



ELSEVIER

Contents lists available at ScienceDirect

## Comptes Rendus Chimie

www.sciencedirect.com



Full paper/Mémoire

## Glass transition temperature of ionic liquids using molecular descriptors and artificial neural networks

José O. Valderrama <sup>a, d, \*</sup>, Richard A. Campusano <sup>b, d</sup>, Roberto E. Rojas <sup>c</sup><sup>a</sup> Faculty of Engineering, Department of Mechanical Engineering, University of La Serena, La Serena, Chile<sup>b</sup> Faculty of Sciences, Department of Physics and Astronomy, University of La Serena, La Serena, Chile<sup>c</sup> Faculty of Sciences, Department of Chemistry, University of La Serena, La Serena, Chile<sup>d</sup> Center for Technological Information (CIT), Monseñor Subercaseaux 667, La Serena, Chile

## ARTICLE INFO

## Article history:

Received 12 October 2016

Accepted 24 November 2016

Available online 20 February 2017

## Keywords:

Ionic liquids

Glass transition temperature

Neural networks

Molecular structure

Molecular descriptors

## ABSTRACT

Glass transition temperature data of ionic liquids (ILs) are analyzed to study the capabilities of artificial neural networks to correlate and predict this property. Molecular descriptors from computational chemistry are considered as independent variables to define the characteristics of an IL molecule. Several network architectures were considered, combinations of different descriptors were analyzed, and results were compared with other values reported in the literature. The independent variables (those that could have influence on the glass transition temperature) considered for training the artificial neural networks were (1) mass connectivity index  $\lambda$ , (2) cation mass  $M(+)$ , (3) anion mass  $M(-)$ , (4) surface area  $S_A$ , (5) van der Waals volume  $V_w$ , (6) connectivity index  $X_0$ , and (7) number of carbon atoms  $nC$ . The mass connectivity index is a parameter previously defined by the authors and is calculated for each IL, whereas the descriptors  $S_A$ ,  $V_w$ ,  $X_0$ , and  $nC$  were determined using the software Dragon7. As a measure of the accuracy of the method, the average relative deviation and the average relative absolute deviation are evaluated. Results of this work and others indicate that appropriate selection of data, good combination of architecture, and variables can lead to acceptable correlation of data but accurate prediction is not yet possible. The lack of a clear definition of the glass transition temperature and the lack of knowledge on what are the properties that most affect liquid–solid transition are the main causes of the present incapability for accurately predicting the glass transition temperature of the IL studied in this work.

© 2016 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Glass transition is the change that happens from solid state to amorphous solid and knowing the temperature at which this change occurs is of interest in various applications: (1) diffusion coefficient conductivity are related to  $T_g$  [1]; (2)  $T_g$  can be used to predict the dependence of viscosity on temperature [2]; (3)  $T_g$  also serves as a cohesive

energy parameter [3]; (4)  $T_g$  is one of the main criteria for the evaluation of the potential options for electrolyte applications [4]; and (5)  $T_g$  is important for phenomena in polymeric materials, amorphous pharmaceutical solids, and semiconductors [5].

Data of glass transition temperature for some ionic liquids (ILs) are available in the literature and databases such as ILs' database of the IUPAC [6], Beilstein database [7], Dortmund Data Bank [8], or the compilation by Zhang et al. [9] are available. An estimate from these sources indicates that there must be around 900 values of  $T_g$  for around 800 ILs. The total amount of data is greater than the amount of

\* Corresponding author. University of La Serena, Faculty of Engineering, Department of Mechanical Engineering, Box 554, La Serena, Chile.

E-mail address: [jvalderr@userena.cl](mailto:jvalderr@userena.cl) (J.O. Valderrama).

ILs because for some ILs several values, reported by different researchers, are available.

Glass transition does not occur at a specified fixed temperature, commonly denoted as  $T_g$ , but in a range of temperature or transition region [10]. However, the assignment of a single value for  $T_g$  is the common practice found in the literature for this property. In differential scanning calorimetry experiments (DSC), the usual way to determine this property is that  $T_g$  is commonly assigned to the onset point, intersection of the initial straight line and the transition region straight line, or to the midpoint of the transition region (inflection point). In this study, the highest value provided in the literature for  $T_g$  was considered as the true value. This is because the main uses of ILs are those in which the phase remains as liquid. Crystal formation or solidification is to be avoided in most applications of ILs. Therefore, by selecting or estimating a higher value for  $T_g$  the range of applicability of the IL is reduced but one can assure that crystal formation or solidification is to be avoided in most applications of ILs.

Schmelzer et al. [11] present a detailed discussion about the glass transition phenomenon and the glass transition temperature, but it is mainly dedicated to the vitreous state of glasses and not to ILs. However, the phenomenon is similar and the fundamentals are also analogous. According to the authors, “the notation glass transformation (or glass transition) temperature, proposed by Tammann, is to some extent misleading. Correct with respect to the indicated mechanism of vitrification is the proposal developed by Simon, to denote  $T_g$  as the freezing-in temperature of the glass”. In the physics of high polymers, the name temperature of vitrification is preferred and more precisely corresponds to the word *Glasteratur* used in the German literature. In the area of IL research the name *glass transition temperature* has been preferred.

Gómez et al. [12] present a thorough analysis of the thermal behavior of pure ILs and define some of the characteristic temperatures that appear during the transition between liquid and solid (or vice versa) of an IL: melting temperature ( $T_m$ ), freezing temperature ( $T_f$ ), cold crystallization temperature ( $T_{cc}$ ), solid–solid transition ( $T_{ss}$ ), and glass transition temperature ( $T_g$ ). They define the glass transition temperature  $T_g$  “as the midpoint of a small heat capacity change upon heating from the amorphous glass state to a liquid state”. However, not all authors use this definition and different values for the same IL are proposed in the literature.

The solid–liquid transition temperatures of ILs are usually less than ambient temperature and can go less than  $-100$  °C, such as in the cases of [C2mim][dca] or [C4mim][C2F5BF3] [13]. As mentioned above, the most common and efficient method for experimentally determining  $T_g$  is by DSC. The thermal behavior of ILs can be relatively complex and some peculiar and particular characteristics have been observed when cooling or heating an IL [12–14]: (1) the cooling from the liquid state may cause glass formation at low temperatures; (2) solidification kinetics is commonly slow; (3) on cooling from the liquid, the low-temperature region is not usually bounded by the phase diagram liquid line; (4) formation of metastable glasses may occur; (5) heating from the glassy state yields an exothermic transition

associated with sample crystallization, followed by subsequent melting; and (6) multiple solid–solid transitions (crystal–crystal polymorphism or plastic crystal phases) may occur. The authors also provided a couple of important recommendations: (1) thermodynamic data should be collected in heating mode to obtain reproducible results; and (2) to obtain reliable transition data, long equilibration times should be allowed and small samples should be taken, to permit rapid cooling. In addition, Gómez et al. [12] state that it is not always possible to correctly identify the different transitions appearing in a thermogram using DSC and additional techniques should be included (crossed polarizing filters, X-ray diffraction, and infrared spectrometry). The authors also state that it is necessary to subject the IL to different heating and cooling rates to have a better interpretation of the thermograms and to define characteristic temperatures such as melting or glass transition. It is not unusual to have ILs with different structures presenting the same thermal behavior and ILs with similar structures presenting different behavior.

These facts may explain the great differences found in reported experimental data of  $T_g$  for the same IL. As shown in Table 1 differences up to 38 °C (20%) are found, such as the case of 1-butyl-3-methylimidazolium trifluoroacetate. Other ILs present lower differences but still of importance for modeling and analysis. For butylammonium formate the difference is 25 °C (16%) and for 1-ethyl-3-methylimidazolium-bis[(trifluoromethyl)sulfonyl]imide the difference is 20 °C (11%).

## 2. Models for $T_g$ presented in the literature

Despite the differences between  $T_g$  data such as those shown in Table 1, some proposals have been presented in the literature for correlating and estimating the glass transition temperature. Mirkhani et al. [16] studied quantitative structure property relationship (QSPR) models for the glass transition temperature of different types of ILs. They claim that a simple predictive model is obtained. Although the absolute average deviation was low (3.8%), deviations more than 10% were found for 10 of the 139 fluids considered in the study. Better results were obtained when the authors considered a specific type of IL such as ammonium-based ILs. In that case, average absolute deviation was 2% and maximum deviations were less than 10%.

Gharagheizi et al. [17] presented a group contribution method to correlate and predict the glass transition temperature of ILs but only for 1,3-dialkylimidazolium-type ILs. For the 190 ILs considered in the study, the authors found an average absolute deviation of 1.9% with maximum deviations of 8.2%. Mousavisafavi et al. [4] also studied the same type of ILs and the same 109 data points using a linear QSPR method and obtained average absolute deviations on the order of 2.7% and maximum deviations of 8.8%. The same group of researchers [18] proposed a nonlinear approach of the QSPR method for obtaining a model that gives average absolute deviation of 1.4% and maximum deviation of 6.7% for the same data set.

Yan et al. [19] also used the QSPR methodology using topological indexes defined by the authors. The QSPR

**Table 1**  
Selected reported values of  $T_g$  showing differences between different literature sources.

N°	IL	Cation	Anion	$T_g$ (K)	$\Delta T$ (K)	
1	1-Butyl-3-methylimidazolium	Tetrafluoroborate	[C4mim]	[BF4]	176.15	0.00
					185.77	9.62
					188.15	12.00
					188.85	12.70
					190.15	14.00
					192.00	15.85
2	1-Butyl-3-methylimidazolium	Hexafluorophosphate	[C4mim]	[PF6]	192.95	16.80
					193.55	17.40
					193.15	0.00
					196.15	3.00
					197.15	4.00
					212.00	18.85
3	1-Butyl-3-methylimidazolium	Trifluoroacetate	[C4mim]	[ta]	195.15	0.00
					203.60	8.45
					233.15	38.00
4	1-Butyl-3-methylimidazolium	Bis[(trifluoromethyl)sulfonyl]imide	[C4mim]	[bti]	169.15	0.00
					184.15	15.00
					186.15	17.00
					187.15	18.00
					187.25	18.10
5	1-Ethyl-3-methylimidazolium	Dicyanamide	[C2mim]	[dca]	169.15	0.00
					183.15	14.00
6	1-Ethyl-3-methylimidazolium	Bis[(trifluoromethyl)sulfonyl]imide	[C2mim]	[bti]	175.15	0.00
					181.15	6.00
					186.00	10.85
					186.28	11.13
					195.15	20.00
7	Alanine methyl ester	Thiocyanate	[AlaC1]	[SCN]	215.00	0.00
8	<i>N</i> -Octyl-isoquinolinium	Bis[(perfluoroethane)sulfonyl]imide	[C8isoq]	[BEI]	241.15	26.15
					193.75	0.00
9	Butylammonium	Formate	[NHHH4]	[HCO2]	218.15	24.40
					153.05	0.00
10	Ethylammonium	Formate	[NHHH2]	[HCO2]	178.15	25.10
					145.65	0.00
					167.15	21.50

All data are from Zhang et al. [9,15]. The lowest reported value was used as a reference to determine the deviations for the other values.

model was applied to five types of ILs (imidazolium, pyridinium, ammonium, sulfonium, and triazolium). The average absolute deviation was 3.3% and maximum average deviation was 20%. Mokadem et al. [5] proposed an enhanced group-interaction contribution method for the prediction of  $T_g$  of ILs. A wide range of ILs (368 data points) were considered in the study obtaining average absolute deviation of 3.1% and maximum deviations of 29%. Eighteen of the 368 data points gave deviations greater than 10%.

### 2.1. Data available

As mentioned in Section 1, data of  $T_g$  for ILs have been presented in the literature and there are some databases, handbooks, and compilations of data for the glass transition temperature of ILs [6–9,15]. New data frequently appear in journals and monographs. However, several of the reported values of  $T_g$  for the same ILs may show great differences, as also explained above (see Table 1).

As explained in Section 1, of all data available for  $T_g$  in the open literature, the highest value of  $T_g$  within a set of data was considered for study and calculations. For instance for 1-ethyl-3-methylimidazolium-bis[(trifluoromethyl)sulfonyl]imide for which five data are available (number 6 in Table 1) the value selected as  $T_g$  is 195.15 K. Table 2 shows

the available amount of data of  $T_g$  for different types of ILs (second column, total data). The third column shows the amount of data actually used in this study. In these final data only one value of  $T_g$  for each IL is included. As observed in Table 2, the imidazolium-type ILs form the group with the greatest number of data. Not only that, but within this big group of imidazolium ILs, there are some specific groups with enough data as to perform more exhaustive analyses.

**Table 2**  
Amount of data of  $T_g$  available in the open literature.

Type of IL	Total data	Final data
Imidazolium	353	248
Ammonium	193	152
Triazolium	48	35
Pyrrolidinium	47	34
Piperidinium	14	11
Pyridinium	60	55
Isoquinolinium	11	9
Sulfonium	9	9
Guanidinium	25	17
Morpholinium	24	23
Oxazolidinium	12	11
Amino acids	25	23
Phosphonium	31	29

Total includes values from different literature sources [6–9,15].

A complete study on the application of artificial neural network (ANNs) for training, testing, and predicting  $T_g$  for ILs requires that following especial aspects are analyzed: (1) the most appropriate architecture must be defined; (2) the most adequate variables that have effect of  $T_g$  must be analyzed; (3) the predictive capabilities of the ANN model must be studied; (4) the effect of cation and anion must be evaluated; (5) different types of ILs must be included in the study; and (6) comparison with literature calculations of  $T_g$  must be done.

## 2.2. Molecular descriptors

Todeschini and Consonni [20] define a molecular descriptor as “the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment”. Of the several possible molecular descriptors that could have effects on defining a value of the glass transition temperature the following were chosen for the study: (1) mass connectivity index  $\lambda$  and connectivity index of order zero  $X_0$  (to account for the type of connections); (2) cation mass  $M(+)$  and anion mass  $M(-)$  to account for the mass of each part of the molecule; (3) surface area  $S_A$  and van der Waals volume  $V_w$  to account for the surface and the volume of the IL; (4) and number of carbon atoms  $nC$  to account for the alternation effect that chain families of IL properties may present. The mass connectivity index  $\lambda$  is a parameter defined by the authors and is calculated for each IL [21]. The descriptors  $S_A$ ,  $V_w$ ,  $X_0$ , and  $nC$  were determined using the software Dragon7 [22].

## 3. Artificial neural networks

An ANN is a mathematical tool that relates the values of a certain dependent function  $F$  (for instance the glass transition temperature) to values of defined independent variables  $x_1, x_2, x_3, \dots, x_n$  (for instance structural parameters of the substance). To find a relation between the function  $F$  and the variables  $x_i$ , the ANN must be trained; this means that the network is provided with values of  $F$  for several values of  $x_i$  to find a relation, a pattern, a dependency of the function  $F$  on the variables  $x_i$ . The form in which the ANN finds the relation is inspired in the behavior of biological neurons [23].

Imaginary units that simulate neurons are organized in layers with a defined number of neurons per layer, forming what is called the “architecture” of the network. An input layer receives the information  $F$  versus  $x_1, x_2, x_3, \dots, x_n$  and makes an initial processing of such data. In fact the network assigns to each variable  $x_i$  a weight and a base value (bias) specific for each neuron. With these weight and bias values the network calculates a value for the exit function  $F$  and compares the result with the known value of  $F$ . In calculating  $F$ , this function is related to the independent variables  $x_i$  through specific mathematical functions named activation functions.

If the deviation between the value of the calculated property  $F$  and the experimental value is greater than a defined value, the weights and biases are reassigned and

the calculations are done again. This method is known as back propagation, and the new values of the weights and biases are determined using an optimization method.

The network stores the values of weights and biases that give the lowest deviation between the calculated values and the experimental data provided as input. These values of weights and biases define the ANN model. Thus the ANN is not an explicit analytical model such as the empirical correlations commonly used in many applications but the model is a structure of weights and biases provided as matrices. Taskinen and Yliruusi [24] presented a complete review on applications of ANNs for the estimation of fluid properties. Properties of ILs were not included in that review.

The architecture of an ANN usually used for correlating and predicting properties of ILs considers a back propagation network with three or four layers: an input layer, and output layer, and one or two hidden layers [25–27]. The optimum number of neurons in each layer is usually found by trial and error because it is not possible to know in advance the appropriate number of neurons for a given application [28]. Also the disadvantages of ANN when they are used for fluid property correlation and prediction have been discussed [28–30] being the following the most important ones: (1) they require a large set of data, depending on the complexity of the relations between dependent and independent variables; (2) it is necessary to know which are the most influential variables  $x_1, x_2, x_3, \dots, x_n$  on the property  $F(x_1, x_2, x_3, \dots, x_n)$  and; (3) the network may suffer from overfitting, a situation in which the network memorizes instead of learning, losing all predictive capacity.

### 3.1. Application of ANN to $T_g$ calculation

To develop an accurate model to correlate and predict the glass transition temperature in the form presented in this work, the following two files were written:

- (1) An Excel file (*Tg\_data.xlsx*) containing all data available for training and testing the network. The file contains several sheets including the data of glass transition temperature and the chosen independent variables, both for training and testing. For the explanation that follows, the independent variables are the following molecular descriptors:  $\lambda$ ,  $M(+)$ ,  $M(-)$ ,  $S_A$ ,  $V_w$ ,  $X_0$ , and  $nC$ .
- (2) A Matlab code for the ANN model (*Tg\_ann.m*) consisting of two parts, a training section and a testing section, as presented in Table 3. In the training section, the program reads the input data (from *Tg\_data.xlsx*), defines the architecture, trains the defined network, generates the weight and bias matrixes, and stores such data for testing. Important to notice are lines 10 and 12 in Table 3 in which the dependent and independent data are read from the Excel file *Tg\_data.xlsx* (sheets *Read\_Tg\_Training* and *Read\_Variables\_Training*, respectively). Line 35 in Table 3 defines the storing of the correlated data and line 37 describes the storing of the weight and bias matrix (file *w\_Tg.mat*).

**Table 3**

Matlab code *Tg\_ann.m* used for training and testing an ANN with glass transition temperature data.

```

1 % Tg_ann.m
2 %*****
3 %
4 % This is the Matlab code for training an ANN with glass
5 % transition temperature data, using as independent
6 % Variables ( $\lambda$ , M+, M-,  $V_w$ )
7 %
8 % Training section
9 %*****
10 % Reading independent variables for training ( $\lambda$ , M+, M-,  $V_w$ )
11 % from Excel file Tg_data.xlsx
12 p = xlsread('Tg_data.xlsx','Read_Variables_Training');p = p';
13 %Reading the dependent variable for training ( $T_g$  listed in
14 % the file variables_training)
15 t = xlsread('Tg_data.xlsx','Read_Tg_Training');t = t';
16 %
17 % Normalization of all data (values between  $-1y + 1$ )
18 [pn,minp,maxp,tn,mint,maxt] = premnmx(p,t);
19 % Definition of ANN:(topology, activation functions,
20 % training algorithm)
21 net = newff(minmax(pn),
22 [4,4,4,1],{'tansig','tansig','tansig','purelin'},'trainlm');
23 % Definition of frequency of visualization of errors during
24 % training
25 net.trainParam.show = 10;
26 % Definition of number of maximum iterations (epochs) and
27 % global error between iterations (goal)
28 net.trainParam.epochs = 1500; net.trainParam.goal = 1e-6;
29 %Network starts: reference random weights and gains
30 w1 = net.LW{1,1}; w2 = net.LW{2,1}; w3 = net.LW{3,2};
31 w4 = net.LW{4,3};
32 b1 = net.b{1}; b2 = net.b{2}; b3 = net.b{3}; b4 = net.b{4};
33 %First iteration with reference values and correlation coefficient
34 before_training = sim(net,pn);
35 corrbefore_training = corrcoeff(before_training,tn);
36 %Training process and results
37 [net,tr]=train(net,pn,tn);
38 after_training = sim(net,pn);
39 % Back-normalization of results, from values between  $-1y + 1$ 
40 % to real values
41 after_training = postmnmx(after_training,mint,maxt);
42 after_training = after_training';
43 Res = sim(net,pn);
44 % Saving results, correlated glass transition temperature in
45 % an Excel file
46 xlswrite('Tg_data.xlsx',after_training,'Results_Correlation', 'D2');
47 %Saving the network (weights and other files)
48 save w_Tg.mat
49 %
50 %Testing Section
51 %*****
52 % This is the Matlab code for testing the ANN determined
53 % above for data not used during training
54 %
55 %Reading weight and other characteristics of the trained ANN
56 % saved in the file W
57 load w_Tg.mat
58 % Reading new data not used during training from Excel file
59 % Tg_data.xlsx
60 pnew = xlsread('Tg_data.xlsx','Read_Variables_Testing');
61 pnew = pnew';
62 % Normalization of all variable (values between  $-1y + 1$ )
63 pnewn = trmnmx(pnew,minp,maxp);
64 anewn = sim(net,pnewn);
65 % Transformation of the normalized exits (between  $-1y + 1$ )
66 % determined by the ANN to real values
67 anew = postmnmx(anewn,mint,maxt); anew = anew';
68 % Saving the predicted glass transition temperature in the
69 % Excel file Tg_data (sheet Results_Testing)
70 xlswrite('Tg_data.xlsx',anew,'Results_Testing', 'D2');
```

In the testing section, the program reads the weight and bias matrixes (line 44) and the Excel sheet *Read\_Variables\_Testing* within the file *Tg\_data.xlsx*, containing the variables for which the glass transition temperature needs to be tested (line 46). The results are stored in the same Excel file in the sheet *Results\_Testing* (line 53). The optimum network is chosen considering the results in both the training and testing sections. It should also be mentioned that the version of the software used in the calculations is Matlab 2012.

To apply the trained network for estimating the glass transition temperature the program *Tg\_prediction.m* including the Excel file *Tg\_prediction.xlsx* and the weight matrix *w\_Tg.mat* must be used. When the Matlab code is run, the variables for those cases for which the glass transition temperature is to be predicted are read from the sheet *Read\_Variab\_Prediction* in the file *Tg\_prediction.xlsx*. The program reads the weight matrix (line 6 in Table 4) and calculates the glass transition temperature for the new variables using the ANN model defined in the matrix *w\_Tg.mat*. The calculated glass transition temperature is stored in the file *Tg\_prediction.xlsx* (sheet *Results\_Prediction*). The three files needed for predicting the glass transition temperature (the Matlab code *Tg\_prediction.m*, the Excel file *Tg\_prediction.xlsx*, and the weight matrix *w\_Tg.mat*) are provided as [Supporting information](#).

#### 4. Results and discussion

As mentioned in the preceding sections, the selection of the best model considered three statistical parameters: the average relative deviation  $\% \Delta T_g$ , the average relative absolute deviation  $\% |\Delta T_g|$ , and the maximum relative absolute deviation  $\% |\Delta T_g|_{\max}$  between calculated and literature data. These parameters are the most significant numbers for evaluating the accuracy and goodness of a model as discussed [31]. These deviations are defined as follows:

$$\% \Delta T_g = \frac{100}{N} \sum_1^N \left[ \frac{T_g^{\text{cal}} - T_g^{\text{lit}}}{T_g^{\text{lit}}} \right]_i \quad (1)$$

$$\% |\Delta T_g| = \frac{100}{N} \sum_1^N \left| \frac{T_g^{\text{cal}} - T_g^{\text{lit}}}{T_g^{\text{lit}}} \right|_i \quad (2)$$

$$\% |\Delta T_g|_{\max} = \text{the maximum value of } 100 \cdot \left| \frac{T_g^{\text{cal}} - T_g^{\text{lit}}}{T_g^{\text{lit}}} \right|_i \quad (3)$$

Three study cases were designed to analyze different aspects of the model development: case 1 in which different architectures are considered to choose the most appropriate one; case 2 in which different descriptors are considered to search for the most appropriate variables; and case 3 in which the predictive capabilities of the ANN model are analyzed by selecting three different data sets: one for training, one for testing, and one for prediction.

Following information from the study by Valderrama et al. [32] around 10% of the data was separated for testing. Thus, of the 248 data, a group of 223 data was selected for training and the remaining 25 were left for testing. In the three study cases, the selection of data for testing was

**Table 4**

Matlab code Tg\_prediction.m used for predicting the glass transition temperature for other cases using the trained ANN (*w\_Tg.mat*).

1	% Tg_prediction.m
2	%
3	%This is the Matlab code for predicting Tg using the trained ANN model ( <i>w_Tg.mat</i> )
4	%
5	%Reading weights and other characteristics of the trained ANN saved in the file <i>w_Tg.mat</i>
6	load <i>w_Tg.mat</i>
7	% Reading Excel file with new independent variables ( $\lambda$ , $M(+)$ , $M(-)$ and $V_v$ ) to predict Tg
8	pnew = xlsread('Tg_Prediction.xlsx','Read_Variab_Prediction'); pnew = pnew';
9	% Normalization of all variables (values between $-1y + 1$ )
10	pnewn = trammx(pnew,minp,maxp);
11	% Obtaining the properties for the variables provided by the Excel file Tg_Prediction.xlsx
12	anewn = sim(net,pnewn);
13	% Transformation of the normalized exits (between $-1y + 1$ ) determined by the ANN to real values
14	anew = postmnmx(anewn,mint,maxt); anew = anew';
15	% Saving the predicted Tg in the Excel file Tg_prediction.xlsx (sheet Results_Prediction')
16	xlswrite('Tg_Prediction.xlsx',anew,'Results_Prediction', 'd3');

randomly done. Random numbers were generated in Excel and the 25 ILS showing the lowest random numbers were chosen for testing.

#### 4.1. Case 1: searching for the appropriate architecture

The whole set of data separated in a training set and in a testing set as explained above was considered for this case (223 for training and 25 for testing). Also, all descriptors chosen as explained in the preceding sections were included in the study ( $\lambda$ ,  $M(+)$ ,  $M(-)$ ,  $S_A$ ,  $V_w$ ,  $X_0$ , and  $nC$ ). In addition, for this study several three- and four-layer architectures were considered and analyzed. To simplify the analysis, the three-layer cases considered followed the structure  $(n,2n,1)$  and the four-layer architectures followed the structure  $(n,n,n,1)$ ,  $n$  being an integer number. The results for selected cases are presented in Table 5. It should be noticed that these two architectures have the same number of total neurons for a given value of  $n$ .

A graphical representation of the result is important to see how the effects of the number of neurons in the hidden layer affect the accuracy of the model. Several authors have found that there are an optimum number of neurons that give the lowest deviation between calculated and experimental data in a given run [25,30,32]. To choose the optimum number of neurons one must decide which of the statistical parameters must be optimum. Two of these parameters could be the lowest average absolute deviation and the lowest maximum deviation. Any of this can be set during training and testing. Because low deviations during testing represent in some way the interpolating or even extrapolating capabilities of the network, we recommend using the result during testing to decide on the best architecture. Also, choosing the lowest value of the maximum absolute deviation as the target parameter would guarantee estimations below that level of error.

Fig. 1 shows the results expressed as the maximum relative absolute deviation for the three- and four-layer

**Table 5**

Results for three- and four-layer networks using all descriptors ( $\lambda$ ,  $M(+)$ ,  $M(-)$ ,  $S_A$ ,  $V_w$ ,  $X_0$ , and  $nC$ ) for 248 data, 223 for training and 25 for testing.

Architecture	Training			Testing		
	% $\Delta T_g$	% $ \Delta T_g $	% $ \Delta T_g _{\max}$	% $\Delta T_g$	% $ \Delta T_g $	% $ \Delta T_g _{\max}$
<i>Case 1 (a): all descriptors (three layers)</i>						
1.2.1	1.3	9.1	33.6	2.4	10.9	24.4
2.4.1	0.6	6.1	24.7	-0.5	8.1	18.8
3.6.1	0.4	4.6	21.8	1.0	7.3	18.9
4.8.1	0.2	3.3	21.9	-1.5	7.0	18.7
5.10.1	0.1	2.2	16.3	-2.5	9.0	26.1
6.12.1	0.1	1.9	15.4	-0.5	7.1	19.9
7.14.1	0.0	0.5	13.3	1.3	17.2	49.1
<i>Case 1 (b): all descriptors (four layers)</i>						
1.1.1.1	1.3	9.0	33.7	2.4	10.8	24.1
2.2.2.1	0.8	6.6	36.4	0.6	8.6	19.8
3.3.3.1	0.4	4.9	21.1	-0.9	9.3	20.7
4.4.4.1	0.3	3.8	26.2	-0.5	6.5	16.9
5.5.5.1	0.2	2.8	12.2	0.6	8.5	23.0
6.6.6.1	0.1	1.6	11.9	1.3	11.7	31.8
7.7.7.1	0.1	1.4	13.4	-1.5	10.1	36.9

architectures. In the three-layer case, an architecture (4,8,1) (that means  $n = 4$ ) gives the lowest deviations in the average relative absolute deviation and in the maximum relative absolute deviation. For the four-layer cases, the best architecture is found to be (4,4,4,1) (that means  $n = 4$ , and the same neurons as in the three-layer case).

#### 4.2. Case 2: searching for the appropriate descriptors

Different descriptors are considered to search for the most appropriate combination of variables. Three

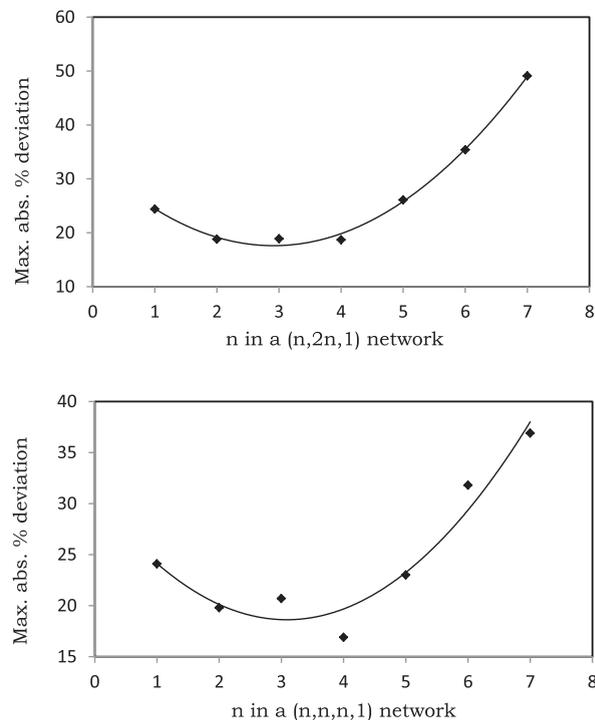


Fig. 1. Maximum relative absolute deviation during testing for three- and four-layer architecture of the type  $(n,2n,1)$  and  $(n,n,n,1)$ .

descriptors are fixed: the mass connectivity index, the cation mass, and the anion mass ( $\lambda$ ,  $M(+)$ , and  $M(-)$ ). In other applications found in the literature these three parameters ( $\lambda$ ,  $M(+)$ , and  $M(-)$ ) have been used in neural network applications and in group contribution methods [25–27]. Others descriptors are added in different combinations to analyze the effect of them in training and testing the ANN. Table 6 presents some selected results for a series of combinations of variables using the four-layer architecture (4,4,4,1).

As observed in Table 6 the inclusion of the carbon number as a parameter (last two lines) does not improve training and testing results. In the data used, just some few compounds have odd number of carbons and the alternation effect occurs for compound with odd and even numbers of carbons in the alkyl chain [33]. The addition of the connectivity index  $X_0$  (line 7) improves the results but not in an appreciable form, considering that six descriptors are used. The inclusion of the whole set of parameters (last line in Table 6) indicates that it is not necessary to include all parameters. So to balance simplicity with accuracy four descriptors are enough to get acceptable results (the three first cases in Table 6).

#### 4.3. Case 3: analyzing the predictive capabilities of the ANN model

The set of data originally available for applying neural networks was divided into three groups: one for training, one for testing, and one for prediction. The testing and prediction sets are not directly used during training, whereas the network learns. However, results obtained during testing for each run of the ANN program are used for deciding which model is the best. This means that the chosen model is that giving low deviations during training and testing at the same time. The predicting set, however, was never used for constructing or for selecting the optimum ANN model.

For all of the imidazolium-type ILs analyzed in this work, the amount of data for each set is as follows: the total number of data is 248, the data for training are 223, the data for testing are 25, and the data for prediction are 18. The results for testing presented in Table 5 and for prediction in Table 7 represent in some way the predictive capabilities of the chosen network, because in both cases

**Table 7**

Predicted  $T_g$  for 18 imidazolium-type ILs using the four-layer model with the following descriptors  $\lambda$ ,  $M(+)$ ,  $M(-)$ , and  $V_w$ .

Cation	Anion	$T_g^{\text{lit}}$	$T_g^{\text{al}}$	$\% \Delta T_g$	$\%  \Delta T_g _{\text{max}}$
[C1im]	[Br]	213.15	189.0	-11.3	11.3
[C2im]	[Cl]	216.15	198.9	-8.0	8.0
[C2mim]	[Br]	218.00	201.2	-7.7	7.7
[C2mim]	[Cl]	234.00	198.4	-15.2	15.2
[C3CNmim]	[Cl]	239.05	225.7	-5.6	5.6
[C4mim]	[Br]	223.15	204.1	-8.5	8.5
[C4mim]	[Br]	190.15	220.2	15.8	15.8
[moemim]	[Cl]	213.15	220.5	3.5	3.5
[C4mim]	[Cl]	204.15	212.7	4.2	4.2
[C3CNmim]	[Cl]	254.25	231.0	-9.1	9.1
[C6mim]	[Br]	224.00	217.3	-3.0	3.0
[C6mim]	[Cl]	198.15	227.0	14.6	14.6
[C3mim]	[BF4]	259.25	189.8	-26.8	26.8
[C8mim]	[Cl]	210.85	236.8	12.3	12.3
[neo-C5mim]	[BF4]	221.00	202.3	-8.4	8.4
[CitronellylC1im]	[Br]	216.15	231.3	7.0	7.0
[C4mim]	[BF4]	193.55	183.7	-5.1	5.1
[CitronellylC12im]	[Br]	203.15	203.0	-0.1	0.1
			Average	-2.9	9.2
			Max	-	26.8

the values of  $T_g$  for the testing set and the prediction test were not used during training, but, as mentioned above, the results of the testing step were used in deciding which was the best network model. In the case shown in Table 6, the 18 values of  $T_g$  were not used in any previous step and therefore represent the true capability of the network to predict  $T_g$  for an IL of the type studied in this work.

#### 4.4. Case 4: analyzing subfamilies of ILs

The results shown in the preceding sections were found for all imidazolium-type ILs for which data of  $T_g$  were available. Within this big family of imidazolium-type ILs, there are subfamilies such as alkyl-methylimidazolium, alkyl-dimethylimidazolium, hydroxyl-alkyl-methylimidazolium, and benzyl-imidazolium, among others. The largest of these families are the *n*-alkyl-3-methylimidazolium for which 105 data are available. Of these data, 95 were left for training, five for testing, and five for prediction. According to previous studies by the authors who explored the concept of homology for the study and calculation of transition properties of IL families [33] better results should be expected. In fact this

**Table 6**

Results for architecture (4,4,4,1) using different types and amount of descriptors.

Case	Case 2: Architecture: 4,4,4,1, different descriptors						
	Descriptors used	Training			Testing		
		$\% \Delta T_g$	$\%  \Delta T_g $	$\%  \Delta T_g _{\text{max}}$	$\% \Delta T_g$	$\%  \Delta T_g $	$\%  \Delta T_g _{\text{max}}$
1	$\lambda$ , $M(+)$ , $M(-)$ , $V_w$	0.5	4.8	25.1	-0.9	7.9	21.5
2	$\lambda$ , $M(+)$ , $M(-)$ , $S_A$	0.4	4.8	19.1	-3.3	8.9	21.2
3	$\lambda$ , $M(+)$ , $M(-)$ , $X_0$	0.5	5.3	28.3	-4.8	8.8	27.1
4	$\lambda$ , $M(+)$ , $M(-)$ , $S_A$ , $V_w$	0.3	4.1	19.6	2.6	9.4	24.6
5	$\lambda$ , $M(+)$ , $M(-)$ , $V_w$ , $X_0$	0.4	4.2	27.2	0.2	10.5	22.6
6	$\lambda$ , $M(+)$ , $M(-)$ , $S_A$ , $X_0$	0.5	5.7	19.5	1.2	9.0	24.2
7	$\lambda$ , $M(+)$ , $M(-)$ , $S_A$ , $V_w$ , $X_0$	0.3	3.7	18.9	-1.4	9.5	20.5
8	$\lambda$ , $M(+)$ , $M(-)$ , $S_A$ , $V_w$ , $nC$	0.3	4.0	17.8	-0.8	6.5	21.1
9	$\lambda$ , $M(+)$ , $M(-)$ , $S_A$ , $V_w$ , $X_0$ , $nC$	0.3	3.8	26.2	-0.5	6.5	16.9

was the case for the family of *n*-alkyl-3-methylimidazolium. Average deviations are less than 3% with maximum deviations less than 10%.

Table 8 presents some results for the 105 *n*-alkyl-3-methylimidazolium-type ILs. It can be observed that, as expected, simpler architectures are needed to obtain similar or better accuracy than that in the preceding cases. Networks with more layers and neurons are not recommended for the amount of data available for training.

Similarly, the predictive capabilities of this reduced family of *n*-alkyl-3-methylimidazolium ILs were also tested. The set of data used for prediction are the nine *n*-alkyl-3-methylimidazolium ILs shown in Table 7. In this case, the results for testing presented in Table 8 and for prediction in Table 9 also present in some way the predictive capabilities of the chosen network for this particular subfamily of *n*-alkyl-3-methylimidazolium ILs.

#### 4.5. Extension to other types of ILs

The detailed analysis presented above for imidazolium-type ILs should be extended to other types of ILs to be able to draw more complete conclusions about the application of neural networks for correlating (and probably predicting)  $T_g$  for any type of IL. Of the other types of ILs (besides imidazoliums), the type of ILs with a reasonable amount of data is the ammonium-type group (152 data points). For this group, the chosen network (4,4,4,1) can be applied and a specific model for ammonium can be found. However, for other groups this type of architecture cannot be applied because the number of data is not enough to evaluate the weights and biases. One should notice that for the network (4,4,4,1) the number of parameters to be determined is 49 (36 weights and 13 biases). We established that the number of data used for training must be at least the double of the number of parameters.

The chosen architecture (4,4,4,1) was applied for ammonium-type ILs and results are as expected. This means that they are similar (and a little higher) than in the case in which all imidazolium-type ILs were studied (see Table 7). In the case of imidazolium, the maximum deviation in prediction was 26.8% (70 °C) for [C3mim][BF4], whereas for ammonium the maximum deviation in prediction was 29%. However, by observing the training and testing results, there are two acceptable models for these ammonium ILs (indicated as model 1 and model 2 in table 10, third column). With the first model (model 1) surprisingly high deviations are found (>1000%) showing again the low predictive capabilities and the erratic behavior of

**Table 8**

Results for several architectures for ILs of the type *n*-alkyl-3-methylimidazolium using three- and four-layer network with the following descriptors  $\lambda$ ,  $M(+)$ ,  $M(-)$ , and  $V_w$ .

Case	Network	Training			Testing			Prediction		
		% $\Delta T_g$	% $ \Delta T_g $	% $ \Delta T_{g max}$	% $\Delta T_g$	% $ \Delta T_g $	% $ \Delta T_{g max}$	% $\Delta T_g$	% $ \Delta T_g $	% $ \Delta T_{g max}$
1	(2.4.1) (7 Neurons)	0.3	3.9	19.4	0.2	1.0	1.8	6.4	6.4	19.4
2	(2.2.2.1) (7 Neurons)	0.3	4.0	16.6	1.8	4.2	6.2	6.1	6.1	16.6
3	(3.6.1) (10 Neurons)	0.1	2.6	15.0	0.2	1.5	3.2	2.5	3.1	5.0
4	(3.3.3.1) (10 Neurons)	0.2	2.6	18.4	1.3	2.6	4.9	4.7	5.6	18.4

**Table 9**

Predicted  $T_g$  using a four-layer architecture (4,4,4,1) with the following descriptors  $\lambda$ ,  $M(+)$ ,  $M(-)$ , and  $V_w$ .

Cation	Anion	$T_g^{lit}$	$T_g^{cal}$	% $\Delta T_g$	% $ \Delta T_g $
[C3mim]	[BF4]	259.3	195.4	-24.6	24.6
[C4mim]	[BF4]	193.6	176.5	-8.8	8.8
[C2mim]	[Br]	218	180.7	-17.1	17.1
[C4mim]	[Br]	223.2	191.6	-14.1	14.1
[C6mim]	[Br]	224	182.4	-18.6	18.6
[C2mim]	[Cl]	234	195.5	-16.4	16.4
[C4mim]	[Cl]	204.2	179.9	-11.9	11.9
[C6mim]	[Cl]	198.2	195.3	-1.4	1.4
[C8mim]	[Cl]	210.9	184.4	-12.6	12.6

the model. The high number of parameters to be determined (69) compared to the number of data for training (122) could be one of the reasons for this behavior. For other types of ILs results are variable, being difficult to draw a general positive recommendation on the use of neural networks.

Table 10 presents selected results for those cases for which neural networks with simple architecture could be applied. For each type of IL results for two runs are presented. The groups with less than 30 data points (piperidinium, isoquinolinium, sulfonium, guanidinium, morpholinium, oxazolidinium, and amino acids) were not analyzed with the ANN model. This restriction is imposed because the number of parameters (weights and biases) to be calculated could be similar or greater than the number of data, a mathematical inconsistency that provides unmeaningful results. As expected, when few data are used the network can still learn and reasonable training can be obtained. However, the predictive capabilities of the network are uncertain; one could get a reasonable prediction with 12.5% deviation as the case of triazolium in Table 10 or 1137% deviation as the case of ammonium in the same Table 10.

#### 4.6. Comparison with literature estimations of $T_g$

As described in Section 1, some other approaches, including QSPR models and group contribution methods, have been presented in the literature. Mousavisafavi et al. [4,18] correlated and predicted glass transition temperature of 1,3-dialkylimidazolium ILs, which could be considered as potential future electrolytes. For this purpose, QSPR method is used to finally produce satisfactory results quantified by following several statistical parameters. Yan et al. [19] also applied a QSPR method based on the general

**Table 10**

Results for  $T_g$  during training, testing, and prediction for various types of ILs using three- and four-layer networks with the following descriptors  $\lambda$ ,  $M(+)$ ,  $M(-)$ , and  $V_w$ .

Type of IL	Network	No. of parameters	Process	No. of data	$\% \Delta T_g$	$\%  \Delta T_g $	$\%  \Delta T_g _{\max}$
Ammonium	(4,4,4,1)	69 (Model 1)	Training	122	1.1	5.8	48.9
			Testing	15	-3.2	15.7	54.8
			Prediction	15	3.1	8.5	29.2
		69 (Model 2)	Training	122	0.4	4.3	25.4
			Testing	15	0.8	21.2	51.9
			Prediction	15	-74.0	89.5	1136.6
Triazolium	(2,4,1)	19 (Model 1)	Training	30	0.0	1.0	4.6
			Testing	3	-2.2	2.2	2.5
			Prediction	2	-5.3	7.2	12.5
	19 (Model 2)	Training	30	0.0	0.6	4.5	
		Testing	3	1.6	1.6	2.6	
		Prediction	2	-6.6	10.5	17.1	
	19 (Model 3)	Training	30	0.0	1.2	8.6	
		Testing	3	2.7	4.3	5.8	
		Prediction	2	-5.4	8.9	14.3	
Pyrrolidinium	(2,4,1)	19 (Model 1)	Training	28	0.0	1.3	7.5
			Testing	3	1.7	7.7	9.0
			Prediction	3	-7.5	11.7	28.8
	19 (Model 2)	Training	28	0.0	0.5	2	
		Testing	3	-0.1	5.2	7.6	
		Prediction	3	-24.8	24.8	39.6	
Pyridinium	(2,4,1)	19 (Model 1)	Training	45	0.1	2.7	9.8
			Testing	5	-3.4	6.1	19.2
			Prediction	5	3.9	8.8	21.4
	19 (Model 2)	Training	45	0.1	2.1	7.2	
		Testing	5	1.8	8.0	19.0	
		Prediction	5	9.1	20.8	70.3	

topological index to predict the glass transition temperatures of ILs. The authors considered five kinds of ILs: imidazolium (Im), pyridinium (Py), ammonium (Am), sulfonium (Su), and triazolium (Tr). Mokadem et al. [5] estimated  $T_g$  with a group-interaction contribution method. Several ILs including imidazolium-, pyridinium-, triazolium-, sulfonium-, pyrrolidinium-, piperidinium-, phosphonium-, oxazolidinium-, ammonium-, morpholinium-, guanidinium-, amino acid-, and caprolactam-based ILs were considered in the study. The authors claim that “this method represents an excellent alternative to previous approaches for the estimation of the glass transition temperature of diverse ionic liquids from the knowledge of their molecular structure”. A comparison of results from

the literature and the results of this work is presented in Table 11.

In addition, Table 12 presents a comparison of predicted  $T_g$  for selected ILs reported in the literature and the values found in this work. As shown in the table, different variations of QSPR techniques, different forms of group contribution methods, and neural network have been explored in the literature, so comparison can be done. Deviations reported in the literature are similar for the different methods.

Results in Tables 11 and 12 are provided as relative percentage deviation with respect to literature values, the form in which most of authors present their results. For other properties (such as density or heat capacity),

**Table 11**

Comparison of results obtained in this work compared with result of other models proposed in the literature.

Model	Type of IL	N	$\% \Delta T_g$	$\%  \Delta T_g _{\max}$	Ref.
QSPR models	Different types of ILs	112 (Training) 27 (Testing)	1.2 4	18 13.7	[34]
Group contribution method	Different types of ILs	396 (Training) 100 (Testing)	3.6 3.7	12 12	[17]
Group contribution method	1,3-Dialkylimidazolium-type ILs	109	1.9	8.6	[17]
A Linear QSPR method	1,3-Dialkylimidazolium-type ILs	88 (Training) 21 (Testing)	2.4 3.2	7.1 7.6	[4]
Nonlinear approach of the QSPR method	1,3-Dialkylimidazolium-type ILs (109)	10 (Validation)	3.4	6.7	[18]
QSPR methodology using topological indexes	Five types of ILs (63 imidazolium, 17 pyridinium, 48 ammonium, 7 sulfonium, and 4 triazolium)	139	3.3	4.2	[19]
ANN with molecular descriptors	Imidazolium-type ILs (223 training and 25 for testing)	18 (Prediction)	9.2	26.8	This work
	Alkyl-3-methylimidazolium (95 for testing and 5 for training)	5 (Prediction)	2.5	5	This work
	Ammonium-type ILs	15 (Prediction)	3.1	8.5	This work
	Triazolium	3 (Testing)	-2.2	2.5	This work
	Pyrrolidinium	3 (Testing)	1.7	7.7	This work

**Table 12**  
Comparison of predicted  $T_g$  for selected ILs.

Ionic Liquids		$T_g^{\text{lit}}$	$T_g^{\text{cal}}$	$ \Delta T_g _{\text{max}}$	Method	Type of IL	Ref.
[BzimBz]	[dca]	227.7	222.4	2.3	GC	[1,3-dialkyl][mim]	[17]
		227.7	221.8	–2.6	ANN	Imidazolium	This work
[C2OHmim]	[bti]	194.2	193.9	0.1	GC	[1,3-dialkyl][mim]	[17]
		194.2	200.7	3.4	ANN	Imidazolium	This work
[C8mim]	[PF6]	191.2	194.8	1.9	GC	[1,3-dialkyl][mim]	[17]
		202.2	196.2	–2.9	ANN	Imidazolium	This work
[BzimBz]	[ta]	236.0	240.2	1.8	GC	[1,3-dialkyl][mim]	[17]
		236.0	235.9	0.0	ANN	Imidazolium	This work
		191.2	197.6	3.4	ANN	Imidazolium	This work
		205.6	200.2	–2.6	ANN	Imidazolium	This work
[C10mim]	[PF6]	202.2	196.2	3.0	GC	[1,3-dialkyl][mim]	[17]
		202.2	193.3	–4.4	ANN	Imidazolium	This work
		189.2	201.0	6.3	ANN	Imidazolium	This work
[PAMPBim]	[PF6]	234.2	234.2	0.0	GC	[1,3-dialkyl][mim]	[17]
		234.2	239.4	2.2	ANN	Imidazolium	This work
[C8mim]	[bti]	189.0	190.2	0.7	GC	[1,3-dialkyl][mim]	[17]
		189.2	197.4	4.4	ANN	Imidazolium	This work
[BzimBz]	[bti]	222.9	228.8	2.7	GC	[1,3-dialkyl][mim]	[17]
		222.9	218.4	–2.0	ANN	Imidazolium	This work
[C10mim]	[Sac]	203.9	203.9	0.0	GC	[1,3-dialkyl][mim]	[17]
		203.9	218.4	7.1	ANN	Imidazolium	This work
[C9mim]	[bti]	190.2	190.9	0.4	GC	[1,3-dialkyl][mim]	[17]
		190.2	195.9	3.0	ANN	Imidazolium	This work
[DPEHIM]	[PF6]	204.4	200.3	2.0	GC	[1,3-dialkyl][mim]	[17]
		204.4	204.8	0.2	ANN	Imidazolium	This work
		238.2	243.0	2.1	ANN	Imidazolium	This work
[acrylateC6mim]	[bti]	198.2	202.9	2.4	GC	[1,3-dialkyl][mim]	[17]
		198.2	196.0	–1.1	ANN	Imidazolium	This work
[DPEOIM]	[PF6]	200.0	201.6	0.8	GC	[1,3-dialkyl][mim]	[17]
		200.0	198.3	–0.8	ANN	Imidazolium	This work
[C2im]	[BF4]	186.2	208.9	12.2	ANN	Any	[16]
		186.2	189.2	1.6	ANN	Imidazolium	This work
[C2OHmim]	[PF6]	201.2	198.5	1.3	GC	[1,3-dialkyl][mim]	[17]
		201.2	213.8	6.3	ANN	Imidazolium	This work
[C2Omim]	[C2F5BF3]	175.2	175.2	0.0	GC	[1,3-dialkyl][mim]	[17]
		175.2	175.3	0.1	ANN	Imidazolium	This work
[DPEHIM]	[BF4]	189.9	198.9	4.8	GC	[1,3-dialkyl][mim]	[17]
		189.9	210.3	10.8	ANN	Imidazolium	This work
[DPEOIM]	[BF4]	188.4	200.3	6.3	GC	[1,3-dialkyl][mim]	[17]
		188.4	197.5	4.9	ANN	Imidazolium	This work
[DPPOIM]	[BF4]	201.5	201.5	0.0	GC	[1,3-dialkyl][mim]	[17]
		201.5	201.7	0.1	ANN	Imidazolium	This work
		183.2	197.9	8.0	ANN	Imidazolium	This work
[Em2im]	[ba]	207.8	201.3	3.1	ANN	Any	[16]
		207.8	213.0	2.5	ANN	Imidazolium	This work
[C2im]	[ClO4]	192.2	213.9	11.3	ANN	Any	[16]
		192.2	205.3	6.8	ANN	Imidazolium	This work

deviations of 10% could be acceptable because the effect of different properties has different effect in different applications [35]. However, 10% in a value of  $T_g$  could mean differences of 15–30 °C, a deviation that we find non-acceptable. Differences of 30 °C considerably reduce the probable liquidus range of the IL, range in which one wants to safely work during a process, without the threatening of crystal formation. We propose here that an acceptable margin to claim success in modeling the glass transition temperature (or any transition temperature) must be on the order of 10 °C as maximum (3%) and average deviations not greater than 3 °C (1%). Therefore, none of the available publications, or this work, can claim success yet in predicting  $T_g$ . However, the efforts of researchers such as those mentioned in this work must continue especially in the area of standardizing the definition of  $T_g$  and its experimental determination.

## 5. Conclusions

ANN models have been used to correlate and predict the glass transition temperature of ILs. The study and the results obtained in this work allow drawing the following main conclusions: (1) simple architectures of three or four layers with maximum of 13 neurons are sufficient to correlate the glass transition temperature of ILs; (2) the descriptors used show to be good representative parameters to distinguish data for the different ILs; (3) better training, testing, and some good predictive capabilities are observed when the family of ILs is reduced so the members of the family have minimal structural differences, such as the case of the *n*-alkyl-3-methylimidazolium family; (4) for reduced families, and if a reasonable amount of data is available,  $T_g$  could be predicted with deviations less than 10%; and (5) the lack of a clear definition of the glass

transition temperature and the lack of knowledge on what are the properties that most affect liquid–solid transition are the main causes of the present incapability for accurately predicting the glass transition temperature of the type of ILs studied in this work.

## Acknowledgments

The authors thank the National Council for Scientific and Technological Research (CONICYT) for providing support through the research grant FONDECYT 1120162 of the Center for Technological Information of La Serena–Chile for computer facilities and of the Direction of Research and Development of the University of La Serena for permanent support and for an especial graduate fellowship PT15142 for one of the authors (R.A.C.).

## Notation

### Symbols

%	percentage
% $\Delta T_g$	average relative deviation of glass transition temperature
% $\Delta T_g$	average absolute relative deviation of glass transition temperature
% $\Delta T_g$   <sub>max</sub>	maximum absolute relative deviation of glass transition temperature
<i>F</i>	general function or IL property
IL	ionic liquid
<i>n</i> C	number of carbon atoms
<i>n</i>	number of neurons in a network
<i>N</i>	number of data in a data set
<i>M</i> (+)	cation mass
<i>M</i> (–)	anion mass
<i>T</i> <sub>cc</sub>	cold crystallization temperature
<i>T</i> <sub>f</sub>	freezing temperature
<i>T</i> <sub>g</sub>	glass transition temperature
<i>T</i> <sub>m</sub>	melting temperature
<i>T</i> <sub>ss</sub>	solid–solid transition
<i>V</i> <sub>w</sub>	van der Waals volume
<i>X</i> <sub>0</sub>	connectivity index of order zero
<i>x</i> <sub>1</sub> , <i>x</i> <sub>2</sub> , <i>x</i> <sub>3</sub> ... <i>x</i> <sub><i>i</i></sub> , <i>x</i> <sub><i>n</i></sub>	general independent variables

### Abbreviations

Am	ammonium
ANNs	artificial neural networks
GC	group contribution
Im	imidazolium
Py	pyridinium
QSPR	quantitative structure–activity relationship
Ref	reference
<i>S</i> <sub>A</sub>	surface area
Su	sulfonium
Tr	triazolium

### Superscripts/subscripts

cal	calculated
g	glass
lit.	literature
max	maximum

### Greek letters

$\Delta$	difference
$\lambda$	connectivity index
$\Sigma$	summation

## Appendix. A. Supporting information

Supporting information related to this article can be found at <http://dx.doi.org/10.1016/j.crci.2016.11.009>.

## References

- [1] T.L. Greaves, C.J. Drummond, *Chem. Rev.* 108 (1) (2008) 206–237.
- [2] S.I. Fletcher, F.B. Sillars, N.E. Hudson, P.J. Hall, *J. Chem. Eng. Data* 55 (2010) 778–782.
- [3] H. Ohno, *Electrochemical Aspects of Ionic Liquids*, Wiley-Interscience, New York, USA, 2005, pp. 5–23.
- [4] S.M. Mousavisafavi, F. Gharagheizi, S.A. Mirkhani, J. Akbari, *J. Therm. Anal. Calorim.* (2013), <http://dx.doi.org/10.1007/s10973-012-2208-7>, 111–1639.
- [5] K. Mokadem, M. Korichi, K. Tumba, *Fluid Phase Equilib.* 425 (2016) 259–268.
- [6] IUPAC, Physical and Biophysical Chemical Division, *Ionic Liquids Database*, 2003. Available at: <http://www.iupac.org/projects/2003/2003-020-2-100.html>.
- [7] The Beilstein Database, MDL Information Systems GmbH, 2016. World Wide Web Available at: [http://www.lib.ncsu.edu/databases/more\\_info.php?database=17407](http://www.lib.ncsu.edu/databases/more_info.php?database=17407).
- [8] The Dortmund Data Bank (DDB), *Ionic Liquids in the Dortmund Data Bank*, 2016. Available at: [http://www.ddbst.de/new/frame\\_ionic\\_liquids.html](http://www.ddbst.de/new/frame_ionic_liquids.html).
- [9] S. Zhang, X. Lu, Q. Zhou, X. Li, X. Zhang, S. Li, *Ionic Liquids. Physicochemical Properties*, 1st ed., Elsevier, Oxford: UK, 2009.
- [10] J. Brandrup, E.H. Immergut, E.A. Grulke (Eds.), *Polymer Handbook*, 4th ed., John Wiley & Sons, Inc., New York, NY, USA, 1999.
- [11] J.W.P. Schmelzer, I.S. Gutzow, O.V. Mazurin, A.I. Priven, S.V. Todorova, B.P. Petroff, *Introduction, in Glasses and the Glass Transition*, Wiley-VCH Verlag GmbH & Co. KGaA, : Weinheim, Germany, 2011.
- [12] E. Gómez, N. Calvar, A. Domínguez, *Thermal Behaviour of Pure Ionic Liquids*, Chapter 8, in: S. Handy (Ed.), *Ionic Liquids – Current State of the Art*, INTECH, 2015, <http://dx.doi.org/10.5772/59271>.
- [13] J.D. Holbrey, R.D. Rogers, *Melting Points and Phase Diagrams*, in: P. Wasserscheid, T. Welton (Eds.), *Ionic Liquids in Synthesis*, Wiley-VCH Verlag GmbH & Co, Weinheim, Germany, 2002.
- [14] L.C. Branco, J.N. Rosa, J.J. Moura-Ramos, C.A. Alfonso, *Chem.—Eur. J.* 8 (2002) 3671–3677.
- [15] S. Zhang, N. Sun, X. He, X. Lu, X. Zhang, *J. Phys. Chem.* 35 (4) (2006) 1475–1517.
- [16] S.A. Mirkhani, F. Gharagheizi, P. Ilani-Kashkouli, N. Farahani, *Fluid Phase Equilib.* 324 (2012) 50–63.
- [17] F. Gharagheizi, M.H. Keshavarz, P. Ilani-Kashkouli, N. Farahani, K. Tumba, *J. Therm. Anal. Calorim.* 114 (3) (2013) 1363–1382, <http://dx.doi.org/10.1007/s10973-012-2907-0>.
- [18] S.M. Mousavisafavi, S.A. Mirkhani, F. Gharagheizi, J. Akbari, *J. Therm. Anal. Calorim.* (2013) 111–235, <http://dx.doi.org/10.1007/s10973-012-2207-8>.
- [19] F. Yan, S. Xia, Q. Wang, Q. Shang, P. Ma, *Fluid Phase Equilib.* 358 (2013) 166–171.
- [20] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, 2nd, Revised and Enlarged Ed., Vol. 41 (2 Volume Set), Wiley-VCH Verlag GmbH & Co, 2009.
- [21] J.O. Valderrama, R.E. Rojas, *Fluid Phase Equilib.* 297 (2010) 107–112.
- [22] Taletè, Software Dragon7, *Molecular Descriptors for Computational Chemistry*. Taletè slr, 2015. Naples, Italy, <http://www.taletè.mi.it>.

- [23] N.K. Bose, P. Liang, *Neural Networks Fundamentals with Graphs, Algorithms, and Applications*, in *Electrical and Computer Engineering*, McGraw-Hill Series, The McGraw-Hill Companies, Inc., USA, 1996.
- [24] J. Taskinen, J. Yliruusi, *Adv. Drug Delivery Rev.* 55 (2003) 1163–1183.
- [25] J.O. Valderrama, J. Makarena, R.E. Rojas, *Korean J. Chem. Eng.* 28 (6) (2011) 1451–1457.
- [26] J.O. Valderrama, A. Toro, R.E. Rojas, *J. Chem. Thermodyn.* 43 (2011) 1068–1073.
- [27] J.O. Valderrama, G. Martinez, R.E. Rojas, *Thermochim. Acta* 513 (2011) 83–87.
- [28] J.O. Valderrama, *Ind. Eng. Chem. Res.* 53 (2014) 1004–1014.
- [29] D.J. Livingstone, D.T. Manallack, I.V. Tetko, *J. Comput. Aided Mol. Des.* 11 (1996) 135–142.
- [30] J.O. Valderrama, C.A. Faúndez, V.J. Vicencio, *Ind. Eng. Chem. Res.* 53 (2014) 10504–10511.
- [31] J.O. Valderrama, V.H. Alvarez, *Can. J. Chem. Eng.* 83 (2005) 1–4.
- [32] J.O. Valderrama, R.A. Campusano, A. Toro, *J. Water Reuse Desal.* 5 (4) (2015) 454–464.
- [33] J.O. Valderrama, R.A. Campusano, *C. R. Chim.* (2016) 654–664, <http://dx.doi.org/10.1016/j.crci.2016.02.002>.
- [34] S.A. Mirkhani, F. Gharagheizi, *Thermochim. Acta* 543 (2012) 88–95.
- [35] M. Harg, *Fluid Phase Equilib.* 14 (1983) 303–313.