

Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume

Ernesto Estrada

Centro de Bioactivos Químicos, Universidad Central de Las Villas, Santa Clara 54830, Villa Clara, Cuba

Received May 15, 1994[®]

Adjacency between edges in molecular graphs containing heteroatoms is considered by using weights in the nondiagonal entries of the edge adjacency matrix. Values of k_{C-X} parameters used in HMO calculation for molecules containing heteroatoms are used as edges weights. One index is derived using the Randić-type graph theoretical invariant. This index is well correlated ($r > 0.99$) with the molar volume of 112 aliphatic organic compounds, including aldehydes, ketones, ethers, thioethers, tertiary amines, and alkyl halides. A parametrization methods is proposed in order to obtain the values of k_{C-X} parameters, based on correlation found between the proposed index and molar volumes.

INTRODUCTION

Chemical graph theory is an emerging branch of theoretical chemistry based on the representation of molecules as simple graphs,^{1–3} where atoms are symbolized by points (vertices) and covalent bonds by lines (edges). One of the principal objectives of this approach is to find a way to represent molecular structures as simple numbers and to reduce the problem of structure–property relationships to a problem of correlation between two series of numbers,^{4–6} one of them representing the property and the other representing the chemical structure. The graph-theoretical approach most employed to represent the chemical structure is the use of several invariants to generate indices commonly known as topological indexes.^{7–9} There are more than 120 of these indices described in the literature, among them we can find the molecular connectivity index χ ,¹⁰ Hosoya Z index,¹¹ Wiener number,¹² Balaban J index,¹³ Randić molecular ID number,¹⁴ electrotopological state indexes,¹⁵ and so forth.

Simple topological indices are those which do not include atom identity in its definition. Two significant absences in these topological indices are the lack of information on heteroatoms and stereochemical features of molecules, which do limit its field of applications. The heteroatom differentiation has been resolved in some topological indices using different empirical and nonempirical approaches.^{15–17}

In a recent paper Randić¹⁶ has stated that these omissions may appear to be severe limitations of this type of approach and several works have been directed to give a solution to these problems; for instance, Kier and Hall¹⁷ introduce the concept of valence connectivities to differentiate heteroatoms using the valence electrons of each atom in the molecule, and, on the other hand, Randić¹⁶ has employed variable parameters as diagonal entries of an adjacency matrix to discriminate some heteroatoms from carbons. In order to introduce three-dimensional features of molecules into molecular descriptors, some promising approaches have appeared in the past few years, such as the topographic distance matrix proposed by Bogdanov et al.,¹⁸ Randić's approach of graph embedded into three-dimensional grids,^{19–21} and the use of graph weighted with quantum-chemical parameters used by Estrada.^{22,23}

Recently, this author²⁴ proposed a graph-theoretical matrix to represent edge-adjacency relationships, and the derived ϵ index was found to correlate very well with molar volumes and related properties of alkanes, such as molar refraction.

In the present paper we will extend ϵ index to molecules with heteroatoms, using as weights parameters introduced by Wheland and Pauling²⁵ for nondiagonal entries of the Hückel matrix for heteroconjugated molecules.

EXTENSION OF E INDEX FOR HETEROATOMS

Edges-adjacency matrix of a simple graph $G = \{V, E\}$, where V is the vertex set and E is the edges set, is a square and symmetric matrix $E = [g_{ij}]_{m \times m}$ where m is the number of edges in graph and elements g_{ij} are defined as follows:²⁴

$$g_{ij} = \begin{cases} 1 & \text{if } \exists v_r \in V/v_r \sim e_i, e_j \\ 0 & \text{otherwise} \end{cases}$$

i.e., edges e_i and e_j are adjacent when there exists one vertex v_r incident with them ($v_r \sim e_i, e_j$).

Edges degree, $\delta(e_r)$, is defined as the sum of elements of r th row or column in E matrix

$$\delta(e_r) = \sum_i g_{ir} = \sum_j g_{rj}$$

For heterographs, i.e., graphs containing atoms others than carbon and hydrogen, it is necessary to differentiate bonds with heteroatoms using elements of a weights set W . Elements of W can be selected from different ways, i.e., empirically or using some parameters with a defined physical meaning (bond length, bond order, etc.).

In the present approach, we propose to use as elements of W set, the k_{C-X} parameters introduced by Pauling²⁵ as nondiagonal elements in the Hückel matrix when there is the presence of heteroatoms. These k_{C-X} parameters are related to the resonance integral β_{cx} associated with the bond between heteroatoms x and the neighboring carbon atom as follows:

$$\beta_{cx} = k_{C-X} \beta_{cc}$$

where β_{cc} is the resonance integral for carbon–carbon bond.

[®] Abstract published in *Advance ACS Abstracts*, July 1, 1995.

Table 1. Values of k_{C-X} Parameters Reported by Pérez²⁶

C-X bond	k_{C-X}	C-X bond	k_{C-X}
C-N	1.0	C-F	0.7
C-O	0.8	C-Cl	0.4
C=O	1.6	C-Br	0.3
C-S	0.7		

Several values have been reported in the literature for k_{C-X} parameters using different methods; in Table 1 we list some of them reported by Pérez.²⁶

The representation of the weighted graph can be expressed mathematically as an ordered 4-tuple $G = (V, E, W, \phi)$, where V and E are the vertex and edge sets, respectively, and ϕ is a surjective mapping (evaluation) of the elements of edges set onto weight set

$$\phi: E \rightarrow W$$

it means that each edge is uniquely evaluated by one element of the weight set W , which elements are parameters k_{C-X} for different bonds.

Elements g_{ij} of the edge weighted adjacency matrix E are defined as follows:

$$g_{ij} = \begin{cases} k_i & \text{if } \exists v_r \in V/v_r \sim e_i, e_j \text{ and } e_j \text{ have weight } k_i \\ 0 & \text{otherwise} \end{cases}$$

Hence, E for heterographs is a square but nonsymmetric matrix, and, as a consequence, edges degree $\delta(e_r)$ is calculated only as the sum of elements of r th row

$$\delta^w(e_r) = \sum_j g_{rj}$$

Edges connectivity index ϵ can be calculated with the same expression used for simple molecular graphs by using the graph-theoretical invariant of Randić²⁷

$$\epsilon = \sum_r [\delta^w(e_i) \delta^w(e_j)]_r^{-1/2}$$

where $\delta^w(e_i)$ are the values of edge-weighted degrees and the summation is over all r -pairs of adjacent edges.

The ϵ index can be calculated using the edge weighted graph with no matrix representation, considering the edge degree $\delta^w(e_i)$ as the sum of weights for all adjacent edges with e_i ; as an example, we illustrate in Figure 1 the weighted edge graph of ethyl isopropyl ether, its edge adjacency matrix, and the calculation of ϵ index.

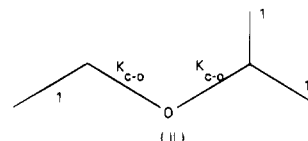
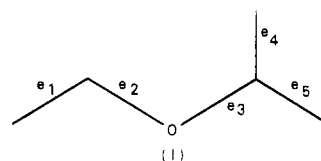
COMPUTATIONS

Values of ϵ index were calculated for a series of 103 aliphatic organic compounds, including 43 aldehydes and ketones, 23 alkyl halides, 20 ethers, 10 thioethers, and 7 tertiary amines. In all cases, standard values of k_{C-X} presented in Table 1 were used as edge weights in calculations of ϵ index. The index was used to describe the molar volumes (MV) of these compounds, calculated as follows

$$MV = \frac{MW}{d}$$

where MW is the molecular weight and d is the density at 20 °C in g/cm³.

The correlation found is illustrated by the following equation



b) The edge adjacency matrix E of weighted graph G :

$$E = \begin{bmatrix} 0 & k_{C-O} & 0 & 0 & 0 \\ 1 & 0 & k_{C-O} & 0 & 0 \\ 0 & k_{C-O} & 0 & 1 & 1 \\ 0 & 0 & k_{C-O} & 0 & 1 \\ 0 & 0 & k_{C-O} & 1 & 0 \end{bmatrix}$$

c) Calculation of edge degrees:

$$\begin{aligned} \delta(e_1) &= k_{C-O} = 0.8 \\ \delta(e_2) &= 1 + k_{C-O} = 1.8 \\ \delta(e_3) &= 2 + k_{C-O} = 2.8 \\ \delta(e_4) &= \delta(e_5) = 1 + k_{C-O} = 1.8 \end{aligned}$$

d) The computation of ϵ index of G :

$$\epsilon = (0.8 \cdot 1.8)^{-1/2} + (1.8 \cdot 2.8)^{-1/2} + 2 \cdot (2.8 \cdot 1.8)^{-1/2} + (1.8 \cdot 1.8)^{-1/2} = 2.7252$$

Figure 1. The computing of the ϵ index for a graph G representing the skeleton of ethyl isopropyl ether: (a) Graph G with labeled edges (I) and weighted edges (II), (b) the edge adjacency matrix E of weighted graph G , (c) calculation of edge degrees, and (d) the computation of ϵ index of G .

$$MV = 32.995 + 33.173\epsilon \quad (1)$$

$$R = 0.9798, \quad S = 4.573, \quad F = 2422$$

where R is the correlation coefficient, S is the standard error estimated, and F is the Fisher ratio. Analysis of the model showed that triethylamine, diethylmethanamine, dimethylbutylamine, and methyl s-pentylketone were statistical outliers. These compounds were identified as outliers because they exceeded the diagnostic cut-off values for the following standard tests: residuals, standardized residuals, and studentized residuals. The residual is the difference between the observed experimental value and the calculated value. A standardized residual is the residual divided by the standard deviation of the regression equation. The studentized residual is the residual for one observation divided by its own standard deviation. One compound is considered as outlier if its residual lies three standard deviations from the mean of the residuals and/or the standardized as well as studentized residual is greater than 3 in absolute value.

In order to avoid outliers in the model we propose to modify the values for k_{C-N} and $k_{C=O}$ parameters using a parametrization method. The values of these parameters were modified empirically in order to reproduce molar volumes of amines and carbonyl compounds (aldehydes and ketones). The values found as optimal were 0.9 for carbon-nitrogen single bond and 1.2 for carbon-oxygen double bond. The regression equation found by using these parameters to describe molar volumes of the 103 compounds is depicted below

$$MV = 31.444 + 33.962\epsilon \quad (2)$$

$$R = 0.9946, \quad S = 2.376, \quad F = 9238$$

Table 2. Experimental Molar Volumes for the Series of Organic Compounds under Study and Predicted Values Based on Sample Linear Regression with ϵ Index Using Equation 1

compound	MV exp	ϵ index	MV calcd	residual	compound	MV exp	ϵ index	MV calcd	residual
methyl ethyl ether	88.736	1.6667	86.704	2.032	1-I,3-Me-butane	130.969	2.8203	124.644	6.325
methyl propyl	100.271	2.1343	102.083	-1.812	2-I-butane	115.578	2.5103	114.449	1.129
methyl isopropyl	102.252	2.1146	101.435	0.817	2-I,2-Me-butane	132.557	2.9528	129.002	3.555
methyl butyl	118.232	2.6231	118.159	0.073	1-I-pentane	130.598	2.8348	125.121	5.477
methyl isobutyl	120.366	2.6240	118.188	2.178	2-I-pentane	131.161	3.0153	131.058	0.103
methyl sec-butyl	118.678	2.5137	114.561	4.117	3-I-pentane	130.469	2.8322	125.036	5.433
methyl tert-butyl	118.839	2.5646	116.235	2.604	trimethyl amine	92.825	1.8750	93.554	-0.729
methyl pentyl	132.986	3.1231	134.636	-1.650	triethyl	138.832	3.2339	138.247	0.585
diethyl	103.670	2.2230	105.000	-1.330	diethyl methyl	123.755	2.7520	122.398	1.357
ethyl propyl	119.144	2.6898	120.352	-1.209	dimethyl isobutyl	142.314	3.3214	141.125	1.189
ethyl isopropyl	124.222	2.7252	121.517	2.705	dimethyl pentyl	154.778	3.8019	156.928	-2.150
ethyl butyl	136.182	3.1787	136.432	-0.250	dimethyl butyl	140.161	3.3019	140.484	-0.329
ethyl isobutyl	135.819	3.1795	136.458	-0.639	tripropyl	189.203	4.7767	188.989	0.214
ethyl sec-butyl	135.946	3.1245	134.649	1.297	methyl ethyl ketone	84.396	1.8079	91.348	-6.952
ethyl tert-butyl	135.656	3.2069	137.359	-1.703	methyl propyl	106.317	2.3513	109.219	-2.902
ethyl pentyl	152.191	3.6786	152.873	-0.682	methyl isopropyl	106.819	2.3408	108.874	-2.055
ethyl tert-pentyl	151.495	3.6598	152.254	-0.760	methyl butyl	123.259	2.8513	125.664	-2.405
dipropyl	138.587	3.1576	135.738	2.849	methyl sec-butyl	123.001	2.8368	125.187	-2.186
propyl isopropyl	138.399	3.1933	136.912	1.487	methyl isobutyl	125.345	2.8880	126.871	-1.526
diisopropyl	140.864	3.2500	138.777	2.087	methyl tert-butyl	124.813	2.8418	125.352	-0.539
methyl ethyl thioether	90.239	1.8336	92.193	-1.954	methyl pentyl	140.550	3.3513	142.108	-1.559
methyl isobutyl	142.065	3.3324	141.487	0.578	methyl hexyl	155.699	3.8513	158.553	-2.854
ethyl propyl	124.253	2.8602	125.957	-1.704	methyl sec-pentyl	137.681	3.3009	140.451	-2.770
methyl pentyl	139.960	3.2543	138.918	1.042	diethyl	105.677	2.2211	104.937	0.740
methyl propyl	106.838	2.2720	106.611	0.227	ethyl propyl	123.183	2.7645	122.809	0.374
methyl isopropyl	108.551	2.2492	105.862	2.689	ethyl isopropyl	120.482	2.7621	122.730	-2.248
dipropyl	140.861	3.2986	140.375	0.486	ethyl butyl	139.313	3.2655	139.287	0.026
diisopropyl	144.928	3.4140	144.171	0.757	ethyl isobutyl	140.915	3.3010	140.454	0.461
methyl butyl	123.427	2.7544	122.477	0.950	ethyl sec-butyl	139.672	3.1844	136.619	3.053
methyl isobutyl	124.775	2.7442	122.142	2.633	ethyl tert-butyl	140.308	3.2689	139.398	0.910
1-Br-butane	107.384	2.2043	104.385	2.999	dipropyl	139.313	3.3079	140.681	-1.368
2-Br-butane	108.859	2.3241	108.325	0.534	propyl isopropyl	140.897	3.3055	140.602	0.295
1-Cl-butane	104.378	2.1499	102.596	1.782	diisopropyl	140.602	3.3081	140.688	-0.086
2-Cl-butane	105.932	2.2456	105.743	0.189	ethyl pentyl	155.699	3.7645	155.698	0.001
1-Br-pentane	123.953	2.7043	120.829	3.124	propyl butyl	157.132	3.8079	157.126	0.006
1-Br,3-methylpentane	139.488	3.1104	134.186	5.302	propyl isobutyl	172.351	4.3254	174.146	-1.795
1-Br,4-methylpentane	141.231	3.2218	137.850	3.382	methyl sec-hexyl	156.479	3.8009	156.895	-0.416
2-Br-pentane	125.052	2.8380	125.227	-0.175	methyl isohexyl	157.036	3.8688	159.129	-2.093
3-Br-pentane	124.546	2.6860	120.227	4.319	methyl tert-hexyl	152.740	3.7794	156.188	-3.448
3-Br,3-methylpentane	139.417	3.1461	135.360	4.057	3,4-dimethylhexanone-2	154.310	3.7212	154.274	0.036
1-Cl-pentane	120.775	2.6499	119.040	1.735	propyl tert-butyl ketone	157.930	3.8124	157.274	0.656
2-Cl-pentane	122.442	2.7631	122.763	-0.321	isopropyl isobutyl	154.780	3.8422	158.254	-3.474
2-Cl-2,4-dimethylpentane	152.388	3.8048	157.024	-4.636	4,4-diMe-hexanone-3	154.496	3.6459	151.798	2.698
3-Cl-pentane	121.979	2.6206	118.077	3.903	3-E,4-M-pentanone-2	157.635	3.7496	155.208	2.467
3-Cl,3-Me-pentane	135.393	3.0777	133.110	2.283	di-tert-butyl ketone	172.330	4.3313	174.340	-2.010
1-Br-propane	90.862	1.7541	89.578	1.284	propanal	71.978	1.3184	75.248	-3.270
2-Br-propane	93.607	2.0095	97.978	-4.371	butanal	88.127	1.8580	92.995	-4.868
1-Cl-propane	88.113	1.6903	87.480	0.633	2-Me-propanal	90.703	1.8496	92.719	-2.016
2-Cl-propane	91.099	1.9095	94.689	-3.590	pentanal	106.238	2.3580	109.440	-3.202
1-F-propane	77.929	1.5338	82.333	-4.404	2-Me-butanal	107.112	2.2625	106.299	0.813
1-F-pentane	113.823	2.5163	114.646	-0.823	3-Me-butanal	107.810	2.3798	110.157	-2.347
1-F-hexane	130.081	3.0163	131.091	-1.010	hexanal	122.865	2.8579	125.881	-3.016
1-F-heptane	146.365	3.5163	147.535	-1.170	2-Et-butanal	123.304	2.6924	120.438	2.866
1-I-propane	97.200	1.9069	94.603	2.596	3-Me-hexanal	138.974	3.3283	141.352	-2.378
2-I-propane	99.806	2.2575	106.134	-6.328	octanal	155.888	3.8579	158.770	-2.882
1-I-butane	113.904	2.3347	108.674	5.230	2-Et-hexanal	149.883	3.7488	155.182	-5.299

A similar parametrization method was used in order to find a value for k_{C-I} parameter, for which no value appears in the literature. The parameter value found using this method was 0.1. A final model including the before mentioned 103 compounds and nine alkyl iodines was found using the newly proposed k_{C-X} parameters. It can be summarized as

$$MV = 31.887 (\pm 1.108) + 32.889 (\pm 0.372)\epsilon \quad (3)$$

$$R = 0.9930, S = 2.635, F = 7792$$

The calculated values and the residuals for the studied series of compounds are shown in Table 2. The uncertainties

of the regression coefficients correspond to 95% confidence intervals. Equation 3 explains almost 99% of the variance in molar volumes for the heterogeneous series of compounds analyzed in this paper, the standard error is about 2.1% (2.635/127.2) of the mean of MV, and the use of the Kolmorov-Smirnov test proved that residuals were normally distributed. A plot of the calculated vs observed MV is shown in Figure 2. The residuals were plotted against the calculated values to check for a relationship between the two. In Figure 3 is shown that the residuals were not a function of the calculated value. In this model, no compounds were identified as statistical outliers using the criteria previously discussed. In Figure 4 we illustrate the histogram for

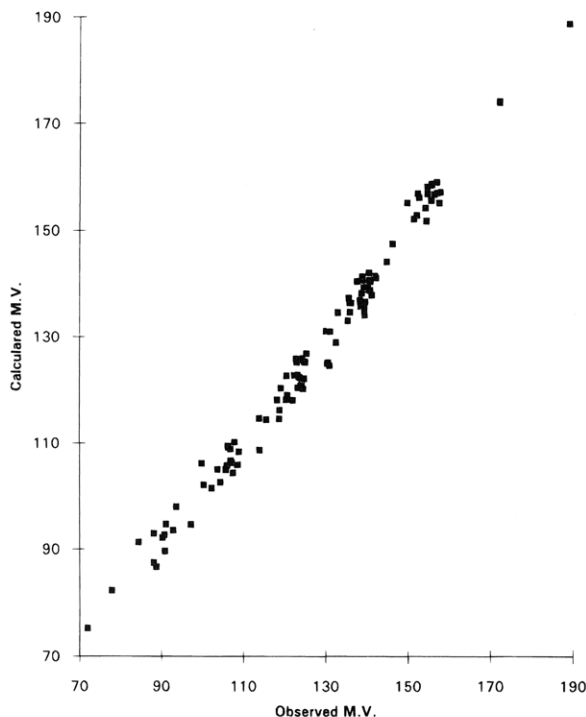


Figure 2. Plot of calculated MV vs observed MV for the 112-compound series.

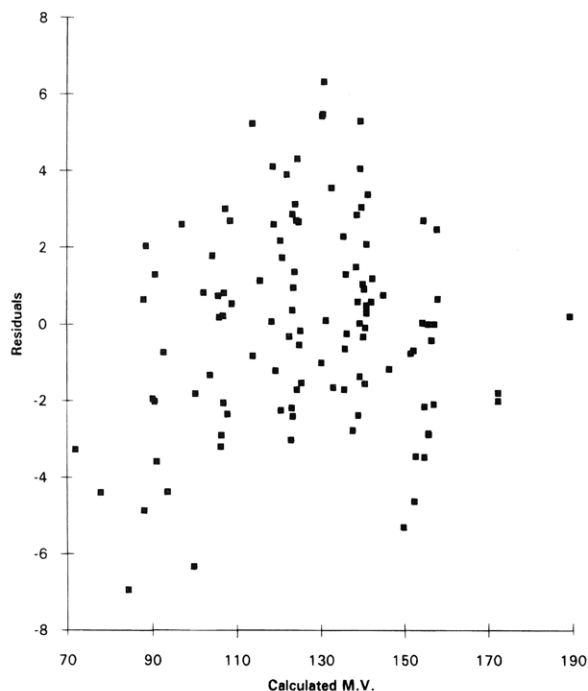


Figure 3. Plot of the residuals vs calculated MV for the 112-compound series.

standardized and studentized residuals, where it can be appreciated that no case appears with an absolute value greater than 3, which is the cut-off-value frequently used for these statistical tests.

In a recent strategy to design quantitative structure–property models (QSPR), Mihalic and Trinajstić⁹ proposed that the quality of this type of approach can be conveniently measured by the correlation coefficient and the standard deviation, and a good QSPR model must have $r > 0.99$, while s depends on the property under study. According to this statement, eq 3 represents a very good model to describe

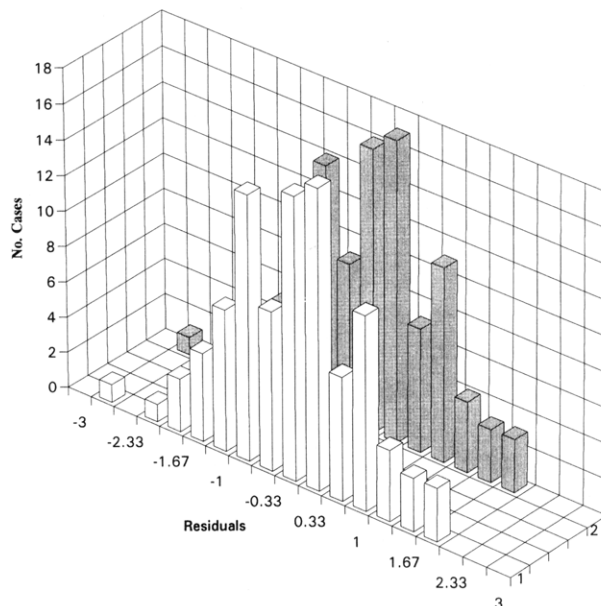


Figure 4. Histogram of the standardized (1) and studentized (2) residuals of the 112-compound series.

the molar volumes of organic aliphatic compounds containing heteroatoms in the structures.

There are several useful properties that are important for molecular descriptors in order to have a practical use in QSPR and QSAR studies. A list of attributes for topological indices has been proposed by Randić²⁹ that represent the very high level of sophistication that is desirable for topological indices.⁹ Among these attributes we can find the direct structural interpretation, to have a good correlation with at least one property, a good discrimination of isomers, to be simple, linearly independent, and so forth.

In a previous paper²⁴ the ϵ index was interpreted as a good measurement of molecular volume of alkanes taking into account the direct relationships between molecular volume and the liquid molar volume of hydrocarbons. For compounds other than alkanes, polarizability, polarity, and/or hydrogen bonding interactions lead to a decrease in molar volumes relative to the values predicted by the correlation mentioned above as a consequence of the closer packing in the pure liquids of compounds with heteroatoms.

Equation 3 represents a good model to describe the molar volumes of compounds containing different heteroatoms and indicate that modified ϵ index contain information not only related to molecular volume but presumably about interactions within the pure liquid too. These interactions are important to explain a process such as partition of solutes between two phases and aqueous solubility,³⁰ giving to ϵ index possibilities to be used in several types of QSPR studies.

One of the weakest points of topological indices is the discrimination of isomers, this property is important in order to use the indices in studies of isomer variations of molecular properties and in general in studies of those aspects of molecular properties which are isomer-sensitive.

The ϵ index shows a good discrimination of isomers for all compounds studied in this paper, differentiating not only the branching in alkyl chain but also the position of heteroatoms and the functional groups of isomeric compounds too. The quality of ϵ index to differentiate isomers can be compared with it for other indices designed to be

Table 3. Regression Equations, Correlation Coefficients (R), and Standard Deviation (s) for Correlation between the Molar Volumes of Alkyl Halides and Selected Topological Indexes

index	intercept	slope	R	S
χ	57.584	30.747	0.9102	7.967
χ^v	43.520	24.166	0.8378	10.504
ϵ	28.285	34.792	0.9875	3.033

used in the presence of heteroatoms, such as the valence molecular connectivity χ^v of Kier and Hall.¹⁴ As an example the values of both indices (ϵ and χ^v) are shown below for two isomeric ethers with different position of heteroatoms:

ethyl butyl ether

$$\chi^v = 2.992, \quad {}^2\chi^v = 1.5505, \quad \epsilon = 3.1787$$

dipropyl ether

$$\chi^v = 2.992, \quad {}^2\chi^v = 1.6124, \quad \epsilon = 3.1576$$

In order to differentiate these two ethers with Kier and Hall indices we need to use "extended" indices such as the second order valence connectivity index.

The linear independence or "orthogonality" of ϵ index can be proven by the fact that the molar volume is a property not successfully analyzed with the existing topological indices;³¹ this is an important feature considering that several topological indexes are interrelated with each other.^{32,33} As an example we compare the correlation found between the molar volumes of alkyl halides studied in this paper with the molecular connectivity χ , valence connectivity χ^v , and ϵ index, obtaining the equations depicted in Table 3. Better correlations can be obtained using higher order valence connectivity indices, such as ${}^3\chi_p^v$ and/or ${}^4\chi_{pc}^v$ and ${}^4\chi_{pc}^v$. However, ϵ index can also be "extended" to higher analogues in order to account for information not contained in its original definition and to produce an improvement in the correlations obtained.

ON THE K_{C-X} PARAMETERS VALUES

Values of k_{C-X} parameters can be obtained using different approaches, and no standard procedure to evaluate these parameters has been proposed. Regarding the problems of the procedure to evaluate parameters used in HMO calculation in presence of heteroatoms, Streitwieser³⁴ stated the following:

"Unfortunately, the ideal procedure has never been followed in toto and has only rarely even been used in part. More often, parameter values have been assigned as guesses partly based on theory or on the basis of applications of assumed validity. The results has been a profusion of parameter values in the literature."

We will analyze some of the proposed parameter values for carbon-oxygen double bond and carbon-nitrogen single bond, for which new values have been proposed in present work using a topological parametrization method.

Assuming a proportionality to the difference between single- and double-bond energies, different k -values have been obtained, which probably should not be weighted highly, these values are $k_{C-N} = 0.3-0.5$, $k_{C-N} = 1-1.2$, and $k_{C=O} = 1.4-2$.

Other authors proposed $k_{C-N} = 0.4-0.55$ using the resonance energies; $k_{C-N} = 0.9-1.1$ and $k_{C=O} = 1.1$ were obtained reproducing the electron-density distribution given

Table 4. Polarographic Reduction Potentials of Aromatic Aldehydes versus Energies of Lowest Unoccupied MO Calculated with Different Values of $K_{C=O}$ Parameter

aldehyde	$E_{1/2}^c$	$-E_{LUMO}$		
		$K_{C=O} = 1.0$	$K_{C=O} = 1.2$	$K_{C=O} = 1.6$
benzaldehyde	1.592	0.291	0.387	0.569
cinnamaldehyde	1.335	0.192	0.256	0.371
1-naphthalenealdehyde	1.476	0.230	0.300	0.417
2-naphthalenealdehyde	1.505	0.267	0.349	0.480
9-anthracenealdehyde	1.207	0.149	0.194	0.268
9-phenanthrenealdehyde	1.449	0.228	0.297	0.412
		0.9837	0.9810	0.9674
		0.027	0.029	0.039

^a Data taken from ref 34.

by molecular orbital (MO) calculations of the self-consistent field (SCF), antisymmetrized molecular orbital (ASMO), or Pariser-Parr type. A most complete analysis for all proposed values for k_{C-N} and $k_{C=O}$ parameters is given in ref 35, where it is concluded that a survey of the various derived values for bond integrals shows a good clustering about values of $k_{C-N} = 0.8-1$; and for oxygen; parameters, where less consistency was shown, the frequently used value of $k_{C=O} = 1.4$ appears to be somewhat high a more adequate value of $k_{C=O} = 1$ was proposed.

The values proposed by us ($k_{C-N} = 0.9$ and $k_{C=O} = 1.2$) using a parametrization method to reproduce the molar volumes of tertiary aliphatic amines and carbonyl compounds, respectively, represent the mean of the values proposed by Streitwieser. In order to show the applicability of the proposed parameter values to describe other properties, we select a study carried out by Streitwieser³⁴ to describe the polarographic reduction potentials of a series of aromatic aldehydes. This author used Hückel molecular orbital (HMO) calculation of the lowest unoccupied molecular orbital (LUMO) energy to correlate the polarographic property. In Table 4 we illustrate the values of LUMO energies for aldehydes calculated with $k_{C=O}$ values of 1.0 used by Streitwieser, 1.2 of the present work, and the value of 1.6 proposed by Pérez in Table 1. In all cases the value of $h_O = 1.6$, suggested by Streitwieser, was used. The statistical parameters in Table 4 depicts that the results obtained by using the value of $k_{C=O}$ parameter proposed by us are similar to those obtained with the ideal value of 1.0 proposed by Streitwieser. On the other hand, there are significative differences between results obtained with the value of 1.6 normally used in HMO studies.

Unfortunately there are no values available (or known to the author) for K_{C-I} parameters in order to compare with the value of 0.1 proposed in the present approach.

We need to consider that the ideal procedure suggested by Streitwieser to obtain the parameter values for HMO calculation in the presence of heteroatoms is based on a correlation between a calculated quantity established for hydrocarbons and some experimental property and later the application of this correlation to compounds with one or more heteroatoms under test. The parametrization method used here to obtain k_{C-X} parameters to be used in calculation of ϵ index fulfill well this philosophy, and it would be interesting to evaluate extensively the new values of k_{C-X} in HMO calculations for heteroconjugated molecules.

K_{C-X} PARAMETERS AS EDGE WEIGHTS

In some cases, weights used in molecular graphs to introduce some structural information lost in simple graph representations of molecules, i.e., differentiation of heteroatoms, are empirically selected. As an example, in his recently proposed approach, Randić¹⁶ use an empirical value of -0.20 in diagonal entries of an adjacency matrix to differentiate chlorine from carbon atoms. A subsequent "reparametrization" using a value of -0.40 produced better correlation, but as the same author stated, this result is also not optimal.

For this type of approach, a question remains open related to the existence of any connection between the empirically determined parameters and those from quantum chemical calculations.³⁶

In the present approach, the use of parameters directly related to resonance integral in HMO calculations solves this problem in a natural way.

In addition, dependence of k_{C-X} parameters on the C-X bond distance in several compounds²⁵ and perhaps its relationships with bond orders opens possibilities to develop indexes containing stereochemical information of molecules and resolves others of the important limitations of graph theoretical approaches in chemistry.

CONCLUDING REMARKS

The proposed methodology of edge adjacency relationships in molecular graphs in order to represent structural features of molecules represents a new important tool in chemical graph theory, and as proof we derived a topological index ϵ that is related to molar volumes in alkane series.

In the present approach we propose a possible solution to the problem of differentiation of heteroatoms in molecular graphs in a natural way, i.e., by using weights in the nondiagonal entries of the edge adjacency matrix. These weights are the same parameters used in HMO calculation of conjugated molecules with heteroatoms, hence the possible connection between the empirically determined parameters based on selected structure-property relationships and those from quantum chemical calculations is solved in a simple way. On the other hand, a parametrization method is proposed to obtain values for these parameters, which is a nonempirical way to fit values of ϵ index in compounds with other types of heteroatoms to produce better correlation. The use of the new proposed parameters for C-N, C=O, and C-I bonds in HMO calculations is an open question for further research. However, we illustrate that the use of $k_{C=O} = 1.2$ proposed here is adequate for HMO calculations of the LUMO's energies to be used in the description of polarographic properties of aromatic aldehydes.

We will finish this work using a remark of M. Randić about the importance of using weighted graphs in chemistry:

"The generalization of graphs to weighted graphs, which then allows one to take a step further and consider any matrix as a "weighted" graph, opens novel horizons for the chemical combinatorics and topology—which graph theory in essence is."

ACKNOWLEDGMENT

I thank the referees for their numerous comments and suggestions on the manuscript which resulted in a significant improvement of the material. I also thank Dr. G. W. A. Milne, the Editor, for his valuable help and interest.

REFERENCES AND NOTES

- (1) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed.; CRC: Boca Raton, FL, 1992.
- (2) Trinajstić, N.; Klein, D. J.; Randić, M. On Some Solved and Unsolved Problems of Chemical Graph Theory. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1986**, *20*, 699-742.
- (3) Balaban, A. T. Application of Graph Theory in Chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334-343.
- (4) Randić, M. Resolution of Ambiguities in Structure-Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1990**, *31*, 311-320.
- (5) Okuyama, T.; Miyashita, Y.; Kanaya, S.; Katsumi, H.; Sasaki, S.; Randić, M. Computer Assisted Structure-Taste Studies on Sulfamates by Pattern Recognition Method Using Graph Theoretical Invariants. *J. Comput. Chem.* **1988**, *9*, 636-646.
- (6) Randić, M. On Computation of Optimal Parameters for Multivariate Analysis of Structure-Property Relationship. *J. Comput. Chem.* **1991**, *12*, 970-980.
- (7) Sabljic, A.; Trinajstić, N. Quantitative Structure-Activity Relationships: The Role of Topological Indices. *Acta Pharm. Jugosl.* **1981**, *31*, 189-214.
- (8) Hansen, P. J.; Jurs, P. C. Chemical Application of Graph Theory. Part I. Fundamentals and Topological Indices. *J. Chem. Educ.* **1988**, *65*, 574-580.
- (9) Mihalic, Z.; Trinajstić, N. A. Graph Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701-712.
- (10) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- (11) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332-2339.
- (12) Wiener, H. Structural Determination of Paraffins Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
- (13) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399-404.
- (14) Randić, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164-175.
- (15) Hall, L. H.; Mohney, B.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76-82.
- (16) Randić, M. Novel Graph Theoretical Approach to Heteroatoms in Quantitative Structure-Activity Relationships. *Chemometrics Intel. Lab. Syst.* **1991**, *10*, 213-227.
- (17) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, Herts, 1986.
- (18) Bogdanov, B.; Nikolic, S.; Trinajstić, N. On the Three-Dimensional Wiener Number. *J. Math. Chem.* **1989**, *3*, 299-309.
- (19) Randić, M. On Characterization of Three-Dimensional Structures. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1988**, *15*, 201-208.
- (20) Randić, M. Molecular Topographic Descriptors. In *MATH/CHEM/COMP 1987*; Lacher, R. C., Ed.; Elsevier: Amsterdam 1988; pp 101-108.
- (21) Randić, M.; Jerman-Blazic, B.; Trinajstić, N. Development of 3-Dimensional Molecular Descriptors. *Computers Chem.* **1990**, *14*, 237-246.
- (22) Estrada, E.; Montero, L. A. Bond Order Weighted Graphs in Molecules as Structure-Property Indices. *Mol. Eng.* **1993**, *2*, 263-273.
- (23) Estrada, E. Three-Dimensional Molecular Descriptors Based on Electron Charge Density Weighted Graphs. *J. Chem. Inf. Comput. Sci.* In press.
- (24) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. Submitted for publication.
- (25) Daudel, R.; Lefebvre, R.; Moser, C. *Quantum Chemistry. Methods and Applications*; Ed. Rev. pp 70-80.
- (26) Ortíz, P. J.; Pérez, C. S. *Elementos de Mecánica Cuántica y Estructura Atómica*; Universidad de La Habana: 1982.
- (27) Randić, M.; Hansen, P. J.; Jurs, P. C. Search for Useful Graph Theoretical Invariants of Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60-68.
- (28) Weast, R. D. *CRC Handbook of Chemistry and Physics*, 67th ed.; 3rd printing; CRC: Boca Raton, FL, 1987.
- (29) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155-168.

- (30) Leahy, D. E. Intrinsic Molecular Volume as a Measure of the Cavity. Term in Linear Solvation Energy Relationships: Octanol-Water Partition Coefficient and Aqueous Solubilities. *J. Pharm. Sci.* **1986**, *75*, 629–636.
- (31) Needham, N. D.; Chien Wei, I.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186–4194.
- (32) Basak, S. C.; Magnuson, V. R.; Niemi, G. I.; Regal, R. R.; Veigh, G. D. Topological Indices: Their Nature, Mutual Relatedness, and Applications. *Math. Modelling* **1987**, *8*, 300–305.
- (33) Razinger, M.; Chrétien, J. R.; Dubois, J. E. Structural Selectivity of Topological Indexes in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23–27.
- (34) Streitwieser, A., Jr. *Molecular Orbital Theory for Organic Chemists*; John Wiley & Sons, Inc.: New York, London, 1962; p 117.
- (35) Streitwieser, A., Jr. *Molecular Orbital Theory for Organic Chemists*; John Wiley & Sons, Inc.: New York, London, 1962; pp 120–125.
- (36) Randić, M. The Nature of Chemical Structure. *J. Math. Chem.* **1990**, *4*, 157–184.

CI940051D