ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/231376571

Prediction of Crystal Lattice Energy Using Enthalpy of Sublimation: A Group Contribution-Based Model

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · JANUARY 2011

Impact Factor: 2.59 · DOI: 10.1021/ie101672j

CITATIONS
25

READS
38

3 AUTHORS, INCLUDING:

Farhad Gharagheizi
Texas Tech University
168 PUBLICATIONS **2,911** CITATIONS

SEE PROFILE

Mehdi Sattari
University of KwaZulu-Natal
40 PUBLICATIONS **325** CITATIONS

SEE PROFILE

Available from: Mehdi Sattari
Retrieved on: 30 January 2016

I&EC research
Industrial & Engineering Chemistry Research

CORRELATION

pubs.acs.org/IECR

Prediction of Crystal Lattice Energy Using Enthalpy of Sublimation: A Group Contribution-Based Model

Farhad Gharagheizi,*,† Mehdi Sattari,† and Behnam Tirandazi‡

†Saman Energy Giti Co., Postal Code: 3331619636, Tehran, Iran

‡Department of Chemical Engineering, Iran University of Science and Technology,Tehran, Iran

Supporting Information

ABSTRACT: In this study, a new group contribution-based model is presented to predict the enthalpy of sublimation of pure compounds. This model can also be used to predict the lattice crystal energy of such compounds. The model is a neural network using the number of occurrences of 172 chemical groups on the chemical structures of pure compounds to predict the enthalpy of sublimation. This comprehensive model is generated using a large data set of pure compounds (1384 pure compounds). The squared correlation coefficient, average percent error, and root-mean-square error of the model over all investigated compounds are 0.9854, 3.54%, and 4.21, respectively.

1. INTRODUCTION

The crystal lattice energy (E_{latt}) is the energy of formation of a crystal from the isolated (gas phase) molecules. The lattice energy of molecular crystals is calculated using the enthalpy of sublimation (ΔH_{sub}) as follows:[1−4]

$$E_{latt} = -\Delta H_{sub} - 2RT \quad (1)$$

where R and T are the universal gas constant and temperature, respectively.

It is clear that the ΔH_{sub} can be experimentally measured, but this is literally a costly and also time-consuming task. Prediction of values of this parameter would be of considerable use, especially without having to synthesize the compound, because it quantifies the strength of the intermolecular interactions in the crystal structure. There are various methods that can empirically and theoretically calculate the crystal lattice energy of a given experimental or hypothetical crystal structure using potential energy functions. The main goal of such calculations is generally the prediction of crystal structure. These methods determine the global minimum of the crystal lattice energy or the global minimum of the sum of the lattice and intermolecular energies for nonrigid molecules, which is assumed to correspond to the most favorable crystal packing.[5−8] Since calculation of the lattice energy by such methods needs both significant computational resources and knowledge of crystal structure, a simpler alternative approach is desired. In many cases, the difference between the calculated lattice energy of the experimental crystal structure and those calculated for several hypothetical alternative structures is not significant.[7] Furthermore, the experimental crystal structure is not often the global minimum of the crystal lattice energy; thus, other polymorphs are experimentally found. Since the observations show that different crystal packings have values within a range of 5−10 kJ/mol,[14] it is logical to predict the crystal lattice energy only based on the enthalpy of sublimation, which is independent of any knowledge about crystal packing. This result leads to development of a model for prediction of the crystal lattice energy of organic compounds through prediction of the enthalpy of sublimation.

So far, several methods have been proposed for estimation of the enthalpy of sublimation of pure components.[8−15] The methods can be divided into two main types, based on their application limit. The first type are those methods that have been generated using data sets containing a few particular chemical families of pure components.[8−13] Therefore, application of the methods is limited to these particular families of pure components. The second type are those methods generated using wide data sets of diverse chemical families of pure components.[14,15] Since the second type covers more chemical families of pure compounds, it is a more general model than the first type.

A comparison among previously presented methods for estimation of the enthalpy of sublimation shows that the more accurate estimations require the more complex computational procedures.

The main aim of this study is to develop a new simple and comprehensive molecular based model for prediction of the enthalpy of sublimation of pure components. The model can be used to calculate the lattice crystal energy of these components.

2. GENERATING MODEL STEPS

Providing Data Set. The first step in developing a new molecular-based model is to identify a data set for the property. The data set should contain a large number of pure components and be comprehensive in the diversity of chemical compositions that are included. These two factors enable the development of a comprehensive model that has a wide range of applicability. For this purpose, the DIPPR 801[16] database was selected to provide a main data set for ΔH_{sub} values of pure components (at triple point condition). The database is the result of a vast literature survey containing more than 20,000 publications.

Using the database, 1348 pure components were found and used as the main data set in this study. The main data set is presented as Supporting Information.

Received: August 5, 2010
Accepted: December 17, 2010
Revised: October 22, 2010
Published: January 24, 2011

ACS Publications © 2011 American Chemical Society

Providing the Collection of Chemical Groups. In this step, the chemical structures of all pure components of the data set were analyzed. Later, 172 chemical groups were selected to predict the ΔH_{sub} values. The chemical groups used in this work are extensively presented in Table S1 as Supporting Information.

The computed number of occurrences of the chemical groups in the investigated components is also presented as Supporting Information.

Developing the Model. Analysis of the structures present in the data set resulted in selection of 172 chemical groups. It should be noted that these 172 chemical groups are not simultaneously present in real pure components. Only a few of these chemical groups are present in a regular pure component at the same time. Thus, the computation of the parameters in a pure compound is very simple.

There are several choices of how to identify a relationship between the 172 chemical groups and the ΔH_{sub}. The simplest way is to assume a multilinear correlation between all of the chemical groups and the ΔH_{sub}. Later, using the well-known least-squares method, we can determine the parameters of the model. These parameters can then be tabulated and used as the group contribution for each chemical group. A similar method is applied in the classic group contribution (GC) method. Although application of this procedure showed satisfactory results for many physical properties, always assuming a linear relation between chemical groups and the property of interest is doubtful. Unsuccessful results in applying the classic GC method to estimate many physical properties demonstrate the idea.

To overcome this inconsistency of the classic GC method, it is assumed that there is a nonlinear relation between the chemical groups and the proposed property. There are many mathematical methods which are able to nonlinearly relate chemical groups to the property. The most widely used methods are artificial neural networks.

Neural networks are extensively used in various scientific and engineering areas, such as predictions of physical and chemical properties.[17] These powerful tools are usually applied to study the complicated systems such as the problem here. Theoretical explanations of neural networks can be found elsewhere.[18]

One of the most common types of neural networks is three-layer feed forward ones. The schematic structure of a three-layer feed forward neural network is presented in Figure 1. The parameters used to train the three-layer neural network are presented in Table 1.

More detailed explanations about the three-layer feed forward neural networks used in this study can be found elsewhere.[19−37]

The Neural Network toolbox in MATLAB software is used in this work. In the first step of the computations, the main data set is divided into three new subdata sets. These sets include the following: a training set, a validation set, and a test set. The training set is used to generate the neural network. The validation set is used to optimize the neural network. The validation process is done to determine the parameters of the neural network using

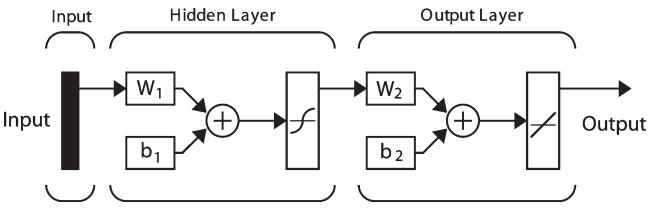

Figure 1. Schematic structure of the three-layer feed forward neural network used in this study.

Table 1. Parameters Used While Training the Three-Layer Feed Forward Neural Network

neural network parameters	status
second layer transfer function	tansig[a]
third layer transfer function	linear
number of epochs	100
performance goal	0
maximum validation failures	5
factor to use for memory/speed trade-off	1
minimum performance gradient	1.00×10^{-10}
initial μ	0.001
μ decrease factor	0.1
μ increase factor	10
maximum μ	1.00×10^{10}

[a] The tansig transfer function is the hyperbolic tangent.

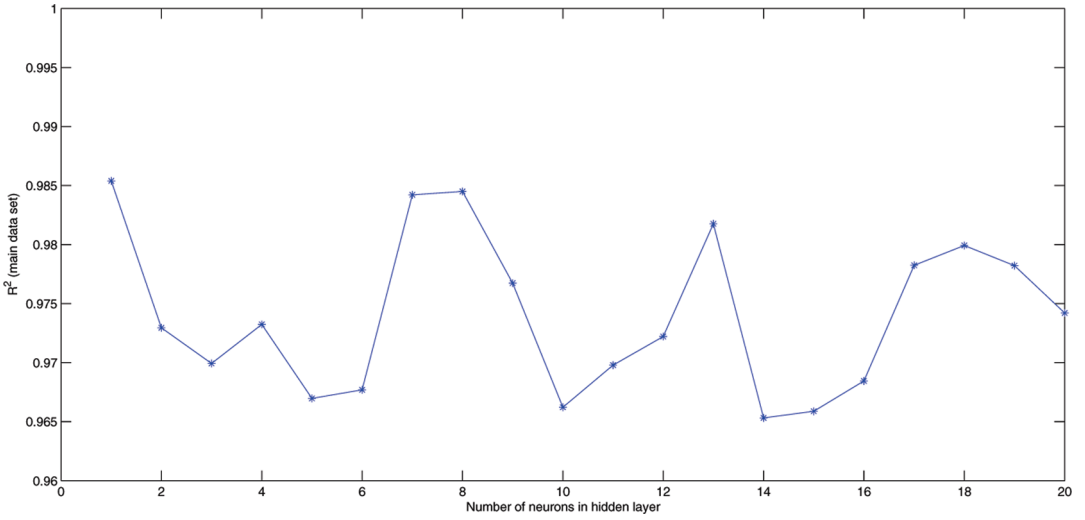

Figure 2. R^2 changes along with hidden layer number of neurons in the three-layer feed forward neural network used in this study.

dx.doi.org/10.1021/ie101672j |*Ind. Eng. Chem. Res.* 2011, 50, 2482–2486

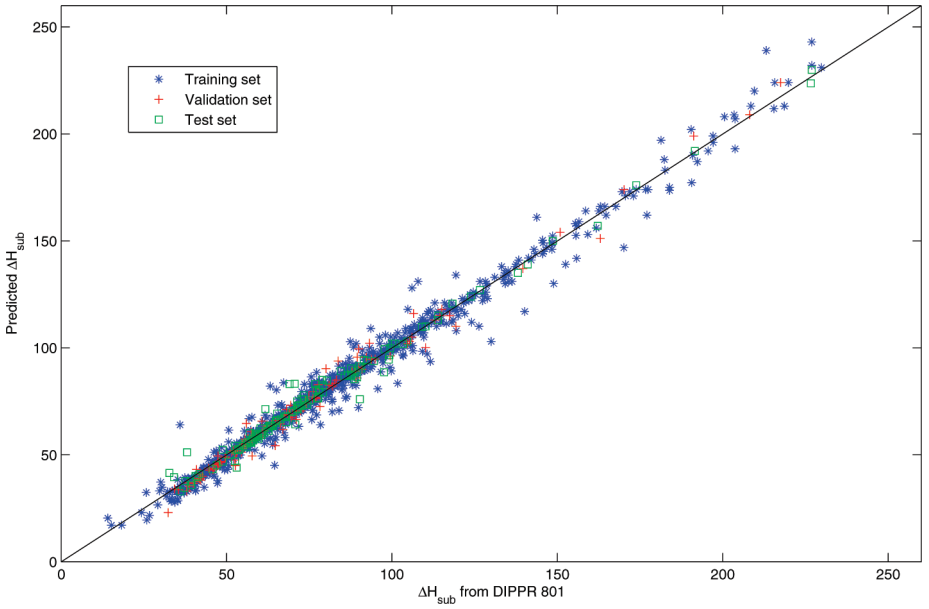

Figure 3. Comparison between the predicted and experimental ΔH_{sub} values.

optimization methods. The test set is used to test the predictive capability of the model obtained. The process of division of the main data set into three new data sets is usually done randomly. For this purpose, 80%, 10%, and 10% of the main data set were assigned to the training set (1080 pure components), the validation set (134 pure components), and the test set (134 pure components). The effect of these percentages on the accuracy of the neural networks has already been studied by the author.[31]

Generating a neural network requires determination of the weight matrices and bias vectors. As shown in Figure 1, there are two weight matrices and two bias vectors in a three-layer feed forward neural network (W_1 and W_2, b_1 and b_2). These parameters should be obtained by minimization of an objective function. The objective function is written as the sum of squares of errors between the outputs of the neural network (the predicted ΔH_{sub}) and the target values (the experimental ΔH_{sub}). This minimization is performed by employing the Levenberg—Marquardt algorithm. This algorithm is rapid and accurate in the process of training neural networks.[17]

3. RESULTS AND DISCUSSION

A three-layer feed forward neural network was obtained for prediction of the ΔH_{sub} using the procedure presented. For determination of the number of neurons in the hidden layer of the neural network, numbers 1 through 20 were tested. Number 1 showed the best results. Therefore, the best three-layer feed forward neural network has a structure of 180-1-1. The accuracy of the three-layer feed forward neural network using hidden layer numbers of neurons ranges from 1 to 20 is shown in Figure 2. As can be found, the best squared correlation coefficient (R^2) would be obtained using a 180-1-1 structured three-layer feed forward neural network. The mat file (MATLAB file format) of the neural network containing all parameters of the obtained model is available as Supporting Information by e-mail from the author.

The predicted values of ΔH_{sub} using the model in comparison with the experimental data are presented in Figure 3. The values of the predicted ΔH_{sub} as well as the status of each pure component in the model (belonging to the training set, the validation set, or the test set) are presented as Supporting Information.

The statistical parameters of the model are presented in Table S2 of the Supporting Information. These results show that the squared correlation coefficient, average percent error, and root-mean-square error of the model over the training set, the validation set, the test set, and the main data set are calculated as 0.9848, 0.9881, 0.9883, 0.9854, 3.53%, 3.95%, 3.31%, 3.54%, 4.34, 3.75, 3.54, and 4.21, respectively.

The average percent deviations of the model for each one of the chemical families of pure compounds used in this study are shown in Table S3 of the Supporting Information.

As can be seen, the largest deviations are related to the cycloaliphatic alcohols, inorganic gases, and elements. This case shows that the model is not able to predict the ΔH_{sub} values for these three chemical categories as well as other chemical ones. The main reason for this weakness is probably due to the number of pure components available in these three data sets. Among the 1348 pure components, only 11 pure compounds belong to these three chemical families. It seems the 11 pure compounds are too small a number to modify the behavior of the model to predict ΔH_{sub} of these three chemical families. It seems the modification of the model needs more experimental data.

Rice et al.[10] presented a molecular-based model to estimate enthalpy of sublimation using the properties associated with quantum mechanically determined electrostatic potentials of isolated molecules. The root mean squares of error and the maximum deviation of their model over 35 pure components are respectively 15 and 52 kJ/mol. Politzer et al.,[9] Matheieu and Simonetti,[10] and Kim et al.[11] used different modifications of van der Waals electrostatic surface potentials and their derived properties to correlate the enthalpy of sublimation. Their models showed good results over 34 pure components used in their studies. Ouvrad and Mitchell[15] presented a simple model using the numbers of occurrences of different atom types as descriptors for prediction of the enthalpy of sublimation. The squared correlation coefficients of their model over 226 pure components used as training set, and 35 pure components as test set, are 0.925 and 0.937. Politzer et al.[13] presented a model to estimate the enthalpy of sublimation of pure components on the basis of the calculated electrostatic potential on the molecular surface. The average

absolute deviation of their correlation over 66 pure components is 11.7 kJ/mol. Recently, Byrd and Rice[14] presented a model to estimate the enthalpy of sublimation using quantum chemical data. The root mean squares of error and the maximum deviation of their model over 35 pure components are respectively 12.5 and 217.7 kJ/mol. Finally, Gharagheizi[16] used a linear quantitative structure property relationship to predict the enthalpy of sublimation using a data set containing 1348 compounds. The squared correlation coefficient, root-mean-square error, and maximum absolute deviation obtained by the model were 0.9746, 5.46, and 27.56 kJ/mol, respectively. To sum up, the model presented in this study shows it is more comprehensive and more accurate than the others.

4. CONCLUSION

A molecular-based model is presented for prediction of the ΔH_{sub} values of pure components using a wide data set of 1348 pure components containing a set of diverse chemical families. The parameters of the model are the numbers of occurrences of 172 chemical groups in the structures investigated. These chemical groups are easily defined by chemical structure. The model is a group contribution-based neural network that can accurately predict the ΔH_{sub} values of pure components.

The data set used in this work covers the majority of chemical families. Therefore, the model has a wide range of applicability; however, application of the model is restricted to those components similar to the investigated ones. Application of the model for components other than those investigated is not recommended.

APPENDIX

The model is very easy to apply. What is needed is to just drag and drop the mat file (freely available upon request) into the MATLAB environment (any version) workspace. One can follow the below example to get a response from the model step by step:

Assume that one is willing to predict the enthalpy of sublimation of 1,1,2-trichloroethane using the developed model. First of all, the group-contribution parameters should be computed from the chemical structure of 1,1,2-trichloroethane. After dragging and dropping the mat file, the following commands should be entered in the MATLAB workspace:

```
GC = [2 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      1 0 0 0 1 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 1 0 0 1 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 2 ...
      1 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0 0 0 0 0 0 1 2 0 0 ...
      0 0 0 0 0 0 0 0 0 0 0 0 ...
      0 0 0];
Enthalpy_S = sim(net,GC')
```

Therefore, one will observe the following estimated ΔH_{sub}: 49.1730.

The experimental value for this ΔH_{sub} is equal to 47.9 kJ/mol. Therefore, the absolute relative deviation of the calculated value from the experimental one is 2.66%. Using this value and eq 1, the crystal lattice energy can be estimated.

ASSOCIATED CONTENT

Supporting Information. Tables S1−S3 and the model parameters and results. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Fax: +98 21 77926580. E-mail: fghara@ut.ac.ir and fghara@gmail.com.

ACKNOWLEDGMENT

The author gratefully thanks Mr Ali Eslamimanesh from CEP/TEP, Mines ParisTech for useful comments on the manuscript.

REFERENCES

(1) Bisker-leib, V.; Doherty, M. F. Modeling the Crystal Shape of Polar Organic Materials: Prediction of Urea Crystals Grown from Polar and Nonpolar Solvents. *Cryst. Growth Des.* **2001**, *1*, 455.

(2) Hammond, R. B.; Pencheva, K.; Roberts, K. J. Molecular Modeling of Crystal-Crystal Interactions between the α- and β-Polymorphic Forms of L-Glutamic Acid Using Grid-Based Methods. *Cryst. Growth Des.* **2007**, *7*, 875.

(3) Bisker-leib, V.; Doherty, M. F. Modeling Crystal Shape of Polar Organic Materials: Applications to Amino Acids. *Cryst. Growth Des.* **2003**, *3*, 221.

(4) Hammond, R. B.; Pencheva, K.; Roberts, K. J. A Structural-Kinetic Approach to Model Face-Specific Solution/Crystal Surface Energy Associated with the Crystallization of Acetyl Salicylic Acid from Supersaturated Aqueous/Ethanol Solution. *Cryst. Growth Des.* **2006**, *6*, 1324.

(5) Pertsin, A. J.; Kitaigorodsky, A. I. In *The Atom-Atom Potential Method*; Springer Verlag: Berlin, 1986.

(6) Gavezzotti, A.; Filippini, G. In *Theoretical Aspects and Computer Modeling*; Gavezzotti, A., Eds.; J. Wiley and Sons: Chichester, 1997; Chapter 3, pp 61−97.

(7) Beyer, T.; Lewis, T.; Price, S. L. Which organic crystal structures are predictable by lattice energy minimisation? *Cryst. Eng. Comm.* **2001**, *3*, 178.

(8) Gavezzotti, A. Ten years of experience in polymorph prediction: what next? *Cryst. Eng. Comm.* **2002**, *4*, 343.

(9) Politzer, P.; Murray, J. S.; Grice, M. E.; Deslvo, M.; Miller, E. Calculation of heats of sublimation and solid phase heats of formation. *Mol. Phys.* **1997**, *91*, 923.

(10) Rice, B. M.; Pai, S. V.; Hare. Predicting Heats of Formation of Energetic Materials Using Quantum Mechanical Calculations *J. Combust. Flame* 1999, *118*, 445

(11) Mathieu, D.; Somonetti, P. Evaluation of solid-state formation enthalpies for energetic materials and related compounds. *Thermochim. Acta* **2002**, *384*, 369.

(12) Kim, C. Y.; Lee, K. A.; Hyun, K. H.; Park, J.; Kwack, I. Y.; Kim, C. K.; Lee, H. W.; Lee, B. S. Prediction of Physicochemical Properties of Organic Molecules Using van der Waals Surface Electrostatic Potentials. *J. Comput. Chem.* **2004**, *25*, 2073.

(13) Politzer, P.; Ma, Y.; Lane, P.; Concha, M. C. Computational Prediction of Standard Gas, Liquid, and Solid-Phase Heats of Formation

and Heats of Vaporization and Sublimation. *Int. J. Quantum Chem.* **2005**, *105*, 341.

(14) Byrd, E. F. C.; Rice, B. M. Improved Prediction of Heats of Formation of Energetic Materials Using Quantum Mechanical Calculations. *J. Phys. Chem. A* **2006**, *110*, 1005.

(15) Ouvrad, C.; Mitchell, J. B. O. Can we predict lattice energy from molecular structure? *Acta Crystallogr., B* **2003**, *59*, 676.

(16) Gharagheizi, F. A new molecular-based model for prediction of enthalpy of sublimation of pure components. *Thermochim. Acta* **2008**, *469*, 8.

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

(18) Taskinen, J.; Yliruusi, J. Prediction of physicochemical properties based on neural network modeling. *Adv. Drug Delivery Rev.* **2008**, *55*, 1163.

(19) Hagan, M.; Demuth, H. B.; Beale, M. H. *Neural Network Design*; Thomson International Publishing: 2002.

(20) Gharagheizi, F. A New Group Contribution-Based Method for Estimation of Lower Flammability Limit of Pure Compounds. *J. Hazard Mater.* **2009**, *170*, 595.

(21) Gharagheizi, F. New Neural Network Group Contribution Model for Estimation of Lower Flammability Limit Temperature of Pure Compounds. *Ind. Eng. Chem. Res.* **2009**, *48*, 7406.

(22) Gharagheizi, F. Prediction of Standard Enthalpy of Formation of Pure Compounds Using Molecular Structure. *Aust. J. Chem.* **2009**, *62*, 376.

(23) 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*, 1678.

(24) Gharagheizi, F.; Mehrpooya, M. Prediction of Some Important Physical Properties of Sulfur Compounds Using QSPR Models. *Mol. Diversity* **2008**, *12*, 143.

(25) Sattari, M.; Gharagheizi, F. Prediction of Molecular Diffusivity of Pure Components into Air: A QSPR Approach. *Chemosphere* **2008**, *72*, 1298.

(26) Gharagheizi, F.; Alamdari, R. F.; Angaji, M. T. A New Neural Network-Group Contribution Method for Estimation of Flash Point. *Energy Fuels* **2008**, *22*, 1628.

(27) Gharagheizi, F.; Fazeli, A. Prediction of Watson Characterization Factor of Hydrocarbon Compounds from Their Molecular Properties. *QSAR Comb. Sci.* **2008**, *27*, 758.

(28) Gharagheizi, F.; Alamdari, R. F. A Molecular-Based Model for Prediction of Solubility of C_{60} Fullerene in Various Solvents, Fullerenes. *Nanotubes Carbon Nanostruct.* **2008**, *16*, 40.

(29) Gharagheizi, F. A New Neural Network Quantitative Structure-Property Relationship for Prediction of (Lower Critical Solution Temperature) of Polymer Solutions. *e-Polym.* **2007**, Article Number 114.

(30) 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*, 165.

(31) Gharagheizi, F. QSPR Analysis for Intrinsic Viscosity of Polymer Solutions by means of GA-MLR and RBFNN. *Comput. Mater. Sci.* **2007**, *40*, 159.

(32) Vatani, A.; Mehrpooya, M.; Gharagheizi, F. Prediction of Standard Enthalpy of Formation by a QSPR Model. *Int. J. Mol. Sci.* **2007**, *8*, 407.

(33) Mehrpooya, M.; Gharagheizi, F. A Molecular Approach for Prediction of Sulfur Compounds Solubility Parameters. *Phosphorus Sulfur Silicon Relat. Elem.* **2010**, *185*, 204.

(34) Gharagheizi, F. A Chemical Structure-Based Model for Estimation of Upper Flammability Limit of Pure Compounds. *Energy Fuels* **2010**, *24*, 3867.

(35) 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*, 10149–10152.

(36) Gharagheizi, F.; Abbasi, R. A Group Contribution-Based Model for Estimation of Upper Flash Point of Pure Components. *Ind. Eng. Chem. Res.* **2010**, *49*, 12685–12695.

(37) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A.; Richon, D. Artificial Neural Network Modeling of Solubilities of 21 Most Commonly Used Industrial Solid Compounds in Supercritical Carbon Dioxide. *Ind. Eng. Chem. Res.* **2011**, *50*, 221–226.