.

The aim of this study is the development of a quantitative structure−property relationship (QSPR) model to predict the C_{pL} values of the ILs without the requirement for any other properties as input parameters. QSPR is a methodology that

Received: June 5, 2013
Revised: July 24, 2013
Accepted: August 6, 2013
Published: August 6, 2013

relates the macroscopic properties to the microscopic structures of the molecules by introducing "descriptors".[10] The next step after calculation of the molecular descriptors is finding a model by statistical and regression analysis with a minimum number of descriptors and the greatest ability to fit the data, as well as predicting a property for a compound that is not in the data set used to develop the model.

2. MATERIALS AND METHODS

2.1. Materials. The NIST Standard Reference Database 103b[11] was used to create a data set consisting of the C_{pL} values of 82 ILs comprised of 39 different cations and 24 anions. Some of the ILs had multiple data sources; therefore, the data were screened, and the most recent source with the lower reported uncertainty was selected. As a result, 3726 experimental data points were extracted that covered the temperature range of 180.06−663 K and C_{pL} values ranging from 259.09 to 1805.7 J mol^{-1} K^{-1}. The structures of the cations and anions as well as the names of the ILs and their temperature and C_{pL} ranges are available as Supporting Information (SI).

The common approach in model development is to divide the data set into two subsets: a *training set* to develop and train the model and a *test set* to determine the prediction capability of the model for compounds that have not been used in model development. *K*-means clustering was used to choose the components for both the training and test sets. *K*-means clustering is a method of cluster analysis that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. Consequently, 80% of the ILs were chosen as the training set and the remainder used in the test set do develop and validate the model.

2.2. Calculation of the Molecular Descriptors. Before finding a relationship between the chemical structures of cation−anion pairs and the desired property, the Dreiding force field, as explained by Chemaxon's *JChem*,[12] was used to optimize the 3D chemical structures of the ions. Then, molecular descriptors were calculated and, thereafter, pair correlation was applied to remove the interrelated descriptors. Thereafter, pairs of descriptors with a correlation coefficient greater than 0.9 were removed, and the remaining descriptors were used to develop the model.

2.3. Developing the Model. In QSPR modeling, it is normally assumed that there is a linear relationship between the target property and one or more descriptors. Therefore, the output is a linear function that can be interpreted and understood easily. This assumption, however, is weak in the case of complex relationships between the behavior and chemical structures of some materials. With the aid of complicated nonlinear modeling tools like ANNs and a support vector machine (SVM), almost all of these problems are solved; however, choosing the most effective descriptors is the big challenge. To find a good predictive and accurate model for certain variables by ANN or SVM, millions of calculations and considerable amounts of computational time are required. Another problem is that the usage of ANN and SVM models requires specialized mathematical software.

In this study, genetic function approximation (GFA) was applied as a tool to select the most effective descriptors, as well as combined double descriptors by multiplication to create new variables. These descriptors were then used in the modeling. GFA is a feature selection approach introduced by Rogers and Hopfinger[13] that is a combination of Friedman's multivariate adaptive regression splines[14] and Holland's genetic algorithm.[15] More details on these methods can be found in previous studies.[16,17]

3. RESULTS AND DISCUSSION

Several models with different sizes were generated to find the most accurate model with an acceptable number of descriptors. Figure 1 shows changes in the accuracy of models in terms of the average absolute relative deviation (AARD %) with respect to the number of descriptors for models found by both the GFA and classical multivariate linear regression methods. It can be easily seen that the GFA models show better accuracy for the same number of parameters. Figure 1 indicates that improvement in the accuracy of linear models becomes insignificant after the 30th descriptor, but in the GFA models, it is constantly improved by the addition of new descriptors or a new combination of previously entered descriptors in the model.

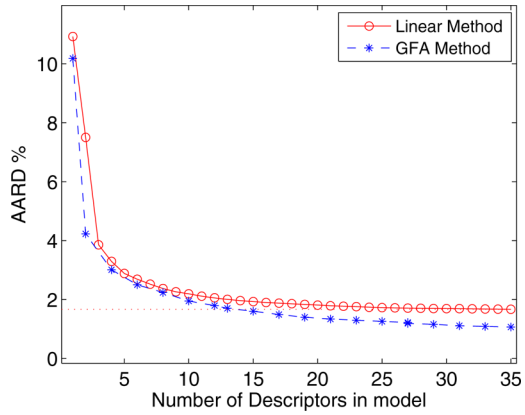

Figure 1. Effect of the number of descriptors on the accuracy of the linear QSPR and GFA models.

To find the most accurate but least complex model, variation in the accuracy of the model versus the number of functional groups was plotted in Figure 2, which shows that, after the 13th descriptor, the accuracy of the model did not change noticeably; however, larger models would have better accuracy. The final model is presented in eq 1, where *T* is the absolute temperature. The other parameters of the model are defined in Table 1.

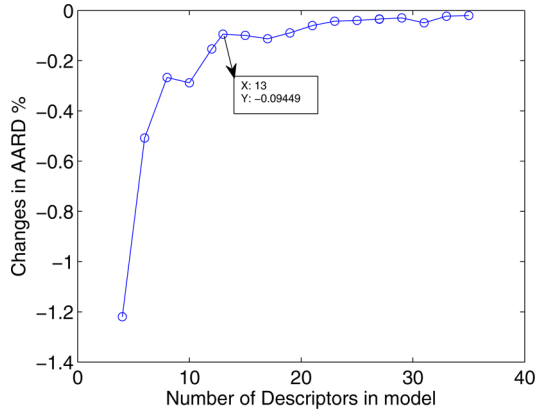

Figure 2. Changes in the accuracy of the model versus the increasing number of descriptors.

Table 1. Definition of the Descriptors Used in the C_{pL} Model

No.	Descriptors	Definitions
1	*nAtoms*	Number of atoms
2	*nSK*	Number of non-h atoms
3	*CH3R*	CH_3—R
4	*CRX3*	R—C(X)(X)X
5	*MATS2m*	Moran autocorrelation - lag 2 / weighted by atomic masses
6	*MATS4m*	Moran autocorrelation - lag 4 / weighted by atomic masses
7	*MATS4e*	Moran autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities
8	*RDF130u*	Radial Distribution Function - 13.0 / unweighted
9	*fHAccSA*	Fractional accessible surface area of hydrogen bond acceptors
10	*DECC*	Eccentric
11	*piPC09*	Molecular multiple path count of order 9
12	*SEigZ*	Eigenvalue sum from Z weighted distance matrix (Barysz matrix)

$$C_{pL} = A + BT \quad (1)$$

where A = 8.404nAtoms$_{cat}$ + 10.246nSK$_{an}$ and B = 0.272 + 5.074 (MATS2m$_{cat}$ × MATS4m$_{cat}$) + 6.594 (MATS4m$_{cat}$ × RDF130u$_{an}$) + 0.108 (CH3R$_{cat}$ × fHAccSA$_{an}$) + 0.448 (MATS4e$_{cat}$ × CRX3$_{an}$) + 2.790 × 10^{-2} (DECC$_{cat}$ × nAtoms$_{an}$) − 2.235 × 10^{-4} (piPC09$_{cat}$ × SEigZ$_{an}$).

The AARD % for eq 1 is 1.55% for the training set, which has 3001 experimental data points for 61 ILs, and 2.32% for the test set consisting of 725 experimental data points for 21 ILs. The coefficients of determination (R^2) for the training and test sets are 0.990 and 0.996, respectively. The statistical parameters for the model for the training and test sets are summarized in Table 4 in the SI. R represents any group linked through carbon. X represents any heteroatom (O, N, S, P, Se, and halogens).

For the training set, 62% of the predicted values of C_{pL} show deviations between 0 and 1%, 25% between 1.01 and 3%, 8% between 3.01 and 5%, 3% between 5.01 and 10%, and 2% over 10%. The calculated values of C_{pL} in the test set show that 41% of the data points are within an AARD range of 0−1%, 33% within the range of 1−3%, 15% within the range of 3−5%, approximately 8% within the range of 5−10%, and 3% greater than 10%. These results are shown in Figure 3.

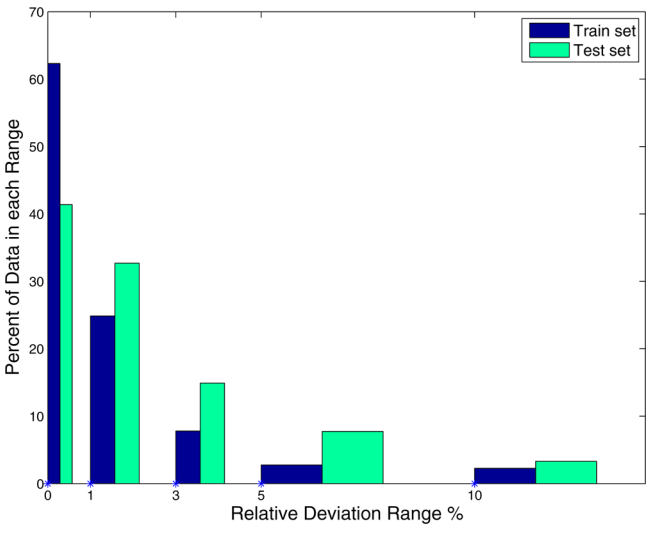

Figure 3. Percentage of the predicted values of C_{pL} in different relative deviation ranges.

Figure 4 shows the experimental data for C_{pL} versus the predicted values. The deviation of the model from the experimental data is shown in Figure 5. In these figures, some ILs show a large deviation, which means that the

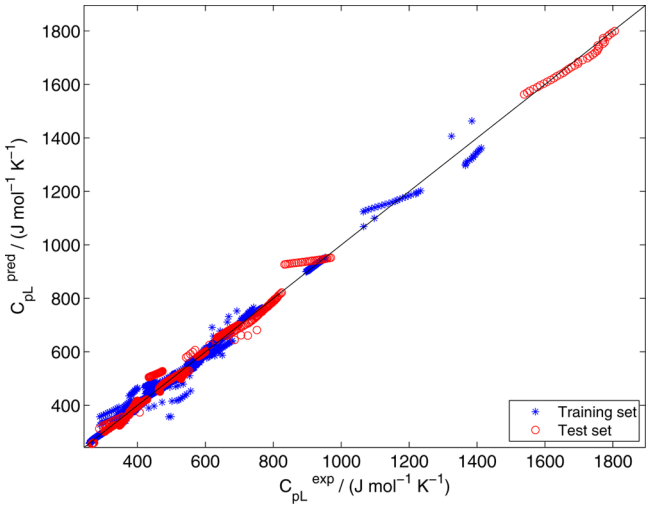

Figure 4. Predicted versus experimental values of C_{pL}.

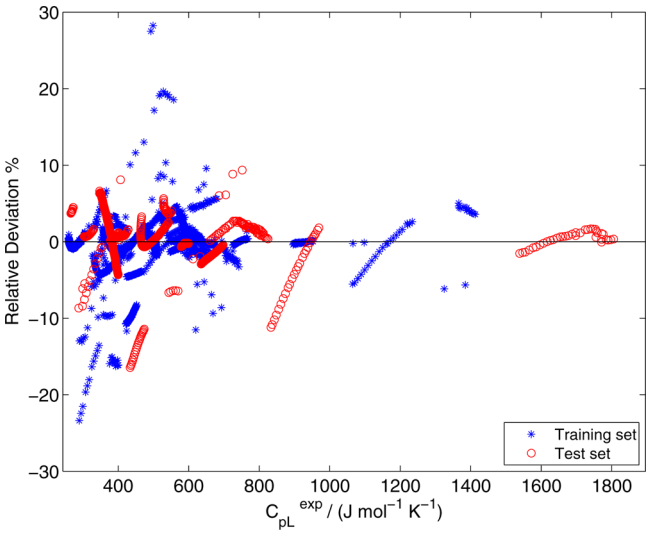

Figure 5. Relative deviation of predicted C_{pL} values from experimental data.

dx.doi.org/10.1021/ie401782n | *Ind. Eng. Chem. Res.* 2013, 52, 13217−13221

proposed model cannot predict these heat capacities well. As indicated in Table 3 in the SI, the largest deviation is observed for 1-ethyl-2,3-dimethylimidazolium bis[(trifluoromethyl)-sulfonyl]amide, with an AARD % of 27.9%. This IL has only two experimental points. Additionally, several other ILs are present in the data set, which have two data points and show high deviations. All of these data points were measured by Crosthwaite et al.,[18] who have produced data for other ILs that are not in good agreement with measurements of other researchers. Such deviations are shown in Figures 6 and 7 for

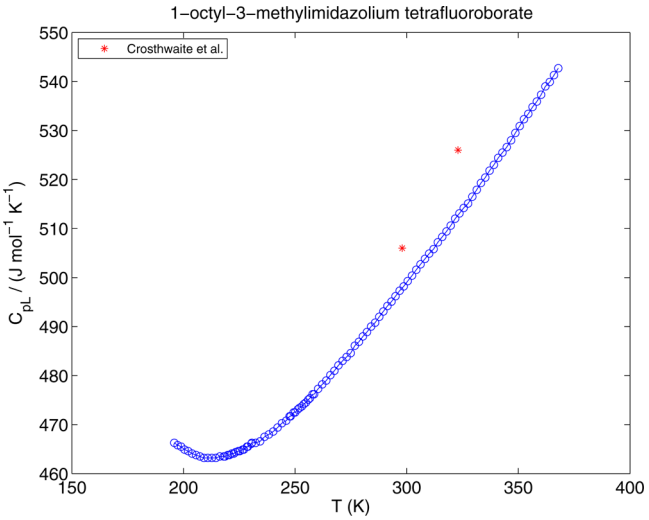

Figure 6. Comparison of the C_{pL} data for 1-octyl-3-methylimidazolium tetrafluoroborate measured by Crosthwaite et al. and other researchers.

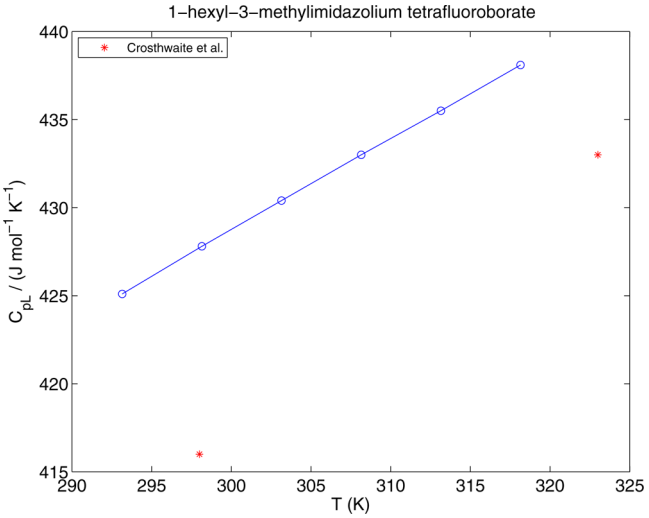

Figure 7. Comparison of the C_{pL} data for 1-hexyl-3-methylimidazolium tetrafluoroborate measured by Crosthwaite et al. and other researchers.

two ILs. It is therefore highly probable that the experimental data for these two-point ILs are not reliable and, consequently, the large deviations with respect to the proposed model.

The second IL with the largest deviation (19.1%) is 1-ethyl-3-methylimidazolium diethylphosphate. The largest model found during this study, which had 35 descriptors, was also used to predict C_{pL} of this IL, and it was observed that it can be predicted well with an AARD % of 0.8%. Consequently, more descriptors are required for better prediction of some of the ILs in the data set, but the aim of this study was to produce the simplest model while still keeping a fairly acceptable prediction capability for most of the ILs in the data set. Similar behavior was observed for most of the ILs with high deviations and more than two data points. The only exception was 1-ethyl-3-methylimidazolium hexafluorophosphate. It seems that the experimental data for this IL is problematic because a similar deviation (26.8%) is reported by Soriano et al.[6] with respect to their accurate model.

The values of the descriptors of the ILs used as well as the other information such as the original source of the data are available as SI.

4. CONCLUSIONS

In this study, a QSPR model was developed using GFA to predict the liquid heat capacity of ILs at atmospheric pressure. For this approach, a binary combination of descriptors was used instead of a single option. The result of this combination is a 13-parameter model with an AARD % of 1.55% for the training set and 2.32% for the test set. As a result, it was observed that binary multiplication of the descriptors was a successful approach in this study, which can predict C_{pL} values of ILs more efficiently than linear QSPR models, which use single descriptors as model parameters.

■ ASSOCIATED CONTENT

Supporting Information

Tables containing the name and structure of cations and anions, name, temperature and C_{pL} ranges, and AARD % of the studied ILs, statistical parameters of the model, and values of the model parameters for all ILs. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ramjuger@ukzn.ac.za (D.R.), a.h.m@irgcp.fr (A.H.M.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Michael Frenkel of the National Institute of Standards and Technology for providing the NIST Standard Reference Database. This work is based on research supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation.

■ NOMENCLATURE

C_p = heat capacity at constant pressure [J mol^{-1} K^{-1}]
C_{pL} = heat capacity of the liquids at constant pressure [J mol^{-1} K^{-1}]
V_m = molar volume [cm^3 mol^{-1}]

Abbreviations

AARD = average absolute relative deviation
ANN = artificial neural network
GCM = group contribution method
GFA = genetic function approximation
QSPR = quantitative structure−property relationship
SVM = support vector machine

REFERENCES

(1) Meindersma, G. W.; Maase, M.; De Haan, A. B. Ionic Liquids. In *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2000.

(2) Preiss, U. P. R. M.; Slattery, J. M.; Krossing, I. In Silico Prediction of Molecular Volumes, Heat Capacities, and Temperature-Dependent Densities of Ionic Liquids. *Ind. Eng. Chem. Res.* **2009**, *48* (4), 2290−2296.

(3) Gardas, R. L.; Coutinho, J. o. A. P. A Group Contribution Method for Heat Capacity Estimation of Ionic Liquids. *Ind. Eng. Chem. Res.* **2008**, *47* (15), 5751−5757.

(4) Paulechka, Y. U.; Kabo, A. G.; Blokhin, A. V.; Kabo, G. J.; Shevelyova, M. P. Heat Capacity of Ionic Liquids: Experimental Determination and Correlations with Molar Volume. *J. Chem. Eng. Data* **2010**, *55* (8), 2719−2724.

(5) Glasser, L.; Jenkins, H. D. B. Ambient Isobaric Heat Capacities, $C_{p,m}$, for Ionic Solids and Liquids: An Application of Volume-Based Thermodynamics (VBT). *Inorg. Chem.* **2011**, *50* (17), 8565−8569.

(6) Soriano, A. N.; Agapito, A. M.; Lagumbay, L. J. L. I.; Caparanga, A. R.; Li, M.-H. A simple approach to predict molar heat capacity of ionic liquids using group-additivity method. *J. Taiwan Inst. Chem. Eng.* **2010**, *41* (3), 307−314.

(7) Valderrama, J. O.; Rojas, R. E. Mass connectivity index, a new molecular parameter for the estimation of ionic liquid properties. *Fluid Phase Equilib.* **2010**, *297* (1), 107−112.

(8) Randic, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97* (23), 6609−6615.

(9) Valderrama, J. O.; Martinez, G.; Faúndez, C. Heat Capacity of Ionic Liquids Using Artificial Neural Networks and the Concept of Mass Connectivity. *Int. J. Thermophys.* **2011**, *32* (5), 942−956.

(10) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.

(11) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C. D.; Kazakov, A. F.; Magee, J. W.; Abdulagatov, I. M.; Kroenlein, K.; Diaz-Tovar, C. A.; Kang, J. W.; Gani, R. ThermoData Engine. *Pure Compounds, Binary Mixtures, and Chemical Reactions*, version 7.0; NIST Standard Reference Database 103b; National Institute of Standards and Technology: Gaithersburg, MD, 2011; http://www.nist.gov/srd/nist103b.cfm.

(12) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897−8909.

(13) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure−Activity Relationships and Quantitative Structure−Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (4), 854−866.

(14) Friedman, J. H. Multivariate Adaptive Regression Splines. *Ann. Stat.* **1991**, *19* (1), 1−67.

(15) Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, 1st ed.; University of Michigan Press: Lansing, MI, 1975; pp xiv, 211.

(16) Gharagheizi, F.; Sattari, M.; Ilani-Kashkouli, P.; Mohammadi, A. H.; Ramjugernath, D.; Richon, D. Quantitative structure—property relationship for thermal decomposition temperature of ionic liquids. *Chem. Eng. Sci.* **2012**, *84* (0), 557−563.

(17) Mirkhani, S. A.; Gharagheizi, F.; Sattari, M. A QSPR model for prediction of diffusion coefficient of non-electrolyte organic compounds in air at ambient condition. *Chemosphere* **2012**, *86* (9), 959−66.

(18) Crosthwaite, J. M.; Muldoon, M. J.; Dixon, J. K.; Anderson, J. L.; Brennecke, J. F. Phase transition and decomposition temperatures, heat capacities and viscosities of pyridinium ionic liquids. *J. Chem. Thermodyn.* **2005**, *37* (6), 559−568.