Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 13 (3) 227-235: Fall 2016
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Quantitative Structure-Property Relationship to Predict Quantum Properties of Monocarboxylic Acids By using Topological Indices

Fatemeh Shafiei

Department of Chemistry, Science Faculty, Arak Branch, Islamic Azad University, Arak, Iran

Received August 2016; Accepted December 2016

ABSTRACT

Topological indices are the numerical value associated with chemical constitution purporting for correlation of chemical structure with various physical properties, chemical reactivity or biological activity. Graph theory is a delightful playground for the exploration of proof techniques in Discrete Mathematics and itsresults have applications in many areas of sciences.A graph is atopological conceptrather than a geometrical concept of fixed geometry, and henceEuclidean metriclengths, angles and three-dimensional spatial configurations have nomeaning. One of the useful indices for examination of structure- property relationship is Randic' index. In this study, the relationship between the Randic'(^1X), Balaban (J) and Szeged (Sz) indices and Harary numbers (H) to the thermal energy (E_{th}), heat capacity (C_v) and entropy(S) of monocarboxylic acids (C_2- C_{20}) are established. The thermodynamic properties are taken from HF level using the ab initio 6-31 G basis sets from the program package Gussian 98. Then, some useful topological indices for examination of the structure-property relationship are presented.

Keywords: Topological indices; Graph theory; Monocarboxylic acid; QSPR

INTRODUCTION

Carboxylic acids are characterized by the carboxyl group which combines the carbonyl group of aldehydes and ketones with the hydroxyl group of alcohols and phenols. Since the carbonyl and hydroxyl groups are directly bonded to each other each affects the properties of the other. The result is that the hydroxyl group of a carboxylic acid is considerably, but not completely, different from its alcohol or phenol sibling; the same can be said when one compares the carbonyl of a carboxylic acid with that of an aldehyde or ketone.

The fact that the properties of a molecule are tightly connected to its characteristics is one of the fundamental concepts in chemistry. In this connection, graph theory has been successfully applied in developing some relationships between topological indices and some thermodynamic properties [1-5].

Chemical graph theory is a branch of mathematical chemistry. It is concerned with handling chemical graphs that represent chemical systems. Hence, chemical graph theory deals with analyses of all consequences of connectivity in a chemical system. In other words, chemical graphs theory is concerned with all aspects of the application of graph theory to

*Corresponding author: f-shafiei@iau-arak.ac.ir

chemistry area.

A graph is a topological concept rather than a geometrical concept of fixed geometry. Therefore, Euclidean metric lengths, angles and three- dimensional spatial configurations have no significance in topological concept.

Chemists employ various types of designations and formulas when they want to communicate information about chemical compounds and their structures. In spite of this fact, most of the names and formulas have no direct, immediate or explicit mathematical meaning.

It has been found to be a useful tool in QSAR (Quantitative Structure- Activity Relationship) and QSPR (Quantitative Structure-Property Relationship)[6-11]. Numerous studies have been made relating to the above mentioned fields by using what are called topological indices (TI). In 1975, Randic' proposed a topological index that has become one of the most widely used in both QSAR and QSPR studies [12-14].

Quantitative structure- activity relationship (QSAR) are mathematical models designed for the correlation of various types of biological activity, chemical reactivity, equilibrium, physical and physicochemical properties with electronic, steric, hydrophobic and other factors of a molecular structure of a given series of compounds such as substitution constants, topological indices (TI) as well as with solvent and other physicochemical parameters.

In this paper, a variant of QSAR studies, the so- called Quantitative structure- property relationship (QSPR), using topological indices as molecular descriptors [10], is used. The incredibly great number of works devoted to this has led to the appearance of hundreds of new indices, which are useful to describe with more or less accuracy specific properties of given compounds.

In the last few years, also the necessity of describing the three- dimensional character of molecular structures has contributed to the development of three-dimensional indices [15]. The classical topological approach [5] relates the chemical structure constitution (the two-dimensional model of a molecule, which is representedby a structural formulae) with a non-dimensional numerical entity, the so-called topological indices.

In this correspondence, each structure has a single associated descriptor, but not vice versa; one index may correspond to more than a graph. Here arises the problem of the degeneracy; so it is desirable that the working indices present low degeneracy.

To translate chemical structures into a single number, the graph theory visualizes chemical structures as mathematical object sets consisting of vertices or points, which symbolize atoms, and vertices or lines, linking a pair of edges, which represent covalent bonds or shared electron pairs of covalently linked atoms situated at a topological distance equal to unit.

In this study, the interesting results of structure-property relationship between the Randic'(^1X)[12], Balaban (J) [16], Szeged (Sz) [17-18]indices and Harary numbers (H) [19]to the thermal energy, heat capacity and entropy of monocarboxylic acids ($C_2H_4O_2$ -$C_{20}H_{40}O_2$) and their the Quantitative structure- property relationship are presented.

DISCUSSIONANDDESIGNINGTHE QSPR MODELS

A graph–theoretical approach to QSPR is based on the use of topological indices for encoding the structural informations. The topological indices term indicates a characterization of a molecule (or a corresponding molecular graph by a single number). The need to represent molecular structure by a single number arises from

the fact that most molecular properties are recorded as single numbers. Therefore, QSPR modeling reduces to a correlation between the two sets of numbers via algebraic expression. (One set of numbers represents the properties, and the other set represents the structures of molecules under study.)

A novel method for computing the new descriptors to construct QSPR is presented. First, molecular topological indices are calculated. The values of topological indices of Randic' (^1X), Harary (H), Balaban (J) and Szeged (Sz) for the 19 compounds of training set are also described in Table 1.

After words, the molecular descriptors, which include the structural information necessary to property describe the system, are employed to derive numerical correlation with property. In this study, the molecular structure of the monocarboxylic acids(C_2- C_{20}) have been considered with ab initio methods by using Gaussian-98 computer program that is implemented on a Pentium-PC computer. Firstly, optimizations were performed at 6-31G basis set at the HF-level in the gas phase.The thermal energy, heat capacity and entropy of monocarboxylic acids (C_2-C_{20}) aretaken from Gussian are shown in Table 2.

Figs. 1-8 were shown, respectively, the plots of E_{th}, C_V, S and their logarithmic values for the compounds of training set against the ^1X, log^1 X, J, log J, H, logH, Sz and logSz. Linear dependence is obtained in each plot of figs 1 to 8.

The coefficient of determination (R^2) for the linear regression is shown in respective figure.

The resultant curves, show highly correlation between the respect values.

It is noticeable that some relationship such as Balaban index (J), against the values of E_{th}, C_V and S for the monocarboxylic acids (C_2- C_{20}) are not linear. Now these nonlinear curves are not included.

Table 1. The values of Randic' (^1X), Balaban (J) and Szeged indices (Sz) and Harary numbers (H) of monocarboxylic acids ($C_2H_4O_2$-$C_{20}H_{40}O_2$)

Name of compounds	Formula	^1X	J	H	Sz
Ethanoic acid	$C_2H_4O_2$	1.73	2.32	4.50	9
propanoic acid	$C_3H_6O_2$	2.27	2.54	6.67	18
butanoic acid	$C_4H_8O_2$	2.77	2.36	9.00	32
pentanoic acid	$C_5H_{10}O_2$	3.27	2.68	11.48	52
hexanoic acid	$C_6H_{12}O_2$	3.77	2.72	14.10	79
pentanoic acid	$C_7H_{14}O_2$	4.27	2.75	16.84	114
octanoic acid	$C_8H_{16}O_2$	4.77	2.77	19.68	158
nonanoic acid	$C_9H_{18}O_2$	5.27	2.80	22.62	212
decanoic acid	$C_{10}H_{20}O_2$	5.77	2.82	25.65	277
undecanoic acid	$C_{11}H_{22}O_2$	6.27	2.84	28.76	354
dodecanoic acid	$C_{12}H_{24}O_2$	6.77	2.85	31.94	444
tridecanoic acid	$C_{13}H_{26}O_2$	7.27	2.87	35.20	548
tetradecanoic acid	$C_{14}H_{28}O_2$	7.77	2.88	38.52	667
pentadecanoic acid	$C_{15}H_{30}O_2$	8.27	2.89	41.91	802
hexadecanoic acid	$C_{16}H_{32}O_2$	8.77	2.90	45.35	954
heptadecanoic acid	$C_{17}H_{34}O_2$	9.27	2.91	48.85	1124
octadecanoic acid	$C_{18}H_{36}O_2$	9.77	2.92	52.40	1313
nonadecanoic acid	$C_{19}H_{38}O_2$	10.27	2.93	56.00	15.22
eicosanoic acid	$C_{20}H_{40}O_2$	10.77	2.94	59.65	1752

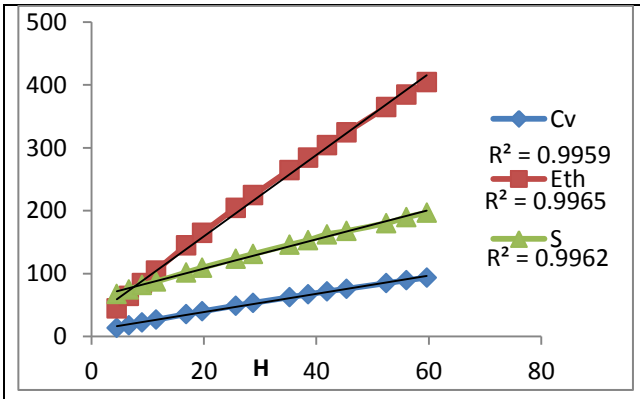

Figure 1 Plots of the Harary number (H) versus C_V, E_{th} and S for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12} , C_{17}).

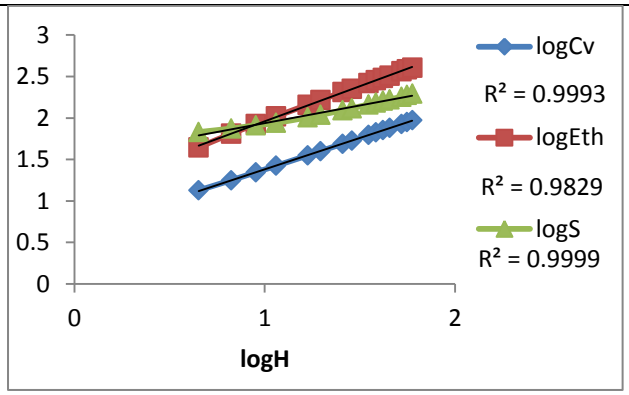

Figure2 Plots of the logH versus log C_V, log E_{th}, and log S for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12} , C_{17}).

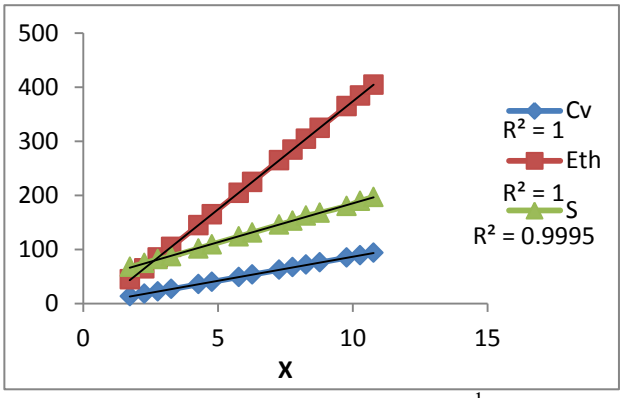

Figure3 Plots of the Randic' index (^{1}X) versus C_V, E_{th} and Sfor (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12} , C_{17}).

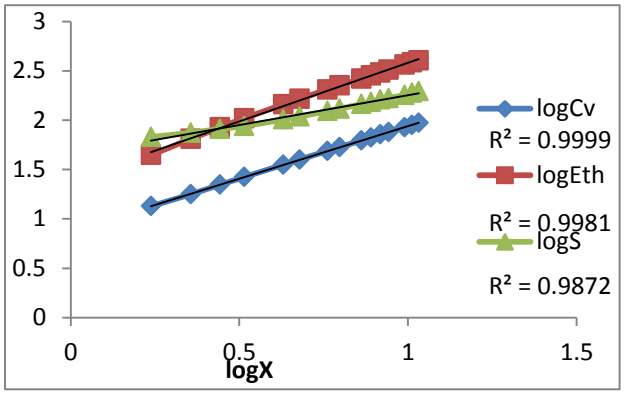

Figure4 Plots of the log X versus log C_V, log E_{th}, and log S for (C_2- C_{20})monocarboxylic acids (without C_6, C_9,C_{12} , C_{17}).

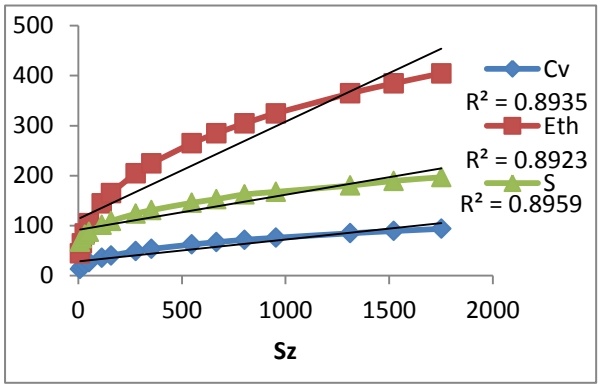

Figure5 Plots of the Szeged index (Sz) versus C_V, E_{th} and S for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12} ,C_{17}).

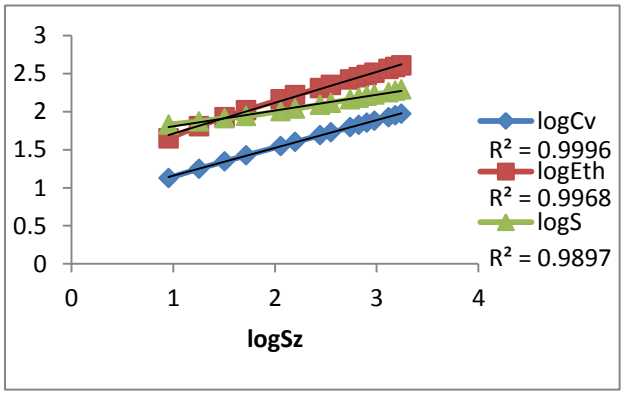

Figure 6 Plots of the log Sz versus log C_V, log E_{th}, and log S for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9,C_{12} , C_{17}).

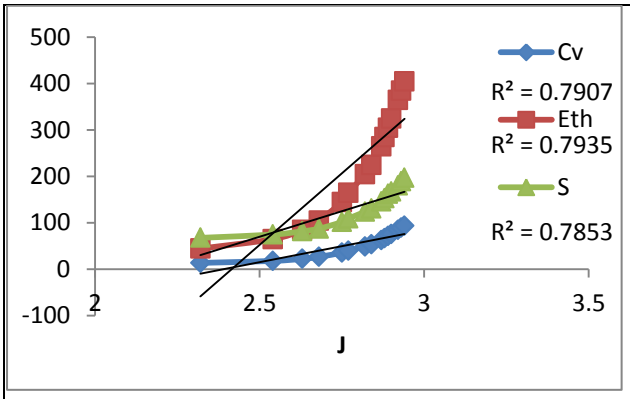

Figure 7 Plots of the Balaban index (J) versus C_V, E_{th} and S for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12} , C_{17}).

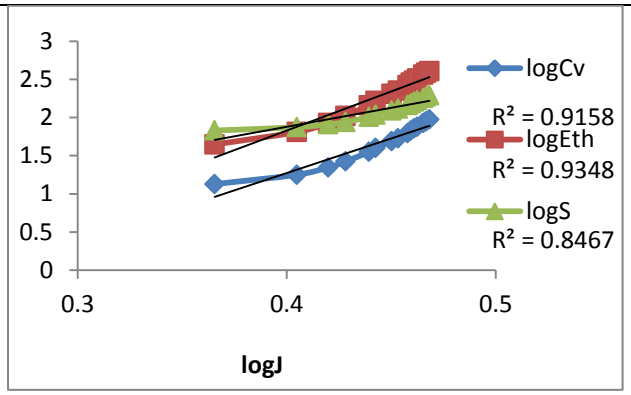

Figure8 Plots of the log J versus log C_V, log E_{th}, and log S for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9,C_{12} , C_{17}).

There are several ways to design the QSPR models. Here we outline one possible strategy which contains five steps:

Step 1.To get a reliable source of experimental data for a given set of molecules. This initial set of molecules is sometimes called the training set. The data in this set must be reliable and accurate. The quality of the selected data is important because it will affect all the following steps.

Step 2. The topological index is selected and computed. This is also an important step because selecting the appropriate topological index (or indices) can facilitate finding the most accurate model.

Step 3. The quality of the QSPR models can be conveniently measured by the

correlation coefficient (R^2). A good QSPR model must have R^2> 0.99. Therefore, step 3 is a central step in the design of the structure – property models.

Step 4. Predictions are made for the values of the molecular property for species that are not part of the training set via the obtained initial QSPR model. The unknown molecules are structurally related to the initial set of compounds.

Step 5. The predictions are tested with unknown molecules by experimental determination of the predicted properties. This step is rather involved because it requires acquiring or preparing the test molecules.

We will apply the procedure from the preceding section, to give an instructive example of the design of the QSPR model for predicting the thermodynamic properties of monocarboxylic acids. As the initial set we will consider monocarboxylic acids with up to 5 carbon atoms (4 molecules). The thermodynamic properties of these monocarboxylic acids are taken from HF level using the ab initio 6-31 G basis sets from the program package Gussian 98. The molecular topological indices such as the Randic' index (^{1}X), Harary number (H), Balaban index (J) and Szeged index (Sz) of the above mentioned are calculated (see table 1).

The following structure–property models are the most successful for predictthe thermodynamic properties.We have separated these models basedon R^2> 0.99.

Models 1 to 12exhibit the applicability of topological and thermodynamics propertiesfor QSPR study of mentioned monocarboxylic acids.

Table 3.shows the variance and highest (positive and negative) differences between Gussian values of thethermal energy and the values obtained using ^{1}X and H for the 4 compounds of training set

(Table 1). The coefficient of determination (R^2) for applying model 1the Gussian data (Table 2) is 1 and it is higher than the respective parameter obtained for the others predictors. Thus, according to this result, the topological ^1X can be used for predicting the thermal energy of monocarboxylic acids.

$$E_{th} = 39.938(^1X) - 25.757 \qquad R^2 = 1 \qquad (1)$$

$$E_{th} = 6.4582(H) + 30.11 \qquad R^2 = 0.9959 \qquad (2)$$

$$\log E_{th} = 0.8442 \log(H) + 1.1155 \qquad R^2 = 0.9993 \qquad (3)$$

$$\log E_{th} = 0.4078 \log Sz + 1.3014 \qquad R^2 = 0.9968 \qquad (4)$$

$$\log E_{th} = 1.1897 \log(^1X) + 1.3934 \qquad R^2 = 0.9981 \qquad (5)$$

$$C_V = 8.9236(^1X) - 2.4912 \qquad R^2 = 1 \qquad (6)$$

$$C_V = 1.4433(H) + 9.9834 \qquad R^2 = 0.9962 \qquad (7)$$

$$\log C_V = 1.0657 \log(^1X) + 0.875 \qquad R^2 = 0.9999 \qquad (8)$$

$$\log C_V = 0.7557 \log(H) + 0.6267 \qquad R^2 = 0.9999 \qquad (9)$$

$$\log C_V = 0.3655 \log(Sz) + 0.7922 \qquad R^2 = 0.9996 \qquad (10)$$

$$S = 14.373(^1X) + 41.365 \qquad R^2 = 0.9995 \qquad (11)$$

$$S = 2.3255(H) + 61.429 \qquad R^2 = 0.9965 \qquad (12)$$

Table 2 The values of thermal energies (E_{th} kcal/mol), constant heat capacities of gas ($C_v$cal/mol K) and entropies of gas (S cal/mol K) are taken from Gussian

compounds	E_{th} $kcal/mol$	C_V $cal/molK$	S $cal/molK$
$C_2H_4O_2$	44.380	13.442	67.558
$C_3H_6O_2$	64.578	17.767	74.612
$C_4H_8O_2$	84.565	22.163	81.713
$C_5H_{10}O_2$	104.607	26.538	87.271
$C_6H_{12}O_2$	124.667	30.927	94.058
$C_7H_{14}O_2$	144.655	35.421	101.646
$C_8H_{16}O_2$	164.637	39.911	109.250
$C_9H_{18}O_2$	184.621	44.399	116.353
$C_{10}H_{20}O_2$	204.604	48.889	123.676
$C_{11}H_{22}O_2$	224.587	53.378	130.987
$C_{12}H_{24}O_2$	244.570	57.867	138.293
$C_{13}H_{26}O_2$	264.554	62.255	145.633
$C_{14}H_{28}O_2$	284.537	66.844	152.932
$C_{15}H_{30}O_2$	304.405	71.627	162.308
$C_{16}H_{32}O_2$	324.571	75.730	167.564
$C_{17}H_{34}O_2$	344.556	80.216	174.860
$C_{18}H_{36}O_2$	364.539	84.704	182.181
$C_{19}H_{38}O_2$	384.522	89.192	189.486
$C_{20}H_{40}O_2$	404.506	93.681	196.113

Table 4.shows the variance and highest (positive and negative) differences between Gussian values of the log arithmic thermal energy and the values obtained using ^1X, H and Gussian values, revealed that models 3to 5 are shown that the topological ^1X is suitable for predicting the $\log E_{th}$ of these carboxylic acids.

Table5.shows comparison of the predicted the heat capacity with Gussian values through the use of topological indices such as ^1X and H (models6&7) for the compounds listed in Table 5. The model (6) establish a relationship between the topological ^1X and C_V by a high squared regression coefficient (R^2=1), which explains the Randic' index for predicting the C_V have better results than the other indices.

Table 3. Comparison between predicted (two model) and Gussian values of thermal energies of respect monocarboxylic acids

compounds	E_{th} $kcal/mol$	E_{th} predicted		Deleted residuals	
	Gussian	model(5)	model(6)	model(5)	model(6)
$C_6H_{12}O_2$	124.667	124.809	121.171	-0.142	3.496
$C_9H_{18}O_2$	184.621	184.716	176.194	-0.095	8.427
$C_{12}H_{24}O_2$	244.570	244.623	236.385	-0.053	8.185
$C_{17}H_{34}O_2$	344.556	344.468	345.593	0.088	-1.037

Table 4. Comparison between predicted (three model) and Gussian values of logarithmic thermal energies of respect monocarboxylic acids

compounds	$\log E_{th}$	$\log E_{th}$ predicted			Deleted residuals		
	Gussian	Model (7)	Model (8)	model (9)	Model (7)	Model (8)	model (9)
$C_6H_{12}O_2$	2.096	2.085	2.075	2.079	0.011	0.021	0.017
$C_9H_{18}O_2$	2.266	2.258	2.250	2.252	0.008	0.016	-0.014
$C_{12}H_{24}O_2$	2.388	2.385	2.380	2.381	0.003	0.008	0.007
$C_{17}H_{34}O_2$	2.537	2.541	2.545	2.544	-0.004	-0.008	-0.007

Table 5. Comparison between predicted (two model) and Gussian values of heat capacities of respect monocarboxylic acids

compounds	C_V $cal/molK$	C_V predicted		Deleted residuals	
	Gussian	Model (10)	Model (11)	Model (10)	Model (11)
$C_6H_{12}O_2$	30.927	31.151	30.334	-0.224	0.593
$C_9H_{18}O_2$	44.399	44.536	42.631	-0.137	1.768
$C_{12}H_{24}O_2$	57.867	57.921	56.082	-0.054	1.785
$C_{17}H_{34}O_2$	80.216	80.230	80.489	-0.014	-0.273

Table 6. shows comparison of the predicted the logarithmic heat capacity with the logarithmic Gussian values through the use of topological indices such as log H, log^{1}X and logSz (models8to 10) for the compounds listed in Table6. The resultants log $C_V$are shown the log (^{1}X),log H and log Sz are suitable for predicting log C_V.

Table 7. shows the variance and highest (positive and negative) differences between Gussian values of the entropy and the values obtained using ^{1}X and H (models 11 & 12) for the 4 compounds of training set (Table 1). According to this result, the topological ^{1}X can be used for predicting the entropy of monocarboxylic acids.

CONCLUSION

Graph theory has provided the chemists with a variety of very useful tools that can be used to predict many interesting physical and chemical properties of considered materials. Regarding this aspect we have presented a structure – property relationship based on topological indices. The instructive example was directed to the design of the structure–property model for predicting the thermal energy (E_{th}), heat capacity (C_V) and entropy(S) of monocarboxylic acids ($C_2H_4O_2$-$C_{20}H_{40}O_2$). Four selected topological indices were tested. The correlation of the Randic index (^{1}X) with the thermal energy (E_{th}), heat capacity (C_V) and entropy(S), appeared to have better results than the other indices. The resultants log C_V are shown, the log (^{1}X), log H and log Sz are suitable for predicting log C_V.

Table 6. Comparison between predicted (three model) and Gussian values of logarithmic heat capacities of respect monocarboxylic acids

compounds	$\log C_V$	log C_v predicted			Deleted residuals		
	Gussian	Model (12)	Model (13)	Model (14)	Model (12)	Model (13)	Model (14)
$C_6H_{12}O_2$	1.490	1.489	1.495	1.486	0.001	-0.005	0.004
$C_9H_{18}O_2$	1.647	1.644	1.650	1.642	0.003	-0.003	0.005
$C_{12}H_{24}O_2$	1.762	1.760	1.763	1.760	0.002	-0.001	0.002
$C_{17}H_{34}O_2$	1.904	1.906	1.903	1.907	-0.002	0.001	-0.003

Table 7. Comparison between predicted (two model) and Gussian values of entropy of respect monocarboxylic acids

compounds	$S\ cal/molK$	$S\ cal/molK$ predicted		Deleted residuals	
	Gussian	Model (15)	Model (16)	Model (15)	Model (16)
$C_6H_{12}O_2$	1.8117	1.8278	2.3832	-1.6×10^{-2}	-0.571
$C_9H_{18}O_2$	1.8823	1.8904	5.2642	-8.1×10^{-3}	-3.382
$C_{12}H_{24}O_2$	1.9292	1.9372	7.4250	-8×10^{-3}	-5.496
$C_{17}H_{34}O_2$	2.0031	1.9959	10.1445	7.2×10^{-3}	-8.141

REFERENCES

[1] A. T. Balaban, Pure. Appl. Chem. 55(1983) 199.
[2] H. Hosoya, Bull. Chem. Soc. Jpn. 44(1971)2332.
[3] M. Randic', Acta. Chim. Slov. 45 (3) (1998) 239.
[4] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17.
[5] A. T. Balaban, T. S. Balaban, Math. Chem. 8 (1991) 383.
[6] J. Devillers, A. T. Balaban, Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach Science Publishers, Amsterdam, the Netherlands, 1999.
[7] M. Randic', J. Chem. 20(1996) 1001.
[8] B. Lučić, N. Trinajstić, SAR QSAR Environ. Res. 7 (1997)45.
[9] L.B. Kier, L. H. Hall and J.W. Frazer, J. Chem. Inf. Comput. Sci. 33 (1993) 143.
[10] M. Randic', S. C. Basak, SAR QSAR Environ. Res., 11 (2000) 1.
[11] M. Thakur, A. Thakur and P. V. Khadikar, Bioorg. Med. Chem. 12 (4) (2004) 789.
[12] M. Randic', J. Am. Chem. 97 (23) (1975) 6609.
[13] A. A. Taherpour, F. Shafiei, J. Mol. Struct. THEOCHEM. 726 (2005) 183.
[14] F. Shafiei, Iranian J. Math. Chem., 6 (2015) 15.
[15] S. C. Basak, G. D. Grunwald and G. J. Niemi, From Chemical Topology to three Dimensional Molecular Geometry in: A. T. Balaban (Ed.), plenum press, (1997).
[16] A.T. Balaban, Chem. Phys. Lett., 89 (1982) 399.
[17] I. Gutman, Formula for Wiener Number of Trees and Its Extensions to Graphs Containing Cycles, Graph Theory Notes, New York. 27 (1994) 9-15.
[18] P.V. Khadikar, N. V. Deshpande and P. P. Dobrynin, J. Chem. Inf. Comput. Sci. 35 (1995) 547.
[19] K. C. Das, B. Zhou and N. Trinajstić, J. Math. Chem. (2009) 1369.