Regular Article

PHYSICAL CHEMISTRY RESEARCH
Published by the
Iranian Chemical Society
www.physchemres.org
info@physchemres.org

***Phys. Chem. Res.**, Vol. 4, No. 1, 61-71, March 2016.*
DOI: 10.22036/pcr.2016.11759

Quantitative Modeling for Prediction of Critical Temperature of Refrigerant Compounds

S. Saaidpour*

Department of Chemistry, Faculty of Science, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran

(Received 29 August 2015, Accepted 8 November 2015)

The quantitative structure-property relationship (QSPR) method is used to develop the correlation between structures of refrigerants (198 compounds) and their critical temperature. Molecular descriptors calculated from structure alone were used to represent molecular structures. A subset of the calculated descriptors selected using a genetic algorithm (GA) was used in the QSPR model development. Multiple linear regressions (MLR) were utilized to construct the linear prediction model. The prediction result agrees well with the experimental value of this property. The comparison results indicate the superiority of the presented model and reveal that it can be effectively used to predict the critical temperatures of refrigerant compounds from the molecular structures alone. The stability and predictivity of the proposed model were validated using internal validation, external validation and Y-scrambling. Application of the developed model to a testing set of 39 organic compounds demonstrates that the new model is reliable with good predictive accuracy and simple formulation. The R^2, RMSEtr and Q^2_{loo} values for the training set were 0.9752, 13.8994 and 0.9742; Q^2_{ext} and RMSEpr for test set were 0.9766 and 12.8654 for GA-MLR model, respectively. The prediction results are in good agreement with the experimental values. In addition, the applicability domain (AD) of the model was analyzed based on the Williams plot.

Keywords: Refrigerants, Critical temperature, QSPR, GA-MLR, Prediction

INTRODUCTION

Halogen-containing organic compounds are an important class of chemicals, with many industrial and laboratory applications. They used as solvents, plastics, anesthetics, foaming agents, refrigerants, and pesticides [1,2]. The refrigerant is a compound, generally a fluid, used in a heat pump and refrigerating cycle. In this cycle, it undergoes phase transitions from a liquid to a gas and back again. Usually, chlorofluorocarbons (CFCs), hydrochloro-fluorocarbons (HCFCs), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) were used as refrigerants. The inert nature of many CFCs and HCFCs, while having the benefit of making them nonflammable and nontoxic, contributes to their stability in the atmosphere, and their corresponding global warming possible and ozone layer depletion possible. Though HFCs and PFCs are non-ozone reducing, many have global warming potential that is thousands of times more than carbon dioxide. Some of the refrigerants such as propane and ammonia are not inert and are flammable or toxic if released. New refrigerants have been developed that are safe to humans and to the environment [3].

The critical point, a point on a phase diagram at which both, the liquid and gas phases of a substance have the same density, and are therefore indistinguishable. The coordinates of this point are called the critical temperature (Tc) and critical pressure (Pc). The temperature above which a gas cannot be liquefied and a substance cannot exhibit distinct gas and liquid phases was called critical temperature (Tc). This property plays an important role in chemical engineering. Because of its importance, it would be useful

*Corresponding author. E-mail: sasaaidpour@iausdj.ac.ir

to develop predictive models, which can be used to predict the Tc of new compounds without experimentation. The Tc is one of the important properties revealing the intermolecular interaction between molecules in the liquid state. The development of a quantitative structure-property relationship (QSPR) model for Tc using genetic algorithm-multiple linear regression (GA-MLR) methodology has been successful. The QSPR approach has been widely used in the prediction of physical and chemical properties of organic compounds [4,5]. QSPR is based on the assumption that the properties of a compound are ultimately determined by its molecular structure. Accordingly, the QSPR approach attempts to establish simple mathematical relationships to describe the correlation of a given property to molecular structures for a set of compounds. Several studies are done to investigations regarding the relationship between Tc of other organic compounds and molecular structure descriptors [6-10].

In the works of Ferri *et al.* and Gasem *et al*., critical properties of diverse organic compounds were studied using Heuristic method and MLR method, respectively [11,12]. In the work of Sobati and Abooali, critical properties of refrigerant compounds were studied [13]. The QSPR can be used to predict physicochemical properties of refrigerant compounds by using theoretical descriptors. So once a reliable QSPR model is established, we can use this model to predict the property of a compound, whether it was synthesized or not. In our previous papers, we reported on the application of quantitative structure-property relationships (QSPR) techniques in developing a new, simplified approach to prediction of organic compounds properties using different models [14-22].

In this study, we present new QSPR model for the prediction of the critical temperature of a diverse set of 198 refrigerant chemicals. Our goal here is to develop an accurate, simple, fast, and less expensive method for calculation of Tc values. A GA-MLR procedure was used for selection of descriptors and modeling. Also, in this work we applied back propagation neural network (BPNN) and support vector machine regression (SVMR) on this data set, but no significant difference between results with MLR method, so we preferred to report of results the MLR method. The predictive power of the resulting model is demonstrated by testing them on unseen data that were not used during model generation. A physicochemical explanation of the selected descriptors is also given.

METHODS AND MATERIALS

To develop a QSPR model for critical temperatures involves several distinct steps, includes (a) data collection, (b) molecular geometry optimization, (c) molecular descriptor generation, (d) descriptor selection, (e) model development and (f) model performance evaluation.

Data Set Selection

All critical temperature data of the present investigation were obtained from the CRC Handbook of Physics and Chemistry and Handbook of Thermophysical Properties of Chemicals and Hydrocarbons [23,24]. Critical temperature range was from 227.5-695 K. The data set consists of 198 compounds, which have been deemed industrially important from the chemical engineering perspective. A complete list of the compound names and corresponding experimental critical temperatures is given as a Supplementary Information.

For evaluating the predictive capability of the proposed model, before model generation both dataset have been split into a training set (~80% of compounds), used for model development, and a prediction set (~20% of compounds), used for external validation. The training set was used to adjust the parameters of the GA-MLR and the test set was used to evaluate its prediction ability.

Molecular Modeling and Descriptor Generation

All numerical calculations have been performed by a computer with intel® Core™ i7 processor and 6Gb RAM characteristics. The ChemDraw Ultra version 13.0 (ChemOffice 2012, CambridgeSoft Corporation; Cambridge, MA) software was used for drawing the molecular structures. The optimizations of molecular structures were done by the HyperChem 8.0 (Hypercube, Inc., Gainesville, 2011) using AM1 method, and descriptors were calculated by Molecular Modeling Pro Plus (MMP+) Version 6.3.3 (ChemSW, Inc.; Fairfield, CA, 2009) software. MMP^{+} software computes six classes of structural descriptors: constitutional; topological; geometrical; electrostatic; quantum chemical and thermodynamic

molecular descriptors [25-27]. Then 72 molecular descriptors were calculated for each compound by the MMP^+ on the minimal energy conformations. In order to reduce redundant and non-useful information, constant or near constant values and descriptors found to be highly correlated pair-wise (one of any two descriptors with a correlation greater than 0.9) were excluded in a pre-reduction step; therefore 43 molecular descriptors underwent subsequent variable selection. Genetic algorithm (GA), included in the QSARINS (version 2.2, 2015), was used for variables selection. MLR was performed by using a routine from the QSARINS [28,29].

Descriptors Selection Based on GA-MLR

In order to perform external validation, the dataset was divided into representative training and prediction set by the Y-ranking method [30]. In this method, the data were sorted firstly according to their Tc values. The training and external prediction sets were chosen from the sorted lists with desired distances from each other. Finally, the training set (159 compounds) was used to establishing QSPR model and the prediction set (39 compounds) was used to evaluate the external predictive ability of the built model. The used molecular descriptors to build model will strongly influence the predictive ability of QSPR model. Among methods of variable selection, genetic algorithm (GA) has been widely used because of its outstanding performance in feature selection [31,32].

Hence, in the present study, we also used GA implemented in QSARINS to perform variable selection. The quality of the model was evaluated by the fitness function, and ranked according to the fitness score. In the work, leave-one-out cross-validation (Q^2_{LOO}) was used as fitness function during the GA process. When increasing the model size does not improve the Q^2 value significantly, the GA selection will be stopped. In this study, genetic algorithm and multiple linear regressions (GA-MLR) the QSARINS software performed using Ordinary Least Square method [28]. The important parameters used in the GA process were set as below: population size 100, maximum allowed descriptors in a model 3 and reproduction/ mutation trade-off 0.5. Finally, we obtained a 3-descriptor subset, which keeps most interpretive information for critical temperatures. Three descriptors were calculated for each compound in the data set. The selected descriptors are enthalpy of vaporization at boiling point (ΔH_{vap}), molar refractivity (MR) and Hansen dispersion forces (DF).

QSPR Modeling and Validation

The models have been developed by multiple linear regression (MLR) using the ordinary least squares (OLS) method, and the genetic algorithm-variable subset selection (GA-VSS) has been applied for variable selection using the in-house software for QSAR modeling, QSARINS. Different parameters have been used to validate the models. The coefficient of determination R^2 has been used as a measure of the total variance of the response explained by the regression model (*i.e.* goodness-of-fit). Model robustness has been verified by the cross-validation coefficient Q^2_{LOO} (leave-one-out) and Q^2_{LMO} (leave many-out, *i.e.* 30% of chemicals excluded in each iteration). To exclude the possibility of chance correlation between modeling descriptors and the response, the Y-Scrambling method has been applied [33], which verifies the fitting of the model developed on randomly re-ordered responses (2000 scrambling iterations); where a low value of the averaged R^2 scrambled (R^2_{ys}) is indicative of a well-founded original model.

It is expected that the resulting QSPR model should generally have low R^2 and low Q^2_{LOO} values. If $Q^2_{yscr} < 0.2$, and $R^2_{yscr} < 0.2$, there is no risk of chance correlation in the developed model. In the present study, y-randomization was performed 2000 times. The external validation of the model has been performed based on four different parameters, *i.e.* $Q^2_{ext\text{-}F1}$, $Q^2_{ext\text{-}F2}$, $Q^2_{ext\text{-}F3}$ and CCC (concordance correlation coefficient) [34,35].

Arbitrary cut-off values have been used to accept models as externally predictive, in the case of small data sets: 0.7 for $Q^2_{ext\text{-}F1}$, $Q^2_{ext\text{-}F2}$ and $Q^2_{ext\text{-}F3}$ and 0.85 for CCC. In addition, the root means squared of errors (RMSE), that summarizes the overall error of the model, has been used to measure and compare prediction accuracy in the training ($RMSE_{tr}$) and in the prediction ($RMSE_{pr}$) sets. Formulas of these validation parameters are given as Supplementary Information.

Applicability Domain (AD)

QSPR model must always be verified for their

applicability domain (AD). In this study, the AD is quantified by applying the leverage approach, which is based on the calculation of the hat matrix for the structural domain. The diagonal values of the hat matrix (h) are used to verify the presence of structural outliers, which are structurally very influential in determining model parameters [36]. Structural outliers are those compounds with h greater than the cut off values h*, in the training set. The h* value is calculated as 3p'/n, where p' is the number of the model variables plus one, and n is the number of training compounds. Response outliers are identified as those compounds with cross-validated standardized residuals greater than 2.5 standard deviation units. Both types of outliers have been detected using the Williams plot, *i.e.* the plot of the diagonal values of the hat matrix (h) versus standardized residuals. The leverage approach has also been applied to evaluate the degree of extrapolation for the predictions obtained for compounds lacking experimental data. Data predicted for high leverage chemicals (h > h*) have been considered less reliable since they are extrapolations from the structural domain of the model. The plot of diagonal hat values *vs*. predicted values, here named Insubria graph, is provided to visualize interpolated and extrapolated predictions.

RESULTS AND DISCUSSION

Result of GA-MLR

After the pre-reduction step of the descriptors calculated by MMP+, totally 72 descriptors were retained for each compound. To select the optimal descriptor subset, 72 descriptors were used as inputs for GA variable selection procedure. When adding another variable did not improve the performance of the model significantly, it means that the optimal subset size was obtained. Finally, a 3-variables model was taken as the optimal MLR model. The regression equation and the statistical items of the optimal model were as follows:

$$Tc = 72.46\,(\pm 27.15) + 7.00\,(\pm 1.26)\Delta H_{vap} + 4.15\,(\pm 1.14)MR + 8.85\,(\pm 2.65)DF$$

n_{tr} = 159, R_2 = 0.9752, *RMSEtr* = 13.8994, F = 2032,

Q^2_{Loo} = 0.9742, $RMSE_{Loo}$ = 14.1889, Q^2_{LMO} = 0.9738,

s = 14.0776, $R^2_{Y-screambling}$ = 0.0185, $Q^2_{Y-screambling}$ = 0.0329

n_{ext} = 39, Q^2_{F1} = 0.9762, Q^2_{F2} = 0.9762, Q^2_{Fa} = 0.9788, CCC_{ext} = 0.9881, $RMSE_{ext}$ = 12.8654

The statistical parameters of the model are satisfying and prove that the MLR model is stable, robust and predictive. In addition, the low value of $R^2_{Y\_scrambling}$ indicating that the obtained model has no chance correlation. The results of coefficients analysis of equation are shown in Table 1.

The results of our work are compared with the previous study on critical temperatures of organic compounds and refrigerants [11-13]. In Table 2 results of the comparison, statistical data are shown. The plot of predicted Tc versus experimental Tc and the residuals (experimental Tc - predicted Tc) *vs*. experimental Tc values, obtained by the GA-MLR modeling, and the random distribution of residuals about zero means are shown in Fig. 1. The predicted values of Tc and residuals obtained by the GA-MLR method are presented in Supplementary Information.

The interpretation of the modeling descriptors in the proposed models selected by GA could provide an effective way to gain some insight into factors that are likely to affect Tc of compounds. Here, three theoretical molecular descriptors were selected and the relative importance of them can be represented by their standardized regression coefficients. As can be seen from Equation, the most significant descriptor (standardized coefficient: 0.5183) is the ΔHvap, which is the Enthalpy of vaporization at the boiling point (kJ mol^{-1}).

Its positive coefficient indicates that the critical temperature capability increases with the increasing of the ΔHvap of compounds. Positive values in the regression coefficients indicate that the indicated descriptor contributes positively to the value of Tc. In other words, increasing the ΔHvap, MR and DF will increase absolute value Tc of the organic compounds.

The predicted values are in good agreement with the experimental values. The robustness of each model was expressed by the cross-validated (leave-one-out technique, LOO) validation coefficient (Q^2_{Loo}) and the root mean square errors of LOO cross-validation ($RMSE_{LOO}$). Successfully validated QSPR model with confirmed

Table1. Results of Coefficients Analysis of GA-MLR Model

				(+/-) Co. int.		t-ratio
Variable	Coeff.	Std. coeff.	Std. err.	95%	p-value	
Intercept	72.4599		13.7451	27.1518	0.0000	5.2717
ΔHvap	7.0039	0.5183	0.6382	1.2608	0.0000	10.9739
MR	4.1509	0.3215	0.5800	1.1458	0.0000	7.1563
Disp	8.8522	0.1738	1.3427	2.6523	0.0000	6.5929

Table 2. Comparison between A. Ferri Model, Kh. A. M. Gasem Model and M.A. Sobati Model and this Work for the Critical Temperature Correlation

	A.Ferri	Kh. A.M. Gasem	M.A. Sobati	This work
Training set	133	1230	159	159
Test set	20	-	39	39
Number of descriptors	8	12	6	3
R^2	0.9856	0.913	0.9651	0.9752
RMSEtr	12.6	16.1	17.5943	13.8994
Q^2_{loo}	-	-	0.9671	0.9742
Q^2_{ext}	-	-	0.9670	0.9766

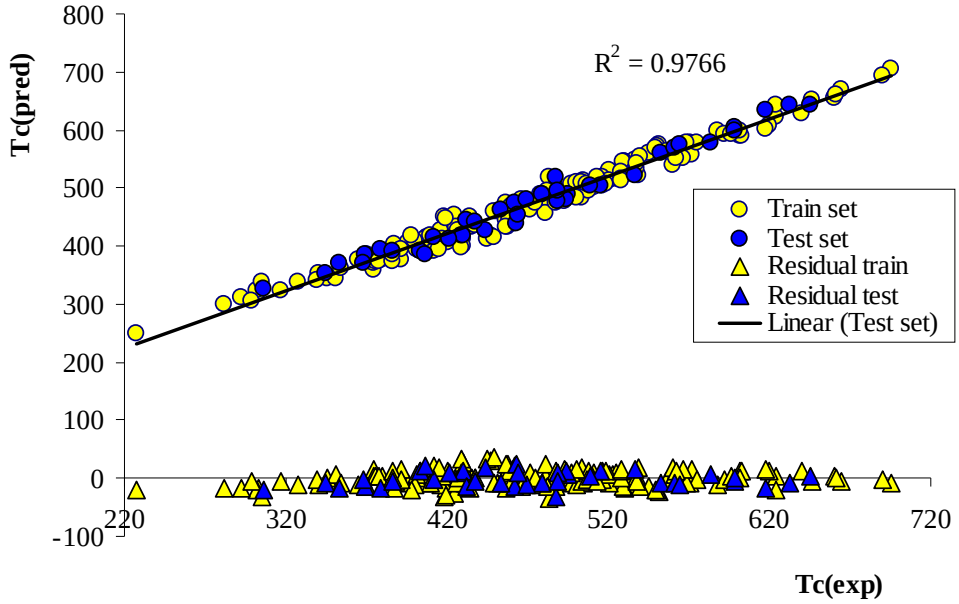

Fig. 1. Scatter plot of the experimental Tc *vs.* predicted and residual T_c values for training, and testing sets of 198 refrigerants compounds.

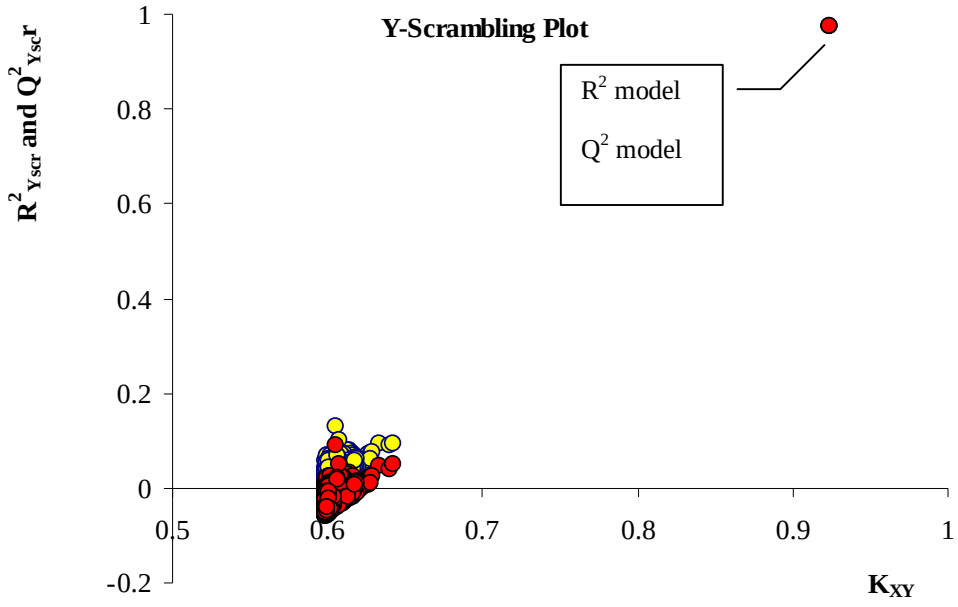

Fig. 2. Y-Scramble plot of Kxy *vs.* R^2 and Q^2 for random models (Kxy: correlations among the block of the descriptors and the experimental data).

predictive abilities was used to predict Tc for all 39 compounds. The internal predictive capability of a model was evaluated by leave-one-out cross-validation (Q^2_{Loo} = 0.9742) on the training set, and the predictive capability of a model on external prediction set can be expressed by Q^2_{ext} = 0.9766.

The model was subsequently validated using the response permutation test, also known as Y-scrambling. This procedure involves fitting several models, on the same dependent variables but on a permutated response. It gave the following results: the random models performed using a scrambled order of the Tc values, found to have significantly lower $R^2_{Y\text{-}scr}$ = 0.0190 and Q^2_{Y-scr} = 0.0324 than the original model corroborating the statistical reliability of the actual model (see Fig. 2). In Fig. 2 correlations among the block of the descriptors and the experimental Tc data *vs.* R^2 and Q^2 for random models are showed. If $Q^2_{yscr} < 0.2$ and $R^2_{yscr} < 0.2$, there is no risk of chance correlation in the developed model. In the present study, y randomization was performed 2000 times.

The Williams plot of the regression allows a graphical detection of both the outliers for the response and the structurally influential chemicals in a model. The leverage (h) of a compound measures its influence on the model. In the standardized residuals plot, all values are within the (-2.5; +2.5) range, which confirms that there are no outliers. To visualize the AD of a QSPR model, the plot of standardized residuals *vs.* leverage values (h) (Williams plot) can be used for an immediate and simple graphical detection of both the response outliers and structurally influential chemicals in a model ($h > h^* = 0.0755$). Samples with high leverages have a stronger influence on the model than other samples; they may or may not be outliers, but they are influential. An influential outlier (high residual + high leverage) is the worst case; it can however easily be detected using an influence plot. Leverages are useful for the detection of samples, which are far from the center within the space described by the model. If a sample has a very large leverage, it may be different from the rest and can be considered to be an outlier. Large leverage indicates a high influence on the model. In the Williams plot for AD , As can be found there is no outlier and leverage chemical in the training and test set used in this study (see Fig. 3). The Insubria graph of this model reported in Fig. 4. This figure shows that GA-MLR model cover a large part of the structural domain occupied by the 198 refrigerants and is similar William plot. Additionally, Figs. 3 and 4 show that the training set is a good representative of the chemical space occupied by the 198-screened compounds.

Interpretation of Descriptors

All descriptors were calculated for the neutral species. The Tc is assumed highly dependent upon the ΔH_{vap}, MR and DF. In the present study, the QSPR model was generated using a training set of 159 molecules. The test set of 39 molecules with regularly distributed Tc values was used to assess the predictive ability of the QSPR model produced in the GA-MLR modeling. The critical temperature of a compound was determined by the different intermolecular interactions between molecules, such as dipole-dipole, dispersion, electronic and hydrogen-bonds interactions. The descriptors in the present model can account for these interactions. The three descriptors involved describe the size and different intermolecular interactions between molecules.

The first descriptor is enthalpy of vaporization (ΔH_{vap}). The change of one phase of a substance to another phase is called phase transition. Therefore, vaporization is a phase transition. The energy that must be supplied as heat at constant pressure per mole of molecules that are vaporized under standard conditions is called the standard enthalpy of vaporization of the liquid. The ΔH_{vap} is a criterion of intermolecular interactions in liquids. The ΔH_{vap} has a high square correlation coefficient with Tc (R^2 = 0.956). The Tc increases with increasing the ΔH_{vap} of a compound.

The second descriptor is molar refractivity (MR). In the following equation, MR has been explained:

$$MR = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{MW}{d} = \frac{\varepsilon - 1}{\varepsilon + 1} \cdot \nabla$$

where *MW* is the molecular weight, *d* the liquid density, and $\overline{V}$ the molar volume, and *n* the refractive index of the liquid referred to the sodium D line, and its square coincides with the dielectric constant ε. MR is the measure of polarizability and molar volume. The MR has a high square correlation coefficient with Tc (R^2 = 0.934). Increasing MR leads to

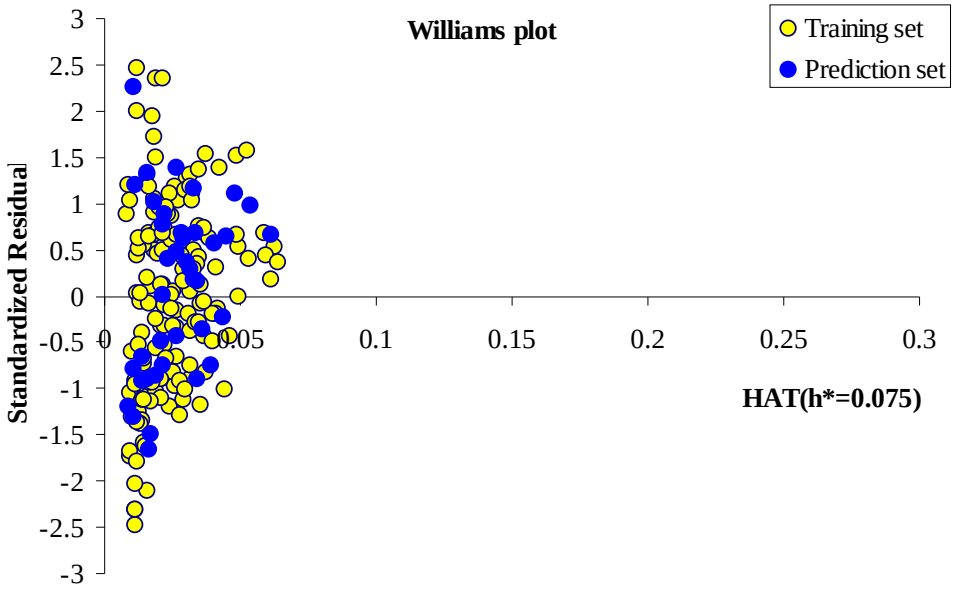

Fig. 3. Williams plot of hat values (h* = 0.075) *vs*. standardized residuals (±2.5σ) for training and testing sets.

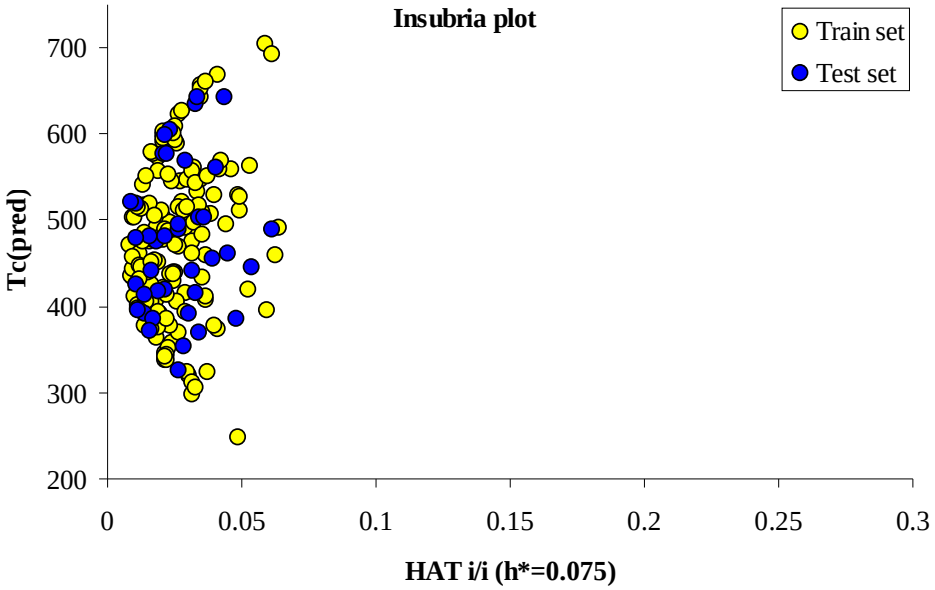

Fig. 4. Insubria graph of hat values (h* = 0.075) *vs*. predicted Tc for training and testing sets.

increasing intermolecular forces. Accordingly, increasing the intermolecular forces, increases extent of Tc related to each refrigerant.

The third descriptor is Hansen dispersion forces (DFs). The DF of attraction, existing between molecules, has no permanent dipole. The van der Waals force is an attractive force between two atoms or nonpolar molecules. This term is attributed to the interaction between two dipoles arising from the dipole fluctuation of one molecule which induces a dipole in the other. With increasing molecular weight, molecular volume and surface area, the van der Waals forces increase. Van der Waals attractive forces exist between all polar and non-polar molecules. The dispersion forces have a high square correlation coefficient with Tc (R^2 = 0.814). Critical temperature increases with increasing dispersion forces.

CONCLUSIONS

In this work, MLR used to construct linear QSPR models to predict critical temperatures of a diverse set of refrigerants. The use of genetic algorithms for descriptor reduction is effective in developing linear-QSPR property model. The results of this study indicate that the use of ΔH_{vap}, MR and DF descriptors provide good estimate for critical temperatures.

The results indicates that the goodness of fit, robustness, and predictivity of GA-MLR model were significant for internal and external validations. In addition, the applicability domain of the optimal GA-MLR model is verified by the leverage approach. Through the mechanism interpretation of the selected three molecular descriptors, we can gain some insights into main molecular structural factors governing the Tc of a compound. In conclusion, the model proposed in this work provides a feasible, effective and practical tool to predict the critical temperatures of refrigerant compounds.

ACKNOWLEDGEMENTS

I wish to thank Prof. Paola Gramatica for their precious help in use of QSARINS software.

Appendix A. Supplementary Information related to this article can be found at ...

REFERENCES

[1] Grigoras, S., A structural approach to calculate physical properties of pure organic substances: the critical temperature, critical volume and related properties. *J. Comput. Chem.* **1990**, *11*, 493-510, DOI: 10.1002/ jcc.540110408.

[2] Egolf, L. M.; Wessel, M. D.; Jurs, P. C., Prediction of boiling points and critical temperatures of industrially important organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1994,** *34*, 947-956, DOI: 10.1021/ ci00020a032.

[3] Haaf, S.; Henrici, H., *Refrigeration Technology*, *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH, 2002.

[4] Katritzky, A. R.; Lobanov, V. S.; Karelson, M., QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure, *Chem. Soc. Rev.*, **1995**, *24*, 279-287, DOI: 10.1039/ CS9952400279.

[5] Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M., Structurally Diverse Quantitative Structure-Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40,* 1-18, DOI: 10.1021/ci9903206.

[6] Hall, L. H.; Story, C. T., Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996,** *36*, 1004-1014, DOI: 10.1021/ ci960375x.

[7] Espinosa, G.; Yaffe, D.; Arenas, A.; Cohen, Y.; Giralt, F., A Fuzzy ARTMAP-Based Quantitative Structure-Property Relationship (QSPR) for Predicting Physical Properties of Organic Compounds, *Ind. Eng. Chem. Res.* **2001**, *40*, 2757-2766, DOI: 10.1021/ie0008068.

[8] Katritzky, A. R.; Mu, L.; Karelson, M., Relationships of Critical Temperatures to Calculated Molecular Properties, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 293-

299, DOI: 10.1021/ci970071q.

[9] Turner, B. E.; Costello, C. L.; Jurs, P. C., Prediction of Critical Temperatures and Pressures of Industrially Important Organic Compounds from Molecular Structure, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 639-645, DOI: 10.1021/ci9800054.

[10] Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, E.; Politzer, P., Relationships of critical constants and boiling points to computed molecular surface properties, *J. Phys. Chem.* **1993**, *97*, 9369-9373, DOI: 10.1021/ j100139a019.

[11] Sola, D.; Ferri, A.; Banchero, M.; Manna, L.; Sicardi, S., QSPR prediction of N-boiling point and critical properties of organic compounds and comparison with a group-contribution method. *Fluid Phase Equilibria* **2008**, *263*, 33-42, DOI:10.1016/ j.fluid.2007.09.022.

[12] Godavarthy, S. S.; Robinson Jr., R. L.; Gasem, K. A. M., Improved structure–property relationship models for prediction of critical properties, *Fluid Phase Equilibria* **2008**, *264*, 122-136, DOI:10.1016/ j.fluid.2007.11.003.

[13] Sobati, M. A.; Abooali, D., Molecular based models for estimation of critical properties of pure refrigerants: Quantitative structure property relationship (QSPR) approach. *Thermochimica Acta* **2015**, *602*, 53-62, DOI: 10.1016/ j.tca.2015.01.006.

[14] Ghasemi, J.; Saaidpour, S., Quantitative structure-property relationship study of n-octanol-water partition coefficients of some of diverse drugs using multiple linear regression. *Anal. Chim. Acta* **2007**, *604*, 99-106, DOI: 10.1016/ j.aca.2007.10.004.

[15] Ghasemi, J.; Saaidpour, S., QSPR prediction of aqueous solubility of drug-like organic compounds. *Chem. Pharm. Bull. (Tokyo)* **2007**, *55*, 669-674, DOI: JST.JSTAGE/cpb/55.669 [pii].

[16] Ghasemi, J.; Saaidpour, S.; Brown, S. D., QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis. *J. Mol. Struct. (Theochem.)* **2007**, *805*, 27-32, DOI: 10.1016/ j.theochem.2006.09.026.

[17] Ghasemi, J.; Saaidpour, S., QSPR modeling of stability constants of diverse 15-crown-5 ethers complexes using best multiple linear regression. *J. Incl. Phenom. Macrocycl. Chem.* **2008**, *60*, 339-351, DOI: 10.1007/s10847-007-9383-3.

[18] Ghasemi, J.; Saaidpour, S., Artificial Neural Network Based Quantitative Structural Property Relationship for Predicting Boiling Points of Refrigerants. *QSAR Comb. Sci.* **2009**, *28*, 1245-1254, DOI: 10.1002/ qsar.200810101.

[19] Ghasemi, J.; Saaidpour, S., QSRR prediction of the chromatographic retention behavior of painkiller drugs. *J. Chromatogr. Sci.* **2009**, *47*, 156-163, DOI: 10.1093/chromsci/47.2.156.

[20] Saaidpour, S., Prediction of drug lipophilicity using back propagation artificial neural network modeling. *Orient. J. Chem.* **2014**, *30*, 793-802, DOI: 10.13005/ ojc/ 300253.

[21] Saaidpour, S., Prediction of the Adsorption Capability onto Activated Carbon of Liquid Aliphatic Alcohols using Molecular Fragments Method. *Iran. J. Math. Chem.* **2014**, *5*, 127-142.

[22] Saaidpour, S.; Zarei, S. A.; Nasri, F., QSPR study of molar diamagnetic susceptibility of diverse organic compounds using multiple linear regression analysis, *Pak. J. Chem.* **2012**, *2*, 6-17, DOI:10.15228/ 2012.v02.i01.p02.

[23] Lide, D. R., *CRC Handbook of Chemistry and Physics*, 90 ed., CRC Press/Taylor and Francis, Boca Raton, Florida, 2010.

[24] Yaws, C. L., *Handbook of Thermophysical Properties of Chemicals and Hydrocarbons*, first ed., William Andrew, New York, 2008.

[25] Huuskonen, J.; Salo, J.; Taskinen, J., Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 450-456, DOI: 10.1021/ ci970100x.

[26] Mitchell, B. E.; Jurs, P. C., Prediction of aqueous solubility of organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 489-496, DOI: 10.1021/ci970117f.

[27] J. M. Sutter, T. A. Peterson, P. C. Jurs, Prediction of gas chromatographic retention indices of alkylbenzenes, *Anal. Chim. Acta* **1997**, *342*, 113-122, DOI: 10.1016/S0003-2670(96)00578-8.

[28] Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S., QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34*, 2121-2132, DOI: 10.1002/jcc.23361.

[29] Gramatica, P.; Cassani, S.; Chirico, N., QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS, *J. Comput. Chem.* (Software News and Updates) **2014**, *35*, 1036-1044, DOI: 10.1002/jcc.23576.

[30] Golbraikh, A.; Tropsha, A., Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol. Divers.* **2002**, *5*, 231-243, DOI: 10.1023/A: 1021372108686.

[31] Leardi, R., *Genetic Algorithms in Feature Selection.* In *Genetic Algorithms in Molecular Modeling*; James, D., Ed. Academic Press: London, 1996.

[32] Haupt, R. L.; Haupt, S. E., *In Practical Genetic Algorithms*; second ed.; Wiley, New Jersey, 2004.

[33] Rücker, C.; Rücker, G.; Meringer, M., y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* **2007**, *47*, 2345-2357, DOI: 10.1021/ci700157b.

[34] Chirico, N.; Gramatica, P., Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.* **2011**, *51*, 2320-2335, DOI: 10.1021/ ci200211n.

[35] Chirico, N.; Gramatica, P., Real external predictivity of QSAR Models. Part2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J. Chem. Inf. Model.* **2012**, *52*, 2044-2058, DOI: 10.1021/ci300084j.

[36] Gramatica, P., Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694 -701, DOI: 10.1002/qsar.200610151.