

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/236029326>

Computation of Upper Hash Point of Chemical Compounds Using a Chemical Structure-Based Model

ARTICLE in INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · APRIL 2012

Impact Factor: 2.59 · DOI: 10.1021/ie202868v

CITATIONS

7

READS

35

4 AUTHORS:



Farhad Gharagheizi

Texas Tech University

168 PUBLICATIONS 2,915 CITATIONS

SEE PROFILE



Poorandokht Ilani-kashkouli

47 PUBLICATIONS 278 CITATIONS

SEE PROFILE



Bahram Mirkhani

The University of Calgary

23 PUBLICATIONS 164 CITATIONS

SEE PROFILE



Amir H. Mohammadi

557 PUBLICATIONS 4,820 CITATIONS

SEE PROFILE

Computation of Upper Flash Point of Chemical Compounds Using a Chemical Structure-Based Model

Farhad Gharagheizi,^{*,†} Poorandokht Ilani-Kashkouli,[†] Seyyed Alireza Mirkhani,[†]
and Amir H. Mohammadi^{*,§,⊥}

[†]Department of Chemical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

[§]MINES ParisTech, CEP/TEP—Centre Énergétique et Procédés, 35 Rue Saint Honoré, 77305 Fontainebleau, France

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

S Supporting Information

ABSTRACT: In this communication, a quantitative structure–property relationship (QSPR) is presented for an estimation of the upper flash point of pure compounds. The model is a multilinear equation that has eight parameters. All the parameters are solely computed based on chemical structure. To develop this model, more than 3000 parameters were evaluated using the Genetic Algorithm Multivariate Linear Regression (GA-MLR) method to select the most statistically effective ones. The maximum average absolute relative deviation (mARD), ARD, squared correlation coefficient, and root mean squares of error of the model from database (DIPPR 801) values for 1294 pure compounds are 25.76%, 3.56%, 0.95, and 17.42 K, respectively.

1. INTRODUCTION

Explosion and fire normally threaten the processes in which combustible or flammable liquids exist. To estimate safe conditions for such processes, accurate knowledge of flammability characteristics is essential. One of the significant flammability characteristics used for this purpose is the upper flammability limit temperature (UFLT) or (upper flash point or upper explosion point). The expression of flammable limits in terms of concentration is not generally appropriate for application in process safety concepts. Therefore, its expression in terms of temperature is preferable.

The UFLT is the (adiabatic) temperature of premixed fuel rich flames at their upper flammability limit with the unburned gas mixture at 20 °C. In other words, the UFLT of a flammable liquid is the temperature related to atmospheric pressure, at which the concentration of a saturated vapor/air mixture is equal to the upper flammability limit. The UFLT of a pure compound can be evaluated using its vapor pressure curve and the upper flammability limit.^{1,2}

The first apparatus used to measure the UFLT property is the Yagyu's³ modified apparatus of that of Jones.^{4,5} However, it was not suitable for practical use because the measurement by this apparatus is a time-consuming process and cleaning the glass part after every test, especially for heavy petroleum residues, is very difficult. Another apparatus used to measure the UFLT is the improved Seta-Flash Closed-Cup Tester.⁶ This apparatus is widely used to measure the UFLT. Of course, the standardizing of the apparatus for measuring the UFLT property is in progress.²

Regarding the aforementioned difficulties of the experimental measurements of the UFLT property, predictive and theoretical methods such as Quantitative Structure Property Relationships (QSPR) are required to estimate flammability characteristics. These methods significantly reduce the amount of required

experimental data. The QSPR technique is now a well-established and useful method to correlate various simple and complex physicochemical properties of a compound based on its molecular structure, through a variety of molecular descriptors (These molecular descriptors are calculated using known mathematical algorithms from the molecular structure of every compound). The basic strategy of QSPR technique is to find optimum quantitative relationships, which can then be used for the prediction of the property from molecular structures. Once a reliable relation has been obtained, it can be applied for prediction of the same property for other structures not yet measured or even not yet prepared. However, these kinds of methods have certain limitations: (I) the family of compounds used to derive the QSPR ("Training set") should be similar according to the chemical structure, and (II) realistic predictions can only be made for compounds that are chemically related to those from which the QSPR model was derived; that is, predictions should be of interpolations or short extrapolations.^{7–51}

In this work, a QSPR study is presented to develop a molecular-based model for prediction of the UFLT of pure compounds using a comprehensive database containing 1294 pure compounds. The application of such a large database leads to extend applicability of the obtained model.

2. MATERIALS AND METHOD

2.1. Data Set. Evaluated databases such as DIPPR 801⁵² are one of the most comprehensive data sets for developing new property prediction models. DIPPR 801⁵² is recommended by

Received: December 7, 2011

Revised: February 22, 2012

Accepted: March 6, 2012

Published: March 6, 2012



the American Institute of Chemical Engineers for physical properties of pure compounds. In this study, 1294 pure compounds that form this database along with their UFLT property are investigated, as pointed out earlier. The studied compounds accompanied with their UFLT values are presented as Supporting Information.

2.2. Determination of Molecular Descriptors. The molecular structures of all 1294 pure compounds are drawn into Hyperchem software⁵³ and optimized using the MM+ molecular mechanics force field in order to determine the molecular descriptors. Thereafter, molecular descriptors are calculated by Dragon software⁵⁴ using these optimized molecular structures. There are many currently used software packages for calculating the molecular descriptors from the chemical structure of chemical compounds. Dragon is the most widely used and most comprehensive one for this purpose because it is capable of computing 20 types of the most important types of molecular descriptors from chemical structure. These types of molecular descriptors, which are treated by the Dragon, include: "Constitutional Descriptors", "Topological Descriptors", "Walk and Path Counts", "Connectivity Indices", "Information Indices", "2D "Autocorrelations", "Edge Adjacency Indices", "Burden Eigenvalue Descriptors", "Topological Charge Indices", "Eigenvalue-Based Indices", "Randic Molecular Profiles", "Geometrical Descriptors", "RDF Descriptors", "3D-Morse Descriptors", "WHIM Descriptors", "GETAWAY Descriptors", "Functional Group Counts", "Atom-Centered Fragments", "Charge Descriptors", and "Molecular Properties". Dragon software is capable of calculating 1664 molecular descriptors for every molecule. For more information about the types of the molecular descriptors and the procedure of calculation of the descriptors, refer to Dragon software user's guide.⁵⁴ These molecular descriptors have been calculated for about 234 000 pure compounds using Dragon software and are freely accessible by Milano Chemometrics and the QSAR research group Web site (http://michem.disat.unimib.it/mol_db).

2.3. GA-MLR Calculations. In general, the next calculation step after calculating molecular descriptors in QSPR studies is to find a linear equation that can predict the desired property with the least number of variables as well as highest accuracy. In other words, the problem is to find a subset of variables (most statistically effective molecular descriptors on UFLT) from all available variables (all molecular descriptors) so that it can predict the UFLT property, with minimum error in comparison with the experimental values. An accepted method for the aforementioned purpose is the Genetic Algorithm-based Multivariate Linear Regression (GA-MLR). In this method, the "genetic algorithm" is used to select the best subset of variables with respect to an objective function. This algorithm was first presented by Leardi et al.⁵⁵

The presented GA-MLR algorithm that implements the RQK fitness function is used for the subset variable selection. The latter was first proposed by Todeschini et al.^{56,57} This method has been well studied by Gharagheizi et al.^{10–12,14–16,18,20,22,26,27,33,36–38,40}

Before starting the computations, the data set must be divided into two new subdata sets; the first one (training set) for developing the model, and the other one (test set) for evaluation of the model. The most accurate model is found by treating the training set data. The predictive capability of the obtained model is evaluated by the test set. In this work, approximately 80% of the database is assigned to the training

set, and 20% is assigned to the test set (1036 compounds are used for the training set, and 258 compounds are used for the test set). The selection is performed randomly.

To obtain the best multivariate linear equation, all molecular descriptors are introduced to the program and the minimum number of possible variables is verified at the starting point. Therefore, running the program is begun with one variable. Later, the best multivariate linear model should be obtained. In the next steps, we increase the number of desired variables to two, three, four, *etc.* It should be noted that the aforementioned calculations should be repeated for the new desired variables. Whenever an increase in the number of variables has no considerable effect on the accuracy of the best-obtained model, the calculations are stopped because the best multivariate linear model is obtained.

3. RESULTS AND DISCUSSION

The most accurate multivariate linear equation contains eight parameters as follows:

$$\begin{aligned} \text{UFLT} = & -19.1651(\pm 5.9629) - 12.6559(\pm 0.8297)n\text{O} \\ & + 5.7444(\pm 0.4647)\text{piPC10} + 102.2474 \\ & (\pm 1.51869)\text{IVDM} + 227.3820(\pm 8.7611)\text{SIC0} \\ & + 42.6139(\pm 1.2148)n\text{ROH} + 24.7320 \\ & (\pm 2.9916)n\text{ArOH} + 6.3012(\pm 0.7039)\text{C} \\ & - 0.25 + 0.99695(\pm 0.0561)\text{TPSA}(\text{Tot}) \end{aligned} \quad (1)$$

$$n_{\text{training}} = 1036; \quad n_{\text{test}} = 258; \quad R^2 = 0.9487$$

$$Q_{\text{LOO}}^2 = 0.9475; \quad Q_{\text{BOOT}}^2 = 0.9471;$$

$$Q_{\text{EXT}}^2 = 0.9537$$

$$s = 17.497; \quad a = 0.0948; \quad F = 2375.72$$

RQK function parameters

$$(\Delta K = 0.059; \quad \Delta Q = 0.000; \quad R^P = 0.009;$$

$$R^N = 0.000)$$

where UFLT is in Kelvin. In the preceding equation (i) "nO" is the number of oxygen atoms. The correlation shows that when the number of oxygen atom in a molecule increases the UFLT property decreases. (ii) "piPC10" denotes molecular multiple path count of order 10. Path counts are atomic and molecular descriptors from hydrogen-depleted molecular graph. These descriptors encode atomic environment, defined as the number of paths of length *m* starting from a vertex to any other vertex in the graph. (iii) The length *m* of the path is the number of edges involved in the path and is called path order. As can be found in eq 1 when the "piPC10" of a molecule increases the UFLT property is also increased. (iv) "IVDM" is the mean information contents on the vertex degree magnitude. This index is proposed as a measure of molecular complexity. In other words, when the molecular complexity of molecule increases its "IVDM" also increases and as a result, its UFLT is enhanced. (v) "SIC0" is the structural information content (neighborhood symmetry of 0-order). This descriptor is a measure of structural complexity per vertex of molecular graph. An increase in "SIC0" causes an increase in mean atomic complexity in a molecule and therefore brings about an increase

in the UFLT property. (vi) “ n_{ROH} ”, “ n_{ArOH} ”, and “C-025” are numbers of hydroxyl groups, aromatic hydroxyls, and tertiary carbon atoms, respectively. These molecular descriptors are functional groups and can be easily calculated from molecular structure. An increase in the number of these functional groups in a desired molecule leads to an increase of the UFLT. (vii) Finally, “TPSA(Tot)” is a measure of the polarity of a molecule. When the polarity of a molecule increases, its UFLT increases, too.⁵⁷ n_{training} and n_{test} are, respectively, the number of compounds in the training set and the test set.

The bootstrapping, y -scrambling, and external validation techniques are used⁵⁶ to evaluate the validity of the model. The bootstrapping is repeated 5000 times. Also y -scrambling is repeated 300 times. As can be seen, the small differences between Q_{LOO}^2 , Q_{BOOT}^2 , Q_{EXT}^2 , and R^2 show that the obtained model is a reliable model and has high predictive capability.⁵⁶ Furthermore, the intercept value of the y -scrambling technique is small ($a = 0.0948$) which reveals the validity of the model (The y -scrambling, bootstrapping, and external validation techniques have been extensively presented by Todeschini et al.⁵⁶) The schematic computation algorithm is presented as Figure-1. The molecular descriptors and their physical meanings are presented in Table 1.

All the validation techniques show that the obtained model is valid and can be reliably used to predict the UFLT of pure compounds. The predicted UFLT values using eq 1 in comparison with the DIPPR 801⁵² data are presented in Figure 2. These values are presented as Supporting Information. Also, the values of the descriptors and status of

Table 1. The Eight Molecular Descriptors of the Model

ID	molecular descriptor	type	definition
1	n_{O}	constitutional descriptors	number of oxygen atoms
2	πPC_{10}	walk and path counts.	molecular multiple path count of order 10
3	IVDM	information indice	mean information content on the vertex degree magnitude
4	SIC0	information indice	structural information content (neighborhood symmetry of 0-order)
5	n_{ROH}	functional group counts.	number of hydroxyl groups
6	n_{ArOH}	functional group counts.	number of aromatic hydroxyls.
7	C-025	atom-centered fragments	number of carbon atoms in R-CR-R groups
8	TPSA(Tot)	molecular properties	topological polar surface area using N, O, S, P polar contributions

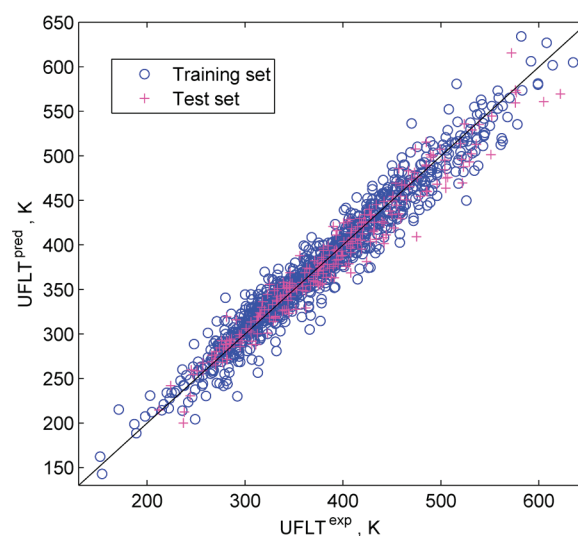


Figure 2. Comparison between the predicted UFLT by eq 1 and DIPPR 801⁵² data. Superscripts pred and exp represent the UFLT values predicted/represented by eq 1 and the UFLT values reported in DIPPR 801⁵² data set, respectively.

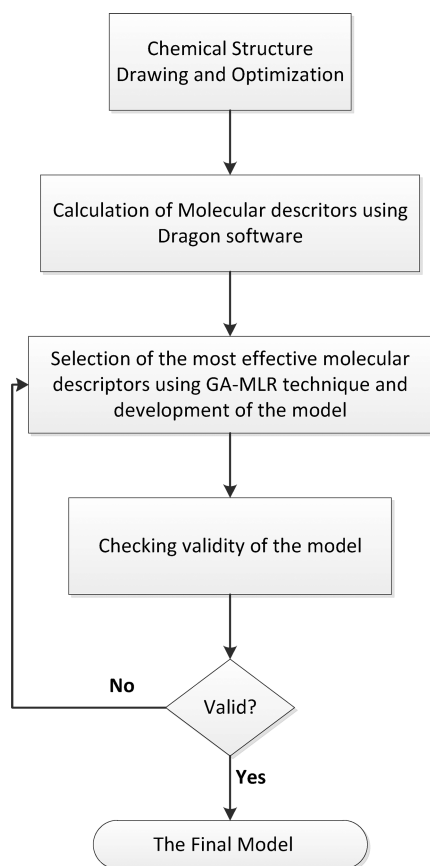


Figure 1. The schematic model development steps.

all the pure compounds (training set or test set) are presented as Supporting Information.

The squared correlation coefficient, root mean squares of error, average absolute relative deviation (ARD), and the maximum ARD (mARD) of the model results from the data set values for all the 1294 pure compounds are, respectively, 0.95, 17.42 K, 3.56%, and 25.76% (flouro-methane (CF_4)). The relative deviation of the model results from data set values is schematically shown in Figure 3

The ARD values of the model predictions from experimental data for each one of the chemical families of pure compounds are presented as Supporting Information. The data demonstrate that the largest errors are related to C, H, F and C, H, I compounds. This case shows that the model is not able to predict the UFLT of halogenated compounds as well as other chemical families. This weakness is probably due to the number of C, H, F and C, H, I compounds in the database. Among all the 1294 pure compounds employed, there are only 3 C, H, F and 4 C, H, I compounds. Therefore, modification of the model for the prediction of UFLT is not urgently necessary. It seems

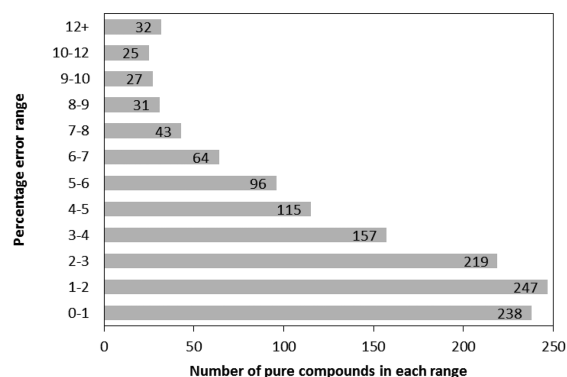


Figure 3. Relative deviation of the model results from data set values.

the modification of the model needs more experimental values for these kinds of compounds.

4. CONCLUSION

In this work, a QSPR model was presented to predict the upper flammability limit temperature of pure compounds. The validity and predictive capability of the model were assessed using several techniques. As a result, the obtained model shows high accuracy. Therefore, it can be reliably used to predict the UFLT of pure compounds.

Since the model has been developed using 1294 pure compounds belonging to diverse chemical families, it can be used to predict the UFLT of common compounds used in process industries. However, application of the model to predict UFLT of compounds which have completely different chemical structures from those used in developing this model is not recommended.

■ ASSOCIATED CONTENT

Supporting Information

Data set containing 1294 pure compounds employed in this study, their UFLTs, the molecular descriptors, their predicted UFLTs, and the AARD% for each chemical family. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*(F.G.) E-mail: fghara@gmail.com. Fax: +98 21 77926580. (A.H.M.) E-mail: amir-hossein.mohammadi@mines-paristech.fr. Phone: + (33) 1 64 69 49 70. Fax: + (33) 1 64 69 49 68.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) Nabert, K.; Schoen, G. *Sicherheitstechnische Kennzahlen brennbarer Gase und Dämpfe (Safety Characteristics of Flammable Gases and Vapors)*, 2nd extended ed. with 6th extension ed.; Braunschweig, Germany, 1990.
- (2) Brandes, E.; Mitu, M.; Pawel, D. The lower explosion point-A good measure for explosion prevention: Experiment and calculation for pure compounds and some mixtures. *J. Loss Prev. Process Ind.* **2007**, *20* (4–6), 536–540.
- (3) Yagyu, S. The graph on the temperatures of flash-points-flammability limits. *J. Jpn Soc. Safety Eng.* **1985**, *24*, 158–152.
- (4) Jones, G. W.; Kennedy, R. E.; Scott, F. E. *Inflammability of Methylene Chloride–Oxygen–Nitrogen Mixtures*; US Bureau of Mines: Pittsburgh PA, 1943; Report of Investigations No. 3727.

(5) Jones, G. W.; Scott, F. E.; Scott, G. S. *Limits of Inflammability and Ignition Temperatures of Acetic Anhydride*. US Bureau of Mines: Pittsburgh PA, 1943; Report of Investigations No. 3741.

(6) Hasegawa, K.; Kashiki, K. A method for measuring upper flash point. *J. Loss Prev. Process Ind.* **1991**, *4*, 176–184.

(7) Mirkhani, S. A.; Gharagheizi, F. Predictive quantitative structure–property relationship model for the estimation of ionic liquid viscosity. *Ind. Eng. Chem. Res.* **2012**, *51*, 2470–2477.

(8) Gharagheizi, F. Determination of diffusion coefficient of organic compounds in water using a simple molecular-based method. *Ind. Eng. Chem. Res.* **2012**, *51*, 2797–2803.

(9) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Artificial neural network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. *Chem. Eng. Sci.* **2011**, *66* (13), 3039–3044.

(10) Gharagheizi, F. QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN. *Comput. Mater. Sci.* **2007**, *40* (1), 159–167.

(11) Gharagheizi, F. Quantitative structure–property relationship for prediction of the lower flammability limit of pure compounds. *Energy Fuels* **2008**, *22* (5), 3037–3039.

(12) Gharagheizi, F. A simple equation for prediction of net heat of combustion of pure chemicals. *Chemom. Intell. Lab. Syst.* **2008**, *91* (2), 177–180.

(13) Gharagheizi, F. A new group contribution-based model for estimation of lower flammability limit of pure compounds. *J. Hazard. Mater.* **2009**, *170* (2–3), 595–604.

(14) Gharagheizi, F. A QSPR model for estimation of lower flammability limit temperature of pure compounds based on molecular structure. *J. Hazard. Mater.* **2009**, *169* (1–3), 217–220.

(15) Gharagheizi, F. Prediction of upper flammability limit percent of pure compounds from their molecular structures. *J. Hazard. Mater.* **2009**, *167* (1–3), 507–510.

(16) Gharagheizi, F.; Abbasi, R.; Tirandazi, B. Prediction of Henry's law constant of organic compounds in water from a new group-contribution-based model. *Ind. Eng. Chem. Res.* **2010**, *49* (20), 10149–10152.

(17) Gharagheizi, F. An accurate model for prediction of autoignition temperature of pure compounds. *J. Hazard. Mater.* **2011**, *189* (1–2), 211–221.

(18) Gharagheizi, F.; Alamdari, R. F. Prediction of flash point temperature of pure components using a quantitative structure–property relationship model. *QSAR Comb. Sci.* **2008**, *27* (6), 679–683.

(19) Gharagheizi, F.; Alamdari, R. F. A molecular-based model for prediction of solubility of C60 fullerene in various solvents. *Fullerenes, Nanotubes, Carbon Nanostruct.* **2008**, *16* (1), 40–57.

(20) Gharagheizi, F.; Babaie, O.; Mazdeyasna, S. Prediction of vaporization enthalpy of pure compounds using a group contribution-based method. *Ind. Eng. Chem. Res.* **2011**, *50* (10), 6503–6507.

(21) Gharagheizi, F.; Eslamimanesh, A.; Tirandazi, B.; Mohammadi, A. H.; Richon, D. Handling a very large data set for the determination of surface tension of chemical compounds using quantitative structure–property relationship strategy. *Chem. Eng. Sci.* **2011**, *66* (21), 4991–5023.

(22) Gharagheizi, F.; Eslamimanesh, A.; Farjood, F.; Mohammadi, A. H.; Richon, D. Solubility parameters of nonelectrolyte organic compounds: Determination using quantitative structure–property relationship strategy. *Ind. Eng. Chem. Res.* **2011**, *50* (19), 11382–11395.

(23) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Phase equilibrium modeling of structure h clathrate hydrates of methane plus water “insoluble” hydrocarbon promoter using QSPR molecular approach. *J. Chem. Eng. Data* **2011**, *56* (10), 3775–3793.

(24) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Artificial neural network modeling of solubilities of 21 commonly used industrial solid compounds in supercritical carbon dioxide. *Ind. Eng. Chem. Res.* **2011**, *50* (1), 221–226.

(25) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Group contribution model for determination of molecular

diffusivity of non-electrolyte organic compounds in air at ambient conditions. *Chem. Eng. Sci.* **2012**, *68*, 290–304.

(26) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Group contribution-based method for determination of solubility parameter of nonelectrolyte organic compounds. *Ind. Eng. Chem. Res.* **2011**, *50* (17), 10344–10349.

(27) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. QSPR approach for determination of parachor of non-electrolyte organic compounds. *Chem. Eng. Sci.* **2011**, *66* (13), 2959–2967.

(28) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Determination of critical properties and acentric factors of pure compounds using the artificial neural network group contribution algorithm. *J. Chem. Eng. Data* **2011**, *56* (5), 2460–2476.

(29) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Representation and prediction of molecular diffusivity of non-electrolyte organic compounds in water at infinite dilution using the artificial neural network-group contribution method. *J. Chem. Eng. Data* **2011**, *56* (5), 1741–1750.

(30) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Use of artificial neural network-group contribution method to determine surface tension of pure compounds. *J. Chem. Eng. Data* **2011**, *56* (5), 2587–2601.

(31) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Empirical method for representing the flash-point temperature of pure compounds. *Ind. Eng. Chem. Res.* **2011**, *50* (9), S877–S880.

(32) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Determination of parachor of various compounds using an artificial neural network-group contribution method. *Ind. Eng. Chem. Res.* **2011**, *50* (9), S815–S823.

(33) Gharagheizi, F.; Gohar, M. R. S.; Vayeghan, M. G. A quantitative structure–property relationship for determination of enthalpy of fusion of pure compounds. *J. Therm. Anal. Calorim.*, in press, DOI: 10.1007/s10973-011-1727-y.

(34) Gharagheizi, F.; Mehrpooya, M. Prediction of standard chemical exergy by a three descriptors QSPR model. *Energy Convers. Manage.* **2007**, *48* (9), 2453–2460.

(35) Gharagheizi, F.; Mirkhani, S. A.; Tofangchi Mahyari, A. R. Prediction of standard enthalpy of combustion of pure compounds using a very accurate group-contribution-based method. *Energy Fuels* **2011**, *25* (6), 2651–2654.

(36) Gharagheizi, F.; Sattari, M. Estimation of molecular diffusivity of pure chemicals in water: A quantitative structure–property relationship study. *SAR QSAR Environ. Res.* **2009**, *20* (3–4), 267–285.

(37) Gharagheizi, F.; Sattari, M. Prediction of the θ (UCST) of polymer solutions: A quantitative structure–property relationship study. *Ind. Eng. Chem. Res.* **2009**, *48* (19), 9054–9060.

(38) Gharagheizi, F.; Sattari, M. Prediction of triple-point temperature of pure components using their chemical structures. *Ind. Eng. Chem. Res.* **2010**, *49* (2), 929–932.

(39) Gharagheizi, F.; Sattari, M.; Tirandazi, B. Prediction of crystal lattice energy using enthalpy of sublimation: A group contribution-based model. *Ind. Eng. Chem. Res.* **2011**, *50* (4), 2482–2486.

(40) Gharagheizi, F.; Tirandazi, B.; Barzin, R. Estimation of aniline point temperature of pure hydrocarbons: A quantitative structure–property relationship approach. *Ind. Eng. Chem. Res.* **2009**, *48* (3), 1678–1682.

(41) Mehrpooya, M.; Gharagheizi, F. A molecular approach for the prediction of sulfur compound solubility parameters. *Phosphorus, Sulfur, Silicon Relat. Elem.* **2010**, *185* (1), 204–210.

(42) Mehrpooya, M.; Gharagheizi, F.; Vatani, A. Thermoeconomic analysis of a large industrial propane refrigeration cycle used in NGL recovery plant. *Int. J. Energy Res.* **2009**, *33* (11), 960–977.

(43) Sattari, M.; Gharagheizi, F. Prediction of molecular diffusivity of pure components into air: A QSPR approach. *Chemosphere* **2008**, *72* (9), 1298–1302.

(44) Eslamimanesh, A.; Gharagheizi, F.; Illbeigi, M.; Mohammadi, A. H.; Fazlali, A.; Richon, D. Phase equilibrium modeling of clathrate hydrates of methane, carbon dioxide, nitrogen, and hydrogen+water

soluble organic promoters using Support Vector Machine algorithm. *Fluid Phase Equilib.* **2012**, *316* (1), 34–45.

(45) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Phase equilibrium modeling of structure H clathrate hydrates of methane + water “insoluble” hydrocarbon promoter using QSPR molecular approach. *J. Chem. Eng. Data* **2011**, *56* (10), 3775–3793.

(46) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D.; Illbeigi, M.; Fazlali, A.; Forghani, A. A.; Yazdizadeh, M. Phase equilibrium modeling of structure H clathrate hydrates of methane + water “insoluble” hydrocarbon promoter using group contribution-support vector machine technique. *Ind. Eng. Chem. Res.* **2011**, *50* (22), 12807–12814.

(47) Gharagheizi, F.; Salehi, G. R. Prediction of enthalpy of fusion of pure compounds using an Artificial Neural Network-Group Contribution method. *Thermochim. Acta* **2011**, *521* (1–2), 37–40.

(48) Gharagheizi, F. Determination of normal boiling vaporization enthalpy using a new molecular-based model. *Fluid Phase Equilib.* **2012**, *317*, 43–51.

(49) Mirkhani, S. A.; Gharagheizi, F.; Sattari, M. A QSPR model for prediction of diffusion coefficient of non-electrolyte organic compounds in air at ambient condition. *Chemosphere* **2012**, *86* (9), 959–966.

(50) Mousavisafavi, S. M.; Gharagheizi, F.; Mirkhani, S. A.; Akbari, J. A predictive quantitative structure–property relationship for glass transition temperature of 1,3-dialkyl imidazolium ionic liquids. Part 2. The nonlinear approach. *J. Therm. Anal. Calorim.*, in press, doi: 10.1007/s10973-012-2208-7.

(51) Mousavisafavi, S. M.; Mirkhani, S. A.; Gharagheizi, F.; Akbari, J. A predictive quantitative structure–property relationship for glass transition temperature of 1,3-dialkyl imidazolium ionic liquids Part 1. The linear approach. *J. Therm. Anal. Calorim.*, in press, doi: 10.1007/s10973-012-2207-8.

(52) Project 801, *Evaluated Process Design Data, Public Release Documentation*; Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE): New York, 2006.

(53) HyperChem Release 7.5 for Windows; Hypercube, Inc.: Gainesville, FL, 2002.

(54) Talete srl, Dragon for Windows (Software for Molecular Descriptor Calculation), version 5.5, 2007.

(55) Leardi, R.; Boggia, R.; Terrile, M. Genetic algorithms as a strategy for feature selection. *J. Chemom.* **1992**, *6*, 267–281.

(56) Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Detecting “bad” regression models: Multicriteria fitness functions in regression analysis. *Anal. Chim. Acta* **2004**, *5*, 208–199.

(57) Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; Wiley-VCH: Weinheim, Germany, 2009.