

A quantitative structure–property relationship for determination of enthalpy of fusion of pure compounds

Farhad Gharagheizi · Mohammad Reza Samiee Gohar ·
Mahsa Ghotbi Vayeghan

Received: 24 May 2011 / Accepted: 3 June 2011 / Published online: 25 June 2011
© Akadémiai Kiadó, Budapest, Hungary 2011

Abstract In this study, the quantitative structure–property relationship method is applied to predict the enthalpy of fusion of pure chemical compounds at their normal melting point. A genetic algorithm-based multivariate linear regression is used to select the most statistically effective molecular descriptors for evaluating this property. To propose a comprehensive and predictive model, 3,846 pure chemical compounds are investigated. The root mean square of error and the average absolute deviation of the model are equal to 2.57 kJ/mol and 9.7%.

Keywords Enthalpy of fusion · QSPR · Genetic algorithm

Introduction

One of most important physical properties is the enthalpy of fusion at normal melting point ($\Delta_{\text{fus}}H_{\text{tm}}$). The property is

defined as the enthalpy change in the transition from the most stable form of solid to liquid state at the normal melting point.

The $\Delta_{\text{fus}}H_{\text{tm}}$ has many important applications. It is an important property applied in energy balances computations when solid–liquid phase change happens in the chemical or petrochemical processes under study. It is also related to the molecular packing in crystals and can be useful in correcting thermochemical data to a standard state when combined with other thermodynamic properties [1]. It is also used to calibrate the commercially manufactured testing equipment such as the differential scanning calorimeters applied for the determination of the temperature and the amount of energy transfers during phase changes [2]. Another important application of $\Delta_{\text{fus}}H_{\text{tm}}$ would be in estimation of other physical properties. There are several reliable methods developed for estimation of solubility of compounds in various solvents based on $\Delta_{\text{fus}}H_{\text{tm}}$ [3].

There are several methods applied for experimentally measuring the $\Delta_{\text{fus}}H_{\text{tm}}$ that can be categorized into two main groups; calorimetric methods and non-calorimetric methods. Of calorimetric methods, we can refer to adiabatic [4–10], isoperibol [11–16], isothermal [17–20], heat conduction [21–27], drop [2] and differential scanning calorimetry [2], and differential thermal analysis [2]. Of non-calorimetric methods, we can refer to cryoscopic [2], vapor pressure, and enthalpy of solution methods [2].

There are several methods for estimation of $\Delta_{\text{fus}}H_{\text{tm}}$. The first attempt to propose a model for estimation of $\Delta_{\text{fus}}H_{\text{tm}}$ was done by Bondi [28]. He used the relation between enthalpy of fusion ($\Delta_{\text{fus}}H_{\text{tm}}$) and the entropy of fusion at the normal melting point ($\Delta_{\text{fus}}S_{\text{tm}}$). These two are related to each other using the Eq. 1.

Electronic supplementary material The online version of this article (doi:10.1007/s10973-011-1727-y) contains supplementary material, which is available to authorized users.

F. Gharagheizi (✉)
Saman Energy Giti Co., 3331619636 Tehran, Iran
e-mail: fghara@ut.ac.ir; fghara@gmail.com

M. R. S. Gohar
Department of Chemical Engineering, Sharif University
of Technology, 16558-65871, Tehran, Iran

M. R. S. Gohar
Process Department, Sazeh Consultants Engineers, Tehran, Iran

M. G. Vayeghan
Department of Chemistry, Islamic Azad University,
North Tehran Branch, Tehran, Iran

$$\Delta_{\text{fus}}H_{\text{tm}} = t_{\text{m}}\Delta_{\text{fus}}S_{\text{tm}} \quad (1)$$

Bondi [28] proposed application of total entropy of fusion at 0 K ($\Delta_{\text{fus}}S_0^{\text{tot}}$) instead of $\Delta_{\text{fus}}S_{\text{tm}}$ in Eq. 1. The equality of $\Delta_{\text{fus}}S_0^{\text{tot}}$ and $t_{\text{m}}\Delta_{\text{fus}}S_{\text{tm}}$ is true just for those compounds that do not have solid–solid transitions. For the compounds, the Eq. 1 is a good idea to give an estimation for $\Delta_{\text{fus}}H_{\text{tm}}$. However, $\Delta_{\text{fus}}S_0^{\text{tot}}$ is much greater than $\Delta_{\text{fus}}S_{\text{tm}}$ for the compounds that have solid–solid transitions. This idea has been recently applied to estimate the total phase change enthalpy of more than 1,000 pure compounds [1]. In another attempt, Marrero and Gani [29] developed several Group Contribution methods (GC). They proposed a first order, a second order, and a third order group contribution methods to estimate the $\Delta_{\text{fus}}H_{\text{tm}}$. Their third order GC methods showed the best results over 741 compounds they studied. The model showed standard deviation, average absolute error, average absolute deviation of 3.65, 2.17 kJ/mol, and 15.7%.

The quantitative structure–property relationship (QSPR) method was also applied to predict the $\Delta_{\text{fus}}H_{\text{tm}}$. The QSPR-based methods were often used to predict the $\Delta_{\text{fus}}H_{\text{tm}}$ of particular chemical families of compounds [30–34]. These methods are not reviewed in this study because they are proposed for especial purposes and cannot be applied for general compounds.

The GC methods have been used for determination of various physical properties [35–49]. Recently, one of the authors of this paper proposed a new GC type method for determination of the $\Delta_{\text{fus}}H_{\text{tm}}$ [48]. The method is a comprehensive and accurate one, however, it needs a large number of parameters to give an estimation for $\Delta_{\text{fus}}H_{\text{tm}}$ [48].

In this study, the QSPR is applied to develop a comprehensive model for estimation of the enthalpy of fusion of pure compounds at their normal melting points. QSPR implements the chemical structure based parameters called molecular descriptors to develop a model.

Materials and methods

Materials

To develop a comprehensive, it is required to have a large experimental data set. The accuracy and reliability of models for estimation of physical properties, especially those dealing with large number of experimental data, directly depends on the quality and comprehensiveness of the applied data set for its development. The aforementioned characteristics of such a model include both diversity in the investigated chemical families and the number of pure compounds available in the data set. In this study,

the database prepared by Yaws [50] was implemented, which is one of the most comprehensive sources of physical property data for chemical species, e.g., $\Delta_{\text{fus}}H_{\text{tm}}$. The $\Delta_{\text{fus}}H_{\text{tm}}$ for 3,864 compounds found in the database and used as main data set in this study.

Computation of molecular descriptors

In QSPR theory, chemical structure of a compound is encoded into some parameters called “molecular descriptors.” The molecular descriptors are basic molecular properties of a compound [49, 51–72]. Each type of molecular descriptors is related to a specific type of interaction between chemical groups in a particular molecule [49, 51–72]. There are many software packages used for the computation of molecular descriptors of any desired chemical structure. A review of these software packages can be found elsewhere [51]. In this study, one of the most widely used software named “Dragon” is used [73]. This software is able to calculate more than 3,000 molecular descriptors for any desired chemical structure. Since the values of many descriptors are related to the bond lengths, bond angles, etc., each chemical structure is optimized before calculating its molecular descriptors. For doing this, chemical structures of all 3,864 pure compounds have been drawn in Hyperchem software [74] and optimized using the MM+ molecular mechanics force field. Finally, the molecular descriptors have been determined using the Dragon software [73].

Generating model

Having calculated the molecular descriptors from the optimized chemical structures of all investigated compounds, a linear equation is presented that is able to represent/predict the desired property with the least number of variables as well as the highest accuracy [49, 52–72]. In other words, the problem is to find an optimal subset of variables (most statistically effective molecular descriptors on $\Delta_{\text{fus}}H_{\text{tm}}$) from all available variables (all molecular descriptors) that are able to calculate the $\Delta_{\text{fus}}H_{\text{tm}}$ with the least possible deviation from the experimental values. A generally accepted method for this problem is genetic algorithm-based multivariate linear regression (GA-MLR) [49, 52–72, 75–77]. In this method, the genetic algorithm is applied to select best subset of variables based on an objective function as performed firstly by Leardi et al. [76]. Fitness functions such as R^2 , adjusted R^2 , Q^2 , “Akaike” information content (measure of the goodness of fit of an estimated statistical model) etc. are generally applied as objective function in GA-MLR technique [75, 77]. The “RQK” fitness function is a novel one for model searching proposed to avoid undesired model properties such as

chance correlation, presence of noisy variables in the models, and other model pathologies causing lack of model prediction capability [49, 52–72, 75, 77]. Besides, RQK is a constrained fitness function based on Q_{LOO}^2 statistics (leave-one-out cross validated variance) and other four tests that must be fulfilled contemporarily. This function is defined as follows [77]:

$$Q_{\text{LOO}}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{ic})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

where y_i is the $\Delta_{\text{fus}}H_{\text{tm}}$ for i th compound, \bar{y} is mean value of $\Delta_{\text{fus}}H_{\text{tm}}$ for all of the investigated compounds, and \hat{y}_{ic} is response of i th object estimated by a model obtained ignoring the value of the related object. Todeschini et al. [77] proposed that the preceding equation should be subjected to the following constraints:

$$\Delta K = K_{XY} - K_X > 0 \text{ (Quick rule)} \quad (3)$$

$$\Delta Q = Q_{\text{LOO}}^2 - Q_{\text{ASYM}}^2 > 0 \text{ (Asymptotic } Q^2 \text{ rule)} \quad (4)$$

$$R^P > 0 \text{ (Redundancy RP rule)} \quad (5)$$

$$R^N > 0 \text{ (Overfitting RN rule)} \quad (6)$$

It should be noted that the RQK function is used in this study as the fitness function. The results of application of GA-MLR with RQK fitness function have been satisfactory in previous studies [43, 49, 52–72].

Of particular interest is the fact that the main data set should be divided into two sub-data sets before performing the GA-MLR computational steps including the “Training” set and the “Test (prediction)” set. In this article, these sets are defined as follows: the “Training set” is used to generate the model. The “Test set” is used to test the prediction capability of the obtained model. The process of division of main data set into three sub-data sets is performed randomly. For this purpose, about 80 and 20% of the main data set are randomly selected for the “Training” set (3,092 compounds), and the “Test” set (772 compounds). The effect of the allocation percent of the two sub-data sets from the data of main data set on the accuracy of the model has been already discussed in previous studies [53].

Several validation techniques are generally used in the QSPR methods to obtain a valid and reliable model. The most widely used techniques have been presented by Todeschini et al. [75]. The bootstrapping, y-scrambling and external validation techniques are used in this study. Using the bootstrapping technique, the original size of the data set (n) is preserved for the “Training” set by the selection of n objects with repetition. In this procedure, the training set usually consists of repeated objects and the evaluation set of the objects left out [49, 52–72, 75, 77]. The model is calculated on the “Training” set and responses are

predicted on the evaluation set [43, 49, 52–72, 75, 77]. All the squared differences between the true response and the predicted response of the objects of the evaluation set are collected “PRESS”. This procedure of building “Training” sets and evaluation sets is repeated thousands of time. “PRESS” is summed and the predictive capability is calculated [43, 49, 52–72, 75, 77].

The y-scrambling technique is adopted to check the obtained models with chance correlation. This test is performed by calculating the quality of the model (usually the Q^2) modifying the sequence of the response vector by assigning to each object a response, randomly selected from the true responses. If the original model has no chance correlation, there is a significant deference in the quality of the original model and that associated with a model obtained with random responses. The procedure is repeated several hundreds of time [43, 49, 52–72, 75, 77].

External technique is a validation method, where a test is retained to perform a further check on the predictive capabilities of a model obtained from a “Training” set and that optimized by an evaluation set [49, 52–72, 75, 77].

Results and discussion

The most accurate multivariate linear equation is obtained following the presented procedure. For obtaining this equation, the best molecular descriptor model is obtained at the first place. Later, the best two molecular descriptors model are determined [49, 52–72, 75, 77]. This procedure is repeated to achieve the most accurate three, four, five, etc., molecular descriptors models. It is found that the most accurate multivariate linear model has seven parameters because further increase in the number of molecular descriptors does not lead to any considerable effects on the accuracy of the model. The final equation and its statistical parameters are presented as follows:

$$\begin{aligned} \Delta_{\text{fus}}H_{\text{tm}} = & -3.04644(\pm 0.13585) \\ & + 1.53111(\pm 0.00444)\text{Sp} \\ & - 2.3318(\pm 0.05299)\text{GGI1} \\ & + 7.88627(\pm 0.24321)\text{SEige} \\ & - 0.48091(\pm 0.03191)\text{RDF030v} \\ & + 9.41134(\pm 0.3597)\text{nRNH2} \\ & + 5.5341(\pm 0.21958)\text{O} - 057 \\ & + 0.04041(\pm 0.00413)\text{TPSA(Tot)} \end{aligned} \quad (7)$$

$$n_{\text{training}} = 3092; n_{\text{test}} = 772;$$

$$R_{\text{training}}^2 = 0.9883; R_{\text{test}}^2 = 0.9854;$$

$$Q_{\text{LOO}}^2 = 0.9877; Q_{\text{EXT}}^2 = 0.9887; Q_{\text{LTO}}^2 = 0.9804;$$

$$s = 2.57; Q_{\text{BOOT}}^2 = 0.9877; a = -0.012; F = 35772.55$$

RQK function parameters

$$(\Delta K = 0.088; \Delta Q = 0.000; R^P = 0.016; R^N = 0.000)$$

where n_{training} and n_{test} are the numbers of compounds available in training set and test set, respectively. “Sp” is a “constitutional descriptor” defined as sum of atomic polarizabilities (scaled on carbon atom). It is a measure of polarity of a molecule. As expected, when it increases, $\Delta_{\text{fus}}H_{\text{tm}}$ increases. “GGI1” is topological charge index of order 1. “topological charge indices” were proposed to evaluate the charge transfer between pairs of atoms. As stated by Todeschini et al. [75], it is a measure of molecular branching in a molecule. So, increase in molecular branching results decrease in $\Delta_{\text{fus}}H_{\text{tm}}$. “RDF030v” is defined as Radial Distribution Function-3.0/weighted by atomic van der Waals volumes. It is a measure of sphericity of a molecule. The more sphericity in a molecule, the lower $\Delta_{\text{fus}}H_{\text{tm}}$. “nNRNH2” is number of primary amine groups (aliphatic amines). It is a group count descriptor. “O-057” is phenol or enol or carboxyl OH group. It is an atomic fragment. “TPSA(Tot)” is defined as topological polar surface area using N, O, S, P polar contributions. In general, the latter three molecular descriptors (“nNRNH2,” “nNRNH2,” and “TPSA(Tot)”) disclose some sort of hydrogen bonding effects in a molecule. The model shows an $\Delta_{\text{fus}}H_{\text{tm}}$ increase when these three descriptors increase in a molecule.

For more information about procedure of calculation of these molecular descriptors from chemical structure of a compound, please refer to the Dragon software user’s guide [73].

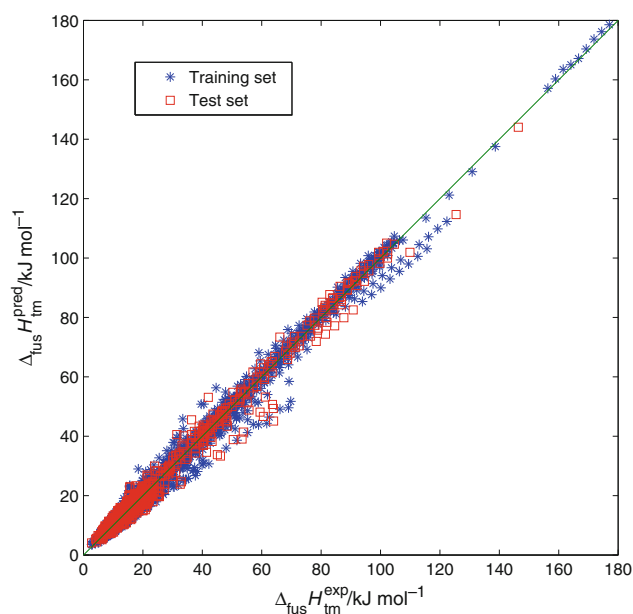


Fig. 1 The comparison between the predicted and experimental $\Delta_{\text{fus}}H_{\text{tm}}$

For testing the validity of the model, bootstrap technique, y-scrambling, and external validation techniques are used [49, 52–72, 75, 77]. The bootstrapping is repeated 5,000 times. Besides, y-scrambling is repeated 300 times. As can be seen, the difference between Q_{LOO}^2 , Q_{BOOT}^2 , Q_{EXT}^2 , and R^2 demonstrates the predictive and accuracy of the proposed model. The intercept value of the y-scrambling technique has low value ($a = -0.011$) that reveals the validity of the model. In addition, the values of four constraints of the model are equal or greater than zero which shows that this model is valid and is not chance correlation.

The predicted $\Delta_{\text{fus}}H_{\text{tm}}$ values by Eq. 7 in comparison with the experimental values [50] are presented in Fig. 1. The predicted $\Delta_{\text{fus}}H_{\text{tm}}$ values for the investigated chemical compounds, the calculated descriptors, and status of all compounds (“Training” or “Test” sets) are presented as supplementary information.

Conclusions

In this study, a QSPR model was presented for determination of the enthalpy of fusion of the chemical compounds at their normal melting points ($\Delta_{\text{fus}}H_{\text{tm}}$). The proposed model is a multivariate linear one consisting seven variables (molecular descriptors), which is developed based on the experimental data of 3,864 chemical compounds. The molecular descriptors were selected using GA-MLR [49, 52–72, 75–77] technique and are calculated based on the chemical structure of molecules. The obtained results show that the presented model is simple, comprehensive, and accurate. Because the model were developed using the largest database of the experimental values [50] of $\Delta_{\text{fus}}H_{\text{tm}}$, the range of application of this model is wide and it may be used for determination of other chemical families excluding those investigated in this study.

References

1. Chickos JS, Acree WE Jr. Total phase change entropies and enthalpies. An update on fusion enthalpies and their estimation. *Thermochim Acta*. 2009;495(1–2):5–13. doi:10.1016/j.tca.2009.05.008.
2. Marsh KN, Hall KR, Frenkel ML. *Thermodynamic properties of organic compounds and their mixtures*. Berlin: Springer; 1995.
3. Jouyban A. *Handbook of solubility data for pharmaceuticals*. Boca Raton: CRC Press; 2010.
4. Xu F, Sun LX, Tan ZC, Liang JG, Zhang T. Adiabatic calorimetry and thermal analysis on acetaminophen. *J Therm Anal Calorim*. 2006;83(1):187–91. doi:10.1007/s10973-005-6969-0.
5. Gonzales C, Sempere J, Nomen DR, Waldram S. Adiabatic calorimetry using directly agitated test cells. *J Therm Anal Calorim*. 1999;58(1):183–91.

6. Maschio G, Feliu JA, Ligthart J, Ferrara I, Bassani C. The use of adiabatic calorimetry for the process analysis and safety evaluation in free radical polymerization. *J Therm Anal Calorim.* 1999;58(1):201–14.
7. Sempere J, Nomen R, Serra R, Gallice F. Determination of activation energies by using different factors Phi in adiabatic calorimetry. *J Therm Anal Calorim.* 1999;58(1):215–23.
8. van Miltenburg JC, Mathot VBF, van Ekeren PJ, Ionescu LD. Adiabatic calorimetry of a very low density polyethylene copolymer. *J Therm Anal Calorim.* 1999;56(3):1017–23.
9. van Ekeren PJ, Ionescu LD, Mathot VBF, van Miltenburg JC. Heat capacities and thermal properties of a homogeneous ethylene-1-octene copolymer by adiabatic calorimetry. *J Therm Anal Calorim.* 2000;59(3):683–97.
10. Song YJ, Meng SH, Wang FD, Sun CX, Tan ZC. A study on the thermodynamic properties of polyimide BTDA-ODA by adiabatic calorimetry and thermal analysis. *J Therm Anal Calorim.* 2002;69(2):617–25.
11. El-Bushra SE. Construction of an isoperibol calorimeter to measure the specific heat capacity of foods between 20 and 90°C. *J Therm Anal Calorim.* 2001;64(1):261–72.
12. Schaffer B, Lorinczy D. Isoperibol calorimetry as a tool to evaluate the impact of the ratio of exopolysaccharide-producing microbes on the properties of sour cream. *J Therm Anal Calorim.* 2005;82(2):537–41. doi:10.1007/s10973-005-0929-6.
13. Wurster DE, Bandopadhyay R. The use of isoperibol calorimetry for the study of age-related changes in glucose uptake by erythrocytes. *J Therm Anal Calorim.* 2005;79(3):737–40.
14. Santos L, Silva MT, Schroder B, Gomes L. Labtermo: methodologies for the calculation of the corrected temperature rise in isoperibol calorimetry. *J Therm Anal Calorim.* 2007;89(1):175–80.
15. Vargas EF, Moreno JC, Forero J, Parra DF. A versatile and high-precision solution-reaction isoperibol calorimeter. *J Therm Anal Calorim.* 2008;91(2):659–62. doi:10.1007/s10973-007-7613-y.
16. Ellaite M, Dalmazzone D. Modernization and validation of an isoperibol rotating bomb calorimeter for the measurement of energies of combustion of sulphur compounds. *J Therm Anal Calorim.* 2010;99(3):939–45. doi:10.1007/s10973-009-0555-9.
17. Baert G, Hoste S, De Schutter G, De Belie N. Reactivity of fly ash in cement paste studied by means of thermogravimetry and isothermal calorimetry. *J Therm Anal Calorim.* 2008;94(2):485–92. doi:10.1007/s10973-007-8787-z.
18. Razafindralambo H, Dufour S, Paquot M, Deleu M. Thermodynamic studies of the binding interactions of surfactin analogues to lipid vesicles. Application of isothermal titration calorimetry. *J Therm Anal Calorim.* 2009;95(3):817–21. doi:10.1007/s10973-008-9403-6.
19. Taheri-Kafrani A, Bordbar AK. Energetics of micellization of sodium *n*-dodecyl sulfate at physiological conditions using isothermal titration calorimetry. *J Therm Anal Calorim.* 2009;98(2):567–75. doi:10.1007/s10973-009-0170-9.
20. Behbehani GR, Saboury AA, Barzegar L, Zarean O, Abedini J, Payehghdr M. A thermodynamic study on the interaction of nickel ion with myelin basic protein by isothermal titration calorimetry. *J Therm Anal Calorim.* 2010;101(1):379–84. doi:10.1007/s10973-009-0596-0.
21. Ageeva T, Utzig E, Golubchikov O, Zielenkiewicz W. Thermokinetic studies of porphyrin complexation in non-aqueous solutions by heat conduction microcalorimetry. *J Therm Anal Calorim.* 1998;54(1):243–8.
22. Socorro F, de Rivera MR. Micro-effects on continuous-injection heat conduction calorimetry. *J Therm Anal Calorim.* 1998;52(3):729–37.
23. Torra V, Tachoire H. Conduction calorimeters heat transmission systems with uncertainties. *J Therm Anal Calorim.* 1998;52(3):663–81.
24. Utzig E. Heat conduction microcalorimeter for thermokinetics and titration experiments. *J Therm Anal Calorim.* 1998;54(1):391–7.
25. Zielenkiewicz W, Kaminski M. A conduction calorimeter for measuring the heat of cement hydration in the initial hydration period. *J Therm Anal Calorim.* 2001;65(2):335–40.
26. O'Neill MAA, Beezer AE, Morris AC, Urakami K, Willson RJ, Connor JA. Solid-state reactions from isothermal heat conduction microcalorimetry—theoretical approach and evaluation via simulated data. *J Therm Anal Calorim.* 2003;73(2):709–14.
27. Shao YH, Ren XN, Liu ZR. Studies of mechanism of silica polymerization reactions in the combination of silica sol and potassium sodium waterglass via isothermal heat conduction microcalorimetry. *J Therm Anal Calorim.* 2010;101(3):1135–41. doi:10.1007/s10973-010-0697-9.
28. Bondi AA. Physical properties of molecular crystals, liquids, and glasses. Wiley Series on the Science and Technology of Materials. New York: Wiley; 1968.
29. Marrero J, Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* 2001;183–184:183–208.
30. Dyekjaer JD, Jonsdottir SO. QSPR models based on molecular mechanics and quantum chemical calculations. 2. Thermodynamic properties of alkanes, alcohols, polyols, and ethers. *Ind Eng Chem Res.* 2003;42(18):4241–59.
31. Puri S, Chickos JS, Welsh WJ. Three-dimensional quantitative structure-property relationship (3D-QSPR) models for prediction of thermodynamic properties of polychlorinated biphenyls (PCBs): enthalpies of fusion and their application to estimates of enthalpies of sublimation and aqueous solubilities. *J Chem Inf Comput Sci.* 2003;43(1):55–62. doi:10.1021/ci0200164.
32. Kim CK, Lee KA, Hyun KH, Park HJ, Kwack IY, Lee HW, et al. Prediction of physicochemical properties of organic molecules using van der Waals surface electrostatic potentials. *J Comput Chem.* 2004;25(16):2073–9. doi:10.1002/jcc.20129.
33. Goodarzi M, Chen T, Freitas MP. QSPR predictions of heat of fusion of organic compounds using Bayesian regularized artificial neural networks. *Chemom Intell Lab Syst.* 2010;104(2):260–4. doi:10.1016/j.chemolab.2010.08.018.
34. Zong L, Ramanathan S, Chen CC. Predicting thermophysical properties of mono- and diglycerides with the chemical constituent fragment approach. *Ind Eng Chem Res.* 2010;49(11):5479–84. doi:10.1021/ie901948v.
35. Gharagheizi F. New neural network group contribution model for estimation of lower flammability limit temperature of pure compounds. *Ind Eng Chem Res.* 2009;48(15):7406–16. doi:10.1021/ie9003738.
36. 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. doi:10.1016/j.jhazmat.2009.05.023.
37. Gharagheizi F. An accurate model for prediction of autoignition temperature of pure compounds. *J Hazard Mater.* 2011;189(1–2):211–21. doi:10.1016/j.jhazmat.2011.02.014.
38. Gharagheizi F, Abbasi R. A new neural network group contribution method for estimation of upper flash point of pure chemicals. *Ind Eng Chem Res.* 2010;49(24):12685–95. doi:10.1021/ie1011273.
39. 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):10149–52. doi:10.1021/ie101532e.
40. 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):6503–7. doi:10.1021/ie2001764.

41. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Representation/prediction of solubilities of pure compounds in water using artificial neural network-group contribution method. *J Chem Eng Data*. 2011;56(4):720–6. doi:10.1021/je101061t.
42. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Determination of parachor of various compounds using an artificial neural network-group contribution method. *Ind Eng Chem Res*. 2011;50(9):5815–23. doi:10.1021/ie102464t.
43. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. QSPR approach for determination of parachor of non-electrolyte organic compounds. *Chem Eng Sci*. 2011;66(13):2959–67. doi:10.1016/j.ces.2011.03.039.
44. Gharagheizi F, Eslamimanesh A, Mohammadi AH, 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):1741–50. doi:10.1021/je101190p.
45. Gharagheizi F, Eslamimanesh A, Mohammadi AH, 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):2460–76. doi:10.1021/je200019g.
46. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Use of artificial neural network-group contribution method to determine surface tension of pure compounds. *J Chem Eng Data*. 2011;56(5):2587–601. doi:10.1021/je2001045.
47. 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):2482–6. doi:10.1021/ie101672j.
48. Gharagheizi F, Salehi GR. Prediction of enthalpy of fusion of pure compounds using an artificial neural network-group contribution method. *Thermochim Acta*. doi:10.1016/j.tca.2011.04.001.
49. Gharagheizi F. Quantitative structure-property relationship for prediction of the lower flammability limit of pure compounds. *Energy Fuels*. 2008;22(5):3037–9. doi:10.1021/ef800375b.
50. Yaws CL. Yaws' handbook of thermodynamic and physical properties of chemical compounds. New York: Knovel; 2003.
51. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics. 2nd ed. Revised and enlarged edition. Weinheim: Wiley-VCH; Chichester: John Wiley [distributor]; 2009.
52. Gharagheizi F. A new accurate neural network quantitative structure-property relationship for prediction of theta (lower critical solution temperature) of polymer solutions. *E-Polymers*; 2007; no. 114.
53. Gharagheizi F. QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN. *Comput Mater Sci*. 2007;40(1):159–67.
54. Gharagheizi F, Mehrpooya M. Prediction of standard chemical energy by a three descriptors QSPR model. *Energy Convers Manag*. 2007;48(9):2453–60.
55. Vatani A, Mehrpooya M, Gharagheizi F. Prediction of standard enthalpy of formation by a QSPR model. *Int J Mol Sci*. 2007; 8(5):407–32.
56. Gharagheizi F. A simple equation for prediction of net heat of combustion of pure chemicals. *Chemom Intell Lab Syst*. 2008; 91(2):177–80.
57. Gharagheizi F. A new molecular-based model for prediction of enthalpy of sublimation of pure components. *Thermochim Acta*. 2008;469(1–2):8–11.
58. Gharagheizi F. QSPR studies for solubility parameter by means of genetic algorithm-based multivariate linear regression and generalized regression neural network. *QSAR Comb Sci*. 2008; 27(2):165–70.
59. Gharagheizi F, Alamdari RF. Prediction of flash point temperature of pure components using a quantitative structure–property relationship model. *QSAR Comb Sci*. 2008;27(6):679–83.
60. Gharagheizi F, Alamdari RF. A molecular-based model for prediction of solubility of C60 fullerene in various solvents. *Fuller Nanotub Carbon Nanostruct*. 2008;16(1):40–57.
61. Gharagheizi F, Fazzeli A. Prediction of the Watson characterization factor of hydrocarbon components from molecular properties. *QSAR Comb Sci*. 2008;27(6):758–67.
62. Gharagheizi F, Mehrpooya M. Prediction of some important physical properties of sulfur compounds using quantitative structure-properties relationships. *Mol Divers*. 2008;12(3–4):143–55.
63. Sattari M, Gharagheizi F. Prediction of molecular diffusivity of pure components into air: a QSPR approach. *Chemosphere*. 2008; 72(9):1298–302.
64. 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):217–20.
65. Gharagheizi F. Prediction of upper flammability limit percent of pure compounds from their molecular structures. *J Hazard Mater*. 2009;167(1–3):507–10.
66. Gharagheizi F. Prediction of the standard enthalpy of formation of pure compounds using molecular structure. *Aust J Chem*. 2009;62(4):376–81.
67. 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):267–85.
68. Gharagheizi F, Sattari M. Prediction of the theta (UCST) of polymer solutions: a quantitative structure-property relationship study. *Ind Eng Chem Res*. 2009;48(19):9054–60.
69. 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): 1678–82.
70. Gharagheizi F. Chemical structure-based model for estimation of the upper flammability limit of pure compounds. *Energy Fuels*. 2010;24(7):3867–71.
71. Gharagheizi F, Sattari M. Prediction of triple-point temperature of pure components using their chemical structures. *Ind Eng Chem Res*. 2010;49(2):929–32.
72. Mehrpooya M, Gharagheizi F. A molecular approach for the prediction of sulfur compound solubility parameters. *Phosphorus Sulfur Silicon Relat Elem*. 2010;185(1):204–10.
73. Talete srl, Dragon for Windows Software for molecular descriptor calculation, version 5.4. 2007.
74. Hyperchem Release 7.5 for Windows; Gainesville: Hypercube, Inc. 2002.
75. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics. 2nd ed. Revised and enlarged edition. Weinheim: Wiley-VCH; Chichester: John Wiley; 2009.
76. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *J Chemom*. 1992;6:267–81.
77. Todeschini R, Consonni V, Mauri A, Pavan M. Detecting “bad” regression models: multicriteria fitness functions in regression analysis. *Anal Chim Acta*. 2004;515(1):199–208.