A simple accurate model for prediction of flash point temperature of pure compounds

Farhad Gharagheizi · Mohammad Hossein Keshavarz · Mehdi Sattari

Abstract In this study, a simple three-parameter linear model is presented for estimation of flash point (FP) of pure compounds. The parameters of the model contain experimental normal boiling point of the compound and two chemical structure-based parameters. A comprehensive database of FPs containing 1472 pure compounds of various chemical structures was used to develop the model. The squared correlation coefficient and average absolute error of the model calculation results for all of the compounds presented in the database are evaluated to be 0.982 and 7.2 K, respectively.

Introduction

Flammability characteristics of substance are required in most scientific and engineering branches [1–18]. Flash

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

M. H. Keshavarz
Department of Chemistry, Malek-ashtar University
of Technology, Shahin-shahr, P.O. Box 83145/115,
Islamic Republic of Iran

M. Sattari
Mapna Generator Engineering & Manufacturing Co (Pars),
Tehran, Islamic Republic of Iran

point (FP) is the best known and most widely used flammability property for evaluation of flammability hazard of combustible liquids. This property is defined as the lowest temperature at which it can form an ignitable mixture with air. At this temperature, the vapor may cease to burn when the source of ignition is removed.

A number of methods have been presented to estimate the FP of pure compounds. These methods can be classified into three main categories based on the type of parameters applied.

The first group includes those correlations using at least one another physical property such as boiling point, density, vapor pressure, critical properties, and enthalpy of vaporization. This class of methods usually focuses on the application of some molecular-based parameters for estimation of the FP property. According to this group, we can refer to the correlations presented by Prugh [19], Fuji and Herman [20], Patil [21], Suzuki [22], Satyarayana and Kakati [23], Satyarayana and Rao [24], Metcalfe and Metcalfe [25], Hshieh [26], Catoire and Naudet [27], and Carroll et al. [28].

The second group of the methods contains correlations called quantitative structure–property relationships (QSPR) [29–74]. These correlations employ only molecular-based parameters to predict the FP property of pure compounds. For this group, we can refer to the models presented by Tetteh et al. [75], Katritzky et al. [76, 77], and Gharagheizi and Alamdari [44].

The last set is actually an especial group of the latter group called group contribution methods, which use the number of occurrences of different functional groups in chemical structure to estimate the FP of pure compounds. In this type, we can refer to the works of Albahri [78], Pan et al. [79], Gharagheizi et al. [46], and Keshavarz and Ghanbarzadeh [80].

Comparison between these three groups of methods is pretty difficult, because there are significant points that should be considered in the comparative study; e.g., simplicity of the model, accuracy of model, simplicity of parameters used in the model, and comprehensiveness of model for covering wider range of investigated chemical compounds.

Based on the simplicity of the model, the first category seems to be more convincing than the others. Based on the accuracy, simplicity of parameters used in the model, and comprehensiveness of model, the third group appeared to be more convincing than the others, neglecting the results of the present study.

The main aim of this study is to present a new simple and very accurate model based on QSPR using the normal boiling point (NBP) as a key property. NBP property has been used as a key property in some of the correlations presented in the first group. However, no unified, easy to use, comprehensive, and accurate model has been presented for estimation of the FP property, so far.

Materials and methods

Materials

DIPPR 801 [81] database has been found especial applications in developing new models for prediction of physical properties, because it contains a large number of pure compounds as well as their evaluated physical properties. This database is recommended by AIChE (American Institute of Chemical Engineers). In order to provide a dataset for calculation of FP, 1472 pure compounds were selected, and the related values of the FP and NBP properties were observed for the calculations. These compounds are presented as supplementary material.

Determination of model parameters

Molecular descriptors are defined as numerical characteristics associated with chemical structures. Each one is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number applied to correlate the physical properties.

A review for software applied to calculate molecular descriptors has been found elsewhere [82]. One of the most widely used software implemented for this propose is the Dragon software [83]. Dragon calculates more than 3000 molecular descriptors for many common chemical structures. Since the values of many descriptors are related to the bonds lengths and bonds angles, etc., the chemical structure of each molecule must be optimized before calculating its molecular descriptors. Owing to this fact, chemical structures of all 1472 pure compounds were drawn in Hyperchem software [84] and optimized using the MM + molecular mechanics force field.

After optimizing the chemical structures, the molecular descriptors were calculated using the Dragon software. The input values to this software have been the optimized chemical structures of the investigated molecules obtained by the MM + optimization strategy.

Development of the model

Using the optimized chemical structures of the compounds obtained in the previous step, molecular descriptors were calculated by the Dragon software. These molecular descriptors were prepared to be applied in the next step of computations. Before beginning of the next computations step, NBP was added to the descriptors as an experimental descriptor. The idea of using the NBP property for prediction of the FP is not a new idea; however, application of NBP coupled with other molecular descriptors in QSPR studies is a new one.

The next problem is to find a linear equation that can predict the FP property with the least number of variables as well as highest accuracy. In other words, the next step is to find a subset of variables (most statistically effective molecular descriptors of FP property) from all available variables (all input variables containing molecular descriptors and NBP property), which is able to predict the FP with the least possible errors in comparison with the existing experimental data.

A generally accepted mathematical method for this problem is genetic algorithm-based multivariate linear regression (GA-MLR). In this method, genetic algorithm is applied for selection of best subset variables with respect to an objective function. This algorithm was firstly presented by Leardi et al. [85].

There are many standard fitness functions such as R^2, adjusted R^2, Q^2, Akaike information content, LOF function, etc., used as objective function in GA-MLR technique [19]. RQK fitness function is a suitable fitness function for model searching proposed to avoid unwanted model properties, such as chance correlation, presence in the models of noisy variables, and other model pathologies resulting in lack of model prediction capability [82, 86]. This fitness function is a constrained fitness function based on Q^2_{LOO} (leave-one-out cross validated variance) statistics and four tests that must be fulfilled contemporarily. The Q^2_{LOO} is defined as:

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

where y_i, $\bar{y}$, and $\hat{y}_{ic}$ are FP property for ith pure compound, mean value of FP for all pure compounds, and $\hat{y}_{ic}$ response of the ith object estimated using a model obtained without using the ith object, respectively.

These four constrains were calculated as follows [82, 86]:

$$\Delta K = K_{XY} - K_X > 0 \text{ (Quick rule)} \tag{2}$$

$$\Delta Q = Q_{\text{LOO}}^2 - Q_{\text{ASYM}}^2 > 0 \left(\text{Asymptotic } Q^2 \text{ rule}\right) \tag{3}$$

$$R^P > 0 \text{ (Redundancy RP rule)} \tag{4}$$

$$R^N > 0 \text{ (Overfitting RN rule)} \tag{5}$$

These four constrains have been extensively explained by Todeschini et al. [82, 86]. Since this algorithm check many problem characteristics during the calculations, we can ensure that the final model is valid and can be regarded as a predictive. In this study, GA-MLR with RQK fitness function was used based on satisfactory results in the author's previous works using the same technique [29–34, 36, 37, 39, 40, 44, 45, 47, 48, 57–60, 63–65, 67, 69–71].

Before performing GA-MLR calculations, the dataset must be divided into two collections. The first one is applied for training and the second one is applied for testing. Using the training set, the best model is found, and then the predictive capability of the obtained model is checked by the test set. In this study, 80% of the database was used for training set (1178 pure compounds) and 20% of the database (294 pure compounds) was used for the test set. It should be noted that these compounds were randomly selected.

Several validation techniques should be used to obtain a valid model. The most widely used techniques have been presented by Todeschini et al. [82, 86]. Among those techniques, the bootstrapping, and the y-scrambling and external validation techniques are used in this study.

Using the bootstrapping technique, the original size of the dataset (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. The model is calculated on the training set and responses are predicted on the evaluation set. 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 average predictive ability is calculated [82].

The y-scrambling technique is adopted to check the models with chance correlation. This test is performed by calculating the quality of the model (usually the Q^2) randomly modifying the sequence of the response vector by assigning a response to each object randomly selected from the true responses. If the original model has no chance correlation, there is a significant difference in the quality of the original model and that associated with a model obtained with random responses. The procedure is repeated hundreds of times [82].

External validation technique is a validation technique in which a test is retained to perform a further check on the predictive capabilities of a model obtained from a training set with predictive ability optimized by an evaluation set [82].

Results and discussion

Following the presented procedure, the most accurate multivariate linear equation was obtained. For obtaining this equation, firstly, the best one-molecular descriptor model was achieved. Later, the best two-molecular descriptor model was evaluated. This procedure was repeated to obtain the best three-, four-, five-, etc. molecular descriptor model. The best multivariate linear model has three parameters because increase in the number of molecular descriptors has no significant effect on the accuracy of the best model. This equation and its statistical parameters are presented as:

$$\begin{aligned}
FP = {} & 43.5120(\pm 1.6290) + 0.8374(\pm 0.0054)\text{NBP} \\
& + 1.3695(\pm 0.0512)\text{Ss} \\
& - 39.1658(\pm 0.8716)\text{VEv1} n_{\text{trainiing}} \\
= {} & 1187; n_{\text{test}} = 294; R_{\text{training}}^2 = 0.9837, R_{\text{test}}^2 \\
= {} & 0.9786 Q_{\text{Loo}}^2 = 0.9835; Q_{\text{BOOT}}^2 = 0.9834; Q_{\text{EXT}}^2 \\
= {} & 0.9821; Q_{\text{LTO}}^2 = 0.9801 s = 9.725; a = -0.018; F \\
= {} & 23610.40 \Delta K = 0.003; \Delta Q = 0.000; R^P = 0.002; R^N \\
= {} & 0.000
\end{aligned} \tag{6}$$

where s is residual mean square error, a is y-scrambling parameter, and F is Fisher function, $n_{\text{trainiing}}$ is number of available pure compounds in the training set and n_{test} is number of available pure compounds in the test set. In addition, the molecular descriptors and their physical meanings are presented in Table 1. More information about procedure of calculation of these molecular descriptors from chemical structure of a compound is indicated in the Dragon software user's guide [83].

Table 1 The molecular descriptors entered to the model

Molecular descriptor	Type	Definition
NBP	Experimental descriptor	Experimental NBP
Ss	Constitutional descriptor	Sum of Kier–Hall electrotopological states
VEv1	Eigenvalue-based indice	Eigenvector coefficient sum from van der Waals weighted distance matrix

"NBP" is the experimental normal boiling point of pure compounds. "Ss" is the sum of Kier–Hall electrotopological states. Those molecules containing atoms of high electronegativity and/or terminal atoms or atoms that lie on the mantle of the molecule, has large "Ss" value, so the "Ss" is a measure of electronic accessibility of an atom and can be interpreted as a probability of interaction with another molecule. Equation 6 indicates this fact that FP property value increases with increase of the "Ss" [82].

"VEv1" is eigenvector coefficient sum from van der Waals weighted distance matrix. It is a measure of average distance between atoms in a molecule. Equation 6 shows the increase in FP values with "VEv1" increases [82].

For checking validity of the model, bootstrap technique, y-scrambling, and external validation techniques were used [82, 86]. The bootstrapping was repeated 5000 times. Also, y-scrambling was repeated 300 times. As can be seen, the small differences among Q_{LOO}^2, Q_{BOOT}^2, Q_{EXT}^2, and R^2 show that the obtained model is a reliable one having convincing predictive capability. In the case that the number of the objects in the dataset is quite large (such as in this study), the predictive ability obtained is too optimistic. This is due to a too small perturbation of the data when only one object is left out. Therefore, in these types of problems, the leave-more-out cross validation technique is used. The leave-10-out cross validation was used for this purpose. This technique was repeated 100 times over 100 random splits of training-test sets. The average the cross validation coefficient was equal 0.9801 ($Q_{LTO}^2 = 0.9801$). Also, the intercept value of the y-scrambling technique has low value ($a = -0.018$) that reveals the model is valid. 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.

All the validation techniques show that the obtained model is a reliable model and it is able to be applied for prediction of the FP property values of pure compounds.

The FPs predicted values using Eq. 6 in comparison with the DIPPR 801 data are presented in Fig. 1. These values in comparison with the DIPPR 801 data are presented as supplementary material. In addition, the values of the descriptors and status of all compounds (training set or test set) are reported as supplementary material.

The absolute percent error obtained by the model for all 1472 pure compounds is shown in Fig. 2. The results indicated in this figure show that the FP values of only 11 of 1472 pure compounds has been predicted by more than 10% using the model. The absolute average error of the calculations of values of this property for 1355 pure compounds have been calculated to be less than 5% in comparison with the existing experimental data. These results confirm the accuracy of the model in prediction of
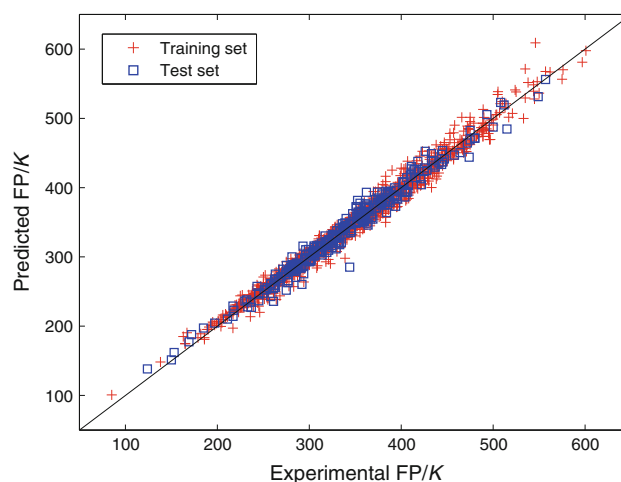

Fig. 1 The comparison between the predicted FP and DIPPR 801 data for training set and test set
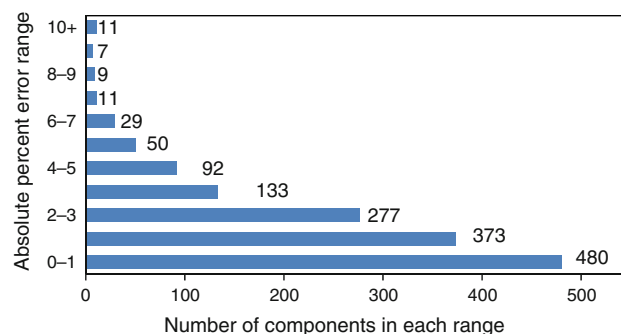

Fig. 2 The absolute percent error of the obtained model over 1472 pure compounds. The absolute percent error is defined as: $100 \times \left| \frac{y_{exp} - y_{calc}}{y_{exp}} \right|$

Table 2 Comparison among the model and the most important previously presented models

Model	R^2	AAE[a]	rms	n
Suzuki et al. [22]	0.9351	10.3	13.5	400
Tetteh et al. [75]	0.9326	10.2	13.1	400
Katritzky et al. [76]	0.902	–	16.1	271
Katritzky et al. [77]	0.878	13.9	–	758
Gharagheizi and Alamdari [44]	0.9669	10.2	12.7	1030
Gharagheizi et al. [46]	0.9757	8.1	11.21	1378
Gharagheizi et al. [52]	0.979	8.1	10.6	1471
This model	0.982	7.2	9.7	1472

[a] Average absolute error

the FP property values for various types of pure compounds.

A comparison between the presented model results and the previously presented ones in the literature is shown in Table 2.

Table 3 The average percent errors of the model over each one of chemical families of compounds used in this study

Family	Average percent error
1-Alkenes	2.52
2,3,4-Alkenes	1.80
Acetates	1.90
Aldehydes	1.92
Aliphatic ethers	1.66
Alkylcyclohexanes	0.82
Alkylcyclopentanes	0.91
Alkynes	2.34
Anhydrides	0.74
Aromatic alcohols	1.33
Aromatic amines	2.22
Aromatic carboxylic acids	1.70
Aromatic chlorides	1.95
Aromatic esters	2.62
C, H, BR compounds	3.12
C, H, F compounds	2.68
C, H, I Compounds	2.35
C, H, multihalogen compounds	3.37
C, H, NO_2 compounds	3.45
C1/C2 aliphatic chlorides	3.80
C3 & higher aliphatic chlorides	2.74
Cycloaliphatic alcohols	2.85
Cycloalkanes	2.14
Cycloalkenes	1.97
Dialkenes	2.62
Dicarboxylic acids	2.12
Dimethylalkanes	1.49
Diphenyl/polyaromatics	1.48
Epoxides	2.33
Ethyl & higher alkenes	1.85
Formates	1.15
Inorganic bases	2.08
Inorganic gases	8.54
Isocyanates/diisocyanates	2.04
Ketones	1.62
Mercaptans	3.35
Methylalkanes	1.20
Methylalkenes	1.40
Multiring cycloalkanes	4.42
N-alcohols	1.85
N-aliphatic acids	1.51
N-aliphatic primary amines	2.24
N-alkanes	3.00
N-alkylbenzenes	2.06
Naphthalenes	1.68
Nitriles	2.65
Nitroamines	1.37

Table 3 continued

Family	Average percent error
Organic salts	2.46
Organic/inorganic compounds	4.81
Other aliphatic acids	1.40
Other aliphatic alcohols	2.85
Other aliphatic amines	3.04
Other alkanes	1.68
Other alkylbenzenes	1.42
Other amines, imines	1.52
Other condensed rings	2.36
Other ethers/diethers	2.74
Other hydrocarbon rings	1.87
Other monoaromatics	1.95
Other polyfunctional C, H, O	2.20
Other polyfunctional organics	2.45
Other saturated aliphatic esters	2.04
Peroxides	2.46
Polyfunctional acids	2.51
Polyfunctional amides/amines	1.74
Polyfunctional C, H, N, Halide, (O)	2.12
Polyfunctional C, H, O, Halide	3.73
Polyfunctional C, H, O, N	1.98
Polyfunctional C, H, O, S	1.69
Polyfunctional esters	2.10
Polyfunctional nitriles	1.96
Polyols	2.27
Propionates and butyrates	1.07
Silanes/siloxanes	2.92
sulfides/thiophenes	2.59
Terpenes	1.64
unsaturated aliphatic esters	0.89

It was found that the model is easiest one to use because it needs only three parameters to predict the FP property values. The least average absolute error and squared correlation coefficient demonstrates that the proposed model would be the most accurate one. The squared correlation coefficient of the model is the largest one among all of the previously proposed models. Furthermore, the model is the most comprehensive one, since it has been checked by most comprehensive FP database (containing 1472 pure compounds).

The average absolute relative error (ARD) of the model for each chemical structure of pure compounds used in this study is shown in Table 3.

It has been proved that the largest deviation is related to the inorganic gases. This case shows that the model is not able to predict the FP property values of the particular chemical structure group. The reason of this drawback is

mainly due to the number of pure compounds of the chemical structure type in the dataset. It seems that the modification of the model needs more experimental data for the chemical structures types especially for the small group of inorganic gases in the database.

Conclusions

In this article, a new simple QSPR model was presented for prediction of the FP property of pure compounds. This model is a three-parameter linear model. These three parameters include experimental NBP of the compounds and two molecular descriptors. These two molecular descriptors are calculated only from chemical structure of the compounds. To develop the model, the most comprehensive database of FPs, containing 1472 pure compounds of various chemical structures were used. The results show that the model is the most accurate and most comprehensive model in comparison with the previously presented models in the literature.

References

1. Wu Q, Bao J, Zhang C, Liang R, Wang B. The effect of thermal stability of carbon nanotubes on the flame retardancy of epoxy and bismaleimide/carbon fiber/buckypaper composites. J Therm Anal Calorim. 2011;103(1):237–42.

2. Šimkovic I, White RH, Fuller AM. Flammability studies of impregnated paper sheets. J Therm Anal Calorim. 2011; 1–8. doi:10.1007/s10973-011-1690-7.

3. Rybiński P, Janowska G, Kucharska-Jastrzabek A, Pajak A, Wójcik I, Wesołek D et al. Flammability of vulcanizates of diene rubbers. J Therm Anal Calorim. 2011; 1–6. doi:10.1007/s10973-011-1728-x.

4. Rybiński P, Janowska G, Jóźwiak M, Pajak A. Thermal properties and flammability of nanocomposites based on diene rubbers and naturally occurring and activated halloysite nanotubes. J Therm Anal Calorim. 2011; 1–7. doi:10.1007/s10973-011-1787-z.

5. Moafi HF, Shojaie AF, Zanjanchi MA. Flame-retardancy and photocatalytic properties of cellulosic fabric coated by nano-sized titanium dioxide. J Therm Anal Calorim. 2011;104(2):717–24.

6. Janowska G, Kucharska-Jastrzabek A, Rybiński P. Thermal stability, flammability and fire hazard of butadiene-acrylonitrile rubber nanocomposites. J Therm Anal Calorim. 2011;103(3):1039–46.

7. Janowska G, Kucharska-Jastrzabek A, Kasiczak A, Rzymski WM. Thermal properties and combustibility of cross-linked XNBR/CSM blends: part I. Influence of the magnesium oxide. J Therm Anal Calorim. 2011;104(3):1107–15.

8. Chang YM, You ML, Lin CH, Wu SY, Tseng JM, Lin CP et al. Fire and explosion hazard evaluation for the acetone aqueous solutions—using weighting analysis of influence factors by grey system theory and 20-L-apparatus flammability investigations. J Therm Anal Calorim. 2011;106(1):179–189.

9. Arora S, Kumar M. Catalytic effect of bases in impregnation of guanidine nitrate on Poplar (Populus) wood—flammability and multiple heating rate kinetic study. J Therm Anal Calorim. 2011; 1–10. doi:10.1007/s10973-011-1779-z.

10. Wu Q, Zhang C, Liang R, Wang B. Combustion and thermal properties of epoxy/phenyltrisilanol polyhedral oligomeric silsesquioxane nanocomposites. J Therm Anal Calorim. 2010; 100(3):1009–15.

11. Stefan CS, Lemordant D, Biensan P, Siret C, Claude-Montigny B. Thermal stability and crystallization of N-alkyl-N-alkyl-pyrrolidinium imides. J Therm Anal Calorim. 2010;102(2):685–93.

12. Rybiński P, Janowska G, Ślusarski L. Influence of cryogenic modification of silica on thermal properties and flammability of cross-linked nitrile rubber. J Therm Anal Calorim. 2010;101(2):665–70.

13. Rybiński P, Janowska G, Kucharska-Jastrzabek A. Influence of surface modification on thermal stability and flammability of cross-linked rubbers. J Therm Anal Calorim. 2010;100(3):1037–44.

14. Mostashari SM, Baie S. TG studies of synergism between red phosphorus (RP)-calcium chloride used in flame-retardancy for a cotton fabric favorable to green chemistry. J Therm Anal Calorim. 2010;99(2):431–6.

15. Lin ML, Lin NK, Chen CC, Liaw HJ, Shu CM. Effects of inert steam on upper explosion limit for oxygenated hydrocarbons. J Therm Anal Calorim. 2010;102(2):541–8.

16. Janowska G, Kucharska-Jastrzabek A, Rybiński P, Wesołek D, Wójcik I. Flammability of diene rubbers. J Therm Anal Calorim. 2010;102(3):1043–9.

17. Janowska G, Kucharska-Jastrzabek A. The effect of chloro-sulphonated polyethylene on thermal properties and combustibility of butadiene-styrene rubber. J Therm Anal Calorim. 2010; 101(3):1093–9.

18. Chang YM, You ML, Tseng JM, Wang YL, Lin CP, Shu CM. Evaluations of fire and explosion hazard for the mixtures of benzene and methanol using rough set method. J Therm Anal Calorim. 2010;102(2):523–33.

19. Prugh RW, XCIX. Estimation of flash point temperature. J Chem Educ. 1973;50(2):A85–9.

20. Fujii A, Hermann ER. Correlation between flash points and vapor pressures of organic compounds. J Saf Res. 1982;13(4):163–75.

21. Patil GS. Estimation of flash point. Fire Mater. 1988;12(3):127–31.

22. Suzuki T, Ohtaguchi K, Koide K. Method for estimating flash points of organic compounds from molecular structures. J Chem Eng Jpn. 1991;24(2):258–61.

23. Satyanarayana K, Kakati MC. Correlation of flash points. Fire Mater. 1991;15(2):97–100.

24. Satyanarayana K, Rao PG. Improved equation to estimate flash points of organic compounds. J Hazard Mater. 1992;32(1):81–5.

25. Metcalfe E, Metcalfe AEM. Communication. On the correlation of flash points. Fire Mater. 1992;16(3):153–4.

26. Hshieh FY. Note: Correlation of closed-cup flash points with normal boiling points for silicone and general organic compounds. Fire Mater. 1997;21(6):277–82.

27. Catoire L, Naudet V. A unique equation to estimate flash points of selected pure liquids application to the correction of probably erroneous flash point values. J Phys Chem Ref Data. 2004;33(4):1083–111.

28. Carroll FA, Lin CY, Quina FH. Calculating flash point numbers from molecular structure: an improved method for predicting the flash points of acyclic alkanes. Energy Fuels. 2010;24(1):392–5.

29. 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.

30. 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.

31. Gharagheizi F. Quantitative structure-property relationship for prediction of the lower flammability limit of pure compounds. Energy Fuels. 2008;22(5):3037–9.

32. Gharagheizi F. A simple equation for prediction of net heat of combustion of pure chemicals. Chemometr Intell Lab Syst. 2008;91(2):177–80.

33. Gharagheizi F. A new molecular-based model for prediction of enthalpy of sublimation of pure components. Thermochim Acta. 2008;469(1–2):8–11.

34. 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.

35. 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.

36. 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.

37. Gharagheizi F. Prediction of upper flammability limit percent of pure compounds from their molecular structures. J Hazard Mater. 2009;167(1–3):507–10.

38. 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.

39. Gharagheizi F. Prediction of the standard enthalpy of formation of pure compounds using molecular structure. Aust J Chem. 2009;62(4):376–81.

40. Gharagheizi F. Chemical structure-based model for estimation of the upper flammability limit of pure compounds. Energy Fuels. 2010;24(7):3867–71.

41. Gharagheizi F. An accurate model for prediction of autoignition temperature of pure compounds. J Hazard Mater. 2011;189(1–2): 211–21.

42. 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.

43. 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.

44. 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.

45. Gharagheizi F, Alamdari RF. A molecular-based model for prediction of solubility of C60 fullerene in various solvents. Fuller Nanotub Carb Nanostruct. 2008;16(1):40–57.

46. Gharagheizi F, Alamdari RF, Angaji MT. A new neural network-group contribution method for estimation of flash point temperature of pure components. Energy Fuels. 2008;22(3): 1628–35.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Empirical method for representing the flash-point temperature of pure compounds. Ind Eng Chem Res. 2011;50(9):5877–80.

53. 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.

54. 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.

55. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Artificial neural network modeling of solubilities of 21 commonly used industrial solid compounds in supercritical carbon dioxide. Ind Eng Chem Res. 2011;50(1):221–6.

56. Gharagheizi F, Eslamimanesh A, Tirandazi B, Mohammadi AH, Richon D. Handling a very large data set for determination of surface tension of chemical compounds using Quantitative Structure-Property Relationship strategy. Chem Eng Sci. 2011; 66(21):4991–5023.

57. Gharagheizi F, Fazzeli A. Prediction of the Watson characterization factor of hydrocarbon components from molecular properties. QSAR Comb Sci. 2008;27(6):758–67.

58. Gharagheizi F, Gohar MRS, Vayeghan MG. A quantitative structure-property relationship for determination of enthalpy of fusion of pure compounds. J Therm Anal Calorim. 2011; 1–6. doi:10.1007/s10973-011-1727-y.

59. Gharagheizi F, Mehrpooya M. Prediction of standard chemical exergy by a three descriptors QSPR model. Energy Convers Manag. 2007;48(9):2453–60.

60. 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.

61. Gharagheizi F, Mirkhani SA, Tofangchi Mahyari AR. Prediction of standard enthalpy of combustion of pure compounds using a very accurate group-contribution-based method. Energy Fuels. 2011;25(6):2651–4.

62. Gharagheizi F, Salehi GR. Prediction of enthalpy of fusion of pure compounds using an artificial neural network-group contribution method. Thermochim Acta. 2011;521(1–2):37–40.

63. 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.

64. Gharagheizi F, Sattari M. Prediction of the θ(UCST) of polymer solutions: a quantitative structure–property relationship study. Ind Eng Chem Res. 2009;48(19):9054–60.

65. 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.

66. 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.

67. 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.

68. Keshavarz MH, Gharagheizi F, Pouretedal HR. Improved reliable approach to predict melting points of energetic compounds. Fluid Phase Equilib. 2011;308(1–2):114–28.

69. 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.

70. Sattari M, Gharagheizi F. Prediction of molecular diffusivity of pure components into air: a QSPR approach. Chemosphere. 2008;72(9):1298–302.

71. 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.

72. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Phase equilibrium modeling of structure h clathrate hydrates of methane + water "Insoluble" hydrocarbon promoter using QSPR molecular approach. J Chem Eng Data (in press). doi: 10.1021/je200444f.

73. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Group contribution-based method for determination of solubility parameter of nonelectrolyte organic compounds. Ind Eng Chem Res (in press). doi:10.1021/ie201002e.

74. Gharagheizi F, Eslamimanesh A, Mohammadi AH, Richon D. Solubility parameters of nonelectrolyte organic compounds: determination using quantitative structure–property relationship strategy. Ind Eng Chem Res (in press). doi:10.1021/ie200962w.

75. Tetteh J, Suzuki T, Metcalfe E, Howells S. Quantitative structure—property relationships for the estimation of boiling point and flash point using a radial basis function neural network. J Chem Inf Comput Sci. 1999;39(3):491–507.

76. Katritzky AR, Petrukhin R, Jain R, Karelson M. QSPR analysis of flash points. J Chem Inf Comput Sci. 2001;41(6):1521–30.

77. Katritzky AR, Stoyanova-Slavova IB, Dobchev DA, Karelson M. QSPR modeling of flash points: an update. J Mol Gr Model. 2007;26(2):529–36.

78. Albahri TA. Flammability characteristics of pure hydrocarbons. Chem Eng Sci. 2003;58(16):3629–41.

79. Pan Y, Jiang J, Wang Z. Quantitative structure-property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network. J Hazard Mater. 2007;147(1–2):424–30.

80. Keshavarz MH, Ghanbarzadeh M. Simple method for reliable predicting flash points of unsaturated hydrocarbons. J Hazard Mater. 2011;193m:335–341.

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

82. Todeschini R, Consonni V. Molecular descriptors for chemo-informatics. 2nd ed. Wiley; 2009.

83. Talete srl, Dragon for windows (software for molecular descriptor calculations), Version 5.5, 2007. http://www.talete.mi.it/.

84. HyperChem Release 7.5 for windows, molecular modeling system, Hypercube, Inc. 2002.

85. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. J Chemom. 1992;6:267–81.

86. 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.