

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

Ovidiu Ivanciuc,* Teodora Ivanciuc, and Alexandru T. Balaban*

University "Politehnica" of Bucharest, Faculty of Chemical Technology, Department of Organic Chemistry,
P.O. Box 12 - 243, 77206 Bucharest, Romania

Received April 1, 1997

Two new approaches are presented for the calculation of atom and bond parameters for heteroatom-containing molecules used in computing graph theoretic invariants. In the first approach, the atom and bond weights are computed on the basis of relative atomic electronegativity, using carbon as standard. In the second system, the relative covalent radii are used to compute atom and bond weights, again with the carbon atom as standard. The new definition of the atom and bond parameters leads to a periodic variation versus the atomic number Z , with a more natural variation when compared with the parameters defined only by Z . The two approaches are used to define and compute topological indices based on graph distance. A quantitative structure–property relationship study is reported for boiling points of 185 acyclic compounds with one or two oxygen or sulfur atoms (devoid of hydrogen bonding), in terms of four or five molecular descriptors.

INTRODUCTION

An important problem in theoretical chemistry is the deduction of molecular properties from the molecular structure considered at various levels of complexity. The first level of molecular structure description uses the connectivity of atoms and does not consider the information regarding the three-dimensional structure of the molecule. Molecular topology determines a large number of molecular properties, ranging from physicochemical and thermodynamic properties to chemical reactivity and biological activity. For the latter type of activity, however, three-dimensional structure descriptors are increasingly being developed.² Among the large class of molecular graph invariants used to describe the chemical structure, we mention here graph theoretic polynomials and spectra, spectral moments, topological indices, and distances, walks, and paths in graphs. The most used molecular graph descriptors in establishing quantitative structure–property relationships (QSPRs) and quantitative structure–activity relationships (QSARs) are topological indices (TIs).^{3–8}

A graph $G = G(V, E)$ is an ordered pair consisting of two sets $V = V(G)$ and $E = E(G)$. Elements of the set $V(G)$ are called vertexes and elements of the set $E(G)$, involving the binary relation between the vertexes, are called edges. In this paper, chemical structures are represented as molecular graphs. By removing all hydrogen atoms from the chemical formula of a chemical compound containing covalent bonds one obtains the hydrogen-depleted (or hydrogen-suppressed) molecular graph of that compound, whose vertexes correspond to non-hydrogen atoms and whose edges correspond to covalent bonds. In this study, the expressions "molecular

graph" and "molecule", "vertex" and "atom", "edge" and "bond" are used interchangeably.

The presence of multiple or aromatic bonds, or of heteroatoms, in molecular graphs requires the development of special parameters. A general approach was developed by Trinajstić and co-workers⁹ by weighting the contributions of atoms and bonds with parameters based on the atomic number Z and the topological bond order. The weight associated with an atom, $A_w Z$, was defined with the following equation:

$$A_w Z_i = 1 - 6/Z_i \quad (1)$$

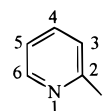
where Z_i is the atomic number Z of the atom i . A few values of the parameter $A_w Z$ for various elements are presented in Table 1, column 4.

The bond weight $B_w Z$, is defined as:

$$B_w Z_{ij} = 36/bZ_i Z_j \quad (2)$$

where b is the bond order, which takes the value 1 for single bonds, 2 for double bonds, 3 for triple bonds, and 1.5 for aromatic bonds. Table 2 presents some bond parameters $B_w Z$.

As an example of the Z weighting scheme, we present in Table 3 the atom- and bond-weighted distance matrix of 2-methylpyridine (see structure).



The Z weighting scheme was applied with good results in various QSAR studies.^{10–12} There exist other methods for computing heteroatom parameters¹³ and weights for chemical

* Corresponding authors: E-mail: o_ivanciuc@chim.upb.ro and balaban@acad.ro, respectively.

Table 1. Atomic Parameters Computed with Different Weighting Schemes^a

Atom	X	Y	A _w Z	A _w X	A _w Y
B	0.851	1.038	-0.200	-0.175	0.037
C	1.000	1.000	0.000	0.000	0.000
N	1.149	0.963	0.143	0.130	-0.038
O	1.297	0.925	0.250	0.229	-0.081
F	1.446	0.887	0.333	0.308	-0.127
Si	0.937	1.128	0.571	-0.067	0.113
P	1.086	1.091	0.600	0.079	0.083
S	1.235	1.053	0.625	0.190	0.050
Cl	1.384	1.015	0.647	0.277	0.015
As	0.946	1.379	0.818	-0.057	0.275
Se	1.095	1.341	0.824	0.087	0.254
Br	1.244	1.303	0.829	0.196	0.233
Te	0.954	1.629	0.885	-0.048	0.386
I	1.103	1.591	0.887	0.093	0.371

^a X represents the relative electronegativity; Y the relative covalent radii; A_wZ the atomic number Z atomic parameter; A_wX the atomic parameters based on relative electronegativity; A_wY the atomic weights defined by relative covalent radii.

Table 2. Bond Parameters B_wZ Computed with the Atomic Number Z

Atom _i	Atom _j	single	double	triple	aromatic
C	C	1.000	0.500	0.333	0.667
C	N	0.857	0.429	0.286	0.571
C	O	0.750	0.375		
C	F	0.667			
C	Si	0.429			
C	P	0.400			
C	S	0.375	0.188		0.250
C	Cl	0.353			
C	Se	0.176			
C	Br	0.171			
C	Te	0.115			
C	I	0.113			
N	N	0.735	0.367		0.490
N	O	0.643	0.321		0.423
O	S	0.281	0.141		

Table 3. The Atom- and Bond-Weighted Distance Matrix of 2-Methylpyridine Computed with Parameters Derived from the Z Weighting Scheme

	1	2	3	4	5	6	7
1	0.143	0.571	1.238	1.905	1.238	0.571	1.571
2	0.571	0.000	0.667	1.333	1.810	1.143	1.000
3	1.238	0.667	0.000	0.667	1.333	1.810	1.667
4	1.905	1.333	0.667	0.000	0.667	1.333	2.333
5	1.238	1.810	1.333	0.667	0.000	0.667	2.810
6	0.571	1.143	1.810	1.333	0.667	0.000	2.143
7	1.571	1.000	1.667	2.333	2.810	2.143	0.000

bonds.¹⁴ A different approach for considering heteroatoms was developed by one of us, which takes into account the periodicity of the two chemical properties electronegativity and covalent radii.¹⁵

We will present a brief overview of the topological indices that are relevant for the present paper. The Wiener index, *W*, is equal to the sum of topological distances between all *N* atoms in the molecular graph, and is expressed by the following equation^{16,17}:

$$W = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N D_{ij} \quad (3)$$

where *D* is the distance matrix of the molecular graph and *D_{ij}* is the *ij*-entry in this matrix.

The distance sum of the vertex *i*, *DS_i*, is defined as the sum of the topological distances between vertex *i* and every vertex in the molecular graph; that is, the sum over row *i* in the *D* matrix^{18,19}:

$$DS_i = \sum_{j=1}^N D_{ij} \quad (4)$$

The highly discriminating TI average distance sum connectivity (*J*) was defined by the following formula^{18,20}:

$$J = \frac{m}{\mu + 1} \sum_{E(G)} (DS_i DS_j)^{-1/2} \quad (5)$$

where *DS_i* and *DS_j* denote the distance sums of the endpoints of an edge in the molecular graph, *m* is the number of edges in the molecular graph, *μ* is the cyclomatic number, and *E(G)* indicates that the summation goes over all edges in the molecular graph.

Because the index *J* contains the factor *m*/(*μ* + 1), it does not increase automatically with increasing number of vertexes and cycles, unlike most TIs. Because of its design, *J* has a very low degeneracy, as was proved analytically²¹ and tested with a computer program.²² For many infinite graphs, *J* attains a finite limit²³; for example, for an infinite linear alkane, *J* tends toward *π* = 3.14159. The definition of *J* was extended for molecular graphs containing heteroatoms and/or multiple bonds^{15,24} and used with success in QSAR studies.^{25,26}

Reciprocal distance invariants were recently defined²⁷ and used for the generation and coding of alkanes²⁸ and to define the reciprocal distance matrix and other molecular graph invariants.^{29,30} In the present paper we will use a slightly modified definition for the reciprocal distance matrix, because a molecular graph that contains heteroatoms has some nonzero diagonal elements. The reciprocal distance matrix *RD(G)* = *RD*, of a graph *G* with *N* vertexes, is a square *N* × *N* symmetrical matrix, whose entries, *RD_{ij}*, are equal to the reciprocal of the distance between vertexes *i* and *j* for nondiagonal elements, and is equal to *D_{ii}* for diagonal elements:

$$RD_{ij} = \begin{cases} D_{ii}, & \text{for } i = j \\ 1/D_{ij}, & \text{for } i \neq j \end{cases} \quad (6)$$

Recently we investigated the precision and error of the computation of molecular graph invariants with the *Z* weighting scheme.³¹ Despite its widespread use, the *Z* weighting scheme gives atomic and bond parameters that monotonically increase or decrease with the atomic number, neglecting the periodic variation of atomic properties. The scope of the present paper is to introduce two new schemes for the calculation of atom and bond parameters used in computing graph theoretic invariants. In the first scheme, the atom and bond weights are computed on the basis of relative atomic electronegativity, and in the second system, the relative covalent radii are used. The two approaches are used to define and compute graph distance topological indices.

PARAMETERS BASED ON ELECTRONEGATIVITY OR COVALENT RADIUS

As already mentioned, one of us¹⁵ introduced two atom-weighting schemes for the index *J*, by using the electrone-

gativity and covalent radii, respectively. The electronegativities of main group atoms recalculated by Sanderson³² on the Pauling's scale, with fluorine, sodium, and hydrogen having the values 4, 0.56, and 2.592, respectively, were used in a biparametric linear regression equation to obtain computed electronegativities, S . The two parameters used in devising this equation are Z , the atomic number, and G , the number of the group in Mendeleev's short form of the Periodic System. Using a similar biparametric equation, the relative electronegativities X , based on the calculated value for carbon $S(\text{carbon}) = 2.629$, were obtained:

$$S_i = 1.1032 - 0.0204Z_i + 0.4121G_i \quad (7)$$

$$X_i = 0.4196 - 0.0078Z_i + 0.1567G_i \quad (8)$$

Using the covalent radii R selected from Sanderson's book, a similar biparametric correlation affords calculated covalent radii data and relative covalent radii Y based on calculated $R(\text{carbon}) = 87.126$ picometers (pm), according to the following equations:

$$R_i = 97.4989 + 1.3898Z_i - 4.6779G_i \quad (9)$$

$$Y_i = 1.1191 + 0.0160Z_i - 0.0537G_i \quad (10)$$

Computed relative electronegativities X and relative covalent radii Y for various elements are presented in Table 1, columns 2 and 3, respectively.

Using the relative electronegativities X and the relative covalent radii Y we now define two new general weighting schemes for molecular graphs. The first one, which uses the relative electronegativity, defines the weight parameter A_wX for atom i :

$$A_wX_i = 1 - 1/X_i \quad (11)$$

A bond weight parameter B_wX is defined as follows:

$$B_wX_{ij} = 1/bX_iX_j \quad (12)$$

Selected values for the parameters A_wX are presented in Table 1, column 5.

The weighting scheme based on atomic electronegativity gives A_wX values that span a more narrow range of values compared with A_wZ . For the elements with an electronegativity lower than that of carbon (Si and Te in Table 1), the corresponding A_wX values are negative, whereas for the elements with a relative electronegativity >1 , the A_wX parameter is positive. From the definition of A_wX , it is clear that this parameter captures the periodicity of electronegativity and can generate molecular descriptors that express both the effect of topology and of electronegativity. Selected bond parameters B_wX are presented in Table 4. The carbon-carbon bonds have weights identical with the corresponding B_wZ parameters. The values of C-Si and C-Te bonds are >1 , whereas all other weights are <1 .

The atom- and bond-weighted distance matrix of 2-methylpyridine computed with the X weighting scheme is presented in Table 5.

The second weighting scheme, which uses the relative covalent radii, defines a weight parameter A_wY for atom i as follows:

Table 4. Bond Parameters B_wX Computed with the Relative Electronegativity X

Atom _i	Atom _j	single	double	triple	aromatic
C	C	1.000	0.500	0.333	0.667
C	N	0.870	0.435	0.290	0.580
C	O	0.771	0.386		
C	F	0.692			
C	Si	1.067			
C	P	0.921			
C	S	0.810	0.405		0.540
C	Cl	0.723			
C	Se	0.913			
C	Br	0.804			
C	Te	1.048			
C	I	0.907			
N	N	0.757	0.379		0.505
N	O	0.671	0.336		0.447
O	S	0.624	0.312		xxx

Table 5. The Atom- and Bond-Weighted Distance Matrix of 2-Methylpyridine Computed with Parameters Derived from the X Weighting Scheme

	1	2	3	4	5	6	7
1	0.130	0.580	1.247	1.914	1.247	0.580	1.580
2	0.580	0.000	0.667	1.333	1.827	1.160	1.000
3	1.247	0.667	0.000	0.667	1.333	1.827	1.667
4	1.914	1.333	0.667	0.000	0.667	1.333	2.333
5	1.247	1.827	1.333	0.667	0.000	0.667	2.827
6	0.580	1.160	1.827	1.333	0.667	0.000	2.160
7	1.580	1.000	1.667	2.333	2.827	2.160	0.000

$$A_wY_i = 1 - 1/Y_i \quad (13)$$

The second weighting scheme defines the bond weight parameter B_wY as follows:

$$B_wY_{ij} = 1/bY_iY_j \quad (14)$$

Selected values for parameters A_wY and B_wY are presented in Table 1, column 6, and Table 6, respectively. The Y weighting scheme applied to the computation of the atom- and bond-weighted distance matrix of 2-methylpyridine gives the matrix shown in Table 7.

All elements with covalent radii lower than that of carbon present negative values for the A_wY parameters; Table 1 presents three such cases, namely N, O, and F. All other heteroatoms from Table 1 present positive A_wY values. The closer the covalent radii of an atom to that of carbon, the closer to zero the corresponding A_wY parameter. The B_wY parameters from Table 6 show the same kind of periodicity of values as the one observed in Table 4 in the case of B_wX . On the other hand, the variation of B_wY parameters does not parallel that of B_wX parameters; for example, B_wX decreases for the bonds C-X ($X=C, N, O, F$), whereas B_wY increases for the same set of bonds.

OPERATOR DEFINITIONS

There exist many molecular graph descriptors, and the development of new molecular matrixes can multiply their number. Also, each structural descriptor can be computed by using one of the three weighting schemes compared in the present study. This situation calls for some kind of rules in the notation of the descriptors; that is, rules that allow the description of the mathematical operation, matrix, and weighting scheme used to compute a particular invariant. In

Table 6. Bond Parameters B_wY Computed with the Relative Covalent Radii Y

Atom _i	Atom _j	single	double	triple	aromatic
C	C	1.000	0.500	0.333	0.667
C	N	1.038	0.519	0.346	0.692
C	O	1.081	0.541		
C	F	1.127			
C	Si	0.887			
C	P	0.917			
C	S	0.950	0.475		0.633
C	Cl	0.985			
C	Se	0.746			
C	Br	0.767			
C	Te	0.614			
C	I	0.629			
N	N	1.078	0.539		0.719
N	O	1.123	0.561		0.748
O	S	1.027	0.513		

Table 7. The Atom- and Bond-Weighted Distance Matrix of 2-Methylpyridine Computed with Parameters Derived from the Y Weighting Scheme

	1	2	3	4	5	6	7
1	−0.038	0.692	1.359	2.026	1.359	0.692	1.692
2	0.692	0.000	0.667	1.333	2.000	1.385	1.000
3	1.359	0.667	0.000	0.667	1.333	2.000	1.667
4	2.026	1.333	0.667	0.000	0.667	1.333	2.333
5	1.359	2.000	1.333	0.667	0.000	0.667	3.000
6	0.692	1.385	2.000	1.333	0.667	0.000	2.385
7	1.692	1.000	1.667	2.333	3.000	2.385	0.000

the present paper we use families of related molecular graph descriptors computed with graph operators.³³ An operator uses some mathematical operations performed on a certain molecular matrix computed with a weighting scheme. The three weighting schemes are denoted by Z for the atomic number scheme, X for the relative electronegativity scheme, and Y for the relative covalent radii scheme. The operatorial notation of the graph invariants offers a simple way to describe similar descriptors and to denote the weighting scheme, and can be extended to the development of new matrixes.

The Wiener operator, $Wi(M,w,G)$, is defined by the following equation:

$$Wi(M, w, G) = \sum_{i=1}^N \sum_{j=1}^N M_{ij} \quad (15)$$

where M represents the molecular matrix of G and w is the weighting scheme. If M is the distance matrix, the Wiener operator gives W , whereas if M is the RD matrix, the Wiener operator gives the Harary index.^{29,30}

By analogy with the hyper-Wiener index,^{34–37} we define the hyper-Wiener operator of a graph G , $HyWi(M,G)$, as follows:

$$HyWi(M,G) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N ([M]_{ij}^2 + [M]_{ij}) \quad (16)$$

If M is the distance matrix, the $HyWi$ operator is identical with the hyper-Wiener index R .^{34–37} Diudea^{36,37} has recently proposed two molecular graph matrixes: the distance-delta matrix, D_Δ , and distance-path matrix, D_p , whose elements are calculated by a combinatorial algorithm from the classical distance matrix D :

$$[D_\Delta]_{ij} = \binom{[D]_{ij}}{2} \quad (17)$$

$$[D_p]_{ij} = \binom{[D]_{ij} + 1}{2} \quad (18)$$

The element $[D_\Delta]_{ij}$ counts the number of “internal” paths (larger than unity) included in the shortest paths between vertexes i and j ; the element $[D_p]_{ij}$ counts all internal paths included in the shortest paths between vertexes i and j in a graph. The matrix D_p allows the direct computation of the hyper-Wiener index R , whereas the matrix D_Δ gives an index related to the “non-Wiener” part of the hyper-Wiener index.

The Atom Sum operator for the vertex i , $AS(M,w,G)_i$ is defined as the sum of the elements in the row i , or column i , of the molecular matrix M computed with the weighting scheme w :

$$AS(M,w,G)_i = \sum_{j=1}^N M_{ij} \quad (19)$$

If M is the adjacency matrix, the AS operator gives the atom degree; if M is the distance matrix, the AS operator gives the distance sum DS ; if M is the RD matrix, the AS operator is identical with the reciprocal distance sum.³⁰

By analogy with the J index we define a more general type of indices, the Ivanciuc-Balaban operator, $IB(M,w,G)$:

$$IB(M, w, G) = \frac{m}{\mu + 1} \sum_{E(G)} (AS_i AS_j)^{-1/2} \quad (20)$$

The IB operator gives the J index when M is the distance matrix.

The two molecular operators just described will be used to compute a series of topological indices. Table 8 presents the Wiener and Ivanciuc-Balaban indices obtained from the distance matrix computed with the three weighting schemes. We have considered nine molecules related to ethanol, with various heteroatoms (Het), with the molecular graph C—C—Het. Inspection of the values in Table 8 shows that whereas $Wi(D,Z)$ decreases and $IB(D,Z)$ increases with increasing atomic number, the $Wi(D,X)$, $Wi(D,Y)$, $IB(D,X)$, and $IB(D,Y)$ indices show a periodicity in their values. For the elements in the first row of the Periodic System, $Wi(D,X)$ and $IB(D,Y)$ decrease, and $Wi(D,Y)$ and $IB(D,X)$ increase from carbon to fluorine. For the halogen series, $Wi(D,X)$ and $IB(D,Y)$ increase, but $Wi(D,Y)$ and $IB(D,X)$ decrease from fluorine to iodine.

In Table 9 the Wiener and Ivanciuc-Balaban operators are applied to the reciprocal distance matrix computed with the Z , X , and Y weighting schemes. Again, the $Wi(RD,X)$, $Wi(RD,Y)$, $IB(RD,X)$, and $IB(RD,Y)$ indices show a periodicity in their values. For the elements in the first row of the Periodic System, $Wi(RD,X)$ and $IB(RD,Y)$ increase, whereas $Wi(RD,Y)$ and $IB(RD,X)$ decrease from carbon to fluorine. For the halogen series, $Wi(RD,X)$ and $IB(RD,Y)$ decrease and $Wi(RD,Y)$ and $IB(RD,X)$ increase.

A scaling factor was defined in order to allow a greater flexibility for the weighting methods using relative electronegativity or covalent radius.²⁴ The scaling factor is used to multiply the atomic parameters and has the effect of enhancing or depressing the effect of heteroatoms. The use

Table 8. Topological Indices Computed with the Wiener and Ivanciuc-Balaban Operators for the Distance Matrix and with Three Weighting Schemes Z, X, and Y

compound	Wi(D,Z)	Wi(D,X)	Wi(D,Y)	IB(D,Z)	IB(D,X)	IB(D,Y)
CH ₃ -CH ₂ -CH ₃	4.000	4.000	4.000	1.633	1.633	1.633
CH ₃ -CH ₂ -SiH ₃	3.429	4.067	3.887	2.148	1.589	1.714
CH ₃ -CH ₂ -NH ₂	3.857	3.870	4.038	1.736	1.726	1.607
CH ₃ -CH ₂ -PH ₂	3.400	3.921	3.917	2.182	1.689	1.692
CH ₃ -CH ₂ -AsH ₂	3.182	4.057	3.725	2.491	1.595	1.845
CH ₃ -CH ₂ -OH	3.750	3.771	4.081	1.823	1.806	1.580
CH ₃ -CH ₂ -SH	3.375	3.810	3.950	2.213	1.774	1.668
CH ₃ -CH ₂ -SeH	3.176	3.913	3.746	2.500	1.694	1.827
CH ₃ -CH ₂ -TeH	3.115	4.048	3.614	2.604	1.601	1.948
CH ₃ -CH ₂ -F	3.667	3.692	4.127	1.897	1.875	1.551
CH ₃ -CH ₂ -Cl	3.353	3.723	3.985	2.242	1.847	1.643
CH ₃ -CH ₂ -Br	3.171	3.804	3.767	2.508	1.779	1.809
CH ₃ -CH ₂ -I	3.113	3.907	3.629	2.608	1.699	1.933

Table 9. Topological Indices Computed with the Wiener and Ivanciuc-Balaban Operators for the Reciprocal Distance Matrix and with Three Weighting Schemes Z, X, and Y

compound	Wi(RD,Z)	Wi(RD,X)	Wi(RD,Y)	IB(RD,Z)	IB(RD,X)	IB(RD,Y)
CH ₃ -CH ₂ -CH ₃	2.500	2.500	2.500	2.309	2.309	2.309
CH ₃ -CH ₂ -SiH ₃	4.605	2.354	2.772	1.417	2.415	2.138
CH ₃ -CH ₂ -NH ₂	2.848	2.813	2.415	2.095	2.114	2.369
CH ₃ -CH ₂ -PH ₂	4.814	2.686	2.696	1.364	2.189	2.183
CH ₃ -CH ₂ -AsH ₂	8.164	2.375	3.233	0.870	2.399	1.899
CH ₃ -CH ₂ -OH	3.155	3.091	2.324	1.936	1.968	2.437
CH ₃ -CH ₂ -SH	5.019	2.978	2.616	1.316	2.025	2.233
CH ₃ -CH ₂ -SeH	8.340	2.704	3.168	0.855	2.178	1.930
CH ₃ -CH ₂ -TeH	11.448	2.394	3.635	0.666	2.385	1.729
CH ₃ -CH ₂ -F	3.433	3.346	2.230	1.811	1.849	2.514
CH ₃ -CH ₂ -Cl	5.220	3.242	2.534	1.272	1.895	2.287
CH ₃ -CH ₂ -Br	8.516	2.995	3.101	0.841	2.016	1.962
CH ₃ -CH ₂ -I	11.618	2.721	3.577	0.659	2.168	1.752

Table 10. Coefficients, Structural Descriptors, and Statistical Indices of the Best MLR Equations with Four and Five Independent Variables for the Computation of the 185 Boiling Points of Oxygen and Sulfur Compounds Described in References 38

eq	a ₀	a ₁	TI ₁	a ₂	TI ₂	a ₃	TI ₃	a ₄	TI ₄	a ₅	TI ₅	r	s	F
1	-104.02	52.94	NoS	-15.38	² χ ^v	-3.05	MinSp(D,X)	75.33	MaxSp(RD,Y)	-15.28	IB(D,X)	0.9899	6.94	1741
2	-104.99	56.77	NoS	-14.10	² χ ^v	-3.02	MinSp(D,Z)	75.24	MaxSp(RD,Y)	-15.61	IB(D,X)	0.9897	6.99	1715
3	-44.23	85.22	NoS	-4.94	Wi(RD,Z)	40.47	HyWi(A,Y)	-0.0025	HyWi(D _p ,Y)	-17.85	IB(D,X)	0.9893	7.13	1646
4	-56.92	50.54	NoS	-15.97	² χ ^v	-49.29	MaxSp(A,Y)	-3.32	MinSp(D,X)	76.01	MaxSp(RD,Y)	0.9892	7.15	1637
5	-54.56	54.39	NoS	-14.59	² χ ^v	-52.47	MaxSp(A,Y)	-3.27	MinSp(D,Z)	76.53	MaxSp(RD,Y)	0.9892	7.18	1625
6	-88.53	15.08	NoO	50.89	NoS	-8.56	² χ ^v	106.00	MaxSp(RD,Y)	-55.16	IB(D,X)	0.9891	7.19	1621
7	-102.61	28.00	NoO	141.86	NoS	49.35	Wi(A,Z)	-9.40	Wi(RD,Z)			0.9886	7.35	1933
8	-110.68	56.13	NoS	-15.91	² χ ^v	-4.22	MinSp(D,X)	60.32	MaxSp(RD,Y)			0.9884	7.41	1905
9	-112.20	61.58	NoS	-14.12	² χ ^v	-4.22	MinSp(D,Z)	59.71	MaxSp(RD,Y)			0.9882	7.48	1869
10	-104.27	42.23	NoO	161.57	NoS	-9.90	Wi(RD,Z)	50.87	HyWi(A,Z)			0.9877	7.63	1791
11	-54.05	88.61	NoS	-5.48	Wi(RD,Z)	38.52	HyWi(A,Y)	-8.99	IB(D,X)			0.9870	7.84	1693
12	-79.58	94.17	NoS	-3.85	³ χ _c ^v	-6.11	Wi(RD,Z)	39.74	HyWi(A,Y)			0.9867	7.94	1652
13	-59.54	20.10	NoO	44.35	NoS	95.56	MaxSp(RD,Y)	-60.81	IB(D,X)			0.9865	7.98	1635

of the scaling factor gave good results in QSAR studies.^{25,26} We propose to use a scaling factor for the Z, X, and Y weighting schemes; this factor is used to multiply the atomic number Z, and X_i and Y_i parameters used to compute the A_w and B_w parameters. For carbon, the scaling factor is equal to 1.

CORRELATIONS WITH THE BOILING POINTS OF ACYCLIC ETHERS, PEROXIDES, ACETALS, AND THEIR SULFUR ANALOGUES

The two weighting schemes defined in this paper, together with the Z weighting scheme, were tested in a QSPR study that computes the boiling points of a set of 185 acyclic ethers, peroxides, acetals, and their sulfur analogues; this series was studied earlier with other descriptors.³⁸ The following five molecular graph matrixes were used to compute TIs: A, D,

RD, D_p,^{36,37} D_Δ.^{36,37} The list of the 78 structural descriptors used in the QSPR study is as follows:

1. Constitutional descriptors: the number of carbon atoms, NoC, the number of oxygen atoms, NoO, the number of sulfur atoms, NoS, and the molecular weight, MW.

2. The Randić, Kier and Hall connectivity indices: ⁰χ^v, ¹χ^v, ²χ^v, ³χ_p^v, and ³χ_c^v.

3. Indices representing the minimum (MinSp) and maximum (MaxSp) value of the molecular graph spectra: MaxSp(A,X), MinSp(A,X), MaxSp(A,Y), MinSp(A,Y), MaxSp(A,Z), MinSp(A,Z), MaxSp(D,X), MinSp(D,X), MaxSp(D,Y), MinSp(D,Y), MaxSp(D,Z), MinSp(D,Z), MaxSp(RD,X), MinSp(RD,X), MaxSp(RD,Y), MinSp(RD,Y), MaxSp(RD,Z), MinSp(RD,Z), MaxSp(D_p,X), MinSp(D_p,X), MaxSp(D_p,Y), MinSp(D_p,Y), MaxSp(D_p,Z), MinSp(D_p,Z), MaxSp(D_Δ,X),

MinSp(D_{Δ}, X), MaxSp(D_{Δ}, Y), MinSp(D_{Δ}, Y), MaxSp(D_{Δ}, Z), MinSp(D_{Δ}, Z).

3. Topological indices obtained from the Wiener operator applied to a set of molecular graph matrixes: Wi(A, X), Wi(A, Y), Wi(A, Z), Wi(D, X), Wi(D, Y), Wi(D, Z), Wi(RD, X), Wi(RD, Y), Wi(RD, Z), Wi(D_p, X), Wi(D_p, Y), Wi(D_p, Z), Wi(D_{Δ}, X), Wi(D_{Δ}, Y), Wi(D_{Δ}, Z).

4. Topological indices obtained from the hyper-Wiener operator applied to a set of molecular graph matrixes: HyWi(A, X), HyWi(A, Y), HyWi(A, Z), HyWi(D, X), HyWi(D, Y), HyWi(D, Z), HyWi(RD, X), HyWi(RD, Y), HyWi(RD, Z), HyWi(D_p, X), HyWi(D_p, Y), HyWi(D_p, Z), HyWi(D_{Δ}, X), HyWi(D_{Δ}, Y), HyWi(D_{Δ}, Z).

5. Topological indices obtained from the Ivanciuc-Balaban operator applied to a set of molecular graph matrixes: IB(A, X), IB(A, Y), IB(A, Z), IB(D, X), IB(D, Y), IB(D, Z), IB(RD, X), IB(RD, Y), IB(RD, Z).

The search for the best multilinear regression (MLR) with n topological indices as independent variables begins by computing all one-parameter correlation equations, selecting for further use only the structural descriptors with a correlation coefficient greater than a threshold, $|r_{\min}| > 0.15$. In the second step, all sets of n structural descriptors with pairwise correlation coefficient r_{ij} lower than a threshold, $|r_{ij}| < 0.8$, are selected and the corresponding MLR equation and statistical indices are computed. In the third step, the best MLR equations are selected and reported. The best MLR equations are reported in Table 10 (eqs 1–6 with four independent variables, and eqs 7–13 with five independent variables). By increasing the number of independent variables to six or more, the correlation coefficient does not significantly increase, and therefore we do not report QSPR equations with more than five independent variables.

The results presented in Table 10 show that the QSPR equations are of good quality, giving better results than those reported in ref 38. Each equation from Table 10 contains the NoS descriptor. Also, the connectivity index ${}^2\chi^v$ appears in a large number of equations. From the set of TIs computed using graph matrixes, MinSp, MaxSp, and IB appear with a greater frequency. There is only one descriptor computed with the D_p matrix, whereas the TIs computed with the D_{Δ} matrix were not selected in the best QSPR models. The best two eqs 1 and 2 with five variables are related to the second best eqs 8 and 9 with four variables. The best eq with four variables (eq 7) does not have a corresponding equation with five variables in Table 10. The standard deviation for most correlations in Table 10 is ~ 7 °C.

Overall, the TIs derived from the X and Y weighting schemes are well represented in the models presented in Table 10, thus demonstrating their utility in computing structural descriptors for QSPR studies.

ACKNOWLEDGMENT

We acknowledge the partial financial support of this research by the Ministry of Research and Technology under Grant 310 TA69 and by the Ministry of National Education under Grant 7001 T34.

REFERENCES AND NOTES

- Previous part: Ivanciuc, O. Design of Topological Indices. Part 9. A New Recurrence Relationship for the Hosoya $Z(\text{Ch})$ Index of a Molecular Graph. *Rev. Roum. Chim.*, in press.
- Balaban, A. T. From Chemical Graphs to 3D Molecular Modeling. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum: New York, 1997; pp 1–24.
- (a) Balaban, A. T. Applications of Graph Theory in Chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334–343. (b) Balaban, A. T. Using Real Numbers as Vertex Invariants for Third Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23–28. (c) Balaban, A. T. Lowering the Intra- and Intermolecular Degeneracy of Topological Invariants. *Croat. Chem. Acta* **1993**, *66*, 447–458. (d) Balaban, A. T. Local versus Global (i.e. Atomic versus Molecular) Numerical Modeling of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398–402. (e) Balaban, A. T. Real-Number Local (Atomic) Invariants and Global (Molecular) Topological Indices. *Rev. Roum. Chim.* **1994**, *39*, 245–257. (f) Balaban, A. T. Local (Atomic) and Global (Molecular) Graph-Theoretical Descriptors. *SAR QSAR Environ. Res.* **1995**, *3*, 81–95.
- Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1992.
- Hosoya, H. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (a) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976. (b) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies: Letchworth, 1986.
- Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
- Barysz, M.; Jashari, G.; Lall, R. S.; Srivastava V. K.; Trinajstić, N. On the Distance Matrix of Molecules Containing Heteroatoms. In: *Chemical Applications of Topology and Graph Theory*; King, R. B. Ed.; Elsevier: Amsterdam, 1983; pp 222–227.
- Medić-Šarić, M.; Nikolić, S.; Matijević-Sosa, J. A QSPR Study of 3-(Phthalimidoalkyl)-pyrazolin-5-ones. *Acta Pharm.* **1992**, *42*, 153–167.
- Nikolić, S.; Medić-Šarić, M.; Matijević-Sosa, J. A QSAR Study of 3-(Phthalimidoalkyl)-pyrazolin-5-ones. *Croat. Chem. Acta* **1993**, *66*, 151–160.
- Nikolić, S.; Trinajstić, N.; Mihalić, Z. Molecular Topological Index: An Extension to Heterosystems. *J. Math. Chem.* **1993**, *12*, 251–264.
- Randić, M. Novel Graph Theoretical Approach to Heteroatoms in Quantitative Structure–Activity Relationships. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213–227.
- Balaban, A. T.; Bonchev, D.; Seitz, W. A. Topological/Chemical Distances and Graph Centers in Molecular Graphs with Multiple Bonds. *J. Mol. Struct. (Theochem)* **1993**, *280*, 253–260.
- Balaban, A. T. Chemical Graphs. Part 48. Topological Index J for Heteroatom-Containing Molecules Taking Into Account Periodicities of Element Properties. *MATCH (Commun. Math. Chem.)* **1986**, *21*, 115–122.
- Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- Wiener, H. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636–2638.
- Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- Seybold, P. G. Topological Influences on the Carcinogenicity of Aromatic Hydrocarbons. I. The Bay Region Geometry. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1983**, *10*, 95–101.
- Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199–206.
- Balaban, A. T.; Quintas, L. V. The Smallest Graphs, Trees, and 4-Trees with Degenerate Topological Index J. *MATCH (Commun. Math. Chem.)* **1983**, *14*, 213–233.
- Balaban, A. T.; Filip, P. Computer Program for Topological Index J (Average Distance Sum Connectivity). *MATCH (Commun. Math. Chem.)* **1984**, *16*, 163–190.
- Balaban, A. T.; Ionescu-Pallas, N.; Balaban, T.-S. Asymptotic Values of Topological Indices J and J' (Average Distance Sum Connectivities) for Infinite Acyclic and Cyclic Graphs. *MATCH (Commun. Math. Chem.)* **1985**, *17*, 121–146.
- Balaban, A. T.; Ivanciuc, O. FORTRAN 77 Computer Program for Calculating the Topological Index J for Molecules Containing Heteroatoms. In *MATH/CHEM/COMP 1988*, Proceedings of an International Course and Conference on the Interfaces Between Mathematics, Chemistry and Computer Sciences, Dubrovnik, Yugoslavia, 20–25 June 1985, Graovac, A., Ed.; Elsevier, Amsterdam, *Stud. Phys. Theor. Chem.* **1989**, *63*, 193–212.
- Balaban, A. T.; Catana, C.; Dawson, M.; Niculescu-Duvaz, I. Applications of Weighted Topological Index J for QSAR of Carcino-

- genesis Inhibitors (Retinoic Acid Derivatives). *Rev. Roum. Chim.* **1990**, *35*, 997–1003.
- (26) Bonchev, D.; Mountain, C. F.; Seitz, W. A.; Balaban, A. T. Modeling the Anticancer Action of Some Retinoid Compounds by Making Use of the OASIS Method. *J. Med. Chem.* **1993**, *36*, 1562–1569.
- (27) Ivanciuc, O. Design on Topological Indices. 1. Definition of a Vertex Topological Index in the Case of 4-Trees. *Rev. Roum. Chim.* **1989**, *34*, 1361–1368.
- (28) Balaban, T. S.; Filip, P. A.; Ivanciuc, O. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and Topological Indices. Derived Canonical Labeling and Coding of Trees and Alkanes. *J. Math. Chem.* **1992**, *11*, 79–105.
- (29) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235–250.
- (30) Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices. *J. Math. Chem.* **1993**, *12*, 309–318.
- (31) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 5. Precision and Error in Computing Graph Theoretic Invariants for Molecules Containing Heteroatoms and Multiple Bonds. *MATCH* (*Commun. Math. Chem.*) **1994**, *30*, 117–139.
- (32) Sanderson, R. T. *Polar Covalence*; Academic: New York; 1983, p 41.
- (33) Ivanciuc, O.; Ivanciuc, T.; Diudea, M. V. Molecular Graph Matrixes and Derived Structural Descriptors. *SAR QSAR Environ. Res.* **1997**, *7*, 63–87.
- (34) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener Matrix: Source of Novel Graph Invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.
- (35) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener Matrix Invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361–367.
- (36) Diudea, M. V. Walk Numbers ${}^{\circ}W_M$: Wiener-type Numbers of Higher Rank. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 535–540.
- (37) Diudea, M. V. Wiener and Hyper-Wiener Numbers in a Single Matrix. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 833–836.
- (38) Balaban, A. T.; Kier, L. B.; Joshi, N. Correlations between Chemical Structure and Normal Boiling Points of Acyclic Ethers, Peroxides, Acetals, and Their Sulfur Analogues. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 237–244.

CI970021L