Prime Scholars Library

Full Length Research Paper

Available online at
https://primescholarslibrary.org/

Journal of Science and Geosciences

Vol. 1 (2), pp. 11 - 14, March, 2013

Study on QSPR strategy for hypothetical calculations of warmth development for some natural compounds

Fadhlan Rio and Aahil Ashim Idris

Department of Chemistry, Faculty of Science, P. N. University, Sari, Iran.

Abstract

Quantitative Structure – Property Relationship (QSPR) models based on molecular descriptors derived from molecular structures have been used for the prediction of heat of formation using a set of 20 organic compounds. The molecular and quantum chemical descriptors used to represent molecular structure include topological indices and constitutional descriptors. Forward stepwise regression was used to construct the QSPR models. Multiple linear regressions are utilized to construct the linear prediction model. The prediction results are in good agreement with the experimental values of these properties.

Keywords: QSPR models, heat of formation, topological indices, linear regression.

INTRODUCTION

The study of the quantitative relationship between property/activity and molecular structure (QSPR/QSAR) is an important research area in computational chemistry and has been widely used in the predication of physicochemical properties and biological activities of organic compounds (Katritzky et al., 2000: Katritzky et al., 2001). This kind of study can not only develop a method for the predication of the property under investigation of new compounds that have not been synthesized but also can identify and describe important structure features of molecules that are relevant to variations in molecular properties, thus gain some insight into structural factors affecting molecular properties. To develop a QSPR model, the following step is usually involved: data collection, molecular geometry optimization, molecular descriptors generation, descriptors selection, model development and finally model performance evaluation. One of the important problems in QSPR is the description of molecular structures using molecular descriptors, which can include structural information as much as possible. At present, there exist a great number of molecular descriptors that encode constitutional, topological, geometry and electronic features of organic compounds (Karelson, 2000; Todeschichini and Consonni, 2000; Devillers and Balaban, 1999). Among various structure descriptors, those derived from molecular structure alone have a particular advantage of the possibility to calculate them based only on molecular structural feature and to be applicable to different families of compounds (Estrada and Uriarte, 2001; Karelson et al., 1996). After the calculation of molecular descriptors, linear methods, such as multiple linear regressions (MLR), can be used in the development of a mathematical relationship between the structural descriptors and the property to be predicted. Physical and thermodynamic property data of organic compounds such as heat of formation are important in the engineering design and operation of industrial che-mical processes. Since the experimental determination of heat of formation is both time – consuming and expen-sive, and there is increased need of reliable physical and thermodynamic data for the optimization of chemical processes, it would be very useful to develop predictive models that can be used to predict these properties of organic compounds that are not synthesized or their properties are unknown. The goal of the present study is to extend our previous investigations (Damian et al., 2003) in order to establish a QSPR model that can predict the heat of formation for a set of organic compounds dependent only upon their molecular structures. MLR is applied to establish quantitative linear relationship between heats

Table 1. The heats of formation for some organic compounds.

Molecule	H_f (experimental)	H_f (MLR)	Residual
Ethanol	-277.6	-259.93	-17.67
Propanol	-215.6	-230.78	15.18
1- Propanol	-302.6	-288.41	-14.19
2- Propanol	-318.1	-311.96	-6.14
1- Butanol	-327.3	-322.99	-4.31
2- Butanol	-342.6	-340.29	-2.31
2- Methyl-1- propanol	-334.7	-326.17	-8.53
2- Methyl-2- propanol	-359.2	-356.96	-2.24
Cyclopentanol	-300.1	-307.88	7.78
1-Pentanol	-351.6	-355.08	3.48
2- Pentanol	-365.2	-361.19	-4.01
3- Pentanol	-368.9	-348.76	-20.14
2-Methyl -1-Butanol	-356.6	-353.75	-2.85
3-Methyl -1-Butanol	-356.4	-358.48	2.08
2-Methyl -2-Butanol	-379.5	-372.56	-6.94
3-Methyl-2-Butanol	-366.2	-370.97	4.77
Cyclohexanol	-348.2	-323.24	-24.96
1-Hexanol	-377.5	-366.07	-11.43
2- Hexanol	-392.0	-348.32	-7.68
1-Heptanol	-403.3	-385.53	-17.77

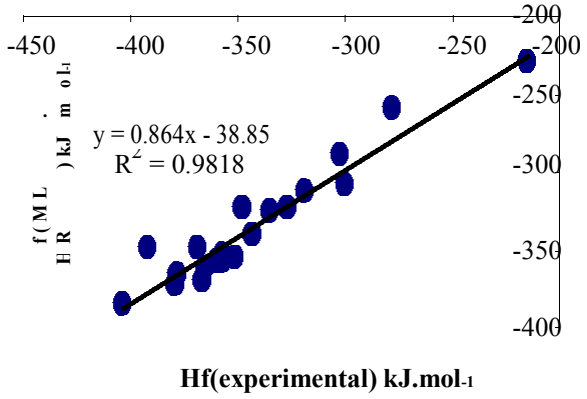

Figure 1. The plot of the residuals versus experimental values.

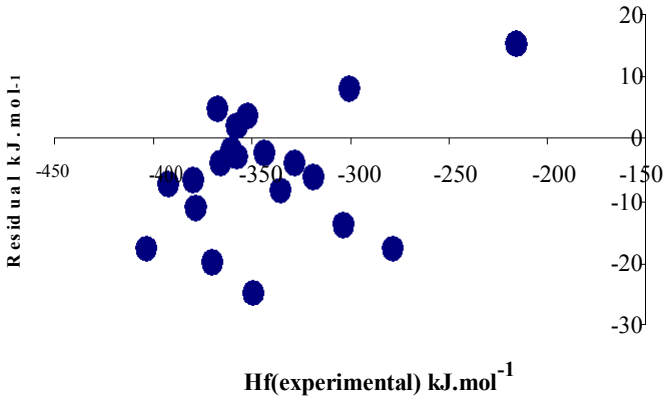

Figure 2. The plot of the residuals versus experimental values errorin using MLR model, because of propagation of residuals on the both sides of zero.

of formation and molecular descriptors.

METHOD

Heat of formation

Over-all heats of formation data in the present investigation were obtained from the CRC (2007). The compounds include a diverse set of substituted alcohols. Compounds and corresponding heats of formation is shown in Table 1. Figure 1 and Figure 2 show related plots for these data.

Figure 1 shows that, there is a good agreement experimental and predicted values using MLR model, and Figure 2 shows that there is no systematical.

Molecular descriptor generation

The calculation of molecular descriptors is described as below: all molecular descriptors were drown into Hyperchem (1994) and pre-optimized using MM+ molecular mechanics force field. A more precise optimization is done with semi–empirical PM3 method in Hyperchem and thereafter quantum chemical descriptors were obtained, which are unusable in this work. The resulted geometry was then transferred into software Dragon (Todeschini, 2000) to calculate constitutional and topological descriptors. Constitutional descriptors are basically related to the number of atoms and bonds in each molecule topological descriptors include valence and non-valence molecular connectivity indices calculated from the hydro-gen suppressed formula of the molecule encoding information

Table 2. Specification of the multiple linear regression models.

Descriptor	Notation	coefficient	Mean effect
Mean information index on atomic composition	AAC	385.45	8.515
Connectivity index chi - 2	X_2	-47.125	-12.842
Average valance Connectivity index chi - 1	X_1 AV	-849.634	-10.9
1st component shape directional WHIM index weighted by atomic electro - topological states	PIS	29.507	7.567
Positivist atom	+X	109.357	4.624
Molecular path count of order 6	MPCO6	-2.154	-2.528

about the size, composition and the degree of branching of a molecule. The quantum chemical descriptors include information about bonding and formation energies, partial atom charge, dipole moment, and molecular orbital energy levels, which are unusable in this work.

Feature selection

Once descriptors were generated, descriptor-screening methods (Yao et al., 2002) are used to select the most relevant descriptor to establish the models to predict the molecular property. Here, the forward stepwise regression method (Yao et al., 2002) was used to choose the subset of molecular descriptors. Forward stepwise regression starts with no model terms and at each step is add the most statistically significant term (the one with the highest F-statistic or lowest p- value) until there is none left. It was determined to be the best model when adding a descriptor no longer improved the cross–validation model (comparison experimental data with MLR method).

Regression analysis

After the descriptor was selected, multiple linear regressions were employed to develop the linear model of the property of interest, which takes the form:

$$Y= b0+b1x1+b2x2+\ldots + b\,n\,x\,n$$

In this equation, Y is the property, that is, the dependent variable, x1 to x n represent the specific descriptor, while b1 to b n represent the coefficient of those descriptor; b0 is the intercept of this equation.

RESULTS AND DISCUSSION

Some of molecular descriptor, which encodes the electronic, geometric and topological features of the molecules, was calculated to describe the molecular structure. Forward stepwise regression routine implemented in SPSS (Marija, 1992) is used to develop the linear model for the prediction of heat of formation using calculated molecular descriptor. The best linear model contains six molecular descriptors. They are listed in Table 2.

Results of MLR model

R = 0.9908, SE = 2.07986. Where R is regression, SE is standard error.

The heat of formation is determined by different descriptors. These descriptors determined by on constitutional descriptor +X and five topological descriptors AAC, X2, X1AV, PIS, and MPCO6 (Table 2) . In order to access the accuracy are predictably of the proposed model, the cross – validation was employed.

Cross-validation test

This test gave the following values: R = 0.974, SE = 10.13706.

Conclusion

QSPR model for prediction of the heat of formation for some organic compounds using MLR based on descriptors calculated from molecular structure have been developed. The correlation coefficient of 0.9908 show a good agreement of MLR values calculated with experimental one (Figure 1), and confirm the ability of MLR model in prediction of heat of formation for some organic compounds. The propagation of residuals on the both sides of zero (Figure 2) indicates that there is non systematic error in using MLR model. Finally, the good agreement between experimental and predicted values using MLR model, confirm its validity.

REFERENCES

CRC (2007) Handbook of Chemistry and Physics, 83rd Ed. Section 3: Constants of Organic compounds.

Damian JG, Marino PJP, German K, Eduardo A, Castro EA (2003) QSPR Modeling of Heat of Formation and Heat of Evaporation of Aliphatic Ketons by Means of Electro topological Indices, Elsevier, Chem. Phys. Lett., 369: 325.

Devillers Y, Balaban AT (1999). Eds. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon & Breach; Amsterdam, Netherlands.

Estrada E,Uriarte E (2001). Recent Advances on the role 0f Topological Indices in Drug Discovery Research, Curr. Med., 8: 1699.

Hyperchem(1994). 4.0, Hypercube, Inc.,Release for Windows, Molecular Modeling System, Hypercube Inc., Canada.

Karelson M (2000). Molecular descriptors in (QSAR /QSPR); John Wiley Sons, New York.

Karelson M, Lobanov VS, Katritzky AR (1996) Quantum Chemical Descriptors in (QSAR /QSPR) Studies; j. Chem. Rev. 96: 1027.

Katritzky AR, Maran U, Lobanov VS, Karelson M (2000). Structurally Diverse Quantitative Structure – Property Relationship Correlation of

Technologically Relevant Physical Properties; J. Chem Inf Comput Sci, p. 4.
KatritzkyAR, Petrukhin R, Tatham D (2001). Interpretation of Quantitative Structure – Property and Activity Relationship J. Chem. Inf. Comput., Sci. 41: 679.
Marija J (1992). SPSS for windows, Norusos/SPSS Inc., Ver. 5.
Todeschichini R, Consonni V (2000). Handbook of Molecular Descriptors; Wiley –VCH; Weinheim, Germany
Todeschini R (2000). Dragon Software for the Calculation of Molecular Descriptors, Rel. 1.1 for Windows. Milan, p. 101
Yao X, Zhang X, Zhang R, Liu M, Hu Z, Fan B (2002). Radial Basis Function Neural Network Based QSPR for prediction of Critical Pressures of Substituted Benzenes, Elsevier, Computers & Chem., 26: 159.