JOURNAL OF CHEMICAL & ENGINEERING DATA

CORRELATION

pubs.acs.org/jced

Representation/Prediction of Solubilities of Pure Compounds in Water Using Artificial Neural Network−Group Contribution Method

Farhad Gharagheizi,[†] Ali Eslamimanesh,[‡] Amir H. Mohammadi,*[,‡,§] and Dominique Richon[‡]

[†]Saman Energy Giti Co., Postal Code 3331619636, 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 work, the artificial neural network−group contribution (ANN−GC) method has been applied to represent/predict the solubilities of pure chemical compounds in water over the (293 to 298) K temperature range at atmospheric pressure. A set of 3585 pure compounds from various chemical families has been investigated to propose a comprehensive and predictive method. The obtained results show a squared correlation coefficient (R^2) value of 0.96 and a root-mean-square error of 0.4 for the calculated/predicted properties with respect to existing experimental values, demonstrating the reliability of the proposed model.

1. INTRODUCTION

Aqueous systems are often found in oil and gas industries, drug production units, wastewater treatment plants, etc.[1−4] For instance, carbon dioxide/methane systems containing water may cause complications such as corrosion and gas hydrate formation in the oil and gas industries, while aqueous solubility is a very significant factor that affects a drug's oral bioavailability.[1−4] On the other hand, natural gas processing, storage, and transportation, CO_2 capture from industrial/flue gases, separations of industrial compounds from aqueous solutions using ionic liquids, and syntheses of new drugs are among the novel technologies dealing with mixtures of these compounds with water.[2,5] Therefore, accurate knowledge of the solubilities of pure compounds in water is necessary for the development of special models for desired processes. Many studies have been done in order to determine the solubilities of pure compounds in water. In general, the presented models fall into two main categories: thermodynamic models, especially equations of state (EoS)-based models, and mathematical/empirical correlations.

Cubic equations of state [e.g., the Peng−Robinson (PR) EoS,[6]] are widely used for calculations of the solubilities of pure compounds in water, especially in the oil and gas industries. However, these approaches may have several drawbacks for predictions of aqueous solubilities, including the following:

1. Critical properties of the pure compounds are required for calculations of the solubilities using these kinds of models. However, many of the investigated compounds, especially those with complex chemical structures, become degraded before reaching the critical conditions. Therefore, measurement of the related properties does not seem to be practical.
2. Many physicochemical factors of the system should be considered in order to modify the thermodynamic model parameters for more accurate prediction of the phase behavior of particular systems. For example, the aqueous solubility of many of pure chemical compounds, especially those applied in the pharmaceutical industry, is the result of a complex interplay of several factors ranging from the hydrogen-bond donor and acceptor properties of the molecule and water to the energetic cost of disrupting the crystal lattice of the solid in order to bring it into solution.[7]
3. The interaction parameters, including the binary interaction parameters of the EoS and/or the activity model (in cases where G^E-based mixing rules are used) should be tuned over the experimental solubility data for each investigated system. This is a time-consuming process, especially for solubility calculations for large numbers of pure compounds.
4. Since some pure compounds (e.g., methane and heavier hydrocarbons) have very low solubilities in water (typically 10^{-3} to 10^{-5} mole fraction), the numerical methods for solving the equations of state should be applied with much attention to avoid divergence.
5. The optimization algorithms used for obtaining the optimal values of the thermodynamic model parameters have a strong influence on the results of the solubility predictions. Therefore, selecting the best optimization strategy is also another factor that should be taken into account.In spite of the aforementioned drawbacks, the thermodynamic models have been successfully applied for predictions of the solubilities of hydrocarbons, CO_2, N_2, and similar gases in water, which are generally encountered in the oil and gas industries under different temperature and pressure conditions.[1−3,8−12] However, the application of cubic equations of state has to date been checked for only a

Special Issue: John M. Prausnitz Festschrift

Received: October 21, 2010
Accepted: January 7, 2011
Published: February 04, 2011

dx.doi.org/10.1021/je101061t | *J. Chem. Eng. Data* 2011, 56, 720–726

limited number of complex pure compounds for the reasons listed above.

The second category of approaches used for solubility predictions focuses on developing correlations, especially mathematical algorithms using the physical properties of the compounds in a way that makes these methods general and comprehensive. The latter methods are capable of representing/predicting the solubilities of many compounds in any state (including solid, liquid, or gas) from various chemical families in water. Abraham et al.[7] proposed a linear solvation energy relationship (LSER) method based on excess molar refractivity, solute hydrogen-bond acidity and basicity, solute dipolarity/polarizability, and McGowan's characteristic volume.[7,13] Although the model is capable of correlating the solubilities of simple organic compounds in water, the parameters of the model are not easily estimated from the chemical structures of the molecules. The same approach has also been applied by other authors.[14−16]

Large aqueous solubility data sets are generally used to present comprehensive methods. Delaney[17] presented a solubility prediction tool for a data set containing 2874 compounds using nine simple descriptors, including molecular weight, aromatic proportion, non-carbon proportion, polar surface area, and so on. Another attempt was made by Votano et al.,[18] who constructed a set of methods using topological structure indices as descriptors for a data set consisting of 4115 aromatic compounds. Although the larger data sets ensure the reliability of the predictive tools, these two methods do not lead to acceptable average deviations of the results from experimental solubility values [e.g., the root-mean-square errors (RMSEs) are larger than 1.0 based on the natural logarithm unit].[4]

Wang and co-workers[4] developed a method that could be optimized to reduce the RMSE by seeking a different combination of descriptors. They also tested a set of robust models to represent/predict aqueous solubilities for 3664 molecules as a set of high-quality solubility data from a variety of sources. They tested the reliability and predictability of the tools through a set of stringent cross-validation experiments.[4]

The most cited works on the application of mathematical correlations for representation/prediction of the solubilities of large pure compounds in water have been done by Yalkowski and Valvani,[19] Yalkowski and Pinal,[20] Abramovitz and Yalkowski,[21] and Myrdal et al.,[22] which have focused on the applications of LogP (the natural logarithm of the octanol/water partition coefficient) and aqueous functional group activity coefficient (AQUAFAC) methods. The most recent approach for this purpose has been proposed by Jain and Yalkowski.[23] They employed the SCRATCH algorithm using predicted melting points and aqueous activity coefficients. As a matter of fact, the method uses two additive, constitutive molecular descriptors (enthalpy of melting and aqueous activity coefficient) and two nonadditive molecular descriptors (symmetry and flexibility). A comparison between the proposed method and a widely used general solubility equation (GSE) was finally made in order to prove the capability of the model to represent/predict the solubilities of 2200 compounds in water under ambient conditions. Other methods based on mathematical correlations can be found in good reviews by Lipinski et al.,[13] Jorgensen and Duffy,[24] Wang et al.,[25] and Hou and Wang.[26] Table 1 shows a brief survey of several methods previously presented in the literature.

Table 1. Brief Review of Several Methods Previously Presented in the Literature for Evaluation of Aqueous Solubilities

author	N^a	RMSE^b
Klopman et al.[27]	1288	0.54
Kühne et al.[28]	—^{c}	0.96
Yan and Gasteiger[29] (MLR^d)	1293	1.06
Yan and Gasteiger[29] (ANN)	1293	0.52
Hou et al.[30]	1290	0.53
Huuskonen[31] (MLR^d)	1297	0.53
Huuskonen[31] (ANN)	1297	0.53
Duchowicz and Castro[32]	148	1.20
Tetko et al.[33]	1297	0.52
Liu and So[34]	1312	0.51
Wegner and Zell[35]	1269	0.53
Delaney[17]	2847	1.01
Votano et al.[18]	5964	0.74−1.04
Wang et al.[25]	3664	0.84
Myrdal et al.[22]	2432	0.448
Klopman and Zhu[36]	1290	0.79
Abraham and Le[37]	594	0.56
Meylan and Howard[38]	3000	0.9
Jorgensen and Duffy[24]	150	0.72
Mitchell and Jurs[39]	1297	0.67

[a] Number of investigated pure compounds. [b] Root-mean-square error. [c] Not determined. [d] Multilinear relationship model.

Regarding the methods reported to date, there are some elements to be considered. First and perhaps the most important is the fact that the failure to select proper solubility data is one of the significant reasons that some of these methods lack high predictive ability.[4] Furthermore, accurate prediction of the solubilities of complex multifunctional compounds is still required, especially in the drug industry, where the reliable values of the solubilities of pure compounds in water are vital.[4] Finally, many of the previously presented methods[4,13−39] have shown RMSEs in the range of 0.6 to 0.8,which has motivated researchers to develop more accurate predictive tools.

In this work, an artificial neural network−group contribution (ANN−GC) algorithm was developed to represent/predict the solubilities of 3585 pure compounds belonging to various chemical families in water under temperature and pressure conditions that are especially significant in the drug industry.

2. METHODS

2.1. Data Set Preparation. The accuracy and reliability of a method for representation/prediction of physical properties, especially one dealing with a large experimental data set, directly depend on the quality and comprehensiveness of the data set applied for its development.[40] The aforementioned characteristics of such a method include diversity in both the investigated chemical families and the pure compounds available in the data set.[41] In this work, we used the data set evaluated by Wang et al.,[4] which is one of the most comprehensive experimental data sets available in the literature. The experimental values of the solubilities are given in the Supporting Information. In this material, S1 refers to the first data set collected by Wang et al.,[4] and S2, S3, and S4 are the second, third, and fourth ones, respectively.

2.2. Generation of Artificial Neural Network-Based Group Contribution. With this definition of the data set, the chemical structures of all 3585 compounds were analyzed. Consequently, 90 functional groups were found to be more efficient for representation/prediction of the solubility of the related pure compounds in water. The functional groups used in this study are presented in the Supporting Information. In addition, their numbers of occurrences in the pure compounds used in this work are also extensively presented in the Supporting Information. These chemical groups are used as the parameters of the proposed method.

The next calculation step, and perhaps the most significant one, was to search for a relationship between the chemical functional groups and the solubility of the pure compounds in water under the investigated conditions. The simplest method for this purpose was to assume the existence of a multilinear relationship between these groups and the desired property (here, the aqueous solubility of the pure compounds).[40] This technique is similar to the method used in most of the classical group-contribution approaches.[26] Several calculations showed that the application of the aforementioned methodology for the current problem yielded poor results. Consequently, the nonlinear mathematical method of an artificial neural network (ANN) was investigated.

ANNs have been used extensively in various scientific and engineering problems,[40−69] including calculations/estimations of physical and chemical properties of different pure compounds.[40,54−68] These capable mathematical tools are generally applied to the study of complicated systems.[41−53] Theoretical explanations of neural networks can be found elsewhere.[70] Using the artificial neural network toolbox of the MATLAB software (The Mathworks Inc.), a three-layer feed-forward artificial neural network (FFANN) was developed for the problem. The typical structure of a three-layer FFANN is schematically presented in Figure 1. The capabilities of this kind of ANN have been demonstrated in previous works.[40−69]

All of the aqueous solubility values of the pure compounds were normalized between −1 and +1 to decrease computational errors. This was performed by using the maximum and minimum values for each functional group as input data and the maximum and minimum values of the aqueous solubilities as output parameters. Later, the main data set was divided into three new data subsets, namely, the "training" set, the "validation (optimization)" set, and the "test (prediction)" set. In this work, the training set was used to generate the ANN structure, the validation (optimization) set to optimize the model, and the test (prediction) set to investigate the predictive capability and validity of the obtained model. The process of dividing the main data set into three data subsets was performed randomly. For this purpose, about (80, 10, and 10) % of the main data set were randomly selected for the training set (2869 compounds), the validation set (358 compounds), and the test set (358 compounds), respectively. The effect of the allocation percentages for obtaining the three data subsets from the main data set on the accuracy of the ANN method was studied in previous work.[70,71]

As a matter of fact, generation of an ANN model involves determination of the weight matrices and bias vectors.[40] As shown in Figure 1, there are two weight matrices ($\mathbf{W}_1$ and $\mathbf{W}_2$) and two bias vectors ($\mathbf{b}_1$ and $\mathbf{b}_2$) in a three-layer FFANN.[40−68] These parameters should be obtained by minimization of an objective function. The objective function used in the present study was the sum of squares of differences between the outputs of the ANN (correlated/predicted solubilities) and the target values (experimental solubility values). This minimization was performed using the Levenberg−Marquardt (LM)[70] optimization strategy. There are also more accurate optimization methods (i.e., methods for which the convergence of the algorithms to the global optimum is more probable) than this algorithm; however, they need much more convergence time. In other words, the more accurate the optimization, the more time is needed for the algorithm to converge to the global optimum. The LM strategy[70] is the most widely used algorithm for training and was robust and accurate enough to deal with the considered system.[40−68]

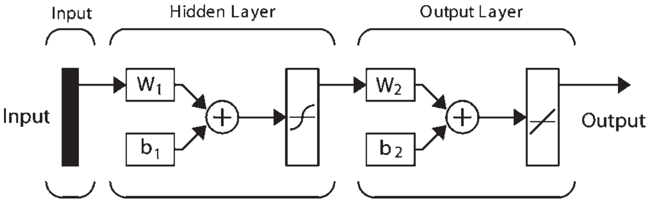

Figure 1. Schematic structure of the three-layer feed-forward artificial neural network method used in this study. Definitions: W, weight; b, bias.

Developing an ANN method capable of predicting the target values as accurately as expected was one of the main steps in the calculations. This step was repeated until the best ANN structure was obtained. Generally, it is more efficient to optimize the numbers of neurons in the hidden layer according to the accuracy of the obtained FFANN than to apply a fixed number of neurons.[40−68]

3. RESULTS AND DISCUSSION

An optimized FFANN for the prediction of the solubilities of 3585 pure compounds in water under the mentioned conditions was obtained using the aforementioned procedure. For this purpose, several three-layer FFANN modules were generated assuming numbers 1 through 50 for the number of neurons in the hidden layer, *n*, using the previously described procedure. The most accurate results were observed for $n = 25$. It should be noted that this value is not the global value, because the optimization method used to train the ANN had a great effect on the obtained value.[40] Therefore, the developed three-layer FFANN has the structure of 90−25−1.

The mat file (MATLAB file format) of the obtained ANN containing all of the parameters of the model is freely available from the authors upon request. In addition, instructions for running the program are presented in the Supporting Information. The represented/predicted solubilities are shown in Figure 2 along with the experimental values for comparison. These values are also reported in the Supporting Information.

The results obtained by the ANN−GC method are reported in Table 2. These results imply that the obtained ANN−GC algorithm is an accurate tool for representing/predicting the solubilities of the pure compounds in water under the specified conditions, which are especially interesting for the drug industry.

To compare this method with techniques previously presented in the literature, some points should be considered. The first of these is the comprehensiveness of the method, which is imperative in the representation/prediction of physical properties of large numbers of pure compounds, especially for application in the drug production industry. The proposed tool is among the most comprehensive ones in comparison with the methods reported in the literature to date because it was developed using a diverse set of 3585 pure compounds from various chemical families. Another element to consider is the RMSE of the presented model results, which is equal to 0.4 for natural-logarithm-based data. This value is considerably lower than the

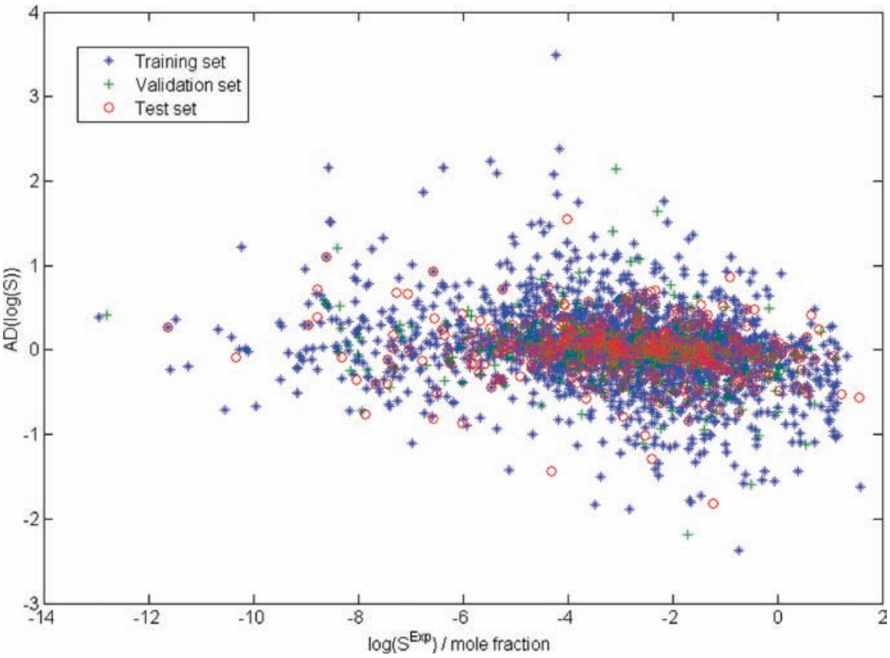

Figure 2. Absolute deviations of the represented/predicted results from the experimental values[4] of the solubilities of pure compounds in water under the investigated conditions. S is the solubility in water ($\mu g \cdot mL^{-1}$); AD is the absolute deviation from the experimental values, defined as $|\log(S^{calcd/pred}) - \log(S^{exp})|$.

Table 2. Statistical Parameters for the Presented Method

statistical parameter	value
Training Set	
R^2 [a]	0.961
standard deviation error	2.0
mean square error	0.2
RMSE	0.4
N[b]	2869
Validation Set	
R^2 [a]	0.963
standard deviation error	1.9
mean square error	0.1
RMSE	0.4
N[b]	358
Test Set	
R^2 [a]	0.971
standard deviation error	2.0
mean square error	0.1
RMSE	0.3
N[b]	358
Training + Validation + Test Set	
R^2 [a]	0.962
standard deviation error	2.0
mean square error	0.2
RMSE	0.4
N[b]	3585

[a] Squared correlation coefficient. [b] Number of data points.

lowest value reported previously by Wang et al.,[4] who used the same data set for representation/prediction of aqueous solubilities.

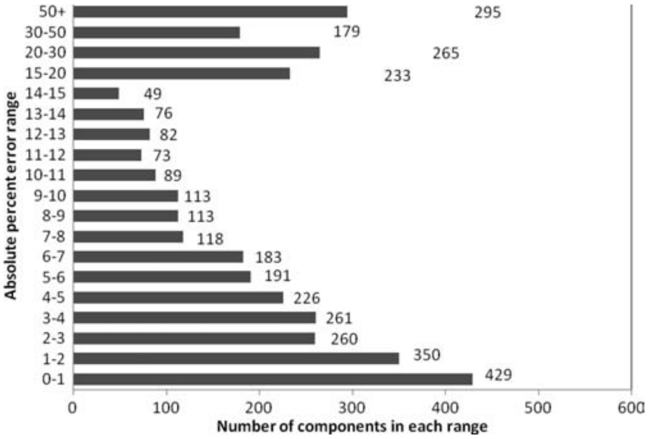

Figure 3. Absolute percent error ranges of the proposed model results from experimental values.[4]

Figure 3 illustrates a better way to understand the accuracy of this method. The absolute percentage deviations of the results from the experimental aqueous solubility values[4] for 1526 compounds are lower than 5 %. To omit from our discussion the fact that there are 295 compounds for which the presented model results lead to an absolute deviation from experimental values of more than 50% would be an oversight. The large deviations of these results from the experimental values[4] may be due mainly to the following reasons:

(a) The presented method may not be the most accurate tool that is possible to develop. We tried our best to develop a technique for calculation/estimation of the aqueous solubilities of pure compounds that is not only more accurate than the previously presented methods reported in the literature but also comprehensive because of the large data set we used.

(b) Particular phenomena such as polymorphism may contribute to the deviations, as the proposed method may not be able to account for compounds that can exist in different crystalline or enantiomeric forms.

(c) It is probable that the experimental solubility values for those compounds showing large deviations from the related results are not accurate or may somehow be erroneous because of the difficulties that existed in the solubility measurements, especially for those compounds with complex chemical structures.[72] It should be noted that the absolute average deviations of the results from the experimental values for the three different data sets used in this study were (22.5, 22.8, and 32.2) % for data sets 1, 2, and 3, respectively.

4. CONCLUSION

In this study, a group-contribution based method for the representation/prediction of the solubilities of pure compounds in water over the (293 to 298) K temperature range at atmospheric pressure has been presented. These conditions are of much interest in industry, especially for drug industry processes. The method is the result of a combination of feed-forward neural networks and group contributions. The required parameters of this tool are the numbers of occurrences of 90 functional groups in each investigated molecule. It should be noted that most of these functional groups are not simultaneously present in a particular molecule. Therefore, computation of the required parameters from the chemical structure of any molecule is simple. For developing this method, the experimental aqueous solubility values from a large data set[4] containing 3585 pure compounds from various chemical families were applied. As a consequence, a comprehensive tool for representing/predicting the aqueous solubilities of many of pure compounds was developed, although it still has some limitations. The method has a wide range of applicability, but its prediction capability is restricted to compounds that are similar to those ones applied in the development of this method. Application of this tool for compounds totally different than the investigated ones is not recommended, although it may be used for a rough estimation of the aqueous solubilities of these kinds of compounds. Finally, a preliminary comparison between the presented method and existing methods in the literature has shown that our approach is comprehensive and accurate. It should be noted that extending our method to different temperature and pressure conditions would require adequate data on solubilities of the investigated compounds in water under those conditions. More meticulous experimental works are required to be done for this concept.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. A spreadsheet listing the investigated 90 functional groups (XLS); a spreadsheet containing the occurrences of all of the functional groups in all of the 3585 pure compounds in the main data set, the values of aqueous solubilities under ambient conditions, and the obtained result (XLS); a document containing the structures of all of the investigated pure compounds; and instructions for running the mat program to apply the method. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: amir-hossein.mohammadi@mines-paristech.fr. Tel.: + (33) 1 64 69 49 65. Fax: + (33) 1 64 69 49 68.

Funding Sources

A.E. thanks MINES ParisTech for providing a Ph.D. scholarship.

■ REFERENCES

(1) Chapoy, A.; Mohammadi, A. H.; Richon, D.; Tohidi, B. Gas solubility measurement and modeling for methane−water and methane−ethane−*n*-butane−water systems at low temperature conditions. *Fluid Phase Equilib.* **2004**, *220*, 113–121.

(2) Chapoy, A.; Mohammadi, A. H.; Chareton, A.; Tohidi, B.; Richon, D. Measurement and modeling of gas solubility and literature review of the properties for the carbon dioxide−water system. *Ind. Eng. Chem. Res.* **2004**, *43*, 1794–1802.

(3) Mohammadi, A. H.; Richon, D. Development of predictive techniques for estimating liquid water−hydrate equilibrium of water −hydrocarbon system. *J. Thermodyn.* **2009**No. 120481.

(4) Wang, J.; Hou, T.; Xu, X. Aqueous solubility prediction based on weighted atom type counts and solvent accessible surface areas. *J. Chem. Inf. Model.* **2009**, *49*, 571–581.

(5) El Abedin, S. Z.; Pölleth, M.; Meiss, S. A.; Janek, J.; Endres, F. Ionic liquids as green electrolytes for the electrodeposition of nonmaterials. *Green Chem.* **2007**, *9*, 549–553.

(6) Peng, D. Y.; Robinson, D. B. A new two-constant equation of state. *Ind. Eng. Chem. Fundam.* **1976**, *15*, 59–64.

(7) Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* **1994**, *83*, 1257–1268.

(8) Zabaloy, M. S.; Mabe, G. D. B.; Bottini, S. B.; Brignole, E. A. Vapor−liquid equilibria in ternary mixtures of water −alcohol−nonpolar gases. *Fluid Phase Equilib.* **1993**, *83*, 159–166.

(9) Gros, H. P.; Bottini, S. B.; Brignole, E. A. A group contribution equation of state for associating mixtures. *Fluid Phase Equilib.* **1996**, *116*, 537–544.

(10) Ferreira, O.; Brignole, E. A.; Macedo, E. A. Modelling of phase equilibria for associating mixtures using an equation of state. *J. Chem. Thermodyn.* **2004**, *36*, 1105–1117.

(11) Pereda, S.; Awan, J. A.; Mohammadi, A. H.; Valtz, A.; Coquelet, C.; Brignole, E. A.; Richon, D. Solubility of hydrocarbons in water: Experimental measurements and modeling using a group contribution with association equation of state (GCA-EoS). *Fluid Phase Equilib.* **2009**, *275*, 52–59.

(12) Sánchez, F. A.; Soria, T. M.; Pereda, S.; Mohammadi, A. H.; Richon, D.; Brignole, E. A. Alkane solubilities in alkanolamine aqueous solutions with group contribution with association equation of state. *Ind. Eng. Chem. Res.* **2010**, *49*, 7085–7092.

(13) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.

(14) Ruelle, P.; Sarraf, E.; Kesselring, U. W. Prediction of carbazole solubility and its dependence upon the solvent nature. *Int. J. Pharm.* **1994**, *104*, 125–133.

(15) Ruelle, P.; Rey-Mermet, C.; Buchmann, M.; Nam-Tran, H.; Kesselring, U. W.; Huyskens, P. L. A new predictive equation for the solubility of drugs based on the thermodynamics of mobile disorder. *Pharm. Res.* **1991**, *8*, 840–850.

(16) Ruelle, P.; Kesselring, U. W. Solubility predictions for solid nitriles and tertiary amides based on the mobile order theory. *Pharm. Res.* **1994**, *11*, 201–205.

(17) Delaney, J. S. ESOL: Estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000–1005.

(18) Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Hall, L. M. Prediction of aqueous solubility based on large datasets using several QSPR models utilizing topological structure representation. *Chem. Biodiversity* **2004**, *1*, 1829–1841.

(19) Yalkowski, S. H.; Valvani, S. C. Solubility and partitioning 1: Solubility of nonelectrolytes in water. *J. Pharm. Sci.* **1980**, *69*, 912–922.

(20) Yalkowski, S. H.; Pinal, R. Estimation of the solubility of complex organic compounds. *Chemosphere* **1993**, *26*, 1239–1261.

(21) Abramovitz, R.; Yalkowski, S. H. Estimation of aqueous solubility and melting point of PCB congeners. *Chemosphere* **1990**, *21*, 1221–1229.

(22) Myrdal, P. B.; Manka, A. M.; Yalkowsky, S. H. AQUAFAC 3: Aqueous functional group activity coefficients; application to the estimation of aqueous solubility. *Chemosphere* **1995**, *30*, 1619–1637.

(23) Jain, N.; Yalkowsky, S. H. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J. Pharm. Sci.* **2001**, *90*, 234–252.

(24) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Delivery Rev.* **2002**, *54*, 355–366.

(25) Wang, J.; Krudy, G.; Hou, T.; Holland, G.; Xu, X. Development of reliable aqueous solubility models and their application in drug-like analysis. *J. Chem. Inf. Model.* **2007**, *47*, 1395–1404.

(26) Hou, T. J.; Wang, J. M. Structure−ADME relationship: Still a long way to go. *Expert Opin. Drug Metab. Toxicol.* **2008**, *4*, 759–771.

(27) Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of aqueous solubility of organic molecules by the group contribution approach. Application to the study of biodegradation. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474–482.

(28) Kühne, R.; Ebert, R. U.; Kleint, F.; Schmidt, G.; Schüürmann, G. Group contribution methods to estimate water solubility of organic chemicals. *Chemosphere* **1995**, *30*, 2061–2077.

(29) Yan, A.; Gasteiger, J. Prediction of aqueous solubility of organic compounds based on a 3D structure representation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 429–434.

(30) Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J. ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 266–275.

(31) Huuskonen, J. Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 773–777.

(32) Duchowicz, P. R.; Castro, E. A. QSPR studies on aqueous solubilities of drug-like compounds. *Int. J. Mol. Sci.* **2009**, *10*, 2558–2577.

(33) Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. Estimation of aqueous solubility of chemical compounds using E-state indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488–1493.

(34) Liu, R.; So, S. S. Development of quantitative structure−property relationship models for early ADME evaluation in drug discovery. 1. Aqueous solubility. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1633–1639.

(35) Wegner, J. K.; Zell, A. Prediction of aqueous solubility and partition coefficient optimized by a genetic algorithm based descriptor selection method. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1077–1084.

(36) Klopman, G.; Zhu, H. Estimation of the aqueous solubility of organic molecules by the group contribution approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 439–445.

(37) Abraham, M. H.; Le, J. The correlation and prediction of the solubility in water using an amended solvation energy relationship. *J. Pharm. Sci.* **1999**, *88*, 868–880.

(38) Meylan, W. M.; Howard, P. H. Estimating log P with atom/fragments and water solubility with log P. *Perspect. Drug Discovery Des.* **2000**, *19*, 67–84.

(39) 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.

(40) 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.

(41) 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*, 221–226.

(42) Chouai, A.; Laugier, S.; Richon, D. Modeling of thermodynamic properties using neural networks: Application to refrigerants. *Fluid Phase Equilib.* **2002**, *199*, 53–62.

(43) Piazza, L.; Scalabrin, G.; Marchi, P.; Richon, D. Enhancement of the extended corresponding states techniques for thermodynamic modelling. I. Pure fluids. *Int. J. Refrig.* **2006**, *29*, 1182–1194.

(44) Scalabrin, G.; Marchi, P.; Bettio, L.; Richon, D. Enhancement of the extended corresponding states techniques for thermodynamic modelling. II. Mixtures. *Int. J. Refrig.* **2006**, *29*, 1195–1207.

(45) Chapoy, A.; Mohammadi, A. H.; Richon, D. Predicting the hydrate stability zones of natural gases using artificial neural networks. *Oil Gas Sci. Technol.* **2007**, *62*, 701–706.

(46) Mohammadi, A. H.; Richon, D. Hydrate phase equilibria for hydrogen + water and hydrogen + tetrahydrofuran + water systems: Predictions of dissociation conditions using an artificial neural network algorithm. *Chem. Eng. Sci.* **2010**, *65*, 3352–3355.

(47) Mohammadi, A. H.; Richon, D. Estimating sulfur content of hydrogen sulfide at elevated temperatures and pressures using an artificial neural network algorithm. *Ind. Eng. Chem. Res.* **2008**, *47*, 8499–8504.

(48) Mohammadi, A. H.; Richon, D. A mathematical model based on artificial neural network technique for estimating liquid water−hydrate equilibrium of water−hydrocarbon System. *Ind. Eng. Chem. Res.* **2008**, *47*, 4966–4970.

(49) Mohammadi, A. H.; Afzal, W.; Richon, D. Determination of critical properties and acentric factors of petroleum fractions using artificial neural networks. *Ind. Eng. Chem. Res.* **2008**, *47*, 3225–3232.

(50) Mohammadi, A. H.; Richon, D. Use of artificial neural networks for estimating water content of natural gases. *Ind. Eng. Chem. Res.* **2007**, *46*, 1431–1438.

(51) Mohammadi, A. H.; Martínez-López, J. F.; Richon, D. Determining phase diagrams of tetrahydrofuran + methane, carbon dioxide or nitrogen clathrate hydrates using an artificial neural network algorithm. *Chem. Eng. Sci.* **2010**, *65*, 6059–6063.

(52) Mehrpooya, M.; Mohammadi, A. H.; Richon, D. Extension of an artificial neural network algorithm for estimating sulfur content of sour gases at elevated temperatures and pressures. *Ind. Eng. Chem. Res.* **2010**, *49*, 439–442.

(53) Mohammadi, A. H.; Belandria, V.; Richon, D. Use of an artificial neural network algorithm to predict hydrate dissociation conditions for hydrogen + water and hydrogen + tetra-*n*-butyl ammonium bromide + water systems. *Chem. Eng. Sci.* **2010**, *65*, 4302–4305.

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

(55) Gharagheizi, F. New neural network group contribution model for estimation of lower flammability limit temperature of pure compounds. *Ind. Eng. Chem. Res.* **2009**, *48*, 7406–7416.

(56) 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*, 267–285.

(57) Gharagheizi, F. Prediction of standard enthalpy of formation of pure compounds using molecular structure. *Aust. J. Chem.* **2009**, *62*, 376–381.

(58) 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*, 1678–1682.

(59) Gharagheizi, F.; Mehrpooya, M. Prediction of some important physical properties of sulfur compounds using QSPR models. *Mol. Diversity* **2008**, *12*, 143–155.

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

(61) Gharagheizi, F.; Alamdari, R. F.; Angaji, M. T. A new neural network−group contribution method for estimation of flash point. *Energy Fuels* **2008**, *22*, 1628–1635.

(62) Gharagheizi, F.; Fazeli, A. Prediction of Watson characterization factor of hydrocarbon compounds from their molecular properties. *QSAR Comb. Sci.* **2008**, *27*, 758–767.

(63) 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*, 40–57.

(64) Gharagheizi, F. A new accurate neural network quantitative structure−property relationship for prediction of θ (lower critical solution temperature) of polymer solutions. *e-Polymers* **2007**No. 114.

(65) 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*, 165–170.

(66) Gharagheizi, F. A chemical structure-based model for estimation of upper flammability limit of pure compounds. *Energy Fuels* **2010**, *27*, 3867–3871.

(67) Vatani, A.; Mehrpooya, M.; Gharagheizi, F. Prediction of standard enthalpy of formation by a QSPR Model. *Int. J. Mol. Sci.* **2007**, *8*, 407–432.

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

(69) Kalogirou, S. A. Artificial neural networks in renewable energy systems applications: A review. *Renewable Sustainable Energy Rev.* **2001**, *5*, 373–401.

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

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

(72) Stefani, V. D.; Baba-Ahmed, A.; Richon, D. A review of experimental methods for solid solubility determination in cryogenic systems. *Cryogenics* **2004**, *44/9*, 631–641.