DESIGN OF TOPOLOGICAL INDICES. Part 14[1]

DISTANCE-VALENCY MATRICES AND STRUCTURAL DESCRIPTORS FOR VERTEX- AND EDGE-WEIGHTED MOLECULAR GRAPHS

Ovidiu IVANCIUC

Department of Organic Chemistry,
Faculty of Chemical Technology, "Politehnica" University
Oficiul 12 C.P. 243, 78100 Bucharest, Roumania
E-mail: o_ivanciuc@chim.upb.ro

Topological indices represent an important class of structural descriptors, widely used in modeling physical, chemical, or biological properties, in similarity and diversity assessment, database mining, and in the virtual screening of combinatorial libraries. Recently, the distance matrix was used as the source of a new family of molecular matrices, the distance-valency matrices. The definition of the distance-valency matrix is extended to vertex- and edge-weighted molecular graphs. Using five new approaches for the calculation of atom and bond parameters in weighted molecular graphs, topological indices derived from various distance-valency matrices are applied in QSPR models for the amines boiling temperatures. The topological indices are computed with the Wiener, hyper-Wiener, and graph spectra operators. The best QSPR models contain only structural descriptors representing the minimum, MinSp, or maximum, MaxSp, eigenvalue for the following molecular matrices: A, D, RD, Dval(–1,1,1), Dval(–2,1,1), Dval(–2,0,0).

INTRODUCTION

A large number of topological indices, geometric, and quantum descriptors are used in modeling physical, chemical, or biological properties, in similarity and diversity assessment, database mining, and in the virtual screening of combinatorial libraries. A significant fraction of the topological indices used in quantitative structure-property relationship (QSPR) and quantitative structure-activity relationship (QSAR) models are derived from molecular graphs. Molecular graphs are non-directed chemical graphs that represent, in different conventions, molecules. In molecular graphs vertices correspond to atoms and edges represent covalent bonds between atoms, while geometrical features of molecules, such as bond lengths or bond angles, are not considered. Using molecular graphs the chemical structure of a chemical compound can be expressed by means of various graph matrices, polynomials, spectra, spectral moments, sequences counting distances, paths, and walks, or topological indices. A topological index (TI) is a numerical descriptor of the molecular structure based on certain topological feature of the molecular graph. Numerous reviews[2–13] were published on the theory and applications of topological indices in QSPR and QSAR models.

The distance-valency atomic descriptors VTI[14] were used to define a new class of matrices, the distance-valency matrices Dval.[15] These new matrices are very flexible in representing the chemical structure in a numerical form suitable for an efficient use in structure-property, structure-activity and molecular similarity studies. We have to mention that the previously defined distance and reciprocal distance graph matrices are particular cases of the distance-valency matrices. In the present paper the

definition of the distance-valency matrix is extended to vertex- and edge-weighted molecular graphs. The atom and bond parameters for heteroatom-containing organic compounds, represented as vertex- and edge-weighted molecular graphs, are computed with five new weighting schemes.[16] The first two weighting schemes compute the atom and bond weights on the basis of relative atomic mass, using carbon as standard. The remaining three systems use the relative atom polarizability, radius, and electronegativity to compute atom and bond weights, again with the carbon atom as standard. We have to mention that the definition introduced here is general, and other weighting schemes can be used to compute the vertex and edge weights.[17-21] Using these five weighting schemes, a large number of topological indices derived from various distance-valency matrices are used to develop QSPR models for the amines boiling temperatures. The topological indices are computed with the Wiener, hyper-Wiener, and graph spectra operators. The best QSPR models contain only structural descriptors representing the minimum, **MinSp**, or maximum, **MaxSp**, eigenvalue for the following molecular matrices: **A**, **D**, **RD**, **Dval**(-1,1,1), **Dval**(-2,1,1), **Dval**(-2,0,0).

VALENCY OF ATOMS IN MOLECULAR GRAPHS

In this paper chemical structures are represented as molecular graphs. By removing all hydrogen atoms from the chemical formula of a compound containing covalent bonds one obtains the hydrogen-depleted (or hydrogen-suppressed) molecular graph of that compound, whose vertices correspond to non-hydrogen atoms and whose edges correspond to covalent bonds. 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 vertices and elements of the set $E(G)$, involving the binary relation between the vertices, are called edges. The number of vertices N represents the number of elements in $V(G)$, $N = |V(G)|$, and the number of edges M represents the number of elements in $E(G)$, $M = |E(G)|$. The graph vertices are labeled from 1 to N, $V(G) = \{v_1, v_2, ..., v_N\}$, and the edge connecting vertices v_i and v_j is denoted by e_{ij}. Two vertices v_i and v_j of a graph G are said to be adjacent if there is an edge e_{ij} joining them. The degree of the vertex v_i from the molecular graph G, denoted by $\deg_i = \deg(G)_i$, is equal to the number of vertices adjacent to vertex v_i.

A generalization of the distance sum **DS** vertex invariant was proposed by introducing the distance-degree **VTI** descriptors.[14] These vertex (atomic) descriptors can be obtained also from the distance-valency matrices.[15] The valency of the vertex v_i, $\mathbf{val}(w)_i = \mathbf{val}(w,G)_i$, is defined as the sum of the weights $Ew(w)_{ij}$ of all edges e_{ij} incident with vertex v_i:

$$\mathbf{val}(w)_i = \sum_{e_{ij} \in E(G)} Ew(w)_{ij} \tag{1}$$

where w is the weighting scheme used to compute the bond parameters Ew. Alternatively, the valency of the vertex v_i can be computed as the sum of the non-diagonal elements in the row i, or column i, of the adjacency matrix $\mathbf{A}(w) = \mathbf{A}(w,G)$, of a molecular graph G with N vertices:

$$\mathbf{val}(w)_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{A}(w)]_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{A}(w)]_{ji} \tag{2}$$

The set of valency values for all vertices in a graph forms the vector $\mathbf{Val} = \mathbf{Val}(G)$ whose ith element represents the valency of the vertex v_i. In alkanes and cycloalkanes the degree of a vertex v_i, $\deg_i$, is identical with the valency of that vertex, $\mathbf{val}_i$, while for molecules containing heteroatoms and/or multiple bonds, represented as vertex- and/or edge-weighted molecular graphs, this equality is not true.

THE DISTANCE-VALENCY MATRIX

The distance-valency matrix of a simple (non-weighted) graph G with N vertices, $\mathbf{Dval}(p,q,r) =$ $= \mathbf{Dval}(p,q,r,G)$, is a square $N \times N$ matrix, whose entries $[\mathbf{Dval}(p,q,r)]_{ij}$ are equal to:[15]

$$[\mathbf{Dval}(p,q,r)]_{ij} = \begin{cases} d_{ij}^{p}\mathbf{val}_i^q\mathbf{val}_j^r & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{3}$$

where d_{ij} is the graph distance between vertices v_i and v_j, and $\mathbf{val}_i$ is the valency of vertex v_i. If the parameters p, q, r are natural numbers then the $\mathbf{Dval}$ matrix has elements natural numbers. In the particular case when $p = 1$ and $q = r = 0$ the $\mathbf{Dval}(p,q,r)$ matrix is identical with the distance matrix $\mathbf{D}$. If the parameters p, q, r are not natural numbers then one can obtain $\mathbf{Dval}$ matrices with elements real numbers. If $p = -1$ and $q = r = 0$ then $\mathbf{Dval}(-1,0,0)$ is identical with the reciprocal distance matrix $\mathbf{RD}$.[22-24] From the definition of the $\mathbf{Dval}$ matrix it is clear that nonsymmetric matrices can be obtained if $q \neq r$. The computation of the distance-valency matrix is presented for the molecular graph of 2-methylhexane 1.
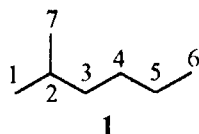
The distance matrix $\mathbf{D}(1)$ of 2-methylhexane 1, identical with the distance-valency matrix $\mathbf{Dval}(1,0,0,1)$, is:

Dval(1,0,0,1) = D(1)						
	1	2	3	4	5	6	7
1	0	1	2	3	4	5	2
2	1	0	1	2	3	4	1
3	2	1	0	1	2	3	2
4	3	2	1	0	1	2	3
5	4	3	2	1	0	1	4
6	5	4	3	2	1	0	5
7	2	1	2	3	4	5	0

The valency $\mathbf{Val}$ vector, identical with the degree $\mathbf{Deg}$ vector, is $\mathbf{Val}(1) = \{1, 3, 2, 2, 2, 1, 1\}$. With the above distance matrix and valency vector one can compute the $\mathbf{Dval}(1,1,1,1)$ matrix:

Dval(1,1,1,1)						
	1	2	3	4	5	6	7
1	0	3	4	6	8	5	2
2	3	0	6	12	18	12	3
3	4	6	0	4	8	6	4
4	6	12	4	0	4	4	6
5	8	18	8	4	0	2	8
6	5	12	6	4	2	0	5
7	2	3	4	6	8	5	0

As an example we present the computation of the element $\mathbf{Dval}(1,1,1,1)_{2,6}$:

$$\mathbf{Dval}(1,1,1,1)_{2,6} = [\mathbf{D}(1)]_{2,6}\mathbf{val}(1)_2\mathbf{val}(1)_6 = 4 \cdot 3 \cdot 1 = 12.$$

When at least one of the three parameters p, q, r is lower than zero one obtains distance-valency matrices with elements real numbers. If $q = r$ the **Dval** matrix is symmetric. Consider the **Dval**$(1,-1,-1,1)$ matrix of 2-methylhexane **1**:

Dval(1,-1,-1,1)

	1	2	3	4	5	6	7
1	0.000	0.333	1.000	1.500	2.000	5.000	2.000
2	0.333	0.000	0.167	0.333	0.500	1.333	0.333
3	1.000	0.167	0.000	0.250	0.500	1.500	1.000
4	1.500	0.333	0.250	0.000	0.250	1.000	1.500
5	2.000	0.500	0.500	0.250	0.000	0.500	2.000
6	5.000	1.333	1.500	1.000	0.500	0.000	5.000
7	2.000	0.333	1.000	1.500	2.000	5.000	0.000

The distance-valency matrices **Dval**(p,q,r) represent a rich family of matrices that display interesting properties. Using combinations of the three parameters p, q, r one can obtain both symmetric and nonsymmetric molecular matrices with elements integer or real numbers. We have to mention here that the computation of structural descriptors from nonsymmetric matrices requires the definition of new equations for certain graph invariants.

THE DISTANCE-VALENCY MATRIX FOR WEIGHTED GRAPHS

Using graph theory, an organic compound containing heteroatoms and/or multiple bonds can be represented as a vertex- and edge-weighted molecular graph. A vertex- and edge-weighted (VEW) molecular graph $G = G(V,E,Sy,Bo,Vw,Ew,w)$ consists of a vertex set $V = V(G)$, an edge set $E = E(G)$, a set of chemical symbols of the vertices $Sy = Sy(G)$, a set of topological bond orders of the edges $Bo = Bo(G)$, a vertex weight set $Vw(w) = Vw(w,G)$, and an edge weight set $Ew(w) = Ew(G)$. The elements of the vertex and edge weight sets are computed with the weighting scheme w. Usually, hydrogen atoms are not considered in the molecular graph, and in a VEW graph the weight of a vertex corresponding to a carbon atom is 0, while the weight of an edge corresponding to a carbon-carbon single bond is 1.

In a weighting scheme w the vertex Vw and edge Ew parameters are computed from a property p_i associated with every vertex v_i from G, $v_i \in V(G)$, and the topological bond order Bo of all edges from the molecular graph. The vertex parameter $Vw(w)_i$ for the vertex v_i is:

$$Vw(w)_i = 1 - p_C/p_i \qquad (4)$$

and the edge parameter $Ew(w)_{ij}$ for the edge between vertices v_i and v_j is:

$$Ew(w)_{ij} = p_C p_C / Bo_{ij} p_i p_j \qquad (5)$$

where p_i is the atomic property of vertex v_i, p_j is the atomic property of vertex v_j, and p_C is the atomic property for carbon atom. Several weighting schemes can be obtained when p represents different atomic properties:[16] A, when p is the atomic mass; P, when p is the atomic polarizability; E, when p is the atomic electronegativity; R, when p is the atomic radius.

The AH weighting scheme[16] uses the following equation to define the vertex parameter $Vw(AH)_i$ for the non-hydrogen atom i:

$$Vw(AH)_i = 1 - A_C/(A_i + NoH_i A_{11}) = 1 - 12.011/(A_i + 1.0079 NoH_i). \qquad (6)$$

The edge parameter $Ew(AH)_{ij}$ for the bond between atoms i and j is defined with the equation:

$$Ew(AH)_{ij} = A_C A_C / Bo_{ij}(A_i + NoH_i A_H)(A_j + NoH_j A_H) =$$

$$= 12.011 \cdot 12.011 / Bo_{ij}(A_i + 1.0079 NoH_i)(A_j + 1.0079 NoH_j)$$

(7)

where $A_C = 12.011$ is the atomic mass for carbon, $A_H = 1.0079$ is the atomic mass for hydrogen, NoH_i is the number of hydrogen atoms bonded to the heavy atom i, and NoH_j is the number of hydrogen atoms bonded to the heavy atom j. The above five weighting schemes are used in this study to compute structural descriptors.

The distance-valency matrix of a vertex- and edge-weighted graph G with N vertices, $Dval(p,q,r,w) = Dval(p,q,r,w,G)$, is a square $N \times N$ matrix, whose entries $[Dval(p,q,r,w)]_{ij}$ are equal to:

$$[Dval(p,q,r,w)]_{ij} = \begin{cases} d(w)_{ij}^p \, val(w)_i^q \, val(w)_j^r & \text{if } i \neq j \\ Vw(w)_i \, val(w)_i^{q+r} & \text{if } i = j \end{cases}$$

(8)

where $Vw(w)_i$ is the weight of vertex v_i, $d(w)_{ij}$ is the graph distance between vertices v_i and v_j, $val(w)_i$ is the valency of vertex v_i, and $val(w)_j$ is the valency of vertex v_j, all computed with the weighting scheme w. The definition of the Dval matrix for weighted graphs was formulated in analogy with the definitions for the distance and reciprocal distance matrices, respectively, with $Dval(1,0,0,w,G) = D(w,G)$, and $Dval(-1,0,0,w,G) = RD(w,G)$.

An example of the computation of the Dval matrices for weighted graphs is presented for n-propylamine 2 that has the molecular graph 3.
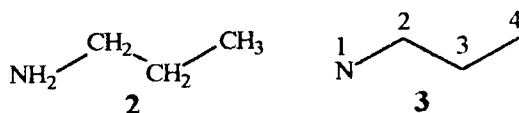
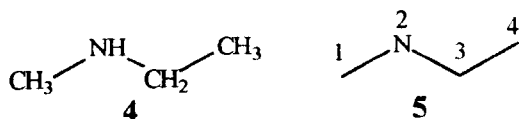The distance matrix $D(A,3)$ computed with the atomic polarizability weighting scheme P is:

Dval(1,0,0,P,3) = D(P,3)

	1	2	3	4
1	-0.600	1.600	2.600	3.600
2	1.600	0.000	1.000	2.000
3	2.600	1.000	0.000	1.000
4	3.600	2.000	1.000	0.000

From the elements of the above distance matrix one can easily compute the $Dval(-2,1,1,P,3)$ matrix:

Dval(-2,1,1,P,3)

	1	2	3	4
1	-1.536	1.625	0.473	0.123
2	1.625	0.000	5.200	0.650
3	0.473	5.200	0.000	2.000
4	0.123	0.650	2.000	0.000

The second example of Dval matrices for weighted graphs is presented for N-methylethylamine 4 that has the molecular graph 5.

From the molecular graph 5 one can easily obtain the **Dval**(–2,0,0,A,5) matrix, computed with the atomic mass weighting scheme A:

Dval(–2,0,0,A,5)			
	1	2	3	4
1	0.000	1.360	0.340	0.136
2	1.360	0.143	1.360	0.290
3	0.340	1.360	0.000	1.000
4	0.136	0.290	1.000	0.000

Another distance-valency matrix computed for the molecular graph 5 is presented below:

Dval(–1,1,1,A,5)			
	1	2	3	4
1	0.000	1.635	0.909	0.310
2	1.635	0.488	3.635	0.899
3	0.909	3.635	0.000	1.817
4	0.310	0.899	1.817	0.000

In the following sections the distance-valency matrices of vertex- and edge-weighted molecular graphs will be used to compute a large number of topological indices. Several QSPR models for the amines boiling temperatures will be developed with these structural descriptors.

STRUCTURAL DESCRIPTORS DERIVED FROM MOLECULAR GRAPHS

There exist many molecular graph descriptors, and the development of new molecular matrices and weighting schemes can multiply their number. This problem can be solved by using operators to denote all topological indices that can be computed with the same mathematical operation. An operator can be used to compute a family of topological indices by applying a certain mathematical operation or algorithm to a molecular matrix **M** calculated with a weighting scheme w from the molecular graph G. In the present paper we use families of related molecular graph descriptors computed with graph operators.

The Wiener operator, **WI**(**M**,w,G), is defined by analogy with the Wiener index:

$$\mathbf{WI}(\mathbf{M}, w, G) = \sum_{i=1}^{N} \sum_{j=i}^{N} [\mathbf{M}(w)]_{ij} \tag{9}$$

where **M** represents the molecular matrix of G, and w is the weighting scheme. If **M** is the distance matrix, the operator gives the Wiener index W,[25,26] while if **M** is the reciprocal distance matrix, this operator is identical with the Harary index.[23,24] The Wiener operator applied to the molecular matrices previously presented in this paper gives the following topological indices:

WI(Dval(1,0,0),1) = 52   **WI**(Dval(1,1,1),1) = 130

WI(Dval(1,–1,–1),1) = 28   **WI**(Dval(1,0,0),P,3) = 11.200

WI(Dval(–2,1,1),P,3) = 8.536   **WI**(Dval(–2,0,0),A,5) = 4.628

WI(Dval(–1,1,1),A,5) = 9.693

By analogy with the hyper-Wiener index[13,27–32] we define the Hyper-Wiener operator **HyWI**(**M**,G) of a graph G with N vertices:

$$\mathbf{HyWI}(\mathbf{M}, w, G) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=i}^{N} ([\mathbf{M}(w)]_{ij}^{2} + [\mathbf{M}(w)]_{ij}) \tag{10}$$

where **M** represents the molecular matrix of G, and w is the weighting scheme. If **M** is the distance matrix, the **HyWI** operator is identical with the hyper-Wiener index WW.

The spectrum operator $Sp(M,w,G) = \{x_i, i = 1, 2, ..., N\}$ represents the eigenvalues of the matrix $M(w) = M(w,G)$, or the roots of the characteristic polynomial $Ch(M,w,G,x)$, $Ch(M,w,G,x) = 0$. The $MinSp(M,w,G)$ and $MaxSp(M,w,G)$ operators are equal to the minimum and maximum values of $Sp(M,w,G)$, respectively. The two molecular spectra operators $MinSp$ and $MaxSp$ were used with success to compute the boiling temperatures of acyclic compounds containing oxygen or sulfur atoms.[21] The indices $MinSp$ and $MaxSp$ derived from the D_p, D, and RD_p matrices were used to model the boiling temperatures, heat of vaporization, molar refraction, molar volume, critical pressure, critical temperature, and surface tension of alkanes.[24,32]

MODELING THE AMINE BOILING TEMPERATURES

Several structural descriptors computed from distance-valency matrices were tested in a QSPR study that models the boiling temperatures of a set of 33 amines. The structure of the amines and their experimental boiling temperatures[33] are presented in Table 1. Five weighting schemes w were used,

Table 1

Amines, structural descriptors, and experimental boiling temperatures (t_b) used in the QSPR study

Amine	MinSp(D,E)	MaxSp(RD,AH)	MinSp(Dval(-2,0,0),A)	t_b (K)
methylamine	-0.7311	1.8958	-1.2906	266.8
ethylamine	-1.7483	2.7388	-1.5137	289.7
isopropylamine	-2.0000	3.3613	-1.6819	305.6
tert-butylamine	-2.0000	3.8184	-1.8180	317.6
n-propylamine	-3.1994	3.3723	-1.5821	321.6
sec-butylamine	-3.7867	3.9049	-1.7242	336.1
isobutylamine	-3.6010	3.8981	-1.6708	340.9
n-butylamine	-5.0517	3.8874	-1.6102	350.6
2-methylbutylamine	-5.0562	4.3717	-1.6996	368.7
n-pentylamine	-7.3037	4.3225	-1.6246	377.6
cyclopentylamine	-3.9453	4.7234	-1.7506	380.2
n-hexylamine	-9.9564	4.6993	-1.6326	404.6
cyclohexylamine	-5.5380	5.1396	-1.7807	407.6
2-aminoheptane	-11.9393	5.0664	-1.7419	416.2
n-heptylamine	-13.0108	5.0316	-1.6375	428.2
n-octylamine	-16.4679	5.3287	-1.6407	452.8
n-nonylamine	-20.3282	5.5973	-1.6428	475.4
n-decylamine	-24.5921	5.8423	-1.6443	493.7
n-dodecylamine	-34.3324	6.2759	-1.6461	532.4
dimethylamine	-1.6346	2.8392	-1.6846	280.0
diethylamine	-4.5509	3.9971	-1.8317	328.6
diisopropylamine	-6.3924	4.8677	-1.9528	357.1
N-methylbutylamine	-7.0060	4.3968	-1.7961	364.2
N-*tert*-butylisopropylamine	-7.1616	5.2009	-2.0123	371.2
N-ethylbutylamine	-9.2905	4.7949	-1.8480	381.2
di-*n*-propylamine	-9.1077	4.8054	-1.8561	382.0
N-methylhexylamine	-12.7407	5.0868	-1.7984	414.2
N-methylcyclohexylamine	-7.5980	5.5455	-1.9034	422.2
diamylamine	-23.0972	5.9355	-1.8633	476.1
trimethylamine	-1.6346	3.5416	-1.9596	276.0
tripropylamine	-9.1077	6.0014	-2.0889	429.7
tri-*n*-butylamine	-15.2908	6.7575	-2.0910	487.2
triamylamine	-23.0972	7.3631	-2.0912	516.2

namely P, E, R, A, and AH.[16] The list of the 204 structural descriptors used in the QSPR study is presented below:

(1) constitutional descriptors: the number of hydrogen atoms attached to the nitrogen atom, **NoHN**, and the molecular weight, **MW**;

(2) the Kier and Hall connectivity indices:[34,35] $^0\chi^v$, $^1\chi^v$, $^2\chi^v$, $^3\chi_p^v$, $^3\chi_c^v$;

(3) indices representing the minimum (**MinSp**) and maximum (**MaxSp**) value of the molecular graph spectra: **MinSp(A,w)**, **MaxSp(A,w)**, **MinSp(D,w)**, **MaxSp(D,w)**, **MinSp(RD,w)**, **MaxSp(RD,w)**, **MinSp(Dval(1,1,1),w)**, **MaxSp(Dval(1,1,1),w)**, **MinSp(Dval(-1,1,1),w)**, **MaxSp(Dval(-1,1,1),w)**, **MinSp(Dval(1,-1,-1),w)**, **MaxSp(Dval(1,-1,-1),w)**, **MinSp(Dval(-1,-1,-1),w)**, **MaxSp(Dval(-1,-1,-1),w)**, **MinSp(Dval(-2,1,1),w)**, **MaxSp(Dval(-2,1,1),w)**, **MinSp(Dval(-2,-1,-1),w)**, **MaxSp(Dval(-2,-1,-1),w)**, **MinSp(Dval(-2,0,0),w)**, **MaxSp(Dval(-2,0,0),w)**;

(4) topological indices obtained from the Wiener operator applied to a set of molecular graph matrices: **WI(D,w)**, **WI(RD,w)**, **WI(Dval(1,1,1),w)**, **WI(Dval(-1,1,1),w)**, **WI(Dval(1,-1,-1),w)**, **WI(Dval(-1,-1,-1),w)**, **WI(Dval(-2,1,1),w)**, **WI(Dval(-2,-1,-1),w)**, **WI(Dval(-2,0,0),w)**;

(5) topological indices obtained from the Hyper-Wiener operator applied to a set of molecular graph matrices: **HyWI(A,w)**, **HyWI(D,w)**, **HyWI(RD,w)**, **HyWI(Dval(1,1,1),w)**, **HyWI(Dval(-1,1,1),w)**, **HyWI(Dval(1,-1,-1),w)**, **HyWI(Dval(-1,-1,-1),w)**, **HyWI(Dval(-2,1,1),w)**, **HyWI(Dval(-2,-1,-1),w)**, **HyWI(Dval(-2,0,0),w)**.

All studies that develop QSPR models from a large set of computed descriptors use a wide range of algorithms for selecting significant descriptors. Because the exhaustive test of all MLR equations requires too large computational resources, we have used a heuristic method for descriptor selection. This heuristic algorithm starts from the set of 204 structural descriptors and develops QSPR models by applying the following steps:

(1) All one-parameter correlation equations are computed and all descriptors with a correlation coefficient greater than a threshold, $|r_{min}| > 0.15$, are selected for further use.

(2) Biparametric regression equations are computed with all possible pairs of descriptors selected in step (1) that are not significantly correlated. Two descriptors are considered to be not significantly correlated if their intercorrelation coefficient r_{ij} is lower than a threshold, $|r_{ij}| < 0.8$. The most significant 30 pairs of molecular descriptors were used in the third step.

(3) To a MLR model containing n descriptors a new descriptor is added to generate a model with $n + 1$ descriptors if the new descriptor is not significantly correlated with the previous n descriptors.

(4) The most significant 30 MLR models containing $n + 1$ descriptors are selected.

Steps (3) and (4) are repeated until MLR models with a certain maximum number of descriptors are obtained.

The best MLR equations with two independent variables for the computation of the boiling temperatures of amines are reported in Table 2, equations 1–10. Similarly, in Table 3 we present the best QSPR models with three independent variables. By increasing the number of independent variables to four or more the correlation coefficients does not significantly increase and the F test decreases, indicating that such equations do not improve the QSPR model.

The best QSPR model is represented by equation (11) from Table 3, with $r = 0.998$, $s = 4.68$, and $F = 2443$; the values of the three topological indices involved in this equation, namely **MinSp(D,E)**, **MaxSp(RD,AH)**, and **MinSp(Dval(-2,0,0),A)**, are presented in Table 1. An inspection of the equations from Tables 2 and 3 reveals an unexpected finding: all structural descriptors selected in equations (1–20) are derived only from the **MinSp** and **MaxSp** operators, representing the minimum and maximum values of matrix spectra. Although a large fraction of the structural descriptors are obtained from the **WI** and **HyWI** operators, no such indices were included in the QSPR models. Also, the molecular connectivity descriptors are absent in the reported models. It is clear that the **MinSp** and **MaxSp** descriptors outperform the connectivity, **WI** and **HyWI** descriptors in modeling the amines boiling temperatures. Topological indices computed with these two spectral operators were used with success to compute the boiling temperatures of acyclic compounds containing oxygen or sulfur

Table 2

Coefficients, structural descriptors, and statistical indices for the best MLR equations with two independent variables for the computation of the boiling temperatures of the 33 amines from Table 1.

The MLR equations have the general form: $t_b = a_0 + a_1 SD_1 + a_2 SD_2$

Eq.	a_0	a_1 / SD_1	a_2 / SD_2	r	s	F
1	261.66	−102.38 MaxSp(A,P)	73.83 MaxSp(RD,AH)	0.997	5.97	2251
2	558.49	369.21 MinSp(RD,E)	75.47 MaxSp(RD,AH)	0.996	6.47	1916
3	348.71	72.13 MaxSp(RD,AH)	170.25 MinSp(Dval(−2,0,0),A)	0.996	6.83	1718
4	321.82	152.73 MinSp(A,R)	79.16 MaxSp(RD,AH)	0.994	7.90	1277
5	588.03	404.49 MinSp(RD,A)	77.67 MaxSp(RD,AH)	0.992	9.07	966
6	307.78	68.87 MaxSp(RD,AH)	130.96 MinSp(Dval(−2,0,0),E)	0.992	9.43	893
7	179.11	102.22 MaxSp(RD,R)	16.63 MinSp(Dval(−2,1,1),A)	0.992	9.46	887
8	177.53	101.99 MaxSp(RD,R)	16.23 MinSp(Dval(−2,1,1),E)	0.991	9.60	862
9	178.70	94.58 MaxSp(RD,R)	21.37 MinSp(Dval(−1,1,1),AH)	0.991	9.91	807
10	360.82	−165.33 MaxSp(A,P)	109.14 MaxSp(RD,E)	0.990	10.13	771

atoms,[21] and to model the boiling temperatures, heat of vaporization, molar refraction, molar volume, critical pressure, critical temperature, and surface tension of alkanes,[24,32] demonstrating their utility in structure-property studies. In order to obtain a definite conclusion, this finding must be tested in other QSPR and QSAR models.

Table 3

Coefficients, structural descriptors, and statistical indices for the best MLR equations with three independent variables for the computation of the boiling temperatures of the 33 amines from Table 1.

The MLR equations have the general form: $t_b = a_0 + a_1 SD_1 + a_2 SD_2 + a_3 SD_3$

Eq.	a_0	a_1 / SD_1	a_2 / SD_2	a_3 / SD_3	r	s	F
11	323.72	−1.41 MinSp(D,E)	61.24 MaxSp(RD,AH)	134.71 MinSp(Dval(−2,0,0),A)	0.998	4.68	2443
12	324.79	−1.40 MinSp(D,A)	61.30 MaxSp(RD,AH)	135.51 MinSp(Dval(−2,0,0),A)	0.998	4.69	2436
13	255.34	−84.79 MaxSp(A,P)	−1.12 MinSp(D,E)	64.82 MaxSp(RD,AH)	0.998	4.71	2415
14	255.86	−85.23 MaxSp(A,P)	−1.11 MinSp(D,A)	64.91 MaxSp(RD,AH)	0.998	4.73	2400
15	244.59	−90.24 MaxSp(A,P)	76.28 MaxSp(RD,AH)	3.13 MinSp(Dval(−2,1,1),E)	0.998	4.74	2385
16	245.76	−90.83 MaxSp(A,P)	76.24 MaxSp(RD,AH)	3.10 MinSp(Dval(−2,1,1),A)	0.998	4.75	2341
17	315.72	75.64 MaxSp(RD,AH)	3.98 MinSp(Dval(−2,1,1),E)	145.98 MinSp(Dval(−2,0,0),A)	0.998	4.82	2309
18	317.54	75.63 MaxSp(RD,AH)	3.96 MinSp(Dval(−2,1,1),A)	147.00 MinSp(Dval(−2,0,0),A)	0.998	4.83	2297
19	317.21	75.68 MaxSp(RD,AH)	4.39 MinSp(Dval(−2,1,1),AH)	146.43 MinSp(Dval(−2,0,0),A)	0.998	4.97	2166
20	317.06	75.10 MaxSp(RD,AH)	3.94 MinSp(Dval(−2,1,1),R)	144.87 MinSp(Dval(−2,0,0),A)	0.998	5.06	2090

From the large number of molecular matrices that we have used to compute topological indices, the following six generated the descriptors in equations (1-20): **A**, **D**, **RD**, **Dval**(-1,1,1), **Dval**(-2,1,1), **Dval**(-2,0,0). The **MinSp** and **MaxSp** structural descriptors derived from the distance-valency matrices are present in eight out of ten QSPR equations with three independent variables from Table 3. The results obtained in this study clearly demonstrate the utility of the distance-valency matrices in computing helpful structural descriptors for QSPR studies.

REFERENCES

1. Part 13: O. Ivanciuc, *Rev. Roum. Chim.*, **2000**, *45*, 000-000.
2. M. V. Diudea and O. Ivanciuc, *Molecular Topology,* Comprex, Cluj, Romania, **1995**.
3. O. Ivanciuc and A. T. Balaban. *Graph Theory in Chemistry.* In: *The Encyclopedia of Computational Chemistry,* Eds.: P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner. John Wiley & Sons, Chichester, 1998, pp. 1169-1190.
4. A. T. Balaban and O. Ivanciuc, *Historical Development of Topological Indices.* In: *Topological Indices and Related Descriptors in QSAR and QSPR,* Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 21-57.
5. O. Ivanciuc and A. T. Balaban, *The Graph Description of Chemical Structures.* In: *Topological Indices and Related Descriptors in QSAR and QSPR,* Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 59-167.
6. O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, *Vertex- and Edge-Weighted Molecular Graphs and Derived Structural Descriptors.* In: *Topological Indices and Related Descriptors in QSAR and QSPR,* Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 169-220.
7. O. Ivanciuc and T. Ivanciuc, *Matrices and Structural Descriptors Computed from Molecular Graph Distances.* In: *Topological Indices and Related Descriptors in QSAR and QSPR,* Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 221-277.
8. A. T. Balaban, I. Moţoc, D. Bonchev, and O. Mekenyan, *Top. Curr. Chem.,* **1983**, *114*, 21-55.
9. A. T. Balaban, *J. Chem. Inf. Comput. Sci.,* **1992**, *32*, 23-28.
10. A. T. Balaban, *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 398-402.
11. A. T. Balaban, *Rev. Roum. Chim.,* **1994**, *39*, 245-257.
12. O. Ivanciuc, T. Ivanciuc, and M. V. Diudea, *SAR QSAR Environ. Res.,* **1997**, *7*, 63-87.
13. M. V. Diudea, and I. Gutman, *Croat. Chem. Acta,* **1998**, *71*, 21-51.
14. O. Ivanciuc, *Rev. Roum. Chim.,* **1989**, *34*, 1361-1368.
15. O. Ivanciuc, *Rev. Roum. Chim.,* **1999**, *44*, 000-000.
16. O. Ivanciuc, *Rev. Roum. Chim.,* **1999**, *44*, 000-000.
17. M. Barysz, G. Jashari, R. S. Lall, V. K. Srivastava, and N. Trinajstić, *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.
18. O. Ivanciuc and A. T. Balaban, *MATCH (Commun. Math. Chem.),* **1994**, *30*, 117-139.
19. A. T. Balaban, *MATCH (Commun. Math. Chem.),* **1986**, *21*, 115-122.
20. A. T. Balaban and O. Ivanciuc, *FORTRAN 77 Computer Program for Calculating the Topological Index J for Molecules Containing Heteroatoms.* "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 1988, A. Graovac, Ed., Studies in Physical and Theoretical Chemistry, Vol. 63, pp. 193-211, Elsevier, Amsterdam, 1989.
21. O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, *J. Chem. Inf. Comput. Sci.,* **1998**, *38*, 395-401.
22. T. S. Balaban, P. A. Filip, and O. Ivanciuc, *J. Math. Chem.,* **1992**, *11*, 79-105.
23. O. Ivanciuc, T.-S. Balaban, and A. T. Balaban, *J. Math. Chem.,* **1993**, *12*, 309-318.
24. M. V. Diudea, O. Ivanciuc, S. Nikolić, and N. Trinajstić, *MATCH (Commun. Math. Comput. Chem.),* **1997**, *35*, 41-64.
25. H. Wiener, *J. Am. Chem. Soc.,* **1947**, *69*, 17-20.
26. H. Wiener, *J. Am. Chem. Soc.,* **1947**, *69*, 2636-2638.
27. M. Randić, X. Guo, T. Oxley, and H. Krishnapriyan, *J. Chem. Inf. Comput. Sci.,* **1993**, *33*, 709-716.
28. M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, and L. Naylor, *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 361-367.
29. D. J. Klein, I. Lukovits, and I. Gutman, *J. Chem. Inf. Comput. Sci.,* **1995**, *35*, 50-52.
30. M. V. Diudea, *J. Chem. Inf. Comput. Sci.,* **1996**, *36*, 535-540.
31. M. V. Diudea, *J. Chem. Inf. Comput. Sci.,* **1996**, *36*, 833-836.
32. O. Ivanciuc, M. V. Diudea, and P. V. Khadikar, *Ind. J. Chem.,* **1998**, *37A*, 574-585.
33. A. R. Katritzky, V. S. Lobanov, and M. Karelson, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28-41.
34. L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research,* Academic Press, New York, **1976**.
35. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis,* Research Studies Press, Letchworth, **1986**.