Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 12 (3) 209-216: Fall 2015
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Application of Graph Theory: Relationship of Topological Indices with the Partition Coefficient (logP) of the Monocarboxylic Acids

Fatemeh Shafiei

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

Received November 2015; Accepted November 2015

ABSTRACT

It is well known that the chemical behavior of a compound is dependent upon the structure of its molecules. Quantitative structure – activity relationship (QSAR) studies and quantitative structure – property relationship (QSPR) studies are active areas of chemical research that focus on the nature of this dependency. 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 its results have applications in many areas of sciences. One of the useful indices for examination of structure- property relationship is Randic' index. In this study is represented the relationship between the Randic', Balaban and Szeged indices and Harary numbers to the octanol-water partition coefficient (logP) of monocarboxylic acids (C_2- C_{20}) are established, and then, some useful topological indices for examination of the structure- property relationship are presented.

Keywords: Topological indices; logP; 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 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 chemistry area.

A graph is a topological concept rather

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

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].

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 relationships (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 [13]. The classical topological approach [5] relates the chemical structure constitution (the two-dimensional model of a molecule, which is represented by 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.

The studies of quantitative structure-activity relationship (QSAR) and the quantitative structure - property relationships (QSPR) involve various chemometric methods in which the chemical behavior of a compound is correlated with its structure represented by the topological indices [1]. For example, QSPR methods were applied for the prediction of an octanol-water partition coefficient [2]

In this study, the interesting results of structure-property relationship between the Randic'(^{1}X), Balaban (J) and Szeged (Sz) indices and Harary numbers (H) to the octanol-water partition coefficient (logP) of monocarboxylic acids ($C_2H_4O_2$ - $C_{20}H_{40}O_2$) are presented.

At first, a brief review on the classical topological indices are calculated. After wards, the molecular descriptors that

include the necessary structural information for properly description of system are employed to derive a numerical correlation with thermodynamic properties. Finally, some useful topological indices for examination of the structure- property relationship are presented.

DATA SOURCES

($C_2H_4O_2$ - $C_{20}H_{40}O_2$) are taken from chemicalize Web Book[20].

TOPOLOGICAL INDICES

All the used topological indices were calculated using all hydrogen suppressed graph by deleting all the carbon hydrogen as well as heteroatomic hydrogen bonds from the structure of the monocarboxylic acids. The calculations of these indices are well documented in the literature and therefore, their detailed calculations are not given here. However, blow we have given the final expression for the calculation of these indices.

The connectivity index (^1X)

The connectivity index $^1X = {}^1X(G)$ of G is defined by Randic' [12] as:

$$^1x = {}^1x(G) = \sum_{\text{alledges}} \left[d_{(i)} d_{(j)} \right]^{-0.5} \qquad (1)$$

where $d_{(i)}$ and $d_{(j)}$ are the valencies of the vertices and j that define the edge ij.

Szeged index (Sz)

The Szeged index, Sz = Sz(G), is calculated [16,17] according to the following expression:

$$Sz = Sz(G) = \sum_{\text{alledges}} n_u \cdot n_v \qquad (2)$$

Edges where n_u is the number of vertices lying closer to one end of the edge e=uv; the meaning of n_v is analogous. Edges equidistance from both the ends of an edges, e=uv are not taken into account.

Balaban index (J)

The Balaban index, J=J(G) of G, was introduced by Balaban [1-5-18] as the average- distance sum connectivity. It is defined as:

$$J = \frac{M}{\mu + 1} \sum_{\text{alledges}} \left[d_{(i)} d_{(j)} \right]^{-0.5} \qquad (3)$$

where M is the number of edges in G; μ is the cyclomatic number of G; and d (i) is the distance sum where i=1,2,…,N. The cyclomatic number M=M(G) of a polycyclic graph G is equal to the minimum number of edges that must be removed from G to transform it to the related acyclic graph. For trees, M=0; for monocycles M=1.

Harary number (H)

The Harary number (H) was introduced in 1991 by Harary [14, 15]. This index is defined from the inverse of the squared elements of the distance matrix according to the expression:

$$H = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(D_{ij} \right)^{-2} \qquad (4)$$

where D^{-2} is the matrix whose elements are the squares of the reciprocal distances.

Graphs

For drawing the graphs of our results, we used the Microsoft Office Excel – 2003 program.

DISCUSSION AND DESIGNING THE 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' (^{1}X), 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. The octanol-water partition coefficient (logP) of monocarboxylic acids Tables 2 are taken from book and web book [19,20] for the compounds of training set.

Figs. 1-8 were shown, respectively, the plots of logP for the compounds of training set against the ^{1}X, log^1 X, J, log J, H, logH, Sz and logSz.

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 octanol-water partition coefficient of monocarboxylic acids. As the initial set we will consider monocarboxylic acids with up to 5 carbon atoms (4 molecules). The octanol-water partition coefficient of these monocarboxylic acids are taken from [19,20], and 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).

Models 5 to 12 exhibit the applicability of topological and logP for QSPR study of mentioned monocarboxylic acids.

The coefficient of determination (R^2) for applying model (5) to the experimental data (Table 2) is 0.9965 and for applying model (7) is 0.9949. Thus, according to this result, the topological ^{1}X and H can be used for predicting the octanol-water partition coefficient (logP) of monocarboxylic acids.

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

Comparison of the predicted logP with the experimental values, revealed that models 5 and 7 are shown that the topological ^1X is quite suitable for predicting the logP of these carboxylic 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 octanol-water partition coefficient 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 (^1X) with logP, appeared to have better results than the other indices. Meanwhile, the discussion of relationship areas become straightforward**.**

$$\log P = 0.9063(^1X) - 1.6449 \qquad R^2 = 0.9965 \qquad (5)$$

$$\log P = 10.637(\log {}^1X) - 3.8932 \qquad R^2 = 0.9450 \qquad (6)$$

$$\log P = 0.1465(H) - 0.3766 \qquad R^2 = 0.9949 \qquad (7)$$

$$\log P = 7.5182(\log H) - 6.3435 \qquad R^2 = 0.9362 \qquad (8)$$

$$\log P = 14.413(J) - 36.138 \qquad R^2 = 0.7905 \qquad (9)$$

$$\log P = 86.424(\log J) - 34.379 \qquad R^2 = 0.7606 \qquad (10)$$

$$\log P = 0.0044(Sz) + 1.6272 \qquad R^2 = 0.8943 \qquad (11)$$

$$\log P = 3.6531(\log Sz) - 4.7304 \qquad R^2 = 0.9491 \qquad (12)$$

Table 1. The values of randic' (X), 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	X	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
heptanoic 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

Table 2. The values of the octanol-water partition coefficient (logP) monocarboxylic acids ($C_2H_4O_2$ - $C_{20}H_{40}O_2$)

compounds	log *P*	**compounds**	log *P*
$C_2H_4O_2$	-0.22	$C_{12}H_{24}O_2$	4.48
$C_3H_6O_2$	0.48	$C_{13}H_{26}O_2$	4.92
$C_4H_8O_2$	0.92	$C_{14}H_{28}O_2$	5.37
$C_5H_{10}O_2$	1.37	$C_{15}H_{30}O_2$	5.81
$C_6H_{12}O_2$	1.81	$C_{16}H_{32}O_2$	6.26
$C_7H_{14}O_2$	2.26	$C_{17}H_{34}O_2$	6.70
$C_8H_{16}O_2$	2.70	$C_{18}H_{36}O_2$	7.15
$C_9H_{18}O_2$	3.14	$C_{19}H_{38}O_2$	7.59
$C_{10}H_{20}O_2$	3.59	$C_{20}H_{40}O2$	8.30
$C_{11}H_{22}O_2$	4.03		

Table 3. Comparison between predicted (three model) and experimental values of enthalpies of combustion and their logarithmic of respect monocarboxylic acids

	log *P*	log *P* predicted		Deleted residuals	
compounds	experimental	Model (5)	Model (7)	Model (5)	Model (7)
$C_6H_{12}O_2$	1.810	1.772	1.689	0.038	0.121
$C_9H_{18}O_2$	3.140	3.131	2.937	0.009	0.203
$C_{12}H_{24}O_2$	4.480	4.491	4.303	-0.011	0.177
$C_{17}H_{34}O_2$	6.700	6.756	6.780	-0.056	-0.08

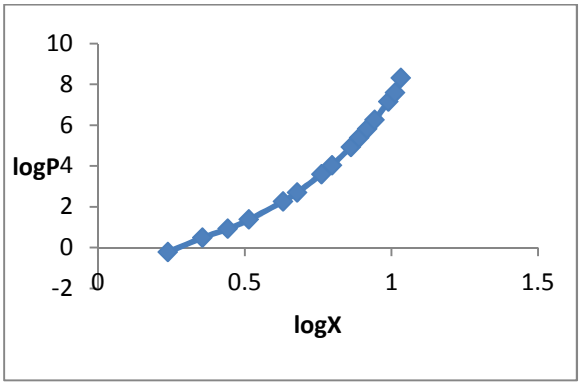

Fig. 1. Plots of the Randic' index (^1X) versus logP for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

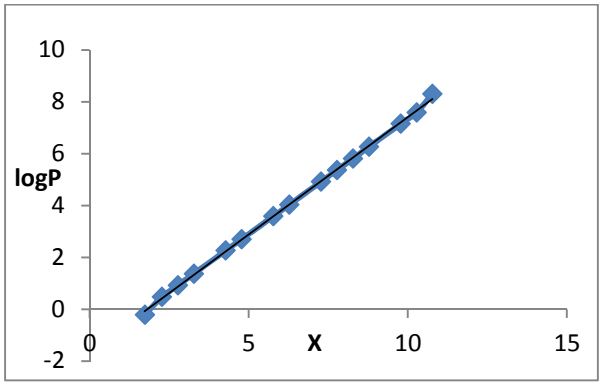

Fig. 2. Plots of the logX versus logP for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

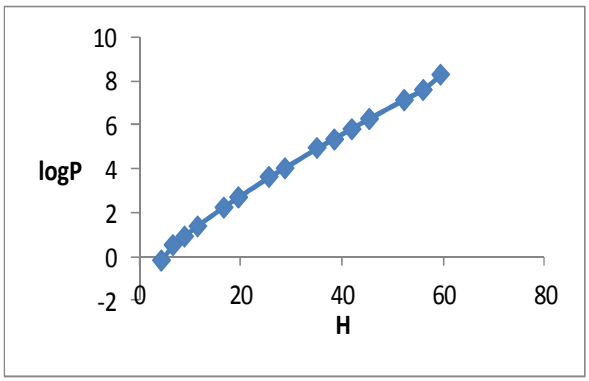

Fig. 3. Plots of the Harary index (J) versus logP for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

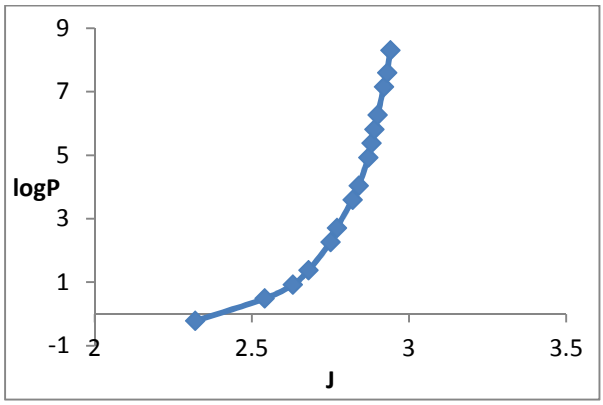

Fig. 6. Plots of the logJ versus for (C_2- C_{20}) for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

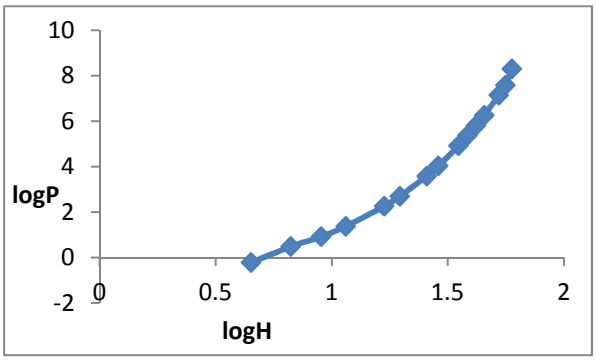

Fig. 4. Plots of the log H versus log P for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

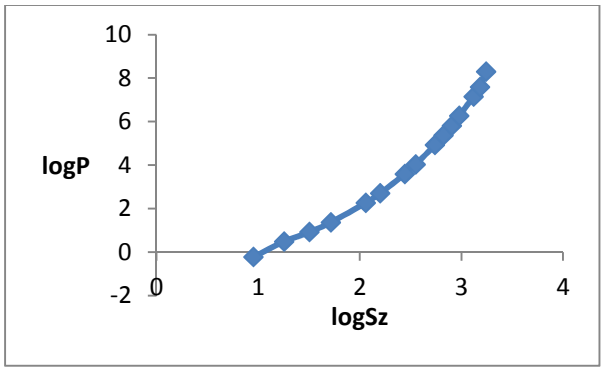

Fig. 7. Plots of the Szeged index (Sz) versus log P monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

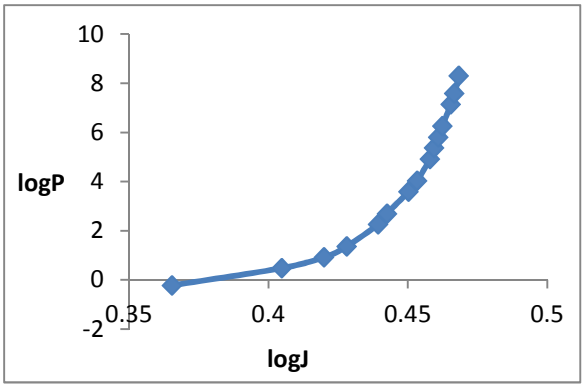

Fig. 5. Plots of the Balaban index (J) versus log P monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

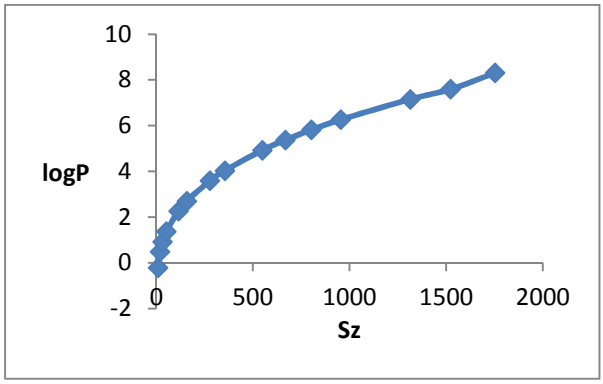

Fig. 8. Plots of the logSz versus logP for (C_2- C_{20}) for (C_2- C_{20}) monocarboxylic acids (without C_6, C_9, C_{12}, C_{17}).

REFERENCE

[1]. A. T. Balaban, topological index based on topological distances in molecular graph, Pure Appl. Chem. 55 (1983) 199-206.

[2]. H. Hosoya, topological index: a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, Bull. Chem. Soc. Jpn. 44 (1971) 2332–2339.

[3]. M. Randic', ON characterization of molecular attributes, Acta. Chim. Slov. 45 (3)(1998) 239-252.

[4]. H. Wiener, Structural Determination of paraffin Boiling Points, J. Am. Chem. Soc., 69 (1947) 17-20.

[5]. A. T. Balaban; T. S. Balaban; New Vertex Invariant and Topological Indices of Chemical Graphs Based on Information on Distance; Math. Chem., 8 (1991) 383-397.

[6]. J. Devillers and A.T. Balaban, Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach Science Publishers, Amsterdam, the Netherlands, 1999.

[7]. M. Randic' Quantitative Structure – property relationship: boiling points of planar benzenoids. New. J. Chem., 20 (1996) 1001-1009.

[8]. B. Lučić, N. Trinajstić, New developments in QSPR/QSAR modeling based on topological index, SAR QSAR Environ. Res., 7 (1997) 45-62.

[9]. L.B. Kier, L. H. Hall and J.W.Frazer. Design of molecules from quantitative structure- activity relationship models. 1. Information transfer between path and vertex degree counts. J. Chem. Inf. Comput. Sci., 33 (1993) 143-147.

[10]. M. Randic', S. C. Basak, Multiple regression analysis with optimal molecular descriptors. SAR QSAR Environ. Res., 11 (2000) 1-23.

[11]. M. Thakur, A. Thakur and P. V. Khadikar, QSAR study on Benzene Sulphonamide Carbonic Anhydrase Inhibitors: Topological Approach Using Balaban Index. Bioorg. Med. Chem., 12(4) (2004)789-793.

[12]. M. Randic', Charactrization of molecular branching. J. Am. Chem., 97(23) (1975) 6609-6615.

[13]. 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).

[14]. S. C. Basak, B. D. Gute and G. D. Grundwald, in: R. Carbo – Dorca, P. G. Mezey (Eds.) Advances in Molecular Similarity vol. 2, JAI Press, Stamford, 1998.

[15]. K. C. Das, B. Zhou and N. Trinajstić, Bounds on Harary index, J. Math. Chem., (2009) 1369–1376.

[16]. I. Gutman, Formula for Wiener Number of Trees and Its Extensions to Graphs Containing Cycles, Graph Theory Notes, New York., 27 (1994) 9-15.

[17]. P. V. Khadikar, N. V. Deshpande and P. P. Dobrynin, The Szeged Index and An Analogy with Wiener Index, J. Chem. Inf. Compt. Sci., 35 (1995) 547-550.

[18]. A. T. Balaban, Highly Discriminating Distance Based Topological Indices; Chem. Phys. Lett., 89 (1982) 399-404.

[19]. K. Baev. Alexei. Specific Intermolecular Interaction of Organic Compounds Springer Heidelberg Dordrecht London New York 2012.

[20]. WWW. Chemicalize.org