Revue Roumaine de Chimie, **1999**, *44* (5), 479–489

DESIGN OF TOPOLOGICAL INDICES. PART 20.[1]

MOLECULAR STRUCTURE DESCRIPTORS COMPUTED WITH INFORMATION ON DISTANCES OPERATORS

Ovidiu IVANCIUC and Alexandru T. BALABAN

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

Received May 5, 1999

The interest of developing new graph descriptors for organic compounds revived in recent years, when topological indices found new applications in database mining, similarity and diversity assessment. The search for new structural descriptors produced new molecular matrices, such as the reciprocal distance, resistance distance, detour, or distance-valency matrices. Using these graph matrices we define new families of topological indices, the information-theory operators **U(M)**, **V(M)**, **X(M)**, and **Y(M)**. The operators are computed from vertex invariants that measure the information content of the matrix elements associated with the respective graph vertex. Structural descriptors derived from the new operators were tested in QSPR studies for alkane densities.

INTRODUCTION

Structural descriptors are used to compute physical, chemical, or biological properties with quantitative structure-property relationship (QSPR) and quantitative structure-activity relationship (QSAR) models. In order to develop good QSPR and QSAR models, a large variety of graph, geometric, and quantum descriptors were defined. Using molecular graphs the chemical structure of a compound can be expressed by means of various graph matrices, polynomials, spectra, spectral moments, sequences counting distances, paths, or topological indices. A topological index (TI) is a numerical descriptor of the molecular structure derived from the molecular graph that offers a simple way of measuring molecular branching, shape, and size. Numerous reviews describe the theory and applications of molecular graph descriptors.[2–13]

The use of TIs in chemistry begins with the Wiener index W,[14] whose graph-theoretical formula was derived from the distance matrix by Hosoya.[15] An analysis of the Wiener index shows that the descriptor is degenerate, i.e. two or more different (nonisomorphic) molecular graphs can have identical W indices. The relatively high degeneracy of W due to the global summation of distances can be reduced by using information theory, as shown by Bonchev and Trinajstić who used distance matrix elements to define the indices I_D^E, $\bar{I}_D^E$, I_D^W, and $\bar{I}_D^W$.[16] Another TI based on graph distances is the Balaban index J,[17,18] that presents the smallest degenerate pairs of 4-trees for graphs with twelve vertices.[19] The summation of distances to give distance sums is the cause of the degeneracy of the index J. Using information theory applied to graph distances, the highly discriminating topological indices U, V, X, and Y were introduced.[20–22]

New molecular matrices were recently defined,[13] such as the reciprocal distance,[23-25] resistance distance,[26] detour,[27] or distance-valency[28] matrices. In the present paper we define the information-theory operators **U(M)**, **V(M)**, **X(M)**, and **Y(M)**, that represent an extension of the U, V, X, and Y indices. Using any symmetric molecular matrix these new operators generate topological indices that can be efficiently used in QSPR and QSAR studies.

THE INFORMATION ON DISTANCES INDICES U, V, X, AND Y

Consider a simple molecular graph G with N vertices and its distance matrix $\mathbf{D} = \mathbf{D}(G)$. The application of the Shannon's formula to the elements of **D** that correspond to a vertex v_i gives the mean local information on the magnitude of distances:

$$u.inf_i = -\sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{[\mathbf{D}]_{ij}}{\mathbf{DS}_i} \log_2 \frac{[\mathbf{D}]_{ij}}{\mathbf{DS}_i} \tag{1}$$

where the summation is done for all non-zero elements in the ith row of the distance matrix **D**. The total information computed from the size of the elements in the ith row of the distance matrix **D** gives the local information on the magnitude of distances:

$$v.inf_i = \mathbf{DS}_i \log_2 \mathbf{DS}_i - u.inf_i \tag{2}$$

where $\mathbf{DS}_i$ represents the distance sum of the vertex v_i.

Using the elements of the distance matrix that correspond to a vertex v_i, two new related vertex invariants were proposed, namely the extended local information on distance magnitude:

$$x.inf_i = \mathbf{DS}_i \log_2 \mathbf{DS}_i - y.inf_i \tag{3}$$

and the mean extended local information on distance magnitude:

$$y.inf_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{D}]_{ij} \log_2 [\mathbf{D}]_{ij} \tag{4}$$

where the summation is done for all non-zero elements in the ith row of the distance matrix **D**.

By analogy with the Randić connectivity index,[29] and with the Balaban index J[17,18] four new TIs were defined on the basis of the four local graph invariants $u.inf_i$, $v.inf_i$, $x.inf_i$, and $y.inf_i$:[20]

$$U(G) = \frac{M}{\mu + 1} \sum_{E(G)} (u.inf_i u.inf_j)^{-1/2} \tag{5}$$

$$V(G) = \frac{M}{\mu + 1} \sum_{E(G)} (v.inf_i v.inf_j)^{-1/2} \tag{6}$$

$$X(G) = \frac{M}{\mu + 1} \sum_{E(G)} (x.inf_i x.inf_j)^{-1/2} \tag{7}$$

$$Y(G) = \frac{M}{\mu + 1} \sum_{E(G)} (y.inf_i y.inf_j)^{-1/2} \tag{8}$$

where M is the number of edges in G, μ is the cyclomatic number of G, and the summation goes over all edges from the edge set $E(G)$. Although by design the information on distances indices U, V, X, and Y are very selective, it was found that certain graphs that have non-equivalent vertices with identical distance degree sequences can generate pairs of graphs with degenerate distance based indices. Such distance degree sequence degenerate graphs were found both for acyclic[30] and cyclic[31] graphs. The

information on distances indices *U*, *V*, *X*, and *Y*, and index *J* were used with good results in structure-property models to estimate the critical volumes, temperatures, and pressures of 49 alkanes, as well as for the coefficients of the Antoine equation that give the saturation pressure versus the absolute temperatures.[22]

THE INFORMATION-THEORY OPERATORS U(M), V(M), X(M), AND Y(M)

The indices *U*, *V*, *X*, and *Y* for information on distances are computed from the elements of the distance matrix of the molecular graph, and these TIs provided good results both for structure discrimination and in structure-property models. Because new graph matrices were defined in recent years,[13] it is possible to extended the definition of these four indices for all dense molecular matrices **M** (a dense matrix is a matrix without zero non-diagonal elements). We define here information-theory operators that can be applied to a matrix with integer value elements, such as the distance matrix **D**, or to a matrix with real value elements, such as the reciprocal distance matrix **RD**. The graph vertex operators **VUinf(M,*G*)**, **VVinf(M,*G*)**, **VXinf(M,*G*)**, and **VYinf(M,*G*)** apply the information theory equations to the non-zero elements of the molecular matrix **M** that correspond to a vertex v_i:

$$\mathbf{VUinf(M)}_i = -\sum_{j=1}^{N} \frac{[\mathbf{M}]_{ij}}{\mathbf{VS(M)}_i} \log_2 \frac{[\mathbf{M}]_{ij}}{\mathbf{VS(M)}_i} \tag{9}$$

$$\mathbf{VVinf(M)}_i = \mathbf{VS(M)}_i \log_2 \mathbf{VS(M)}_i - \mathbf{VUinf(M)}_i \tag{10}$$

$$\mathbf{VXinf(M)}_i = \mathbf{VS(M)}_i \log_2 \mathbf{VS(M)}_i - \mathbf{VYinf(M)}_i \tag{11}$$

$$\mathbf{VYinf(M)}_i = -\sum_{j=1}^{N} [\mathbf{M}]_{ij} \log_2 [\mathbf{M}]_{ij} \tag{12}$$

where **M** is a molecular graph matrix, $\mathbf{VS(M)}_i$ represents the vertex sum of the vertex v_i, and the summations in equations (9) and (12) are done for the non-zero elements of the molecular matrix **M**, $[\mathbf{M}]_{ij} \neq 0$. In a molecular graph *G* with *N* vertices, the vertex sum operator for the vertex v_i, $\mathbf{VS(M},G)_i$, is defined as the sum of the elements in the column *i*, or row *i*, of the molecular matrix **M**:[1,13]

$$\mathbf{VS(M},G)_i = \sum_{j=1}^{N} [\mathbf{M}]_{ij} = \sum_{j=1}^{N} [\mathbf{M}]_{ji} \tag{13}$$

For a general dense molecular graph matrix **M**, the matrix elements $[\mathbf{M}]_{ij}$ may have values lower than 1, giving negative terms for certain vertex structural descriptors computed with the graph vertex operators **VUinf(M,*G*)**, **VVinf(M,*G*)**, **VXinf(M,*G*)**, and **VYinf(M,*G*)**. The Randić-like formula used in the case of the indices *U*, *V*, *X*, and *Y* is therefore replaced by the following equation:

$$f(x,y) = \begin{cases} (xy)^{-1/2} & \text{if } xy > 0 \\ -(|xy|)^{-1/2} & \text{if } xy < 0 \end{cases} \tag{14}$$

The operators **U(M)**, **V(M)**, **X(M)**, and **Y(M)**, representing information on matrix elements, are computed with the equations:

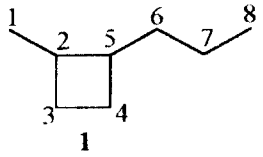
$$\mathbf{U(M},G) = \frac{M}{\mu + 1} \sum_{E(G)} f(\mathbf{VUinf(M)}_i, \mathbf{VUinf(M)}_j) \tag{15}$$

$$\mathbf{V(M},G) = \frac{M}{\mu + 1} \sum_{E(G)} f(\mathbf{VVinf(M)}_i, \mathbf{VVinf(M)}_j) \tag{16}$$

$$\mathbf{X(M},G) = \frac{M}{\mu + 1} \sum_{E(G)} f(\mathbf{VXinf(M)}_i, \mathbf{VXinf(M)}_j) \tag{17}$$

$$\mathbf{Y(M},G) = \frac{M}{\mu + 1} \sum_{E(G)} f(\mathbf{VYinf(M)}_i, \mathbf{VYinf(M)}_j) \tag{18}$$

An example for the computation of the information operators **U(M)**, **V(M)**, **X(M)**, and **Y(M)** is presented for the molecular graph of 1-methyl-2-propylcyclobutane **1**.



1

The first example considers the distance matrix **D(1)**:

D(1)

	1	2	3	4	5	6	7	8
1	0	1	2	3	2	3	4	5
2	1	0	1	2	1	2	3	4
3	2	1	0	1	2	3	4	5
4	3	2	1	0	1	2	3	4
5	2	1	2	1	0	1	2	3
6	3	2	3	2	1	0	1	2
7	4	3	4	3	2	1	0	1
8	5	4	5	4	3	2	1	0

The distance sum vector is computed as the sum over rows or columns in the distance matrix: **DS(1)** = {20, 14, 18, 16, 12, 14, 18, 24}. The vertex structural descriptors **VUinf(D,1)**, **VVinf(D,1)**, **VXinf(D,1)**, and **VYinf(D,1)** are computed from the elements of the distance matrix **D(1)** and that of the distance sum vector **DS(1)**:

VUinf(D,1) = {2.66596, 2.61058, 2.59412, 2.65564, 2.68872, 2.69951, 2.64160, 2.66937}

VVinf(D,1) = {83.77260, 50.69239, 72.46453, 61.34436, 40.33083, 50.60346, 72.41705, 107.36973}

VXinf(D,1) = {53.31915, 36.54808, 46.69412, 42.49022, 32.26466, 37.79319, 47.54888, 64.06493}

VYinf(D,1) = {33.11942, 16.75489, 28.36453, 21.50978, 10.75489, 15.50978, 27.50978, 45.97417}

The values of the indices **U(D,1)**, **V(D,1)**, **X(D,1)**, **Y(D,1)** are readily computed from the above vertex descriptors:

U(D,1) = 12.07298 **V(D,1)** = 0.55624

X(D,1) = 0.78314 **Y(D,1)** = 1.69203

The second example considers the reciprocal distance matrix of 1-methyl-2-propylcyclobutane **1**, **RD(1)**:

RD(1)

	1	2	3	4	5	6	7	8
1	0	1	1/2	1/3	1/2	1/3	1/4	1/5
2	1	0	1	1/2	1	1/2	1/3	1/4
3	1/2	1	0	1	1/2	1/3	1/4	1/5
4	1/3	1/2	1	0	1	1/2	1/3	1/4
5	1/2	1	1/2	1	0	1	1/2	1/3
6	1/3	1/2	1/3	1/2	1	0	1	1/2
7	1/4	1/3	1/4	1/3	1/2	1	0	1
8	1/5	1/4	1/5	1/4	1/3	1/2	1	0

The row or column sums give the vertex sum vector **VS(RD,1)**:

VS(RD,1) = {3.11667, 4.58333, 3.78333, 3.91667, 4.83333, 4.16667, 3.66667, 2.73333}

Equations (9–12) give the vertex graph descriptors:

VUinf(RD,1) = {2.60932, 2.63894, 2.57852, 2.62239, 2.69267, 2.67249, 2.57174, 2.53252}

VVinf(RD,1) = {2.50203, 7.42788, 4.68418, 5.09198, 8.29358, 5.90624, 4.30132, 1.43262}

VXinf(RD,1) = {8.13237, 12.09514, 9.75541, 10.27101, 13.01458, 11.13537, 9.42970, 6.92223}

VYinf(RD,1) = {–3.02103, –2.02832, –2.49271, –2.55664, –2.02832, –2.55664, –2.55664, –2.95709}

As explained above, certain vertex invariants may have negative values, for example that of **VYinf(RD,1)** in this example. The formulas (15–18) give the information indices derived from the reciprocal distance matrix:

U(RD,108) = 12.18610 **V(RD,108)** = 6.52666

X(RD,108) = 3.05367 **Y(RD,108)** = 13.48377

INFORMATION-THEORY OPERATORS FOR WEIGHTED GRAPHS

Chemical compounds containing heteroatoms and/or multiple bonds are represented as vertex and edge-weighted (VEW) molecular graphs.[6] The computation of structural descriptors and topological indices from vertex- and edge-weighted molecular graphs requires the development of special parameters. A widely used set of parameters for VEW graphs was developed by Trinajstić and coworkers[32] by weighting the contributions of atoms and bonds with parameters based on the atomic number Z and the topological bond order. A different approach for considering heteroatoms was developed for the Balaban index J, considering electronegativity or covalent radii.[33,34] Using the parameters developed for J, two general weighting schemes for the molecular matrices were proposed: X, the relative electronegativity scheme, and Y, the relative covalent radii scheme.[35] Five weighting schemes for VEW graphs were recently proposed[36] and used with success in QSPR models.[37,38]

The indices U, V, X, and Y using information on distances were defined for the distance matrix of simple graphs representing alkanes. The extension of such indices to vertex- and edge-weighted graphs considers the possibility that the molecular matrix of a VEW graph may contain negative elements or elements with values between 0 and 1. The graph vertex operators **VUinf(M**,w,G), **VVinf(M**,w,G), **VXinf(M**,w,G), and **VYinf(M**,w,G) apply the information theory equations to the absolute values of the elements of the molecular matrix **M**(w,G). All three weighting schemes X, Y, and Z have atomic weights Vw with negative values for certain elements: $Vw(Z,\mathrm{B}) = -0.200$, $Vw(X,\mathrm{B}) = -0.175$, $Vw(X,\mathrm{Si}) = -0.067$, $Vw(X,\mathrm{As}) = -0.057$, $Vw(X,\mathrm{Te}) = -0.048$, $Vw(Y,\mathrm{N}) = -0.038$, $Vw(Y,\mathrm{O}) = -0.081$, $Vw(Y,\mathrm{F}) = -0.127$. Because the logarithm is defined only for positive arguments, the four graph vertex operators are computed from the elements of a positive matrix **P**(w) = **P**(w,G) whose element $[\mathbf{P}(w)]_{ij}$ is equal to the absolute value of the corresponding element from the **M**(w) matrix, *i.e.* $[\mathbf{P}(w)]_{ij} = |[\mathbf{M}(w)]_{ij}|$. The graph vertex operators are defined with the following equations:

$$\mathbf{VUinf(M},w)_i = -\sum_{j=1}^{N} \frac{[\mathbf{P}(w)]_{ij}}{\mathbf{VS(P},w)_i} \log_2 \frac{[\mathbf{P}(w)]_{ij}}{\mathbf{VS(P},w)_i} \tag{19}$$

$$\mathbf{VVinf(M},w)_i = \mathbf{VS(P},w)_i \log_2 \mathbf{VS(P},w)_i - \mathbf{VUinf(M},w)_i \tag{20}$$

$$\mathbf{VXinf(M},w)_i = \mathbf{VS(P},w)_i \log_2 \mathbf{VS(P},w)_i - \mathbf{VYinf(M},w)_i \tag{21}$$

$$\mathbf{VYinf(M},w)_i = \sum_{j=1}^{N} [\mathbf{P}(w)]_{ij} \log_2 [\mathbf{P}(w)]_{ij} \tag{22}$$

where $\mathbf{VS(P},w)_i$ is the vertex sum of the vertex v_i computed from the matrix **P**, w is the weighting scheme, and the summations in formulas (19) and (22) are done for the absolute values of the non-zero elements of the molecular matrix **P**, $[\mathbf{P}(w)]_{ij} \neq 0$. For the notation of the four graph vertex operators **VUinf(M**,w,G), **VVinf(M**,w,G), **VXinf(M**,w,G), and **VYinf(M**,w,G) we have maintained the molecular matrix **M** to indicate the source of the invariants. The information on matrix elements operators **U(M**,w), **V(M**,w), **X(M**,w), and **Y(M**,w) are computed with the equations:

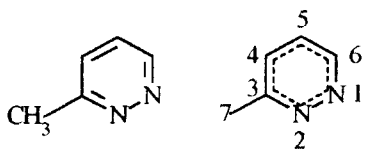
$$\mathbf{U(M},w,G) = \frac{M}{\mu+1} \sum_{E(G)} f(\mathbf{VUinf(M},w)_i, \mathbf{VUinf(M},w)_j) \tag{23}$$

$$\mathbf{V(M},w,G) = \frac{M}{\mu+1} \sum_{E(G)} f(\mathbf{VVinf(M},w)_i, \mathbf{VVinf(M},w)_j) \tag{24}$$

$$\mathbf{X(M},w,G) = \frac{M}{\mu+1} \sum_{E(G)} f(\mathbf{VXinf(M},w)_i, \mathbf{VXinf(M},w)_j) \tag{25}$$

$$\mathbf{Y(M},w,G) = \frac{M}{\mu+1} \sum_{E(G)} f(\mathbf{VYinf(M},w)_i, \mathbf{VYinf(M},w)_j) \tag{26}$$

The computation of the information operators **U(M**,w), **V(M**,w), **X(M**,w), and **Y(M**,w) is presented for 3-methyl-1,2-diazine **2** represented by the molecular graph **3**:



2 **3**

The distance matrix of graph **3** computed with the *Z* weighting scheme is:

D(*Z*,3)

	1	2	3	4	5	6	7
1	0.143	0.490	1.061	1.728	1.238	0.571	2.061
2	0.490	0.143	0.571	1.238	1.728	1.061	1.571
3	1.061	0.571	0.000	0.667	1.333	1.633	1.000
4	1.728	1.238	0.667	0.000	0.667	1.333	1.667
5	1.238	1.728	1.333	0.667	0.000	0.667	2.333
6	0.571	1.061	1.633	1.333	0.667	0.000	2.633
7	2.061	1.571	1.000	1.667	2.333	2.633	0.000

From the distance matrix **D**(*Z*,3) one obtains the vertex sum vector and the four vertex vectors **VUinf(D,*Z*,3)**, **VVinf(D,*Z*,3)**, **VXinf(D,*Z*,3)**, and **VYinf(D,*Z*,3)**:

VS(D,*Z*,3) = {7.293, 6.803, 6.265, 7.299, 7.966, 7.898, 11.265}

VUinf(D,*Z*,3) = {2.507, 2.547, 2.496, 2.491, 2.445, 2.396, 2.523}

VVinf(D,*Z*,3) = {18.396, 16.271, 14.091, 18.441, 21.404, 21.152, 36.836}

VXinf(D,*Z*,3) = {18.283, 17.323, 15.639, 18.186, 19.478, 18.923, 28.426}

VYinf(D,*Z*,3) = {2.620, 1.494, 0.948, 2.747, 4.370, 4.624, 10.933}

The molecular graph **3** has one cycle, seven vertices and seven edges, which leads to a factor $M/(\mu + 1) = 7/(1 + 1) = 3.5$. The values of the indices **U(D,*Z*,3)**, **V(D,*Z*,3)**, **X(D,*Z*,3)**, **Y(D,*Z*,3)** are readily computed from the above vertex descriptors:

U(D,*Z*,3) = 9.863 **V(D,*Z*,3)** = 1.322

X(D,*Z*,3) = 1.339 **Y(D,*Z*,3)** = 10.762

The next example considers the reciprocal distance matrix of 3-methyl-1,2-diazine **3**, **RD**(*Z*,3), computed with the weighting scheme *Z*:

RD(*Z*,3)

	1	2	3	4	5	6	7
1	0.143	2.042	0.942	0.579	0.808	1.750	0.485
2	2.042	0.143	1.750	0.808	0.579	0.942	0.636
3	0.942	1.750	0.000	1.500	0.750	0.613	1.000
4	0.579	0.808	1.500	0.000	1.500	0.750	0.600
5	0.808	0.579	0.750	1.500	0.000	1.500	0.429
6	1.750	0.942	0.613	0.750	1.500	0.000	0.380
7	0.485	0.636	1.000	0.600	0.429	0.380	0.000

The reciprocal distance matrix **RD**(Z,**3**) and the vertex sum vector **VS**(**RD**,Z,**3**) are used to compute the four vertex vectors **VUinf**(**RD**,Z,**3**), **VVinf**(**RD**,Z,**3**), **VXinf**(**RD**,Z,**3**), and **VYinf**(**RD**,Z,**3**):

VS(**RD**,Z,**3**) = {6.748, 6.900, 6.555, 5.736, 5.565, 5.935, 3.530}

VUinf(**RD**,Z,**3**) = {2.485, 2.509, 2.489, 2.469, 2.438, 2.412, 2.505}

VVinf(**RD**,Z,**3**) = {16.104, 16.717, 15.291, 11.988, 11.343, 12.835, 3.919}

VXinf(**RD**,Z,**3**) = {16.767, 17.313, 16.315, 14.161, 13.567, 14.312, 8.841}

VYinf(**RD**,Z,**3**) = {1.822, 1.913, 1.465, 0.296, 0.214, 0.935, -2.418}

Because some elements of the **RD**(Z,**3**) matrix have values lower than 1, the **VYinf**(**RD**,Z,**3**) vector has one negative element. In such conditions the Randić-like formula cannot be used, and the extended equation (14) is used to compute the four information theory indices:

U(**RD**,Z,**3**) = 9.916 **V**(**RD**,Z,**3**) = 1.977

X(**RD**,Z,**3**) = 1.665 **Y**(**RD**,Z,**3**) = 31.826

MODELING THE ALKANE DENSITIES

The information-theory indices computed with the graph operators defined in this paper are tested in a QSPR study for modeling alkane densities. The experimental densities at 25°C (kg m^{-3}) of 69 isomers of decane, taken from the literature,[39] were used to develop multilinear regression (MLR) models. The list of the 19 structural descriptors used in the QSPR study is presented below:

(1) the Kier and Hall connectivity indices[40,41] $^0\chi$, $^1\chi$, $^2\chi$, $^3\chi_p$, $^3\chi_c$;

(2) the Wiener indices[1] **Wi(D)** and **Wi(RD)**;

(3) the hyper-Wiener indices[1] **HyWi(D)** and **HyWi(RD)**;

(4) the Ivanciuc-Balaban indices[35] **IB(D)** and **IB(RD)**;

(5) the information-theory indices **U(D)**, **V(D)**, **X(D)**, **Y(D)**, **U(RD)**, **V(RD)**, **X(RD)**, **Y(RD)**.

In Table 1 we present the statistical indices (r, correlation coefficient, s, standard deviation, and F, Fisher test) for all monoparametric regression equations for the 19 structural descriptors. The best results are obtained with the connectivity index $^3\chi_p$, with r = 0.912, s = 6.16, F = 333. The second place is occupied by an information index, **Y(D)**, with r = 0.903, s = 6.46, F = 296. The connectivity index $^3\chi_p$ reflects the molecular branching and is a weighted sum of n-butane-like subgraphs containing four atoms, while **Y(D)** is a measure of the magnitude of graph distances.

In the second step of the QSPR analysis we have generated all biparametric regression equations computed with all pairs of descriptors 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 coefficients and statistical indices for the best MLR equations with two independent variables are presented in Table 2. The best biparametric QSPR model, Eq. (1) from Table 2, is obtained with the two structural descriptors that produce the best monoparametric equations, namely $^3\chi_p$ and **Y(D)**. The statistical indices of Eq. (1) from Table 2, r = 0.973, s = 3.49, F = 588, show a remarkable improvement over those obtained in monoparametric QSPR models. The connectivity index $^3\chi_p$ appears in all biparametric QSPR models from Table 2, showing its importance in the modeling of alkane densities. The combination between this connectivity index and an information index, either **Y(D)** or **V(D)**, is essential in modeling the alkane densities, as indicated by Eqs. (1) and (2) from Table 2. We have to mention that all ten QSPR models from Table 2 are of comparable statistical quality, and a larger set of compounds is needed in order to identify the best graph descriptors for modeling alkane densities.

Table 1

Statistical indices for the 19 structural descriptors (**SD**) in monoparametric equations for the computation of the alkane densities for 69 decane isomers

SD	r	s	F
$^0\chi$	0.518	12.87	25
$^1\chi$	-0.272	14.48	5
$^2\chi$	0.033	15.04	<1
$^3\chi_p$	0.912	6.16	333
$^3\chi_c$	0.159	14.85	2
Wi(D)	-0.836	8.25	156
Wi(RD)	0.813	8.75	131
HyWi(D)	-0.826	8.48	144
HyWi(RD)	0.801	9.01	120
IB(D)	0.875	7.29	219
IB(RD)	-0.744	10.05	83
U(D)	-0.871	7.39	211
V(D)	0.883	7.07	236
X(D)	0.866	7.53	200
Y(D)	0.903	6.46	296
U(RD)	0.868	7.47	205
V(RD)	-0.672	11.14	55
X(RD)	-0.772	9.56	99
Y(RD)	0.297	14.36	7

Table 2

Coefficients, structural descriptors, and statistical indices for the best MLR equations with two independent variables for the computation of alkane densities for 69 decane isomers. The MLR equations have the general form: $\rho = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2$

Eq.	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	r	s	F
1	667.1524	17.0923	$^3\chi_p$	12.5378	**Y(D)**	0.9731	3.493	588
2	645.2185	18.2054	$^3\chi_p$	62.1405	**V(D)**	0.9682	3.792	494
3	626.4406	18.5965	$^3\chi_p$	19.3875	**IB(D)**	0.9666	3.887	469
4	620.4653	19.0322	$^3\chi_p$	62.6559	**X(D)**	0.9649	3.982	445
5	1227.1714	18.9089	$^3\chi_p$	-19.4688	**U(RD)**	0.9618	4.150	407
6	1345.4753	18.7549	$^3\chi_p$	-24.1873	**U(D)**	0.9598	4.254	386
7	769.0333	20.2598	$^3\chi_p$	-0.5995	**Wi(D)**	0.9593	4.282	381
8	514.7781	21.0416	$^3\chi_p$	8.2175	**Wi(RD)**	0.9580	4.345	369
9	726.5099	20.6407	$^3\chi_p$	-0.1206	**HyWi(D)**	0.9569	4.402	358
10	514.8161	21.4405	$^3\chi_p$	9.9408	**HyWi(RD)**	0.9568	4.407	357

In the third step of the QSPR analysis we have generated QSPR models with all sets of three structural descriptors that are not significantly correlated. The coefficients and statistical indices for the best MLR equations with three independent variables are presented in Table 3. The QSPR models with three independent parameters do not show a significant increase in r, indicating that new descriptors would be needed in order to improve the structure-property models for alkane densities. The best tri-parametric QSPR model, Eq. (11) from Table 3, containing the descriptors $^0\chi$, $^3\chi_p$, and **Y(D)**, has good statistical indices (r = 0.977, s = 3.28, F = 450), but not significantly different from those of Eq. 1 from Table 2. The importance of information-theory indices in modeling alkanes is indicated by the presence of these indices in the first nine equations from Table 2; the best information index is **Y(D)**, that appears in the first five equations from Table 2. All but one equations from Table 3 contain two connectivity indices, and the index $^3\chi_p$ appears in all QSPR models from Table 3. 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.

Table 3

Coefficients, structural descriptors, and statistical indices for the best MLR equations with three independent variables for the computation of alkane densities for 69 decane isomers. The MLR equations have the general form

$$\rho = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$$

Eq.	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
11	767.0527	-12.8885	$^0\chi$	14.7799	$^3\chi_p$	16.7029	**Y(D)**	0.9767	3.275	450
12	617.1066	10.6289	$^1\chi$	14.5026	$^3\chi_p$	15.6037	**Y(D)**	0.9761	3.318	437
13	679.4784	-2.8601	$^2\chi$	13.9241	$^3\chi_p$	15.3027	**Y(D)**	0.9759	3.330	434
14	669.5537	14.2307	$^3\chi_p$	-2.4264	$^3\chi_c$	15.4323	**Y(D)**	0.9759	3.334	33
15	704.0001	16.1051	$^3\chi_p$	13.9107	**Y(D)**	-1.5624	**Y(RD)**	0.9744	3.434	407
16	746.6863	-14.2272	$^0\chi$	15.7395	$^3\chi_p$	86.7956	**V(D)**	0.9721	3.583	372
17	585.7328	11.3052	$^1\chi$	15.5089	$^3\chi_p$	79.6478	**V(D)**	0.9713	3.631	362
18	641.8269	15.1866	$^3\chi_p$	-2.6025	$^3\chi_c$	78.8356	**V(D)**	0.9712	3.642	359
19	652.6164	-3.0144	$^2\chi$	14.9170	$^3\chi_p$	77.7914	**V(D)**	0.9711	3.643	359
20	715.0981	-13.4633	$^0\chi$	16.3371	$^3\chi_p$	26.8034	**IB(D)**	0.9700	3.713	345

CONCLUSIONS

As a result of the considerable interest for structural descriptors, new information-theory operators were defined both for simple and weighted molecular graphs. The equations of the **U(M)**, **V(M)**, **X(M)**, and **Y(M)** operators can be applied to any symmetric molecular matrix, and produce structural descriptors with good correlational power, as demonstrated in a QSPR study of alkane densities. The new operators can generate four families of topological indices for any organic compound; they can find application in QSPR, QSAR or molecular similarity/diversity studies.

ACKNOWLEDGMENT. We acknowledge the partial financial support of this research by the Roumanian Ministry of National Education under Grant 7001 T34.

REFERENCES

1. Part 19: 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. Motoc, 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. A. T. Balaban, *SAR QSAR Environ. Res.*, **1995**, *3*, 81-95.

13. O. Ivanciuc, T. Ivanciuc, and M. V. Diudea, *SAR QSAR Environ. Res.*, **1997**, *7*, 63–87.
14. H. Wiener, *J. Am. Chem. Soc.*, **1947**, *69*, 17–20.
15. H. Hosoya, *Bull. Chem. Soc. Japan*, **1971**, *44*, 2332–2339.
16. D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, **1977**, *67*, 4517–4533.
17. A. T. Balaban, *Chem. Phys. Lett.*, **1982**, *89*, 399–404.
18. A. T. Balaban, *Pure Appl. Chem.*, **1983**, *55*, 199–206.
19. A. T. Balaban and L. V. Quintas, *MATCH (Commun. Math. Chem.)*, **1983**, *14*, 213–233.
20. A. T. Balaban and T.-S. Balaban, *J. Math. Chem.*, **1991**, *8*, 383–397.
21. A. T. Balaban and T.-S. Balaban, *J. Chim. Phys.*, **1992**, *89*, 1735–1745.
22. A. T. Balaban and V. Feroiu, *Rep. Mol. Theor.*, **1990**, *1*, 133–139.
23. O. Ivanciuc, *Rev. Roum. Chim.*, **1989**, *34*, 1361–1368.
24. O. Ivanciuc, T.-S. Balaban, and A. T. Balaban, *J. Math. Chem.*, **1993**, *12*, 309–318.
25. M. V. Diudea, O. Ivanciuc, S. Nikolić, and N. Trinajstić, *MATCH (Commun. Math. Comput. Chem.)*, **1997**, *35*, 41–64.
26. D. J. Klein and M. Randić, *J. Math. Chem.*, **1993**, *12*, 81–95.
27. O. Ivanciuc and A.T. Balaban, *MATCH (Commun. Math. Chem.)*, **1994**, *30*, 141–152.
28. Part 11: O. Ivanciuc, *Rev. Roum. Chim.*, **1999**, *44*, 519–528.
29. M. Randić, *J. Am. Chem. Soc.*, **1975**, *97*, 6609–6615.
30. O. Ivanciuc, T.-S. Balaban, and A. T. Balaban, *J. Math. Chem.*, **1993**, *14*, 21–33.
31. O. Ivanciuc, T. Laidboeur, and D. Cabrol-Bass, *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 485–488.
32. 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.
33. A. T. Balaban, *MATCH (Commun. Math. Chem.)*, **1986**, *21*, 115–122.
34. 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.
35. O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 395–401.
36. Part 12: O. Ivanciuc, *Rev. Roum. Chim.*, **2000**, *45*.
37. Part 14: O. Ivanciuc, *Rev. Roum. Chim.*, **2000**, *45*.
38. Part 15: O. Ivanciuc, *Rev. Roum. Chim.*, **2000**, *45*.
39. A. A. Gakh, E. G. Gakh, B. G. Sumpter, and D. W. Noid, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 832–839.
40. L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.
41. L. B. Kier and L. H. Hall, "Molecular Connectivity in Structure-Activity Analysis", Research Studies Press, Letchworth, 1986.