

A Novel Topological Descriptor Based on the Expanded Wiener Index: Applications to QSPR/QSAR Studies

AFSHAN MOHAJERI^{1,*}, POUYA MANSOUR² AND MAHBOUBEH MOUSAEI¹

¹Department of Chemistry, College of Sciences, Shiraz University, Shiraz 71454, Iran

²Department of Physics, Persian Gulf University, Bushehr 75169, Iran

ARTICLE INFO

Article History:

Received 6 April 2015

Accepted 3 March 2016

Published Online: 10 March 2017

Academic Editor: Ali Reza Ashrafi

Keywords:

Topological index

Graph theory

Expanded Wiener index

QSPR

QSAR

ABSTRACT

In this paper, a novel topological index, named M -index, is introduced based on expanded form of the Wiener matrix. For constructing this index the atomic characteristics and the interaction of the vertices in a molecule are taken into account. The usefulness of the M -index is demonstrated by several QSPR/QSAR models for different physico-chemical properties and biological activities of a large number of diversified compounds. Moreover, the applicability of the proposed index has been checked among isomeric compounds. In each case the stability of the obtained model is confirmed by the cross-validation test. The results of present study indicate that the M -index provides a promising route for developing highly correlated QSPR/QSAR models. On the other hand, the M -index is easy to generate and the developed QSPR/QSAR models based on this index are linearly correlated. This is an interesting feature of the M -index when compared with quantum chemical descriptors which require vast computational cost and exhibit limitations for large sized molecules.

© 2017 University of Kashan Press. All rights reserved

1. INTRODUCTION

Graph theory is one of the most useful tools for studying systems in various fields such as chemistry, physics, computer science, economy, and biology [1–3]. This powerful concept, which introduced originally by great mathematician, Leonhard Euler [4], has been a useful approach to predict some key features of such systems. Chemical graph theory is a branch of graph theory that is concerned with analyses of all consequences of connectivity in a chemical graph. Chemical graph serves as a convenient model for any real or abstracted

* Corresponding author (Email: amohajeri@shirazu.ac.ir)

DOI: 10.22052/ijmc.2017.27307.1101

chemical system. It can represent different chemical objects as molecules, reactions, crystals, polymers, and clusters [5–8].

Successful use of chemical graph to quantitative structure–property and structure–activity relationships (QSPR/QSAR) has led to the emergence of several molecular descriptors [9–13]. The molecular descriptors derived there from are commonly named topological indices [5]. Regardless of the descriptors used in the development of QSPR/QSAR models, all of them share in common a basic approach; molecules are represented by vectors constructed in turn by molecular parameters, which are supposed to contain relevant information about molecular structure. So far, hundreds of topological indices have been proposed in chemical literature [14]. Among these, topological indices such as the molecular connectivity indices of the Wiener [15], Balaban [16], Randić [17], and Hosoya [18] indices have received greater attention due to their application in chemistry.

The Wiener index, W , is one of the most frequently used graph descriptors in QSPR/QSAR models. Its applicability for predicting physico–chemical and pharmacological properties of organic compounds is well documented and was outlined in quite a few reviews [19–21]. In the past decades, a remarkably large number of modifications and extensions of the Wiener index was put forward and studied by mathematical chemists [22,23]. These indices are generally based on the adjacency matrix or on the distance matrix.

However, most of introduced indices lack certain information such as the features and interaction of vertices (atoms). Recent attentions of graph theoretical chemists have been focused on resolving this problem [24]. In this respect, Yang et al. [25–28] used the electronegativity, the energy, the length between vertices, and principal quantum number to reform the distance matrix of Wiener so that more information of the distance matrix were included in the molecular graph. They have applied these modified Wiener in prediction of the retention indices of gas chromatography, the standard formation enthalpy and gaseous solubility.

However, owing to the complexity of the molecular structure, it seems to be impossible to expect that a single set of descriptors would contain all the relevant structural information. Hence, introducing topological indices which can predict a wide range of physico–chemical properties requiring a minimum number of inputs is the goal of many studies. As far as we know the Wiener index and its modification have been widely used to account for many physico–chemical properties. In the present study, we intend to propose a new topological index based on the reciprocal form of the expanded distance matrix. The expanded form of the Wiener index was suggested by Tratch et al. [29] for characterization of molecular graphs and structure–property correlations. This index is more sensitive to the structural characteristic of alkane molecule as compared with normal Wiener index and also can differentiate several graphs having just the same value of the very powerful

Balaban index. On the other hand, the inclusion of the interactions among vertices (atoms) as well as the vertex properties is a crucial issue in describing the whole properties of a molecule.

Accordingly, based on the above-mentioned issues, in the present study we have introduced a new topological index and named it *M*-index. This newly proposed index includes topological properties of the vertices (atom parameter) and the interaction of such vertices in a molecular graph (bond parameter). The applicability of *M*-index for the estimation of physical, chemical, and pharmacological properties such as boiling point, enthalpy of formation, refractive index, retention index, toxicity, Gibbs free energy, heat capacity, and etc. has been investigated. We have checked the appropriateness of this index for simple molecular compounds such as un-branched alkanes or cycloalkanes as well as more complicated systems with various functional groups and isomers.

2. DEFINITION OF THE M-INDEX

2.1 EXPANDED WIENER NUMBER

In 1971, Hosoya proposed a modified Wiener number which can be applied to chain and cyclic molecules [18], as follows:

$$W = \frac{1}{2} \sum_{i,j}^N d_{ij} \quad (1)$$

where *N* is the total number of the atoms in a molecule and d_{ij} 's are the elements of a matrix, called *distance matrix*. d_{ij} is defined as the shortest distance between any two given atoms *i* and *j* in a molecule which is clearly equal to zero for all diagonal elements of $i = j$. To obtain a higher discriminating ability of the Wiener number, Tratch et al. [29] proposed a novel topological index, called expanded Wiener number which is defined as

$$\tilde{W} = \sum_{i<j} \tilde{d}_{ij} = \sum_{i<j} \mu_{ij} \nu_{ij} d_{ij}. \quad (2)$$

In general, the vertices *i* and *j* may be connected by several, μ_{ij} , shortest paths and for each of these paths a set of ν_{ij} shortest super-paths of the length equal or greater than d_{ij} must be taken into account. It may be easily shown that the number of shortest superpaths is just the same for each of the shortest paths connecting *i* with *j*. However, because of computationally extensive nature of the Eq. 2, an alternative method was proposed to compute the expanded Wiener number,

$$\tilde{W} = \sum_{i<j} \tilde{f}_{ij} = \sum_{i<j} \mu_{ij} \frac{d_{ij}(d_{ij}+1)(d_{ij}+2)}{6} \quad (3)$$

in which the total length of all subpaths for every pair of vertices is taken into account [29]. Note that the resulting expanded Wiener numbers calculated from Eqs. 2 and 3 are the same.

2.2 MODIFIED ADJACENCY MATRIX

The Wiener index considers the length of shortest paths only, and the properties of vertices as well as their interactions are not included. In this article, we try to introduce a novel topological index by focusing on some information about the structural details in the molecular graph such as the features and interaction of vertices. For this purpose, we modify the conventional form of adjacency matrix A , by inclusion of the bond parameters. The elements of modified adjacency matrix A^* are defined as: $A_{ij}^* = 0$, in the cases where $i = j$ and where two vertices are not connected (non-adjacent vertices), otherwise $A_{ij}^* = b_{ij} X_{ij}$. The parameter b_{ij} represents the bond order between atoms i and j , and is 1, 2, and 3, respectively for the single, double, and triple bonds. Moreover, we have used the definition of Yang et al. [25–28] for the bonding characteristics, $X_{ij} = (1 + \Delta I_{ij}) / R_{ij}$. In this formula, $\Delta I_{ij} = |I_i - I_j|$ stands for the electronegativity difference between atoms i and j , and $R_{ij} = n_i + n_j$, where n_i and n_j are the maximum principal quantum numbers of the atoms i and j . According to its definition, X_{ij} is a measure for the bonding ability between vertices i and j , i.e., the smaller the value of X_{ij} , the weaker the bonding ability between atoms i and j .

2.3 ELEMENTS OF THE \tilde{M} -MATRIX

The modified version of adjacency matrix has been utilized for constructing the M -index in which not only the characteristics of an individual atom but also the role of that atom in establishing the connection with other atoms in a molecular graph are taken into account. The diagonal elements of the \tilde{M} -matrix contain the electronegativity of atom i as the characteristic of that atom and the sum of the i -th row of the A^* matrix divided by k_i (the number of neighboring atoms of atom i) as the average role of atom i in establishing connections with other atoms in the molecule.

$$\tilde{M}_i = I_i + \frac{\sum_j A_{ij}^*}{k_i}. \quad (4)$$

If the vertices i and j are adjacent, then \tilde{M}_{ij} is obtained by employing the inverse of the off-diagonal elements of the expanded Wiener index. Therefore, the elements of \tilde{M} -matrix are summarized as below

$$\tilde{M}_{ij} = \begin{cases} \tilde{M}_i & i = j \\ \frac{1 + I_i I_j A_{ij}^*}{\tilde{f}_{ij}} & i, j \text{ are adjacent} \\ \frac{1}{\tilde{f}_{ij}} & i, j \text{ are non - adjacent.} \end{cases} \quad (5)$$

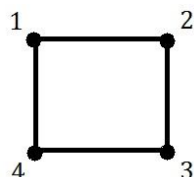
This matrix includes not only the characteristics of atoms but also the pattern of their connections in the molecular graph.

2.4 THE M-INDEX

Taking into account the symmetric property of the \tilde{M} matrix, we introduce our proposed M -index as the sum of the upper triangular and diagonal elements of the \tilde{M} matrix:

$$M = \sum_{i=1}^N \tilde{M}_i + \sum_{i < j}^N \tilde{M}_{ij}. \quad (6)$$

Now, we take an example to illustrate how to get M -index for a simple cyclic graph containing four vertices



The expanded distance matrix of such a graph is obtained by use of Eq. 3:

$$\tilde{f}_{ij} = \begin{pmatrix} 0 & 1*(1*2*3)/6 & 2*(2*3*4)/6 & 1*(1*2*3)/6 \\ 1*(1*2*3)/6 & 0 & 1*(1*2*3)/6 & 2*(2*3*4)/6 \\ 2*(2*3*4)/6 & 1*(1*2*3)/6 & 0 & 1*(1*2*3)/6 \\ 1*(1*2*3)/6 & 2*(2*3*4)/6 & 1*(1*2*3)/6 & 0 \end{pmatrix}.$$

Then, we can find A^* ,

$$A^* = \begin{pmatrix} 0 & b_{12}X_{12} & 0 & b_{14}X_{14} \\ b_{21}X_{21} & 0 & b_{23}X_{23} & 0 \\ 0 & b_{32}X_{32} & 0 & b_{34}X_{34} \\ b_{41}X_{41} & 0 & b_{43}X_{43} & 0 \end{pmatrix}.$$

The diagonal elements, \tilde{M}_i s, can be obtained through Eq. 4.

$$\begin{cases} \tilde{M}_1 = I_1 + (1/2) * (b_{12} X_{12} + b_{14} X_{14}) \\ \tilde{M}_2 = I_2 + (1/2) * (b_{21} X_{21} + b_{23} X_{23}) \\ \tilde{M}_3 = I_3 + (1/2) * (b_{32} X_{32} + b_{34} X_{34}) \\ \tilde{M}_4 = I_4 + (1/2) * (b_{41} X_{41} + b_{43} X_{43}) \end{cases}$$

Finally, the whole \tilde{M} matrix is expressed as

$$\tilde{M} = \begin{pmatrix} \tilde{M}_1 & 1 + I_1 I_2 b_{12} X_{12} & 1/8 & 1 + I_1 I_4 b_{14} X_{14} \\ 1 + I_2 I_1 b_{21} X_{21} & \tilde{M}_2 & 1 + I_2 I_3 b_{23} X_{23} & 1/8 \\ 1/8 & 1 + I_3 I_2 b_{32} X_{32} & \tilde{M}_3 & 1 + I_3 I_4 b_{34} X_{34} \\ 1 + I_4 I_1 b_{41} X_{41} & 1/8 & 1 + I_4 I_3 b_{43} X_{43} & \tilde{M}_4 \end{pmatrix}$$

If we suppose that all vertices of the above graph are carbon the resulting graph represents a cyclobutane molecule in which $b_{12} = b_{23} = b_{34} = b_{41} = 1$, $I_1 = I_2 = I_3 = I_4 = 2.55$, and $n_1 = n_2 = n_3 = n_4 = 2$. These quantities give the value of $M \approx 21.96$ for the cyclobutane.

3. RESULTS AND DISCUSSION

We have carried out comprehensive studies on the physico-chemical properties of a large number of diversified compounds. In this respect, the applicability of the proposed index has been checked for a wide range of properties including: partition coefficient, molar refraction, molar volume, parachor, polarizability, standard enthalpy of formation, toxicity, boiling point, heat capacity, refractive index, and Gibbs free energy. On the other hand, our analysis was based on different category of compounds such as: alkanes, cycloalkanes, silicon/titanium halides, methyl halides, alcohols, aldehydes and ketones, carboxylic acids, as well as isomeric systems. The values of M -index were computed for each compound with a view to study their correlation potential in developing QSPR/QSAR models.

3.1. UN-BRANCHED ALKANES/CYCLOALKANES

In the first test, the proposed M -index is evaluated for the prediction of some physico-chemical properties of a series of simple un-branched alkanes and cycloalkanes. Many properties of alkanes vary in a regular manner with molecular mass and because the alkanes are nonpolar, complexities due to polarity, polarizability, and hydrogen bonding are avoided. Thus, the physico-chemical properties of alkanes are dominated by their inherent structural features, such as molecular dimension or shape. Here, we are mainly concerned with the size effect and consider hydrogen-depleted graphs, i.e. we do not take into account the hydrogen atoms as vertices of the graph.

Table 1 collects the calculated M -index for 19 un-branched alkanes together with the experimental data for the logarithm of partition coefficient in octanol/water ($\log p$), molar refraction (MR), molar volume (MV), parachor (PR), and polarizability (α) [30]. For all studied properties, we have reported the correlation coefficient (R) and the standard error (s). Moreover, the resultant models were validated for generalization and productivity by leave-one-out cross-validation (LOO-CV) method, and calculated R_{cv} and s_{sv} are also given. The statistical significance of the obtained models was confirmed by a high R with a close R_{cv} in value and a small s with a close s_{sv} in value.

Table 1. The calculated M -index and the experimental values of five physico-chemical properties for un-branched alkanes.^a

Compd.	M	$\log p$	MR	MV	PR	α
Methane	0	1.53	11.31	61.50	111.70	4.48
Ethane	8.23	2.06	15.94	78.00	151.50	6.32
Propane	13.90	2.60	20.58	94.50	191.20	8.15
Butane	19.68	3.14	25.21	111.00	231.00	9.99
Pentane	25.50	3.67	29.84	127.50	270.80	11.83
Hexane	31.36	4.21	34.47	144.00	310.60	13.66
Heptane	37.23	4.74	39.11	160.50	350.40	15.50
Octane	43.11	5.28	43.74	177.00	390.20	17.34
Nonane	49.01	5.82	48.37	193.60	430.00	19.17
Decane	54.90	6.35	53.01	210.10	469.70	21.01
Undecane	60.81	6.89	57.64	226.60	509.50	22.85
Dodecane	66.71	7.42	62.27	243.10	549.30	24.28
Tridecane	72.62	7.96	66.90	259.60	589.10	26.52
Tetradecane	78.53	8.50	71.54	276.10	628.90	28.36
Pentadecane	84.45	9.03	76.17	292.60	668.70	30.19
Hexadecane	90.36	9.57	80.80	309.10	708.40	32.03
Heptadecane	96.28	10.10	85.44	325.60	748.20	33.87
Octadecane	102.19	10.64	90.07	342.10	788.00	35.70
Nonadecane	108.11	11.18	94.70	358.60	825.80	37.40

^aExperimental data were taken from [30].

The correlation results for listed properties in Table 1 as well as corresponding statistical quantities are given in Eqs. 7–11. As obvious from these equations, for all five

properties, the R values are near to 1 with very small s . Moreover, the values of R_{cv} and s_{sv} are very close to R and s , and s_{sv} are slightly larger than s indicating that the resulted models are statistically significant and validated for physico-chemical properties of alkanes. Further, there is a good correlativity between M -index and the molecular structure. Inspection of the results in Table 1 reveals that while the considered properties increase with the increase in the alkane size, the values of M -index increase as well.

$$\log p = 1.382 + 0.090M \quad (7)$$

$$R = 0.9999, s = 0.042, R_{cv} = 0.9999, s_{cv} = 0.049 \quad N = 19$$

$$MR = 10.050 + 0.783M \quad (8)$$

$$R = 0.9999, s = 0.354, R_{cv} = 0.9999, s_{cv} = 0.413 \quad N = 19$$

$$MV = 57 + 2.788M \quad (9)$$

$$R = 0.9999, s = 1.264, R_{cv} = 0.9999, s_{cv} = 1.477 \quad N = 19$$

$$PR = 101.100 + 6.714M \quad (10)$$

$$R = 0.9999, s = 3.055, R_{cv} = 0.9999, s_{cv} = 3.563 \quad N = 19$$

$$\alpha = 3.986 + 0.310M \quad (11)$$

$$R = 0.9998, s = 0.168, R_{cv} = 0.9998, s_{cv} = 0.189 \quad N = 19$$

Table 2. The comparison between correlation parameters for modeling physico-chemical properties of alkanes using Sz , PI , and M -indices.^a

	R			s		
	Sz	PI	M	Sz	PI	M
$\log p$	0.8586	0.9397	0.9999	1.957	1.305	0.042
MR	0.8586	0.9397	0.9999	16.914	11.283	0.354
MV	0.8395	0.9402	0.9999	60.194	40.100	1.264
PR	0.8586	0.9397	0.9999	145.251	96.895	3.055
α	0.8592	0.9401	0.9998	6.691	4.457	0.168

^aThe correlation parameters for Sz , PI were taken from [30] and those for M -index were computed in the present work.

The quality of the obtained correlations will be confirmed by comparison of our results with other indices. For this reason, we compared the correlation coefficients and the standard errors of the considered physico-chemical properties obtained by M -index with

those previously reported using Szeged (*Sz*) and Padmakar–Ivan (*PI*) indices in Table 2 [30]. The quality parameters presented in Table 2 confirm the superiority of the *M*–index over *Sz* and *PI* for modeling property/activity of alkanes. Inspection of the reported data in Table 2 reveals that for all five properties, the models base on the *M*–index provides significantly less standard errors.

Now, we extend our QSPR models to normal (un–branched) cycloalkanes. Listed in Table 3 are the values of *M*–index and similar experimental physico–chemical properties for considered cycloalkanes [30]. The corresponding linear correlation results are as follows:

$$\log p = 0.148 + 0.090M \quad (12)$$

$$R = 1.0000, s = 0.012, R_{cv} = 1.0000, s_{cv} = 0.013 \quad N = 17$$

$$MR = 1.226 + 0.770M \quad (13)$$

$$R = 1.0000, s = 0.122, R_{cv} = 1.0000, s_{cv} = 0.133 \quad N = 17$$

$$MV = 4.754 + 2.963M \quad (14)$$

$$R = 1.0000, s = 0.452, R_{cv} = 1.0000, s_{cv} = 0.503 \quad N = 17$$

$$PR = 10.960 + 6.684M \quad (15)$$

$$R = 1.0000, s = 0.905, R_{cv} = 1.0000, s_{cv} = 0.994 \quad N = 17$$

$$\alpha = 0.494 + 0.305M \quad (16)$$

$$R = 1.0000, s = 0.045, R_{cv} = 1.0000, s_{cv} = 0.049 \quad N = 17$$

The Eqs.12–16 can outstandingly reproduce the physico–chemical properties of cycloalkanes which in turn imply that *M*–index can successfully be applied for cyclic structures as well as non-cyclic alkanes.

3.2 INORGANIC COMPOUNDS OF SILICON/TITANIUM HALIDES

To assess the applicability of *M*–index for inorganic compounds, we considered standard enthalpy of formation for a series of silicon/titanium halides with the general formula of $Si(Ti)X_m$ ($X=F, Cl, Br, I$ and $m=1,2,3,4$). The calculated *M*–index and the corresponding experimental standard enthalpy of formation for studied systems are given in Table 4 [31,32].

Table 3. The calculated M -index and the experimental values of five physico-chemical properties for un-branched cycloalkanes.^a

Compd.	M	$\log p$	MR	MV	PR	α
Cyclopropane	16.28	1.61	13.83	53.20	120.10	5.48
Cyclobutane	21.96	2.14	18.44	70.90	160.10	7.31
Cyclopentane	28.38	2.68	23.05	88.70	200.20	9.14
Cyclohexane	34.20	3.22	27.67	106.40	240.20	10.96
Cycloheptane	40.43	3.75	32.28	124.10	280.20	12.79
Cyclooctane	46.31	4.29	36.69	141.90	320.30	14.62
Cyclononane	52.43	4.82	41.50	159.60	360.30	16.41
Cyclodecane	58.33	5.36	46.11	177.40	400.40	18.28
Cycloundecane	64.40	5.90	50.72	195.10	440.40	20.11
Cyclododecane	70.30	6.43	55.34	212.80	480.40	21.93
Cyclotridecane	76.34	6.97	59.95	230.60	520.50	23.76
Cyclotetradecane	82.25	7.50	64.54	248.30	560.50	25.59
Cyclopentadecane	88.26	8.04	69.17	266.00	600.60	27.42
Cyclohexadecane	94.18	8.58	73.78	283.80	640.60	29.25
Cycloheptadecane	100.05	9.11	78.39	301.50	680.60	31.07
Cyclooctadecane	106.09	9.65	83.01	319.30	720.70	32.90
Cyclononadecane	112.07	10.18	87.67	337.60	760.70	34.73

^aExperimental data were taken from [30].

The linear correlation equations are given in Eqs. 17 and 18. For both SiX_m and TiX_m series, there are good linear correlations with R equals to 0.9614 and 0.9631, respectively.

$$\Delta_f H_{Si}^0 = 658.100 - 45.720M \quad (17)$$

$$R = 0.9531, s = 156, R_{cv} = 0.9374, s_{cv} = 160 \quad N = 16$$

$$\Delta_f H_{Ti}^0 = 583.900 - 51.290M \quad (18)$$

$$R = 0.9586, s = 146, R_{cv} = 0.9489, s_{cv} = 146 \quad N = 16$$

Table 4. The calculated M -index and the experimental standard enthalpy of formation, $\Delta_f H^0$ (kJ/mol), for silicon/titanium halides.^a

Compd.	M	$\Delta_f H^0$	Compd.	M	$\Delta_f H^0$
SiF	12.77	-20.92	TiF	11.18	-66.90
SiF ₂	23.27	-589.94	TiF ₂	20.50	-688.30
SiF ₃	34.03	-999.98	TiF ₃	30.07	-1188.70
SiF ₄	45.03	-1625.90	TiF ₄	39.88	-1551.40
SiCl	9.07	154.81	TiCl	8.27	154.40
SiCl ₂	16.12	-167.78	TiCl ₂	14.88	-282.40
SiCl ₃	23.42	-334.72	TiCl ₃	21.73	-539.30
SiCl ₄	30.97	-662.75	TiCl ₄	28.84	-763.20
SiBr	8.10	196.65	TiBr	7.48	212.50
SiBr ₂	14.26	-46.02	TiBr ₂	13.38	-179.10
SiBr ₃	20.67	-158.99	TiBr ₃	19.52	-374.90
SiBr ₄	27.33	-415.47	TiBr ₄	25.91	-550.20
SiI	7.11	259.41	TiI	6.63	274.10
SiI ₂	12.35	92.05	TiI ₂	11.75	-57.70
SiI ₃	17.85	58.58	TiI ₃	17.11	-149.80
SiI ₄	23.59	-110.46	TiI ₄	22.72	-287.00

^aExperimental data were taken from [31,32].

Similarly, for the whole set containing all halides of Table 4 there is a good correlation result as well. By taking into account the 32 compounds as a whole, the obtained QSPR equation for the correlation between M -index and $\Delta_f H^0$ is as follow

$$\Delta_f H_{Si+Ti}^0 = 601.200 - 47.380M \quad (19)$$

$$R = 0.9362, s = 176, R_{cv} = 0.9281, s_{cv} = 168 \quad N = 32$$

The overall indication of these results is that the M -index can be applied not only for organic compounds but also for inorganic compounds containing transition metals.

3.3 METHYL HALIDES

The usefulness of newly constructed topological indices was demonstrated by correlating standard enthalpy of formation of methyl halides. These sets of compounds contain only five atoms. For such small molecules, hydrogen atoms have great impact on their properties and they cannot be neglected. Thus, the hydrogen atoms must be treated as vertices in the molecular graph. Table 5 contains the calculated M -index and the experimental values of the standard enthalpy of formation for 41 methyl halides [33]. By using the linear regression analysis, we obtain

$$\Delta_f H^0 = 1171 - 40.120M \quad (20)$$
$$R = 0.9940, s = 30.000, R_{cv} = 0.9933, s_{cv} = 30.660 \quad N = 41$$

Again, the obtained statistical quantities indicate that there is a good correlation between M -index and $\Delta_f H^0$ for methyl halides compounds.

3.4 SATURATED ALCOHOLS, KETONES, DIOLS, CARBOXYLIC ACIDS

In this part, we intend to study the applicability of our proposed index to predict relative toxic potency of aliphatic compounds. Here we consider the population growth inhibition of the ciliate *Tetrahymena pyriformis* to develop such a toxicity-based QSAR. *T. pyriformis* is one of the generally used ciliated protozoa [34,35] in which diverse endpoints can be used to originate the cytotoxic effects. The experimental $\log(1/IGC_{50})$ values for the four groups of aliphatic compounds [36] under consideration as well as the resulted QSAR models are provided in Table 6. The statistical parameters reported in Table 6 demonstrate very good consistency between R and R_{cv} . The small s values indicate that the M -index can be successfully used as a way for quantifying toxicity of aliphatic compounds even when they have not exhibit a common skeleton requirement of QSAR analysis.

The suitable quality criteria to judge present results can be set up through the comparison with other theoretical predictions for the toxicity of these molecular sets. Roy et. al. [37] reported results for QSAR calculations on these aliphatic compounds using electrophilicity as a possible descriptor. Their statistical parameters for different molecular set (for instance; diols: $R=0.899$, $s=0.486$ and ketones: $R=0.882$, $s=0.612$) are clearly inferior with respect to present results (diols: $R=0.9864$, $s=0.192$ and ketones: $R=0.9850$, $s=0.225$, Table 6).

Table 5. The calculated M -index and the experimental standard enthalpy of formation, $\Delta_f H^0$ (kJ/mol), for methyl halides.^a

Compd.	M	$\Delta_f H^0$	Compd.	M	$\Delta_f H^0$
CH ₂ F ₂	40.43	-452.9	CH ₃ Br	28.94	-37.7
CCl ₂ F ₂	42.17	-493.3	CBrClF ₂	41.04	-471.5
CH ₂ Cl ₂	30.94	-95.4	CH ₂ BrCl	29.81	-50.2
CHCl ₂ F	36.56	-284.9	CHBrClF	35.43	-295.0
CBr ₂ F ₂	39.91	-429.7	CH ₂ BrI	27.59	50.2
CBr ₂ Cl ₂	30.42	-29.3	CHF ₃	46.05	-693.3
CH ₂ Br ₂	28.68	-14.8	CF ₃ I	44.70	-589.9
CHBr ₂ F	34.30	-223.4	CCl ₃ F	37.43	-284.9
CHBr ₂ Cl	29.55	-20.9	CHF ₂ Cl	41.30	-483.7
CBr ₂ ClF	35.17	-231.8	CF ₃ Cl	46.92	-707.9
CHBr ₃	28.42	16.7	CF ₃ Br	45.79	-648.9
CBr ₃ F	34.04	-190.0	CH ₃ F	34.82	-237.7
CBr ₃ Cl	29.29	12.6	CH ₂ I ₂	26.49	118.4
CHI ₃	25.14	210.9	CH ₂ FCI	35.69	-264.4
CF ₄	51.67	-933.0	CH ₂ CII	28.72	12.6
CCl ₄	32.68	-95.8	CH ₃ I	27.84	13.8
Cl ₄	23.79	262.9	CH ₂ FBr	34.56	-252.7
CBr ₄	28.16	79.5	CH ₄	29.20	-74.9
CHBrF ₂	40.17	-463.6	CBrCl ₃	31.55	-37.2
CHBrCl ₂	30.68	-58.6	CHCl ₃	31.81	-102.9
CBrCl ₂ F	36.30	-269.4			

^aExperimental data were taken from [33].

An alternative manner for predicting the toxicity is utilizing topological parameters derived from the electron density, as previously done in our research group [12]. As shown in Ref. [12], predictions improve significantly with respect to the results obtained by Roy et

al. [37]. However, despite its successful in predicting the toxicity, such approach is very computational demanding because it is necessary to perform quantum chemical calculation on each molecule and to derive the topological properties of the electron densities from the calculated wave functions. On the contrary, the most important advantage of the proposed M -index is its computational ease with no need to complicated calculations.

Table 6. The calculated M -index and the experimental values of the toxicity, $\log 1/IGC_{50}$, for aliphatic compounds.^a

Compd.	M	$\log 1/IGC_{50}$	Correlation Equation ^b
<i>Saturated alcohols</i>			
1-Propanol	83.61	-1.7464	
2-Propanol	83.67	-1.8819	
1-Butanol	104.64	-1.4306	
(±)-2-Butanol	104.74	-1.5420	
2-Methyl-1-propanol	104.87	-1.3724	
2-Pentanol	125.93	-1.1596	
3-Pentanol	125.97	-1.2437	
3-Methyl-2-butanol	126.26	-0.9959	
2-Methyl-1-butanol	126.16	-0.9528	
3-Methyl-1-butanol	126.13	-1.0359	$\log 1/IGC_{50} = -3.721 + 0.022M$
2,2-Dimethyl-1-propanol	126.58	-0.8702	(21)
2-Methyl-2-propanol	104.99	-1.7911	$R = 0.9932, s = 0.164,$
1-Hexanol	147.09	-0.3789	$R_{cv} = 0.9914, s_{cv} = 0.176 \quad N = 21$
3,3-Dimethyl-1-butanol	148.06	-0.7368	
1-Heptanol	168.44	0.1050	
1-Octanol	189.83	0.5827	
1-Nonanol	211.26	0.8551	
1-Decano	232.72	1.3354	
1-Undecanol	254.20	1.9547	

1-Dodecanol	275.70	2.1612	
1-Tridecanol	297.22	2.4497	
Ketones			
Acetone	64.52	-2.2036	
2-Butanone	85.27	-1.7457	
2-Pentanone	106.29	-1.2224	
3-Pentanone	106.19	-1.4561	
4-Methyl-2-pentanone	127.75	-1.2085	
2-Heptanone	148.73	-0.4872	
5-Methyl-2-hexanone	149.08	-0.6459	$\log 1/IGC_{50} = -3.446 + 0.020M$
4-Heptanone	148.51	-0.6690	(22)
2-Octanone	170.07	-0.1455	$R = 0.9850, s = 0.225$
2-Nonanone	191.46	0.6598	$R_{cv} = 0.9779, s_{cv} = 0.25 \quad N = 15$
2-Decanone	212.89	0.5822	
3-Decanone	212.69	0.6265	
2-Undecanone	234.35	1.5346	
2-Dodecanone	255.84	1.6696	
7-Tridecanone	276.90	1.5214	
Alcohols: diols			
(±)-1,2-Butanediol	118.23	-2.0482	
(±)-1,3-Butanediol	118.22	-2.3013	
1,4-Butanediol	118.12	-2.2365	$\log 1/IGC_{50} = -4.709 + 0.021M$
1,2-Pentanediol	139.46	-1.6269	(23)
1,5-Pentanediol	139.33	-1.9344	$R = 0.9864, s = 0.192$
(±)-1,2-Hexanediol	160.77	-1.2669	$R_{cv} = 0.9653, s_{cv} = 0.269 \quad N = 9$
1,6-Hexanediol	160.62	-1.4946	
1,2-Decanediol	246.46	0.7640	
1,10-Decanediol	246.30	0.2240	
Carboxylic acids			

Propanoic acid	78.00	-0.5123	
Butyric acid	98.91	-0.5720	
Valeric acid	120.01	-0.2674	
Hexanoic acid	141.23	-0.2083	
Heptanoic acid	162.53	-0.1126	
Octanoic acid	183.90	0.0807	
Nonanoic acid	181.25	0.3509	
Decanoic acid	226.75	0.5063	$\log 1/IGC_{50} = -1.181 + 0.007M$ (24)
Undecanoic acid	248.22	0.8983	$R = 0.9429, s = 0.124$
Iso-Butyric acid	99.04	-0.3334	$R_{cv} = 0.9197, s_{cv} = 0.139 \quad N = 20$
Isovaleric acid	120.26	-0.3415	
Trimethylacetic acid	120.52	-0.2543	
3-Methylvaleric acid	141.62	-0.2331	
4-Methylvaleric acid	141.55	-0.2724	
2-Ethylbutyric acid	141.52	-0.1523	
2-Propylpentanoic acid	184.33	0.0258	
2-Ethylhexanoic acid	184.29	0.0756	
Crotonic acid	81.91	-0.5448	
trans-2-Pentenoic acid	102.55	-0.2774	
trans-2-Hexenoic acid	123.50	-0.1279	

^aExperimental data were taken from [36].

^bFor each series, the correlation equations and statistical quantities were presented in the last column.

3.5. KETONES AND ALDEHYDES

Molar refraction (MR) is a particularly useful physical parameter in chemistry, biological chemistry, and pharmaceutical science because it is closely related to the bulkiness and polarizability of a molecule. We have checked the ability of M -index to predict the molar refraction of a set containing 22 aldehydes and 24 ketones listed in Table 7 [38]. The relationship between MR and M -index is give below

$$MR = 2.223 + 0.217M \quad (25)$$

$$R = 0.9997, s = 0.262, R_{cv} = 0.9997, s_{cv} = 0.266$$

This linear equation indicates an outstanding correlation with high R and small s , implying the reliability of M -index for prediction of molar refraction.

Table 7. The calculated M -index and the experimental molar refraction for aldehydes and ketones.^a

Compd.	M	MR	Compd.	M	MR
Acetaldehyde	44.24	11.5829	2-Butanone	85.27	20.6039
Propionaldehyde	64.74	16.1632	2-Pentanone	106.29	25.2926
Butyl aldehyde	85.60	20.8011	3-Pentanone	106.19	25.2487
2-Methyl propanal	87.31	20.8219	3-Methyl-2-butanone	102.74	25.2603
Pentaldehyde	106.66	25.4983	2-Hexanone	127.46	29.9308
2-Methyl butanal	106.80	25.3943	3-Hexanone	127.31	29.7251
3-Methyl butanal	106.90	25.5327	3-Methyl-2-pentanone	127.75	29.9453
Hexanal	127.85	30.9280	4-Methyl-2-pentanone	127.75	29.9877
2-Methylpentanal	128.02	29.8497	3,3-Dimethyl-2-butanone	128.14	29.6748
2-Ethylbutanal	128.07	29.9981	2-Heptanone	148.73	34.5663
2,3-Dimethylbutanal	128.36	30.0640	3-Heptanone	148.56	34.4230
Heptanal	149.14	34.7004	4-Heptanone	148.51	34.3083
2,2-Dimethylpentanal	149.84	34.7537	5-Methyl-2-hexanone	149.08	34.5773
Octanal	170.49	39.4396	2-Octanone	170.07	39.1959
2-Ethylhexanal	170.78	39.2395	5-Octanone	169.81	39.0616
2-Ethyl-3-	171.38	38.9423	6-Methyl-3-	170.25	38.9478

methylpentanal			heptanone		
Nonanal	191.89	44.2669	2-Nonanone	191.46	43.3542
3,5,5-	193.64	43.9887	5-Nonanone	191.15	43.8710
Trimethylhexanal					
Decanal	213.33	48.6737	2,6-Dimethyl-4-	191.88	43.8902
			heptanone		
2-Methyldecanal	235.00	53.0003	2-Decanone	212.89	48.5304
Dodecanal	256.28	58.0913	2-Undecanone	234.35	52.7129
2-Methylundecanal	256.49	57.9284	6-Undecanone	233.97	53.2109
Acetone	64.52	16.2963	2-Methyl-4-	255.83	57.7027
			undecanone		

^aExperimental data were taken from [38].

3.6. ISOMERIC SYSTEMS

One of the main drawbacks of the most topological indices is their poor discrimination of isomers and the index has the same value for different isomeric compounds. It is well-known that this degeneracy increases when the number of atoms in the molecule increases, even for simple molecules such as alkanes. In previous studies the capacity of one index to discriminate isomers was measured by using a discrimination index, D , which has been calculated as the number of isomers having different values of the index divided by the total number of isomers [39].

Similar to other descriptors, the M -index introduced in the present work requires some modifications to be applicable for isomers. The proposed M -index in Eq. 6 was constructed by parameters which are only dependent on the number of atoms, bond strength and the property of individual atoms. Thus, the calculated values of the M -index for the structural isomers of a specific molecular formula are very close together and cannot well discriminate the isomers. To increase the discriminative power of the M -index, we introduce a quantity in which the effect of different configurations due to structural isomers is taken into account. For the isomeric systems, we first calculate q as

$$q_i = \sum_{d=1} \frac{n_i(d)}{d}, \quad (26)$$

where d is the distance and $n_i(d)$ is the number of neighboring nodes in the distance d of the vertex i . Finally, we defined the discrimination parameter, \tilde{D} used in the present study as

$$\tilde{D} = \left(\frac{\sum_{i=1}^N q_i}{N} \right), \quad (27)$$

where N is the total number of vertices in the molecular graph. Finally, the appropriate form of our proposed index designed for isomeric systems is $\tilde{D}^\beta M$, where β is a free parameter dependent on the considered property.

The predictive ability of M -index modified by discrimination parameter has been checked within two sets of isomeric compounds. First, a large set of 88 aldehydes and ketones and their boiling points [40–42] (Table 8) was taken to construct model. For such a data set, the correlative model was obtained by setting $\beta = -1$ as following equation

$$Bp = -172.800 + 15.590 \tilde{D}^{-1} M \quad (28)$$

$R = 0.9922, s = 7.557, R_{cv} = 0.9918, s_{cv} = 7.647 \quad N = 88.$

Table 8. The calculated M -index and the experimental boiling points ($^{\circ}\text{C}$) for aldehydes and ketones. ^a

Compd.	Bp	$\tilde{D}^\beta M$	Compd.	Bp	$\tilde{D}^\beta M$
Acetaldehyde	20.8	12.39	5-Methyl-2-hexanone	144.0	20.02
Propionaldehyde	48.8	14.07	2-Methyl-3-hexanone	135.0	20.10
Butyl aldehyde	75.7	15.69	4-Methyl-3-hexanone	134.0	19.99
2-Methyl propanal	64.4	15.87	5-Methyl-3-hexanone	135.0	20.08
Pentaldehyde	103.0	17.24	2,2-Dimethyl-3-pentanone	125.0	19.75
2-Methyl butanal	92.5	17.09	2,4-Dimethyl-3-pentanone	125.0	19.89
3-Methyl butanal	92.5	17.05	4,4-Dimethyl-3-pentanone	126.0	19.75
2,2-Dimethylpropanal	77.5	16.88	2-Octanone	172.5	21.59
Hexanal	128.0	18.71	3-Octanone	167.5	21.65

2-Methylpentanal	117.0	18.55	4-Octanone	165.5	21.68
3-Methylpentanal	118.0	18.46	2-Methyl-4-heptanone	154.0	21.47
2-Ethylbutanal	117.0	18.50	3-Methyl-4-heptanone	153.0	21.37
Heptanal	152.8	20.13	3-Methyl-2-heptanone	164.0	21.31
3-Methylhexanal	143.0	19.85	6-Methyl-2-heptanone	167.0	21.38
2,2-Dimethylpentanal	126.5	19.66	2-Methyl-3-heptanone	158.0	21.48
Octanal	171.0	21.50	3,3-Dimethyl-2-hexanone	151.5	20.92
2-Ethylhexanal	160.0	21.26	2,2-Dimethyl-3-hexanone	146.0	21.03
2-Propylpentanal	160.0	21.23	2,5-Dimethyl-3-hexanone	147.5	21.27
Nonanal	191.0	22.83	4,4-Dimethyl-3-hexanone	148.0	20.95
3,5,5-Trimethylhexanal	170.5	21.96	2,2,4-Trimethyl-3-pentanone	135.1	20.91
Decanal	208.5	24.12	2-Nonanone	195.0	22.91
Undecanal	233.0	25.38	3-Nonanone	190.0	23.28
2-Methyldecanal	229.0	24.33	4-Nonanone	187.5	23.31
Dodecanal	254.0	26.61	5-Nonanone	188.4	23.02
2-Methylundecanal	246.0	26.48	7-Methyl-3-octanone	182.5	22.77
Tridecanal	267.0	27.82	3-Methyl-4-octanone	174.0	22.70
Tetradecanal	287.0	29.00	7-Methyl-4-octanone	178.0	22.80
Pentadecanal	304.0	30.17	2,6-Dimethyl-4-heptanone	169.4	22.61
Acetone	56.2	14.18	3,5-Dimethyl-4-heptanone	162.0	22.37
2-Butanone	79.6	15.81	2,2,4,4-Tetramethyl-3-pentanone	152.0	21.89
2-Pentanone	102.0	17.35	2-Decanone	210.0	24.20
3-Pentanone	101.7	17.39	3-Decanone	211.0	24.26
3-Methyl-2-butanone	93.5	16.57	4-Decanone	206.5	24.29
2-Hexanone	127.6	18.82	2-Undecanone	231.5	25.45
3-Hexanone	123.5	18.87	3-Undecanone	227.0	25.51

3-Methyl-2-pentanone	118.0	18.57	5-Undecanone	227.0	25.57
4-Methyl-2-pentanone	117.0	18.61	6-Undecanone	226.0	25.58
2-Methyl-3-pentanone	115.0	18.75	2-Dodecanone	246.5	26.68
3,3-Dimethyl-2-butanone	106.0	18.33	2-Tridecanone	263.0	27.88
2-Heptanone	151.4	20.23	7-Tridecanone	261.0	28.02
3-Heptanone	147.0	20.29	2-Methyl-3-tridecanone	267.0	28.99
4-Heptanone	144.0	20.31	7-Ethyl-2-methyl-4-undecanone	252.5	28.29
3-Methyl-2-hexanone	143.5	19.95	2-Pentadecanone	294.0	30.23
4-Methyl-2-hexanone	139.0	19.93	8-Pentadecanone	291.0	30.37

^aExperimental data were taken from [40–42].

In this model, there is a high $R(0.9922)$ and a small $s(7.557)$ value. The values of $R_{cv}(0.9918)$ and $s_{cv}(7.647)$ are very close to the values of R and s , and s_{cv} are only slightly larger than s . The correlation results and the cross-validation results demonstrate that the obtained model is statistically significant and validated.

In the second attempt, six physico-chemical properties of 77 hydrocarbons [43] (Table 9) including isomers have been studied. For heat capacity and boiling point β is zero, while for density, refractive index, Gibbs free energy, and the standard enthalpy of formation the obtained values for β are, respectively, 2, 2, 3.5, and -0.9 . The correlation equations for these six properties are given below

$$C_p = 20 + 3.851M \quad (29)$$

$$R = 0.9899, s = 3.886, R_{cv} = 0.9894, s_{cv} = 3.926 \quad N = 77$$

$$Bp = -66.810 + 4.068M \quad (30)$$

$$R = 0.9849, s = 5.035, R_{cv} = 0.9838, s_{cv} = 5.145 \quad N = 77$$

$$\rho = 633.5 + 0.111\tilde{D}^2M \quad (31)$$

$$R = 0.9495, s = 8.291, R_{cv} = 0.9460, s_{cv} = 8.455 \quad N = 77$$

$$RI = 1.360 + 5.752e - 5\tilde{D}^2M \quad (32)$$

$$R = 0.9584, s = 3.866e - 3, R_{cv} = 0.9553, s_{cv} = 4.010e - 3 \quad N = 77$$

$$\Delta G = -7.522 + 5.171e - 3\tilde{D}^{3.5}M \quad (33)$$

$$R = 0.9610, s = 3.675, R_{cv} = 0.9586, s_{cv} = 3.735 \quad N = 77$$

$$\Delta_f H^0 = -24.180 + 4.305\tilde{D}^{-0.9}M \quad (34)$$

$$R = 0.9819, s = 0.923, R_{cv} = 0.9810, s_{cv} = 0.935 \quad N = 77$$

As can be seen, all equations have acceptable quality and the propose index used in these equations can explain more than 95% of the variance in the considered physico-chemical properties. The high correlation coefficients of cross validation show that the constructed models are statistically significant.

Table 9. The calculated M -index and the experimental heat capacity (C_p , J/mol.K), boiling point (Bp , °C), density (ρ , kg/m³), refractive index (RI), Gibbs free energy (ΔG , kJ/mol), and the standard enthalpy of formation ($\Delta_f H^0$, kJ/mol) for hydrocarbons.^a

Compd.	C_p	Bp	$\tilde{D}^\beta M$	ρ	RI	$\tilde{D}^\beta M$	ΔG	$\tilde{D}^\beta M$	$\Delta_f H^0$	$\tilde{D}^\beta M$
3-Methylpentane	140.88	63.28	31.63	659.76	1.3739	289.95	-2.12	1527.59	26.32	11.67
2,2-Dimethylbutane	142.26	49.74	31.98	644.46	1.3660	320.67	-7.42	1807.01	25.40	11.33
2,3-Dimethylbutane	140.21	57.99	31.83	657.02	1.3723	308.06	-1.77	1690.50	24.77	11.46
3-Methylhexane	164.50	91.85	37.53	682.88	1.3861	413.46	6.60	2500.07	30.71	12.75
3-Ethylpentane	166.80	93.48	37.60	693.92	1.3911	423.81	12.70	2606.91	31.71	12.64
2,2-Dimethylpentane	167.70	79.17	37.90	669.48	1.3800	451.77	2.10	2898.00	29.50	12.43
2,3-Dimethylpentane	161.80	89.75	37.80	690.81	1.3895	444.39	7.60	2821.19	28.62	12.47
2,4-Dimethylpentane	171.70	80.47	37.70	668.23	1.3788	430.98	4.90	2679.30	29.58	12.60
3,3-Dimethylpentane	166.70	86.04	38.00	689.16	1.3884	465.55	4.80	3048.35	29.33	12.31
2,2,3-Trimethylbutane	164.20	80.86	38.20	685.64	1.3869	487.29	6.30	3288.92	28.28	12.15
N-Octane	188.70	125.68	43.11	698.54	1.3951	508.91	17.67	3240.89	38.12	14.20
2-Methylheptane	188.20	117.65	43.35	693.87	1.3926	538.66	13.37	3564.95	35.82	13.95
3-Methylheptane	186.82	118.93	43.43	701.73	1.3961	552.51	13.79	3721.67	35.31	13.83
2,4-Dimethylhexane	193.35	109.43	43.69	696.17	1.3929	586.01	13.07	4107.42	33.76	13.58
2,5-Dimethylhexane	186.52	109.11	43.59	689.37	1.3900	570.22	11.40	3921.96	33.39	13.71

3,3-Dimethylhexane	191.96	111.97	43.96	705.95	1.3978	620.91	15.13	4523.98	33.43	13.35
3,4-Dimethylhexane	182.72	117.73	43.81	715.15	1.4018	605.15	18.43	4336.01	32.47	13.44
3-Ethyl-2-Methylpentan	193.05	115.66	43.83	715.20	1.4017	609.52	20.68	4389.42	34.31	13.41
2,2,3-Trimethylpentane	186.77	109.84	44.23	712.03	1.4007	657.01	19.45	4971.27	32.13	13.13
2,3,3-Trimethylpentane	188.20	114.77	44.28	722.30	1.4052	664.88	20.04	5071.69	32.17	13.08
2,3,4-Trimethylpentane	192.72	113.47	44.03	715.09	1.4020	633.00	20.76	4673.57	32.55	13.27
2,2,3,3-Tetramethylbutane	188.28	106.29	44.68	729.88	1.4057	714.87	24.04	5718.96	31.84	12.83
2-Methyloctane	210.90	143.28	49.25	709.60	1.4008	689.31	21.60	4988.20	40.42	15.02
3-Methyloctane	209.70	144.23	49.33	716.70	1.4040	706.26	22.00	5197.81	39.92	14.89
4-Ethylheptane	214.30	141.20	49.49	722.30	1.4067	738.47	26.80	5606.79	40.50	14.66
2,2-Dimethylheptane	212.40	132.82	49.72	706.60	1.3995	756.29	19.50	5824.82	38.83	14.61
2,3-Dimethylheptane	207.70	140.50	49.66	722.00	1.4064	755.25	23.50	5816.87	37.82	14.59
2,4-Dimethylheptane	217.10	133.20	49.63	711.50	1.4011	751.95	20.80	5774.96	38.16	14.60
2,5-Dimethylheptane	208.20	136.00	49.58	713.60	1.4015	742.75	18.20	5655.51	37.53	14.67
2,6-Dimethylheptane	210.40	135.22	49.49	704.50	1.3985	724.43	19.80	5421.28	37.99	14.79
3,3-Dimethylheptane	214.00	137.02	49.89	721.60	1.4063	787.88	22.00	6241.86	38.20	14.41
3,4-Dimethylheptane	206.80	140.40	49.76	727.50	1.4091	775.63	24.90	6084.78	37.02	14.46
3,5-Dimethylheptane	214.60	135.70	49.69	716.60	1.4046	762.93	22.00	5917.94	38.07	14.54
3-Ethyl-3-Methylhexane	214.10	140.60	50.06	736.00	1.4134	821.89	30.50	6703.42	37.36	14.21
4-Ethyl-2-Methylhexane	219.70	133.80	49.72	724.20	1.4054	770.66	24.50	6020.30	39.25	14.48
2,2,4-Trimethylhexane	210.70	129.91	50.09	711.80	1.4010	817.86	23.60	6643.10	36.61	14.25
2,3,3-Trimethylhexane	213.30	137.69	50.26	733.50	1.4119	847.96	29.40	7058.75	36.28	14.09
2,3,4-Trimethylhexane	214.00	138.96	50.06	735.10	1.4120	821.89	28.60	6703.42	36.86	14.21
2,3,5-Trimethylhexane	212.50	131.36	49.92	717.90	1.4037	795.75	22.20	6348.35	36.02	14.36
2,4,4-Trimethylhexan	213.50	130.66	50.16	720.05	1.4052	831.08	26.60	6824.87	36.44	14.18

3-Ethyl-2,2-Dimethylpentane	205.00	133.84	50.31	731.00	1.4101	857.89	37.50	7199.46	36.16	14.04
2,2,3,3-Tetramethylpentane	213.34	140.29	50.81	752.97	1.4214	929.70	39.00	8225.65	35.86	13.73
2,2,3,4-Tetramethylpentane	208.50	133.03	50.51	735.22	1.4125	884.63	36.70	7574.18	35.06	13.92
3-Ethylheptane	235.80	166.50	55.37	735.90	1.4136	903.92	34.90	7341.39	45.31	15.76
4-Ethylheptane	236.50	163.64	55.42	734.30	1.4131	917.34	33.40	7528.24	45.10	15.67
2,2-Dimethylheptane	235.10	156.90	55.63	720.80	1.4060	928.65	27.70	7669.23	43.43	15.67
2,5-Dimethylheptane	231.80	158.50	55.53	726.40	1.4089	924.58	26.90	7620.98	41.92	15.66
3,4-Dimethylheptane	229.30	163.40	55.69	741.80	1.4159	956.90	33.00	8075.93	41.80	15.49
3,5-Dimethylheptane	238.30	159.40	55.64	732.90	1.4115	948.37	29.10	7955.70	42.47	15.53
3,6-Dimethylheptane	229.60	160.80	55.59	732.90	1.4115	935.26	28.90	7769.81	41.63	15.61
4,4-Dimethylheptane	239.30	157.50	55.87	731.20	1.4122	983.33	31.90	8449.58	42.30	15.37
4,5-Dimethylheptane	230.10	162.13	55.72	743.20	1.4167	965.17	35.30	8194.87	41.51	15.44
4-N-Propylheptane	237.70	157.50	55.46	732.10	1.4113	927.67	38.20	7673.21	44.85	15.61
4-Isopropylheptane	239.20	158.90	55.77	735.40	1.4132	979.10	37.90	8397.48	43.10	15.36
2-Methyl-3-Ethylheptane	238.50	161.20	55.73	739.80	1.4151	967.48	35.70	8228.62	43.30	15.43
2-Methyl-4-Ethylheptane	243.40	156.20	55.69	732.20	1.4114	962.17	31.60	8153.95	43.64	15.45
3-Methyl-4-Ethylheptane	236.20	162.20	55.84	746.60	1.4183	992.87	36.90	8597.02	42.47	15.29
3-Methyl-5-Ethylheptane	240.90	158.20	55.74	736.80	1.4141	970.77	33.10	8276.38	43.35	15.41
2,2,3-Trimethylheptane	232.50	157.60	56.13	738.50	1.4145	1021.76	34.80	9004.99	41.30	15.21
2,3,3-Trimethylheptane	235.10	160.20	56.21	748.80	1.4202	1039.31	37.30	9267.17	41.00	15.12
2,3,4-Trimethylheptane	237.60	159.90	56.03	748.50	1.4195	1016.82	37.20	8940.44	40.96	15.20
2,3,5-Trimethylheptane	233.90	160.70	55.94	754.50	1.4169	997.74	30.30	8659.63	40.12	15.30
2,3,6-Trimethylheptane	228.50	156.00	55.82	734.70	1.4131	972.26	28.50	8289.11	39.75	15.43

2,4,5-Trimethylheptane	234.10	156.50	55.96	737.30	1.4160	1002.85	36.90	8734.94	39.98	15.27
3,3,5-Trimethylheptane	234.10	155.68	56.19	739.00	1.4170	1038.98	34.10	9264.22	40.46	15.12
3,4,4-Trimethylheptane	235.60	161.10	56.33	753.50	1.4235	1067.62	40.30	9697.26	40.08	14.99
3,4,5-Trimethylheptane	235.10	162.50	56.11	751.90	1.4229	1034.31	39.70	9201.17	41.14	15.12
2-Methyl-3-Isopropylhexane	231.80	166.70	56.09	743.60	1.4172	1033.98	46.80	9198.24	40.46	15.12
2,3-Dimethyl-3-Ethylhexane	238.20	163.70	56.42	759.98	1.4247	1085.61	45.00	9974.15	40.71	14.91
2,3-Dimethyl-4-Ethylhexane	243.00	160.90	56.14	751.60	1.4203	1042.95	42.10	9332.13	42.43	15.08
2,4-Dimethyl-4-Ethylhexane	235.00	160.10	56.29	751.40	1.4202	1061.98	42.30	9612.80	40.29	15.01
3,3-Dimethyl-4-Ethylhexane	228.20	162.90	56.44	759.80	1.4246	1090.98	50.00	10057.7 9	39.92	14.89
3,4-Dimethyl-4-Ethylhexane	235.50	162.10	56.49	759.60	1.4244	1100.24	47.60	10200.8 9	40.42	14.85
2,2,3,3-Tetramethylhexane	238.20	160.31	56.82	760.89	1.4260	1143.72	48.80	10869.4 2	40.00	14.71
2,3,3,4-Tetramethylhexane	241.50	164.59	56.69	765.60	1.4298	1129.30	49.10	10648.7 2	40.04	14.75
2,3,4,4-Tetramethylhexane	231.80	161.60	56.64	758.60	1.4267	1119.89	49.20	10500.9 2	38.87	14.79
2,3,4,5-Tetramethylhexane	243.10	156.20	56.34	745.60	1.4204	1071.09	42.70	9751.18	40.71	14.97

^aExperimental data were taken from [43].

4. CONCLUSIONS

We have employed the expanded form of the Wiener index to introduce a novel topological descriptor, named M -index, which includes both the atom parameter such as electronegativity and principle quantum number and the bond parameter. In fact, we modified the adjacency matrix for constructing the M -index in such a way that not only the characteristics of an individual atom but also the role of that atom in establishing the connection with other atoms in a molecular graph is considered. The proposed M -index was used to correlate with a wide range of properties in various data sets, including; logarithm of partition coefficient in octanol/water, molar refraction, molar volume, parachor, and polarizability for alkanes and cycloalkanes; standard enthalpy of formation for silicon, titanium, and methyl halides; toxicity of saturated alcohols, ketones, diols, and

carboxylic acids; molar refraction, and boiling point of aldehydes and ketones, as well as heat capacity, boiling point, density, refractive index, Gibbs energy, and the standard enthalpy of formation of isomeric compounds. The predictive ability of the developed models has been assessed by leave-one-out-cross-validation test. All the constructed models have favorable statistical parameters and demonstrate satisfactory predictability.

Finally, it is important to note that the M -index is easy to generate and the developed QSPR/QSAR models based on this index are linearly correlated. This is an interesting feature of the M -index when compared with quantum chemical descriptors which require vast computational cost and exhibit limitations for large size molecules.

5. SUPPLEMENTARY INFORMATION

Illustrative examples for computing M -index for the compounds of different tables are presented in Supplementary Information.

ACKNOWLEDGMENTS. Shiraz University is gratefully acknowledged for its support. P. Manshour would also like to acknowledge support of Persian Gulf University Research Council.

REFERENCES

1. A.-L. Barabási, *Linked: The New Science of Networks*, Perseus Publishing, Cambridge, 2003.
2. A.T. Balaban, *Chemical Applications of Graph Theory*, Academic Press, London, 1976.
3. A.T. Balaban, From chemical graphs to 3D molecular modeling, in: A. T. Balaban (Ed.), *From Chemical Topology to Three-Dimensional Geometry*, Plenum Press, New York, 1997, pp. 1–24.
4. L. Euler, Solutio problematis ad geometriam situs pertinentis, *Commentarii Academiae Scientiarum Petropolitanae* **8** (1741) 128–140.
5. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Florida, USA, 1992.
6. Q. Ivanciuc, QSAR comparative study of Wiener descriptors for weighted molecular graphs, *J. Chem. Inf. Comput. Sci.* **40** (2000) 1412–1422.
7. A. Mohajeri, M. Alipour, M. B. Ahmadi, A graph theory study on $(ZnS)_n$ ($n= 3-10$) nanoclusters, *Chem. Phys. Lett.* **503** (2011) 162–166.
8. A. Kurup, R. Garg, C. Hansch, Comparative QSAR study of tyrosine kinase inhibitors, *Chem. Rev.* **101** (2001) 2573–2600.

9. R. García-Domenech, J. Gálvez, J. V. de Julián-Ortiz, L. Pogliani, Some new trends in chemical graph theory, *Chem. Rev.* **108** (2008) 1127–1169.
10. C. Cao, Y. Hua, Topological indices based on vertex, distance, and ring: On the boiling points of paraffins and cycloalkanes, *J. Chem. Inf. Comput. Sci.* **41** (2001) 867–877.
11. H. Yuan, A. L. Parrill, QSAR development to describe HIV-1 integrase inhibition, *J. Mol. Struct. (THEOCHEM)* **529** (2000) 273–282.
12. A. Mohajeri, M. H. Dinpajoo, Structure-toxicity relationship for aliphatic compounds using quantum topological descriptors, *J. Mol. Struct. (THEOCHEM)* **855** (2008) 1–5.
13. B. Hemmateenejad, A. Mohajeri, Application of quantum topological molecular similarity descriptors in QSPR study of the O-methylation of substituted phenols, *J. Comput. Chem.* **29** (2008) 266–274.
14. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley VCH, Weinheim, 2000.
15. H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.
16. A. T. Balaban, Highly discriminating distance-based topological index, *Chem. Phys. Lett.* **89** (1982) 399–404.
17. M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
18. 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.
19. D. H. Rouvray, *The role of the topological distance matrix in chemistry*, in: N. Trinajstić (Ed.), *Mathematics and Computational Concepts in Chemistry*, Horwood, Chichester, 1986, pp. 295–306.
20. I. Gutman, J. H. Potgieter, Wiener index and intermolecular forces, *J. Serb. Chem. Soc.* **62** (1997) 185–192.
21. D. H. Rouvray, Should we have designs on topological indices? *Chemical Applications of Topology and Graph Theory*, *Stud. in Phys. & Theoret. Chem.*, Elsevier, Amsterdam, 1983, pp. 159–177.
22. I. Lukovits, Wiener-type graph invariants, in: M. V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science, Huntington, 2001, pp. 31–38.
23. I. Gutman, J. Žerovnik, Corroborating a modification of the Wiener index, *Croat. Chem. Acta.* **75** (2002) 603–612.
24. O. Ivanciuc, T. Ivanciuc, A. T. Balaban, Design of topological indices, part 10.1: Parameters based on electronegativity and covalent radius for the computation of

- molecular graph descriptors for heteroatom-containing molecules, *J. Chem. Inf. Comput. Sci.* **38** (1998) 395–401.
25. F. Yang, Z. D. Wang, Y. P. Huang, X. R. Ding, Modification of Wiener index and its application, *J. Chem. Inf. Comput. Sci.* **43** (2003) 753–756.
 26. F. Yang, Z. D. Wang, Y. P. Huang, P. J. Zhou, Modification of the Wiener index 2, *J. Chem. Inf. Comput. Sci.* **43** (2003) 1337–1341.
 27. F. Yang, Z. D. Wang, Y. P. Huang, H. L. Zhu, Novel topological index F based on incidence matrix, *J. Comput. Chem.* **24** (2003) 1812–1820.
 28. F. Yang, Z. D. Wang, Y. P. Huang, Modification of the Wiener index 4, *J. Comput. Chem.* **25** (2004) 881–887.
 29. S. S. Tratch, M. I. Stankevitch, N. S. Zefirov, Combinatorial models and algorithms in chemistry. The expanded Wiener number—A novel topological index, *J. Comput. Chem.* **11** (1990) 899–908.
 30. P. V. Khadikar, S. Karmarkar, A novel PI index and its applications to QSPR/QSAR studies, *J. Chem. Inf. Comput. Sci.* **41** (2001) 934–949.
 31. R. Walsh, Thermochemistry of Silicon-containing compounds, *J. Chem. Soc. Faraday Trans.* **79** (1983) 2233–2248.
 32. X. Zhi-hong, W. Le-shan, *The Data Base of Inorganic Chemical Thermodynamics*, Science Press, Beijing, 1987.
 33. J. G. Stark, H. G. Wallace, *Chemistry Data Book*, John Murray, London, 1982.
 34. G. Gini, M. V. Craciun, C. König, Combining unsupervised and supervised artificial neural networks to predict aquatic toxicity, *J. Chem. Inf. Comp. Sci.* **44** (2004) 1897–1902.
 35. Y. Xue, H. Li, C. Y. Ung, C. W. Yap, Y. Z. Chen, Classification of a diverse set of Tetrahymena pyriformis toxicity chemical compounds from molecular descriptors by statistical learning methods, *Chem. Res. Toxicol.* **19** (2006) 1030–1039.
 36. T. W. Schultz, Tetratox: Tetrahymena pyriformis population growth impairment endpoint a surrogate for fish lethality, *Toxicol. Mech. Method* **7** (1997) 289–309.
 37. D. R. Roy, R. Parthasarathi, B. Maiti, V. Subramanian, P. K. Chattaraj, Electrophilicity as a possible descriptor for toxicity prediction, *Bioorg. Med. Chem.* **13** (2005) 3405–3412.
 38. C. L. Yaws, *Chemical Properties Handbook*, McGraw–Hill, New York, 1999.
 39. E. Estrada, L. Rodriguez, A. Gutiérrez, Matrix algebraic manipulations of molecular graphs. 1. Distance and Vertex–Adjacency Matrices, *MATCH Commun. Math. Comput. Chem.* **35** (1997) 145–156.
 40. R. C. Weast, *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 1989.
 41. D. R. Lide, G. W. A. Milne, *Handbook of Data on Common Organic Compounds*, CRC Press, Boca Raton, FL, 1995.

42. J. A. Dean, *Lange's Handbook of Chemistry*, McGraw-Hill, New York, 1999.
43. A. A. Gakh, E. G. Gakh, B. G. Sumpter, D. W. Noid, Neural network-graph theory approach to the prediction of the physical properties of organic compounds, *J. Chem. Inf. Comput. Sci.* **34** (1994) 832-839.

A novel topological descriptor based on the expanded wiener index: applications to QSPR/QSAR studies

AFSHAN MOHAJERI¹, POUYA MANSOUR², AND MAHBOUBEH MOUSAEI¹

¹Department of Chemistry, College of Sciences, Shiraz University, Shiraz 71454, Iran.

²Department of Physics, Persian Gulf University, Bushehr 75169, Iran.

ارائه توصیفگر توپولوژیکی نوین بر پایه شاخص تعمیم یافته وینر:

کاربردهایی در مطالعات QSPR/QSAR

ادیتور (ابطا): علیرضا اشرفی

چکیده

در این مقاله، یک شاخص جدید توپولوژیکی به نام شاخص M بر پایه شکل تعمیم یافته ماتریس وینر معرفی می‌شود. برای ساخت این شاخص، خصوصیات اتمی و برهم‌کنش بین راس‌ها در یک ملکول در نظر گرفته می‌شوند. مفید بودن شاخص M به وسیله مدل‌های QSPR/QSAR گوناگون برای خصوصیات مختلف فیزیکی-شیمیایی و فعالیت‌های زیستی تعداد زیادی از ترکیبات مختلف نشان داده می‌شود. به علاوه، کاربردی بودن شاخص پیشنهادی بین ترکیبات ایزومری نیز بررسی شده است. در هر مورد، پایداری مدل بدست آمده به وسیله‌ی آزمون اعتبار سنجی متقابل تایید می‌شود. نتایج مطالعه‌ی پیش رو نشان می‌دهد که شاخص M یک راه امید بخش برای توسعه‌ی مدل‌های شدیداً همبسته QSPR/QSAR را فراهم می‌آورد. از سوی دیگر، شاخص M به سادگی تولید شده و توسعه مدل‌های QSPR بر پایه‌ی این شاخص به صورت خطی همبسته هستند. این یک خاصیت جالب شاخص M در مقایسه با توصیف کننده‌های کوانتومی شیمیایی است که نیازمند هزینه‌های محاسباتی وسیع بوده و محدودیت‌هایی را برای ملکول‌های بزرگ از خود نشان می‌دهند.

لغات کلیدی: شاخص توپولوژیکی، نظریه‌ی گراف، شاخص وینر تعمیم یافته، QSPR، QSAR.