Internet Electronic Journal of Molecular Design

September 2002, Volume 1, Number 9, Pages 467–487

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Haruo Hosoya on the occasion of the 65th birthday
Part 1

Guest Editor: Jun–ichi Aihara

Quantitative Structure–Property Relationship Evaluation of Structural Descriptors Derived from the Distance and Reverse Wiener Matrices

Ovidiu Ivanciuc,[1] Teodora Ivanciuc,[2] and Alexandru T. Balaban[3]

[1] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157
[2] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[3] Texas A&M University at Galveston, Department of Oceanography and Marine Sciences, 5007 Avenue U, Galveston, Texas 77551

Received: July 9, 2002; Accepted: September 3, 2002; Published: September 30, 2002

Citation of the article:
O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Quantitative Structure–Property Relationship Evaluation of Structural Descriptors Derived from the Distance and Reverse Wiener Matrices, *Internet Electron. J. Mol. Des.* **2002**, *1*, 467–487, http://www.biochempress.com.

Internet Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Quantitative Structure–Property Relationship Evaluation of Structural Descriptors Derived from the Distance and Reverse Wiener Matrices[#]

Ovidiu Ivanciuc,[1,*] Teodora Ivanciuc,[2] and Alexandru T. Balaban[3]

[1] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157
[2] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[3] Texas A&M University at Galveston, Department of Oceanography and Marine Sciences, 5007 Avenue U, Galveston, Texas 77551

Received: July 9, 2002; Accepted: September 3, 2002; Published: September 30, 2002

Internet Electron. J. Mol. Des. **2002**, *1* (9), 467–487

Abstract

Structural descriptors derived from the molecular graph are widely used in developing QSPR and QSAR models, in chemical database searching, drug design, toxicology, virtual screening of combinatorial libraries, similarity and diversity assessment. As a consequence of the significant interest in defining additional structural descriptors for QSPR and QSAR models, we present new molecular descriptors computed from the reverse Wiener **RW** and reciprocal reverse Wiener **RRW** matrices. The graph structural descriptors computed with the distance **D**, reciprocal distance **RD**, **RW**, and **RRW** matrices are used to develop quantitative structure–property models for the boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of 134 alkanes C_6–C_{10}.

Keywords. QSAR, quantitative structure–activity relationships; QSPR, quantitative structure–property relationships; molecular matrices; structural descriptors; topological indices.

1 INTRODUCTION

The physical, chemical, and biological properties of chemical compounds are ultimately determined by the molecular structure. Quantitative structure–property relationships (QSPR) and quantitative structure–activity relationships (QSAR) models represent well established computational tools for the molecular design of new compounds with desired properties. All QSPR and QSAR models are statistically–based and designed to extract the maximum information from experimental data on compounds of known structure. All structure–property equations use atomic,

---

[#] Dedicated to Professor Haruo Hosoya on the occasion of the 65th birthday.
* Correspondence author; E–mail: ivanciuc@netscape.net.

bond, and molecular descriptors; these structural descriptors are numbers obtained from the chemical structure with the aid of various mathematical formulas or algorithms. The most efficient software used in QSPR or QSAR studies integrate the computation of structural descriptors with the generation of structure–property models [1,2]. Several programs from this category, such as ADAPT [3], OASIS [4,5], PRECLAV [6], SciQSAR [7], and CODESSA [8–10], were used with success in developing a large number of QSPR and QSAR models. These programs compute more than one thousand structural descriptors from five classes: constitutional, graph theoretic and topological indices, geometrical, electrostatic, quantum–chemical, and grid (field) descriptors. Using statistical methods, such as multilinear regression, PCA, PLS, or neural networks, the best descriptors are selected in the final structure–property model. A survey of the QSPR and QSAR models developed with the above programs shows that molecular graph descriptors and topological indices are used with success to model various properties, and demonstrates that they are valuable descriptors of chemical structure [11–27]. The interest of developing new graph descriptors for organic compounds revived in recent years, when topological indices found new applications in similarity and diversity assessment, database mining, and in the virtual screening of combinatorial libraries [28–30].

A molecular graph descriptor or topological index (TI) is a numerical representation of the molecular structure derived from the corresponding molecular graph; they are used with success as structural descriptors for pattern recognition, qualitative and quantitative structure–property models, or to measure the molecular similarity and diversity. The graph description of a molecule contains information on the atom–atom connectivity in the molecule, and encodes the size, shape, and branching features that determine the molecular properties; molecular graph descriptors represent valuable descriptors that complement (and not substitute) the structural information encoded in other classes of descriptors (constitutional, geometric, electrostatic, quantum, or field descriptors). Molecular graphs are non–directed chemical graphs that represent, in different conventions, molecules. Usually, only non–hydrogen atoms are taken into account in molecular graphs. In molecular graphs, vertexes 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. Because TIs are global descriptors of the molecular graph, they do not contain explicit information regarding the number of functional groups, pharmacophores, volume, surface area, interatomic distances, stereochemistry, charge distribution, orbital energy, or electrostatic potential; such information must be provided by other structural descriptors.

Molecular matrices, encoding in various ways the topological information, are an important source of structural descriptors for QSPR and QSAR models [31]. A large number of molecular matrices were defined in the chemical literature, such as the adjacency **A**, distance **D**, reciprocal distance **RD** [32], distance–path $\mathbf{D_p}$ [33,34], distance–delta $\mathbf{D_\Delta}$ [33,34], reciprocal distance–path $\mathbf{RD_p}$ [33,34], resistance distance matrix $\mathbf{\Omega}$ [35], electrical conductance [36], detour Δ [37], detour–

distance **Δ–D** [37], edge Szeged $\mathbf{Sz_e}$ [38–41], path Szeged $\mathbf{Sz_p}$ [38–41], reciprocal Szeged $\mathbf{RSz_p}$ [38–41], edge Cluj $\mathbf{Cj_e}$ [42], path Cluj $\mathbf{Cj_p}$ [42], distance–valency **Dval** [43,44], distance complement **DC** [45], reverse Wiener **RW** [46], complementary distance **CD** [47], and reciprocal complementary distance **RCD** [47] matrices. Certain features of the chemical structure that are encoded in an implicit form in the molecular graph are represented explicitly in molecular matrices; however, molecular matrices are not structural invariants, unless a canonical form is calculated.

In the present study we define new structural descriptors based on the recently introduced reverse Wiener **RW** [46] matrix and its reciprocal matrix **RRW**. The scope of this paper is to evaluate in QSPR models the topological indices derived from four molecular matrices, namely distance **D**, reciprocal distance **RD**, reverse Wiener **RW**, and reciprocal reverse Wiener **RRW**. Structural descriptors computed with these four molecular matrices are used to develop structure–property models for the normal boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of 134 alkanes C_6–C_{10}.

2 MATERIALS AND METHODS

2.1 The Reverse Wiener Matrix and Related Molecular Matrices

For simple molecular graphs, representing alkanes or cycloalkanes, the value of the *ij*–th element in the distance matrix **D** is equal to the number of bonds between two graph vertices v_i and v_j on the shortest path between them. In this way, the more distant two vertices, the larger the corresponding element in the distance matrix. Therefore, the largest contribution to the numerical value of the structural descriptors computed from the distance matrix arises from pairs of distant vertices, such as in the Wiener index W [11,12,16]. In four recently introduced molecular matrices, namely the reciprocal distance **RD** [32], distance complement **DC** [45], complementary distance **CD** [47], and the reverse Wiener **RW** [46] matrices, the value of the matrix elements corresponding to pairs of vertices decreases when the distance between the vertices increases. In this section we present the definition of **RW**, the new reciprocal reverse Wiener matrix **RRW**, and several related matrices.

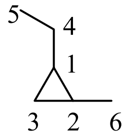
The diameter $d_{\max}$ of a graph is the largest topological distance between any two vertices v_i and v_j, i. e. the largest d_{ij} value in the distance matrix:

$$d_{\max} = \mathbf{max}\{d_{ij}, v_i, v_j \in V(G), v_i \neq v_j\} \tag{1}$$

The reverse Wiener matrix $\mathbf{RW} = \mathbf{RW}(G)$ of a graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are obtained by subtracting from $d_{\max}$ each d_{ij} value in the distance matrix [46]:

$$[\mathbf{RW}]_{ij} = \begin{cases} d_{\max} - [\mathbf{D}]_{ij} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases} \tag{2}$$

where $[\mathbf{D}]_{ij}$ is the *ij*–th element of the distance matrix **D** which is equal to the graph distance between vertices v_i and v_j. An example for the computation of the reverse Wiener matrix is presented for the molecular graph of 1–ethyl–2–methylcyclopropane **1**.



1

The first step is represented by the computation of the distance matrix **D(1)**:

D(1)

	1	2	3	4	5	6
1	0	1	1	1	2	2
2	1	0	1	2	3	1
3	1	1	0	2	3	2
4	1	2	2	0	1	3
5	2	3	3	1	0	4
6	2	1	2	3	4	0

Using the definition of the reverse Wiener matrix from Eq. (2) one obtains the reverse Wiener matrix **RW**(**1**):

RW(1)

	1	2	3	4	5	6
1	0	3	3	3	2	2
2	3	0	3	2	1	3
3	3	3	0	2	1	2
4	3	2	2	0	3	1
5	2	1	1	3	0	0
6	2	3	2	1	0	0

From the above example one can recognize a property of the reverse Wiener matrix, i.e. the matrix elements that correspond to $d_{\max}$ in the distance matrix **D** have zero values in **RW**, such as $\mathbf{RW(1)}_{5,6}$ and $\mathbf{RW(1)}_{6,5}$.

The reciprocal reverse Wiener matrix **RRW** = **RRW**(*G*) of a molecular graph *G* with *N* vertices is the square *N*×*N* symmetric matrix with real elements (rational numbers) defined with the elements of the **RW** matrix:

$$[\mathbf{RRW}(G)]_{ij} = \begin{cases} 0 & \text{if} \quad [\mathbf{RW}(G)]_{ij} = 0 \\ 1/[\mathbf{RW}(G)]_{ij} & \text{if} \quad i \neq j \\ [\mathbf{RW}(G)]_{ii} & \text{if} \quad i = j \end{cases} \tag{3}$$

The application of the above formula gives the reciprocal reverse Wiener matrix **RRW**(**1**) from the elements of **RW**(**1**):

RRW(1)

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

Recently we have defined a matrix related to **RW**, the complementary distance matrix. The complementary distance matrix **CD** = **CD**(G) of a graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are defined as [47]:

$$[\mathbf{CD}]_{ij} = \begin{cases} d_{\max} + d_{\min} - [\mathbf{D}]_{ij} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases} \tag{4}$$

where $[\mathbf{D}]_{ij}$ is the ij–th element of the distance matrix **D** which is equal to the graph distance between vertices v_i and v_j, and $d_{\min}$ is the minimum distance between two distinct graph vertices (equal to 1 for alkanes and cycloalkanes):

$$d_{\min} = \mathbf{min}\{d_{ij}, v_i, v_j \in V(G), v_i \neq v_j\} \tag{5}$$

It can be observed that for alkanes and cycloalkanes all entries in the complementary distance matrix **CD** are higher by 1 than those in the reverse Wiener matrix **RW**.

The reciprocal complementary distance matrix **RCD** = **RCD**(G) of a molecular graph G with N vertices is the square $N \times N$ symmetric matrix with real elements (rational numbers) defined with the equation [47]:

$$[\mathbf{RCD}(G)]_{ij} = \begin{cases} 1/[\mathbf{CD}(G)]_{ij} & \text{if} \quad i \neq j \\ [\mathbf{CD}(G)]_{ii} & \text{if} \quad i = j \end{cases} \tag{6}$$

Another matrix that has a certain similarity with **RW** and **CD** is the distance complement matrix introduced by Randić [45]. The distance complement matrix **DC** = **DC**(G) of a graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are defined as:

$$[\mathbf{DC}]_{ij} = \begin{cases} N - [\mathbf{D}]_{ij} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases} \tag{7}$$

2.2 Structural Descriptors Derived From The Reverse Wiener Matrix

The molecular graph operators were recently introduced as an extension of topological indices; a graph operator uses a mathematical equation to compute a family of related molecular graph descriptors with different molecular matrices and various sets of parameters for atoms and bonds [48–52]. The use of molecular graph operators introduces a systematization of topological indices by putting together all descriptors computed with the same mathematical formula or algorithm. As a consequence, when new molecular matrices are introduced there is no need to invent new names and symbols for the topological indices derived from them; the notation of graph operators is simple and general, and can accommodate new matrices, weighting schemes, and any parameter used in the definition of a family of topological indices. In this section we present molecular graph operators that we use to compute the structural descriptors. Because the graph operators are newly introduced, we present several examples for the computation of the structural descriptors used in the QSPR models from this paper.

Characteristic Polynomial Operator. The characteristic polynomial operator **Ch**(**M**) = **Ch**(**M**,*G*,*x*) of the molecular matrix **M** = **M**(*G*) is defined with the following equation [51]:

$$\mathbf{Ch}(\mathbf{M},G,x) = \mathbf{det}(x\mathbf{I}-\mathbf{M}) = \sum_{n=0}^{N} c_n x^{N-n} \tag{8}$$

where **I** is the unit matrix of order N and c_n is the n–th coefficient of the characteristic polynomial **Ch**(**M**). The characteristic polynomial of 1–ethyl–2–methylcyclopropane **1** computed from the **RW** and **RRW** matrices are:

$\mathbf{Ch}(\mathbf{RW},\mathbf{1}) = x^6 - 77x^4 - 368x^3 - 397x^2 + 684x + 1196$
$\mathbf{Ch}(\mathbf{RRW},\mathbf{1}) = x^6 - 4.91667x^4 - 4.12963x^3 + 0.61883x^2 + 0.98045x + 0.16459$

Hosoya Operator. The Hosoya operator **Ho**(**M**) = **Ho**(**M**,*G*) is defined as the sum of the absolute values of the coefficients c_n of the characteristic polynomial of the matrix **M** [50]:

$$\mathbf{Ho}(\mathbf{M}) = \sum_{n=0}^{N} |c_n| \tag{9}$$

For alkanes and if **M** is the adjacency matrix **A** the **Ho** operator is identical with the Hosoya index Z [12]. The Hosoya indices of 1–ethyl–2–methylcyclopropane **1** computed from the **RW** and **RRW** matrices are **Ho**(**RW**,**1**) = 2723 and **Ho**(**RRW**,**1**) = 11.81016.

Spectral Operators. The matrix spectrum operator **Sp**(**M**,*G*) = $\{x_i, i = 1, 2, ..., N\}$ represents the eigenvalues of a matrix **M** or the roots of the characteristic polynomial **Ch**(**M**,*G*,*x*), **Ch**(**M**,*G*,*x*) = 0 [51]. The spectral operators **MinSp**(**M**,*G*) and **MaxSp**(**M**,*G*) are equal to the minimum and maximum values of **Sp**(**M**,*G*), respectively:

$$\mathbf{MinSp}(\mathbf{M},G) = \mathbf{min}\{\mathbf{Sp}(\mathbf{M},G)\} \tag{10}$$

$$\mathbf{MaxSp}(\mathbf{M},G) = \mathbf{max}\{\mathbf{Sp}(\mathbf{M},G)\} \tag{11}$$

Structural descriptors derived from these operators were used with good results to develop QSPR models for the normal boiling temperature, heat of vaporization, molar refraction, molar volume, critical pressure, critical temperature, and surface tension of alkanes, to estimate the boiling points of acyclic compounds containing oxygen or sulfur atoms, to model the boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of alkanes, and to predict the retention index of alkylphenols [51]. The spectra of the molecular graph **1** computed from the **RW** and **RRW** matrices are:

Sp(**RW**,**1**) = {–3.45421, –3.37619, –3.11578, –2.16500, 1.42233, 10.68885}
Sp(**RRW**,**1**) = {–1.40631, –1.01528, –0.33187, –0.27460, 0.50044, 2.52762}

Wiener Operator. The Wiener operator **Wi**(**M**) = **Wi**(**M**,*G*) of a molecular graph *G* with *N* vertices is computed from the symmetric $N \times N$ molecular matrix **M** = **M**(*G*) [51]:

$$\mathbf{Wi}(\mathbf{M},G)=\sum_{i=1}^{N}\sum_{j=i}^{N}[\mathbf{M}(G)]_{ij} \tag{12}$$

The Wiener operator **Wi**(**M**) is an extension of the topological index *W* introduced by Wiener for alkanes [11], and extended to cycloalkanes by Hosoya [12]. While *W* is computed from the distance matrix, the Wiener operator **Wi**(**M**) can be applied to any molecular matrix, derived either from the molecular graph or from the three–dimensional structure of a chemical compound. From the definition of the reverse Wiener matrix from Eq. (2) and the formula of the Wiener operator, a simple relationship is revealed between the Wiener indices **Wi**(**D**) and **Wi**(**RW**):

$$\mathbf{Wi}(\mathbf{RW},G)=\frac{1}{2}d_{\max}N(N-1)-\mathbf{Wi}(\mathbf{D},G) \tag{13}$$

A similar relationship exists between the Wiener indices computed from the distance and the complementary distance matrices, namely **Wi**(**D**) and **Wi**(**CD**):

$$\mathbf{Wi}(\mathbf{CD},G)=\frac{1}{2}(d_{\max}+d_{\min})N(N-1)-\mathbf{Wi}(\mathbf{D},G) \tag{14}$$

For alkanes and cycloalkanes, when $d_{\min}$ = 1, one obtains a simple relationship between the Wiener indices computed from the reverse Wiener and the complementary distance matrices:

$$\mathbf{Wi}(\mathbf{CD},G)=\frac{1}{2}N(N-1)+\mathbf{Wi}(\mathbf{RW},G) \tag{15}$$

From the molecular matrices **RW** and **RRW** of molecule **1** one obtains **Wi**(**RW**,**1**) = 31 and **Wi**(**RRW**,**1**) = 7.5.

Hyper–Wiener Operator. The hyper–Wiener operator **HyWi**(**M**) = **HyWi**(**M**,*G*) of a molecular graph *G* with *N* vertices is computed from the symmetric $N \times N$ molecular matrix **M** = **M**(*G*) [51]:

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

The hyper–Wiener index *WW* was defined for alkanes by Randić [53] and extended to cycloalkanes by Klein, Lukovits, and Gutman [54]. Diudea proposed an alternative method for computing the hyper–Wiener from the distance–path matrix $\mathbf{D_p}$ [33]. Equation (16) extends the computation of the hyper–Wiener indices to molecular graph matrices or matrices derived from the three–dimensional molecular structure. The molecular matrices **RW** and **RRW** of 1–ethyl–2–

methylcyclopropane give **HyWi**(**RW**,**1**) = 54 and **HyWi**(**RRW**,**1**) = 6.20833.

Vertex Sum Operator. In a molecular graph G with N vertices, the vertex sum operator for the vertex v_i, $\mathbf{VS}(\mathbf{M},G)_i$, is defined as the sum of the elements in the column i, or row i, of the molecular matrix **M** [51]:

$$\mathbf{VS}(\mathbf{M},G)_i = \sum_{j=1}^{N}[\mathbf{M}]_{ij} = \sum_{j=1}^{N}[\mathbf{M}]_{ji} \tag{17}$$

Using the above equation for the molecular matrices **RW** and **RRW** of molecule **1** one obtains the following vertex sum vectors:

VS(**RW**,**1**) = {13, 12, 11, 11, 7, 8}
VS(**RRW**,**1**) = {2.00000, 2.50000, 2.66667, 2.66667, 2.83333, 2.33333}

Ivanciuc–Balaban Operator. Using the formula for Balaban's index J [14], the Ivanciuc–Balaban operator of a graph G, **IB**(**M**) = **IB**(**M**,G), of the symmetric $N \times N$ molecular matrix **M** = **M**(G) is [15]:

$$\mathbf{IB}(\mathbf{M},G) = \frac{M}{\mu+1}\sum_{e_{ij}\in E(G)}\left(\mathbf{VS}(\mathbf{M})_i \times \mathbf{VS}(\mathbf{M})_j\right)^{-1/2} \tag{18}$$

where $\mathbf{VS}(\mathbf{M})_i$ and $\mathbf{VS}(\mathbf{M})_j$ denote the vertex sums of the two adjacent vertices v_i and v_j that are incident with an edge e_{ij} in the molecular graph G, M is the number of edges in the molecular graph, μ is the cyclomatic number (the number of cycles in the graph, $\mu = M - N + 1$, where N is the number of atoms the molecular graph), and the summation goes over all edges from the edge set $E(G)$. The application of formula (18) to the vertex sum vector **VS**(**RW**,**1**) gives the value of the Ivanciuc–Balaban index **IB**(**RW**,**1**):

$$\mathbf{IB}(\mathbf{RW},\mathbf{1}) = 4[(13\cdot12)^{-1/2} + (12\cdot11)^{-1/2} + (13\cdot11)^{-1/2} + (13\cdot11)^{-1/2} + (11\cdot7)^{-1/2} + (12\cdot8)^{-1/2}] = 1.65112$$

Analogously, from the reciprocal reverse Wiener matrix **RRW**(**1**) one obtains **IB**(**RRW**,**1**) = 7.43514.

Information Theoretic Operators U, V, X, and Y. The indices U, V, X, and Y for information on distances are computed from the elements of the distance matrix of the molecular graph [55], and these TIs provided good results both for structure discrimination and in structure–property models [56]. Because new graph matrices were defined in recent years, it is possible to extend the definition of these four indices for all molecular matrices **M**. We have recently introduced four 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 [57]:

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

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

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

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

where **M** is a molecular graph matrix, $\mathbf{VS}(\mathbf{M})_i$ represents the vertex sum of the vertex v_i, and the summations in equations (19) and (22) are done for the non–zero elements of the molecular matrix **M**, $[\mathbf{M}]_{ij} \neq 0$. 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:

$$f(x,y) = \begin{cases} (xy)^{-1/2} & \text{if } xy > 0 \\ -(|xy|)^{-1/2} & \text{if } xy < 0 \end{cases} \tag{23}$$

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

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

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

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

$$\mathbf{Y}(\mathbf{M}, G) = \frac{M}{\mu + 1} \sum_{E(G)} f(\mathbf{VYinf}(\mathbf{M})_i, \mathbf{VYinf}(\mathbf{M})_j) \tag{27}$$

The values of the local invariants **VUinf**, **VVinf**, **VXinf**, and **VYinf**, computed for all vertices in the molecular graph **1** from the reverse Wiener matrix are:

VUinf(**RW**,**1**) = {2.29547, 2.22957, 2.23127, 2.23127, 1.84237, 1.90564}
VVinf(**RW**,**1**) = {45.81025, 40.78998, 35.82248, 35.82248, 17.80911, 22.09436}
VXinf(**RW**,**1**) = {29.84105, 26.75489, 24.54397, 24.54397, 12.89660, 15.24511}
VYinf(**RW**,**1**) = {18.26466, 16.26466, 13.50978, 13.50978, 6.75489, 8.75489}

The above vertex invariants give the molecular information indices: **U**(**RW**,**1**) = 8.25738, **V**(**RW**,**1**) = 0.51470, **X**(**RW**,**1**) = 0.76211, **Y**(**RW**,**1**) = 1.32385. For the reciprocal reverse Wiener matrix one obtains the following values for the four vectors of local invariants:

VUinf(**RRW**,**1**) = {2.29248, 2.15591, 2.18628, 2.18628, 1.86544, 1.87739}
VVinf(**RRW**,**1**) = {–0.29248, 1.14891, 1.58716, 1.58716, 2.39165, 0.97486}
VXinf(**RRW**,**1**) = {4.58496, 5.38978, 5.83007, 5.83007, 5.28541, 4.38057}
VYinf(**RRW**,**1**) = {–2.58496, –2.08496, –2.05664, –2.05664, –1.02832, –1.52832}

As explained above, certain vertex invariants may have negative values, such is observed for **VVinf**(**RRW**,**1**) and **VYinf**(**RRW**,**1**); in these cases the modification of the Randić–like formula (23) is necessary to compute the four molecular information indices: **U**(**RRW**,**1**) = 8.38802, **V**(**RRW**,**1**) = –7.38540, **X**(**RRW**,**1**) = 3.45701, **Y**(**RRW**,**1**) = 9.08672.

2.3 Structure–Property Models

Data. The QSPR models were developed for a data set consisting of 134 alkanes between C_6 and C_{10}, for the following six physical properties [58]: t_b, boiling temperature at normal pressure (°C); C_p, molar heat capacity at 300 K (J K^{-1} mol^{-1}); $\Delta_f G°_{300}$ (g), standard Gibbs energy of formation in the gas phase at 300 K (kJ mol^{-1}); $\Delta_{vap}H_{300}$, vaporization enthalpy at 300 K (kJ mol^{-1}); n_D^{25}, refractive index at 25 °C; ρ, density at 25 °C (kg m^{-3}). The value of the refractive index of 2,2,3,3–tetramethylbutane is missing, while the reported density of this compound, 821.70 kg m^{-3}, is too high when compared with the density of similar alkanes and it was not considered in the computation of the density QSPR models. As it is known, there are 142 constitutional isomers for these alkanes, but data for all six properties are missing for the following eight of them: *n*–hexane, *n*–nonane, *n*–decane, 2–methylnonane, 3–methylnonane, 4–methylnonane, 5–methylnonane, 3–ethyl–2,4–dimethylhexane.

Molecular matrices. Molecular matrices represent an important source of structural descriptors computed from molecular graphs. Usually, a small set of matrices is used to characterize the molecular topology, namely the adjacency, the distance, and sometimes, the Laplacian matrix. The structural descriptors defined on molecular matrices, with the exception of the Hosoya index, were computed from the following four graph matrices: distance **D**, reciprocal distance **RD**, reverse Wiener **RW**, and reciprocal reverse Wiener **RRW** matrices. In previous studies we have pointed that the Hosoya indices for certain molecular matrices can have too large values to be useful as structural descriptors in QSPR and QSAR models [51]. In this study the Hosoya indices were computed from the adjacency **A**, **RD**, and **RRW** matrices.

Structural descriptors. The 45 structural descriptors used in the QSPR study are: (1) molecular weight, **MW**; (2) five Kier and Hall connectivity indices $^0\chi$, $^1\chi$, $^2\chi$, $^3\chi_p$, $^3\chi_c$ [59,60]; (3) three Hosoya indices **Ho**(**M**); (4) four Wiener indices computed with the Wiener operator **Wi**(**M**); (5) four hyper–Wiener indices computed with the hyper–Wiener operator **HyWi**(**M**); (6) eight spectral operators **MinSp**(**M**) and **MaxSp**(**M**); (7) four Ivanciuc–Balaban indices **IB**(**M**); (8) sixteen information–theory indices **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**).

QSPR model. The QSPR models were obtained by selecting the best combination of structural descriptors that correspond to certain conditions. This algorithm starts from the set of 45 structural descriptors and develops QSPR models by applying the following steps: (1) All one–parameter correlation equations are computed. All descriptors with a correlation coefficient greater than a threshold, $|r_{\min}| > 0.15$, are selected for further use. (2) Multiple linear regression (MLR) regression equations are computed with all possible groups of k 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 ten MLR equations are reported. (3) Step (2) is performed for k from 2 to 4.

For all six alkane properties the best results are obtained with MLR models containing three structural descriptors, when a maximum is identified for the values of the Fisher test F. Due to this finding, only these QSPR equations are presented in the next section.

3 RESULTS AND DISCUSSION

Normal boiling temperature. In Table 1 we present the coefficients, confidence interval, structural descriptors, and statistical indices for the best ten MLR equations with three independent variables that model the alkane boiling temperature.

Table 1. Coefficients, confidence interval, structural descriptors $\mathbf{SD}_i$ (i = 1–3), and statistical indices for the best ten MLR equations with three independent variables that model the alkane boiling temperature at normal pressure, t_b (°C). The MLR equations have the general form: $t_b = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Eq.	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
1	-1.4678×10^2	8.833	$^3\chi_p$	4.548×10	**MaxSp(RD)**	6.535	**V(RD)**	0.9939	2.97	3495.3
	±8.22	±0.495		±2.55		±0.366				
2	-7.561×10	1.1880	**MW**	1.2531×10	$^3\chi_p$	2.753	**V(RD)**	0.9926	3.25	2912.8
	±4.64	±0.0729		±0.769		±0.169				
3	-8.387×10	1.811×10	$^0\chi$	1.3691×10	$^3\chi_p$	4.863	**V(RD)**	0.9925	3.27	2869.5
	±5.18	±1.12		±0.846		±0.300				
4	-3.471×10	1.1313×10	$^3\chi_p$	6.176	**HyWi(RD)**	4.568	**V(RD)**	0.9923	3.33	2767.0
	±2.18	±0.712		±0.389		±0.287				
5	-8.799×10	1.2424×10	$^3\chi_p$	2.165×10	**MaxSp(RD)**	2.009×10	**X(RD)**	0.9921	3.36	2726.1
	±5.58	±0.788		±1.37		±1.27				
6	-2.216×10	1.2379	**MW**	1.5319×10	$^3\chi_p$	-3.911×10	**V(D)**	0.9920	3.40	2661.9
	±1.42	±0.0794		±0.983		±2.51				
7	-2.830×10	1.1275×10	$^3\chi_p$	4.825	**Wi(RD)**	4.495	**V(RD)**	0.9918	3.43	2609.3
	±1.83	±0.731		±0.313		±0.291				
8	-6.611×10	3.943×10	$^1\chi$	5.009	$^2\chi$	1.0221×10	$^3\chi_p$	0.9915	3.50	2508.0
	±4.37	±2.61		±0.331		±0.676				
9	-8.468×10	1.1345×10	$^3\chi_p$	1.632×10	**IB(D)**	2.612×10	**X(RD)**	0.9914	3.51	2491.0
	±5.62	±0.752		±1.08		±1.73				
10	-3.823×10	1.2469	**MW**	1.511×10	$^3\chi_p$	−6.650	**Y(D)**	0.9910	3.59	2376.8
	±2.59	±0.0847		±1.03		±0.452				

The best MLR equation with two independent variables, with r = 0.9939, s = 2.97, F = 3495.3, contains the connectivity index $^3\chi_p$, the maximum eigenvalue **MaxSp**(**RD**), and the information index **V**(**RD**) computed with the reciprocal distance matrix. In this equation, the index $^3\chi_p$

represents the weighted contribution of butane–like subgraphs and is a measure of molecular branching, **MaxSp**(**RD**) is mainly a shape descriptor free from size contribution, and **V**(**RD**) is an information index related to the size of the elements in the **RD** matrix. Out of the 14 descriptors in Table 1 computed from molecular matrices, 11 are derived from the reciprocal distance matrix **RD** and 3 from the distance matrix **D**. All ten QSPR equations from Table 1 contain the connectivity index ${}^3\chi_p$. Other important descriptors for the modeling of alkane boiling temperature are **V**(**RD**) selected in 5 equations, **MW** selected in 3 equations, **MaxSp**(**RD**) and **X**(**RD**) selected in 2 equations each. The QSPR models from Table 1 do not contain structural descriptors computed from the reverse Wiener **RW** and reciprocal reverse Wiener **RRW** matrices, showing that these indices are not important in modeling this property. One of the most widely used topological index in QSPR models, mainly for boiling temperature, is the Wiener index **Wi**(**D**); however, this index is missing from the equations reported in Table 1, indicating that the new graph descriptors **MaxSp**(**RD**) and **V**(**RD**) are able to offer better structure–property models.

Molar heat capacity. The best ten QSPR models with three independent variables that model the alkane molar heat capacity are presented in Table 2.

Table 2. Coefficients, confidence interval, structural descriptors **SD**$_i$ (i = 1–3), and statistical indices for the best ten MLR equations with three independent variables that model the alkane molar heat capacity at 300 K, C_p (J K^{-1} mol^{-1}). The MLR equations have the general form: $C_p = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Eq.	a_0	a_1	SD$_1$	a_2	SD$_2$	a_3	SD$_3$	r	s	F
11	7.053	1.660	**MW**	3.083×10^{-2}	**Ho**(**RRW**)	-4.472×10^{-1}	**Y**(**RRW**)	0.9883	3.92	1817.4
	±0.548	±0.129		±0.00239		±0.0347				
12	3.389	1.769	**MW**	5.295	**MinSp**(**RRW**)	−1.1624	**X**(**RRW**)	0.9883	3.92	1816.4
	±0.263	±0.137		±0.411		±0.0903				
13	1.0194	1.738	**MW**	4.711×10^{-1}	**MinSp**(**D**)	3.913	**MinSp**(**RRW**)	0.9882	3.94	1800.5
	±0.0795	±0.136		±0.0368		±0.305				
14	8.784	1.595	**MW**	5.091×10^{-1}	**Wi**(**RRW**)	5.094	**MinSp**(**RRW**)	0.9881	3.95	1788.9
	±0.687	±0.125		±0.0398		±0.399				
15	1.0493×10	3.038×10	**MaxSp**(**RD**)	5.361	**IB**(**RD**)	-2.194×10^{-1}	**Y**(**RRW**)	0.9881	3.95	1787.6
	±0.821	±2.38		±0.420		±0.0172				
16	1.0910×10^{2}	9.474×10	**MinSp**(**RD**)	4.136×10	**MaxSp**(**RD**)	4.560	**IB**(**RD**)	0.9881	3.96	1782.1
	±8.55	±7.43		±3.24		±0.358				
17	-6.334×10^{-1}	1.778	**MW**	4.822	**MinSp**(**RRW**)	-3.218×10^{-1}	**IB**(**RRW**)	0.9880	3.96	1778.3
	±0.0497	±0.140		±0.378		±0.0253				
18	−1.422	4.022×10	${}^1\chi$	8.293	${}^2\chi$	5.391	**IB**(**D**)	0.9880	3.96	1777.4
	±0.112	±3.16		±0.651		±0.423				
19	4.419	1.736	**MW**	1.496	**MinSp**(**RW**)	4.236	**MinSp**(**RRW**)	0.9880	3.96	1774.9
	±0.347	±0.136		±0.118		±0.333				
20	1.2091×10	4.673×10	${}^1\chi$	8.059	${}^2\chi$	−1.794	**V**(**RD**)	0.9880	3.96	1774.9
	±0.950	±3.67		±0.633		±0.141				

The first QSPR model, with r = 0.9883, s = 3.92, F = 1817.4, contains as descriptors the molecular weight **MW**, the Hosoya index **Ho**(**RRW**), and the information–theoretic index **Y**(**RRW**). In modeling this property the **RRW** matrix, defined in this paper, gives topological indices with a better correlational power than those derived from the distance matrix **D**, which until recently was the main source of graph invariants. In the set of ten equations from Table 2, **MW** and

MinSp(**RRW**) appear 6 times each, while $^1\chi$, $^2\chi$, **Y**(**RRW**), **MaxSp**(**RD**), and **IB**(**RD**) appear 2 times each. The majority of TIs are computed from the reciprocal matrices **RRW** and **RD**, namely 11 and 6 descriptors, respectively, while only one descriptor is derived from the reverse Wiener matrix, namely **MinSp**(**RW**). The QSPR models from Table 2 indicate that TIs derived from the **RRW** and **RW** matrices are important descriptors in QSPR of alkane molar heat capacity, giving better correlations than those obtained from the **D** and **RD** matrices. The size descriptor **MW** is highly significant in this models, while connectivity indices have a small importance in modeling alkane molar heat capacity.

Standard Gibbs energy of formation. In Table 3 we give the best ten QSPR models with three structural descriptors that model the alkane standard Gibbs energy of formation. The best MLR equation, with r = 0.9564, s = 4.35, F = 464.6, contains the Ivanciuc–Balaban indices **IB**(**D**) and **IB**(**RW**), and the information–theoretic descriptor **V**(**RD**).

Table 3. Coefficients, confidence interval, structural descriptors $\mathbf{SD}_i$ (i = 1–3), and statistical indices for the best ten MLR equations with three independent variables that model the alkane standard Gibbs energy of formation in the gas phase at 300 K, $\Delta_fG°_{300}$ (g) (kJ mol^{-1}). The MLR equations have the general form: $\Delta_fG°_{300} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Eq.	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
21	-1.463×10^{2}	3.118×10	**IB**(**D**)	3.688	**IB**(**RW**)	4.809	**V**(**RD**)	0.9564	4.35	464.6
	±22.5	±4.79		±0.566		±0.738				
22	-1.420×10^{2}	–6.010	**MinSp**(**RRW**)	3.234×10	**IB**(**D**)	4.180	**V**(**RD**)	0.9548	4.42	447.5
	±22.2	±0.940		±5.06		±0.654				
23	-1.343×10^{2}	4.089×10^{-2}	**Ho**(**RRW**)	3.257×10	**IB**(**D**)	4.172	**V**(**RD**)	0.9531	4.51	429.6
	±21.4	±0.00653		±5.20		±0.666				
24	-1.538×10^{2}	3.426×10	**IB**(**D**)	4.915	**V**(**RD**)	6.37	**X**(**RW**)	0.9530	4.51	428.8
	±24.6	±5.48		±0.786		±1.02				
25	-1.395×10^{2}	4.231×10	**IB**(**D**)	4.833	**V**(**RD**)	–1.479	**Y**(**RD**)	0.9528	4.52	427.2
	±22.3	±6.78		±0.774		±0.237				
26	-1.027×10^{2}	3.851×10	**IB**(**D**)	1.333×10	**X**(**RD**)	–3.205	**Y**(**RD**)	0.9510	4.60	410.2
	±16.8	±6.29		±2.18		±0.524				
27	-6.62×10	2.387	**MaxSp**(**D**)	3.988×10	**IB**(**D**)	–4.219	**Y**(**RD**)	0.9502	4.64	403.0
	±10.9	±0.394		±6.58		±0.696				
28	-4.707×10	–7.34	$^2\chi$	-1.841×10^{-1}	**Wi**(**RW**)	7.96	**HyWi**(**RD**)	0.9501	4.64	402.4
	±7.77	±1.21		±0.0304		±1.31				
29	-1.720×10^{2}	-1.804×10^{-1}	**Ho**(**A**)	4.098×10	**IB**(**D**)	5.829	**V**(**RD**)	0.9497	4.66	398.8
	±28.5	±0.0299		±6.79		±0.966				
30	-1.659×10^{2}	4.186×10	**IB**(**D**)	–1.328	**IB**(**RD**)	6.09	**V**(**RD**)	0.9497	4.66	398.2
	±27.5	±6.94		±0.220		±1.01				

A comparison of the statistical indices r, s, and F show that all equations from Table 3 are similar from a statistical point of view, although they are obtained with different sets of structural descriptors. This happens because QSPR and QSAR equations establish a statistical (and not causal) mathematical relationship between a set of structural descriptors and a physical, chemical, or biological property. Whenever these models are generated from a large set of descriptors, one obtains a group of statistically equivalent QSPR and QSAR equations that contain different structural descriptors. These similar statistical models offer information on the important structural descriptors that determine the investigated property. An analysis of the TIs from Table 3 shows that structural descriptors computed with the **RW** and **RRW** matrices are significant parameters for modeling the alkane standard Gibbs energy of formation, because each of the first four QSPR models contain one descriptor computed with these matrices. Overall, 13 descriptors are computed from **RD**, 10 from **D**, 3 from **RW**, and 2 from **RRW**. The structural descriptors **IB**, **V**, and **Y** computed from the distance and reciprocal distance matrices were selected with a higher frequency: **IB**(**D**) 9 times, **V**(**RD**) 7 times, and **Y**(**RD**) 3 times. This result clearly indicates the superiority of the new generation topological indices, represented by the Ivanciuc–Balaban and information theoretic operators, over those computed with simpler mathematical operators.

Vaporization enthalpy. The best ten QSPR models with three topological indices that model the alkane vaporization enthalpy are presented in Table 4. The first equation, with $r = 0.9895$, $s = 0.63$, F = 2033.4, contains the spectral descriptor **MaxSp**(**RD**), and the information–theoretic indices **V**(**RD**) and **X**(**RW**).

Table 4. Coefficients, confidence interval, structural descriptors $\mathbf{SD}_i$ (i = 1–3), and statistical indices for the best ten MLR equations with three independent variables that model the alkane vaporization enthalpy at 300 K, $\Delta_{vap}H_{300}$ (kJ mol^{-1}). The MLR equations have the general form: $\Delta_{vap}H_{300} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Eq.	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
31	-1.0177×10	6.610	**MaxSp**(**RD**)	1.684	**V**(**RD**)	7.131×10^{-1}	**X**(**RW**)	0.9895	0.63	2033.4
	±0.747	±0.485		±0.124		±0.0523				
32	–9.285	6.380	**MaxSp**(**RD**)	3.112×10^{-1}	**IB**(**RW**)	1.677	**V**(**RD**)	0.9895	0.63	2023.4
	±0.683	±0.469		±0.0229		±0.123				
33	–9.938	6.641	**MaxSp**(**RD**)	1.675	**V**(**RD**)	7.098×10^{-1}	**V**(**RW**)	0.9894	0.63	2016.0
	±0.733	±0.490		±0.124		±0.0523				
34	–8.565	6.069	**MaxSp**(**RD**)	4.962×10^{-1}	**MaxSp**(**RRW**)	1.660	**V**(**RD**)	0.9894	0.63	2008.3
	±0.633	±0.448		±0.0367		±0.123				
35	–9.063	1.0304	$^{2}\chi$	4.918	**IB**(**D**)	2.234	**V**(**RD**)	0.9893	0.63	1999.2
	±0.671	±0.0763		±0.364		±0.165				
36	–7.328	9.207×10^{-2}	**HyWi**(**RRW**)	5.981	**MaxSp**(**RD**)	1.636	**V**(**RD**)	0.9893	0.63	1996.6
	±0.543	±0.00682		±0.443		±0.121				
37	–9.505	6.661	**MaxSp**(**RD**)	1.652	**V**(**RD**)	1.390×10^{-1}	**Y**(**RW**)	0.9893	0.64	1985.0
	±0.706	±0.495		±0.123		±0.0103				
38	–9.704	-3.571×10^{-1}	$^{3}\chi_{p}$	7.248	**MaxSp**(**RD**)	1.555	**V**(**RD**)	0.9891	0.64	1959.3
	±0.726	±0.0267		±0.542		±0.116				
39	–3.535	1.279×10^{-1}	**MW**	2.366	**IB**(**D**)	1.422	**V**(**RD**)	0.9889	0.65	1916.6
	±0.267	±0.00967		±0.179		±0.108				
40	–8.325	1.752×10^{-1}	$^{2}\chi$	6.501	**MaxSp**(**RD**)	1.581	**V**(**RD**)	0.9888	0.65	1906.2
	±0.631	±0.0133		±0.493		±0.120				

Several structural descriptors computed from the reverse Wiener and reciprocal reverse Wiener matrices are present in these structure–property models, such as **X**(**RW**), **IB**(**RW**), **V**(**RW**), **MaxSp**(**RW**), **HyWi**(**RRW**), and **Y**(**RW**). This finding demonstrates that although all four matrices are computed from graph distances, the novel **RW** and **RRW** matrices reflect some structural features that are absent from the **D** and **RD** matrices. All QSPR equations from Table 4 contain the information index **V**(**RD**), while the maximum eigenvalue of the **RD** matrix, **MaxSp**(**RD**), was selected in eight models. The structural descriptors computed from the reciprocal distance matrix were selected with a higher frequency: **RD** 18 times, **RW** 4 times, **D** and **RRW** twice each.

Refractive index. In Table 5 we present the best ten MLR equations with three structural descriptors that model the alkane refractive index. The first MLR equation has good statistical indices, r = 0.9840, s = 0.0025, F = 1309.4, and contains a connectivity index, $^3\chi_p$, the Ivanciuc–Balaban index computed from the distance matrix, **IB**(**D**), and the information index **X**(**D**).

Table 5. Coefficients, confidence interval, structural descriptors $\mathbf{SD}_i$ (i = 1–3), and statistical indices for the best ten MLR equations with three independent variables that model the alkane refractive index at 25 °C, n_D^{25}. The MLR equations have the general form: $n_D^{25} = a_0 + a_1\mathbf{SD}_1 + a_2\mathbf{SD}_2 + a_3\mathbf{SD}_3$.

Eq.	a_0	a_1	$\mathbf{SD}_1$	a_2	$\mathbf{SD}_2$	a_3	$\mathbf{SD}_3$	r	s	F
41	1.377 ±0.126	8.563×10^{-3} ±0.000783	$^3\chi_p$	3.545×10^{-2} ±0.00324	**IB**(**D**)	-9.069×10^{-2} ±0.00830	**X**(**D**)	0.9840	0.0025	1309.4
42	1.285 ±0.132	9.145×10^{-3} ±0.000939	$^3\chi_p$	2.254×10^{-2} ±0.00231	**MaxSp**(**RD**)	8.335×10^{-4} ±0.000086	**V**(**RD**)	0.9799	0.0027	1039.8
43	1.306 ±0.134	1.000×10^{-2} ±0.00103	$^3\chi_p$	2.368×10^{-2} ±0.00243	**MaxSp**(**RD**)	-1.521×10^{-2} ±0.00156	**X**(**D**)	0.9799	0.0027	1039.5
44	1.291 ±0.136	4.317×10^{-3} ±0.000454	$^1\chi$	8.742×10^{-3} ±0.000920	$^3\chi_p$	1.922×10^{-2} ±0.00202	**MaxSp**(**RD**)	0.9790	0.0028	990.1
45	1.277 ±0.135	6.604×10^{-3} ±0.000697	$^3\chi_p$	-2.641×10^{-3} ±0.000279	$^3\chi_c$	2.888×10^{-2} ±0.00305	**MaxSp**(**RD**)	0.9788	0.0028	982.1
46	1.292 ±0.136	9.51×10^{-3} ±0.00101	$^3\chi_p$	1.991×10^{-2} ±0.00211	**MaxSp**(**RD**)	2.328×10^{-3} ±0.000246	**X**(**RD**)	0.9788	0.0028	982.0
47	1.308 ±0.139	6.416×10^{-4} ±0.000068	**MW**	1.110×10^{-2} ±0.00118	$^3\chi_p$	-6.673×10^{-5} ±0.000007	**Wi**(**RW**)	0.9787	0.0028	976.5
48	1.287 ±0.138	9.230×10^{-3} ±0.000986	$^3\chi_p$	2.250×10^{-2} ±0.00240	**MaxSp**(**RD**)	-1.451×10^{-3} ±0.000155	**MinSp**(**RW**)	0.9783	0.0029	960.4
49	1.299 ±0.139	9.44×10^{-3} ±0.00101	$^3\chi_p$	1.886×10^{-2} ±0.00202	**MaxSp**(**RD**)	6.930×10^{-4} ±0.000074	**IB**(**RD**)	0.9781	0.0029	951.8
50	1.298 ±0.140	9.60×10^{-3} ±0.00103	$^3\chi_p$	3.665×10^{-4} ±0.000039	**MaxSp**(**D**)	1.961×10^{-2} ±0.00211	**MaxSp**(**RD**)	0.9780	0.0029	946.8

Although obtained with different sets of structural descriptors, all QSPR models are similar from a statistical point of view. Two structural descriptors computed from the reverse Wiener matrix **RW** were selected in the ten MLR equations, namely **Wi**(**RW**) and **MinSp**(**RW**). All QSPR equations from this table contain the connectivity index $^3\chi_p$, while the maximum eigenvalue of the reciprocal distance matrix, **MaxSp**(**RD**), appears in seven equations; this shows the importance of the weighted count of butane–like subgraphs, represented by $^3\chi_p$, and of the spectral index **MaxSp**, a shape index free from size contribution, in modeling the alkane refractive index.

Density. The best ten MLR equations with three topological indices that model the alkane density are presented in Table 6. The first QSPR model, with r = 0.9902, s = 3.73, F = 2156.0, contains the same descriptors from Eq. (41) from Table 5, namely $^3\chi_p$, **IB**(**D**), and **X**(**D**). There is a significant similarity between the descriptors selected for the alkane refractive index and density QSPR models. An inspection of the QSPR models from Tables 5 and 6 reveals that the following equations have identical structural descriptors: Eqs. (41) and (51); Eqs. (42) and (55); Eqs. (43) and (60); Eqs. (44) and (58); Eqs. (45) and (54).

Table 6. Coefficients, confidence interval, structural descriptors $\mathbf{SD}_i$ (i = 1–3), and statistical indices for the best ten MLR equations with three independent variables that model the alkane density at 25 °C, ρ (kg m^{-3}). 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	**SD**$_1$	a_2	**SD**$_2$	a_3	**SD**$_3$	r	s	F
51	6.553×10^2	1.922×10	$^3\chi_p$	6.289×10	**IB**(**D**)	-1.563×10^2	**X**(**D**)	0.9902	3.73	2156.0
	±46.7	±1.37		±4.48		±11.1				
52	5.973×10^2	2.079×10	$^3\chi_p$	5.288×10	**IB**(**D**)	-1.153×10^2	**V**(**D**)	0.9895	3.85	2020.3
	±43.9	±1.53		±3.89		±8.49				
53	4.619×10^2	−8.326	$^2\chi$	1.257×10	$^3\chi_p$	6.424×10	**MaxSp**(**RD**)	0.9862	4.42	1520.9
	±39.2	±0.707		±1.07		±5.45				
54	4.828×10^2	1.613×10	$^3\chi_p$	−5.192	$^3\chi_c$	5.128×10	**MaxSp**(**RD**)	0.9861	4.44	1510.2
	±41.1	±1.37		±0.442		±4.37				
55	5.004×10^2	2.093×10	$^3\chi_p$	3.942×10	**MaxSp**(**RD**)	1.330	**V**(**RD**)	0.9851	4.59	1411.3
	±44.1	±1.84		±3.47		±0.117				
56	4.630×10^2	1.486×10	$^3\chi_p$	4.494×10	**IB**(**D**)	6.418	**V**(**RD**)	0.9850	4.60	1403.5
	±40.9	±1.31		±3.97		±0.567				
57	5.268×10^2	2.237×10	$^3\chi_p$	4.018×10	**MaxSp**(**RD**)	-1.979×10	**V**(**D**)	0.9847	4.64	1375.5
	±47.0	±2.00		±3.59		±1.77				
58	5.110×10^2	7.053	$^1\chi$	2.029×10	$^3\chi_p$	3.394×10	**MaxSp**(**RD**)	0.9846	4.66	1365.0
	±45.8	±0.632		±1.82		±3.04				
59	5.202×10^2	2.254×10	$^3\chi_p$	3.988×10	**MaxSp**(**RD**)	−3.507	**Y**(**D**)	0.9846	4.67	1359.7
	±46.7	±2.02		±3.58		±0.315				
60	5.327×10^2	2.213×10	$^3\chi_p$	4.130×10	**MaxSp**(**RD**)	-2.251×10	**X**(**D**)	0.9845	4.67	1357.0
	±47.9	±1.99		±3.71		±2.02				

All QSPR equations from Table 6 contain the connectivity index $^3\chi_p$, and the maximum eigenvalue **MaxSp**(**RD**) appears in seven equations. Seven models contain an information–theory index computed mainly from the distance matrix **D**, and the Ivanciuc–Balaban index **IB**(**D**) is present in three equations. We have to mention that the graph descriptors computed from the reverse Wiener **RW** and reciprocal reverse Wiener **RRW** matrices were not selected in the QSPR models from Table 6, showing that these structural descriptors are not very important for the modeling of the alkane density.

Frequency of matrices and descriptors in QSPR models. The scope of this paper is to evaluate the correlational ability of structural descriptors derived from two new matrices, the reverse Wiener **RW** and reciprocal reverse Wiener **RRW** matrices. QSAR and QSPR models developed with topological indices used mainly two types of descriptors, namely the Kier and Hall connectivity indices and the Wiener index W (denoted in this paper **Wi**(**D**)). New molecular descriptors were recently introduced with the aid of graph operators, such as Hosoya **Ho**(**M**), Wiener **Wi**(**M**), hyper–Wiener **HyWi**(**M**), minimum matrix eigenvalue **MinSp**(**M**), maximum matrix eigenvalue **MaxSp**(**M**), Ivanciuc–Balaban **IB**(**M**), information–theory **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**); in the above operators **M** represents a molecular matrix. We have used all these operators with the aim to evaluate the utility of the graph descriptors from the new generation. Four molecular matrices were employed in the computation of TIs, namely distance **D**, reciprocal distance **RD**, reverse Wiener **RW**, and reciprocal reverse Wiener **RRW** matrices; previously, TIs were computed mainly from **D**, and this comparison can assess the value of the newly introduced matrices. For each alkane property we have reported in Tables 1–6 the best ten QSPR models, demonstrating that all six properties can be modeled with the set of 45 structural descriptors. However, each property is best modeled by a particular combination of molecular descriptors, indicating that the QSAR and QSPR modeling needs descriptors that cover the diversity of the chemical space. All QSAR and QSPR equations represent statistical models, and their results have to be interpreted as statistical (and not causal) relationships. In this respect, a trend observed from a set of QSPR equations, as we have obtained from the QSPR models presented in Tables 1–6, is more significant than the conclusions obtained from a single equation, even the "best" one. It is clear that, for the same property, with a different data base of chemical compounds or other combination of structural descriptors the "best" statistical model can significantly change. The existence of several QSPR models of comparable statistical quality whenever the models are generated from a large set of descriptors is clearly demonstrated in this study, but this problem was not considered in previous QSPR studies, and usually only the "best" model was reported. In general, for the same database, one can find several combinations of descriptors that provide models with similar statistical indices; owing to the errors in the experimental data that are modeled, small statistical differences between QSPR equations are not significant.

Table 7. Frequency of the structural descriptors, operators, and molecular matrices in the QSPR models reported in Tables 1–6.

Descriptor	t_b	C_p	$\Delta_f G°_{300}$	$\Delta_{vap} H_{300}$	n_D^{25}	ρ	Total
MW	3	6	0	1	1	0	11
$^0\chi$	1	0	0	0	0	0	1
$^1\chi$	1	2	0	0	1	1	5
$^2\chi$	1	2	1	2	0	1	7
$^3\chi_p$	10	0	0	1	10	10	31
$^3\chi_c$	0	0	0	0	1	1	2
Wi	1	1	1	0	1	0	4
HyWi	1	0	1	1	0	0	3
Ho	0	1	2	0	0	0	3
MinSp	0	8	1	0	1	0	10
MaxSp	2	2	1	9	9	7	30
IB	1	4	11	3	2	3	24
U	0	0	0	0	0	0	0
V	6	1	7	11	1	4	30
X	2	1	2	1	3	2	11
Y	1	2	3	1	0	1	8
Matrix							
A	0	0	1	0	0	0	1
D	3	2	10	2	4	8	29
RD	11	6	13	18	11	9	68
RW	0	1	3	4	2	0	10
RRW	0	11	2	2	0	0	15

An inspection of the QSPR models reported in Tables 1–6 reveals that some descriptors appear with a greater frequency, while others are rarely present in these equations. In order to obtain an indication of the importance of each descriptor in modeling the six alkane properties, for each set of QSPR models we have computed the counts of each descriptor type; these frequencies are presented in Table 7, together with the counts for each matrix type. Overall, the size descriptor **MW** appears in 11 equations, mainly for C_p and t_b, indicating that for the remaining four properties the molecular size is incorporated into other structural descriptors. From the group of connectivity indices, all five indices were selected in the QSPR models, with ${}^3\chi_p$ appearing more frequently (31 equations) mainly for t_b, n_D^{25}, and ρ. The connectivity index ${}^3\chi_p$ is followed by ${}^2\chi$ (7 equations) and ${}^1\chi$ (5 equations). This results suggests the importance of χ indices in modeling the six alkane properties, and that subgraph–counting descriptors cannot be substituted with global molecular descriptors computed with graph operators. Maximum matrix eigenvalue descriptors **MaxSp**, were selected in 30 equations, mainly for $\Delta_{vap}H_{300}$, n_D^{25}, and ρ. The related descriptors representing the minimum matrix eigenvalue **MinSp** was chosen with a lower frequency, in only 10 models, principally for C_p. Equally important is the information–theory operator **V** that depends on the magnitude of matrix elements; **V** was selected in 30 equations, mainly for $\Delta_{vap}H_{300}$, $\Delta_fG°_{300}$, and t_b. Two other information–theory descriptors were found important, namely **X** (11 selections) and **Y** (8 selections), in contrast with **U** that was not chosen in the QSPR equations from Tables 1–6.

The Ivanciuc–Balaban indices **IB**, present for 24 times, are mainly useful in modeling $\Delta_fG°_{300}$ and C_p. An unexpected result is the absence of the original Wiener index **Wi**(**D**), although this index was extensively used in developing QSAR and QSPR models. This finding suggests that the structural descriptors from the new generation, computed with graph operators from a large number of molecular matrices, have superior correlational abilities and explore new dimensions of the chemical structural space. Three indices computed with the Wiener operator **Wi** were selected, namely **Wi**(**RD**), **Wi**(**RW**), and **Wi**(**RRW**), in a total of four equations. The related indices computed with the hyper–Wiener operator **HyWi** were selected in 3 equations, and represent descriptors computed from the reciprocal matrices, namely **HyWi**(**RD**) and **HyWi**(**RRW**). Hosoya indices computed from matrices **A** and **RRW** were selected in 3 equations, for the computation of C_p and $\Delta_fG°_{300}$. The counts for the molecular matrices show that reciprocal matrices are more frequently selected than the original matrices, since **RD** was selected 68 times and **D** 29 times, while **RRW** was selected 15 times and **RW** 10 times. It is clear that **D** and **RD** matrices are more often selected than the newly introduced **RW** and **RRW** matrices, but descriptors computed from these latter two matrices give the best models for C_p, $\Delta_fG°_{300}$, and $\Delta_{vap}H_{300}$. This finding demonstrates that although all four matrices are computed from graph distances, the novel **RW** and **RRW** matrices reflect some structural features that are absent from the **D** and **RD** matrices.

4 CONCLUSIONS

Molecular graph indices represent valuable structural descriptors that can be used with success in developing QSPR and QSAR models; in such structure–property studies, graph descriptors can be used in conjunction with other classes of structural descriptors, such as constitutional, geometrical, electrostatic, and quantum descriptors. In addition to the two well–known books by Kier and Hall

[59,60], several more recent books and monographs in encyclopedias have been published on topological molecular descriptors and QSAR [61–67]. In the present study we have investigated two new molecular matrices derived from graph distances, namely the reverse Wiener **RW** and reciprocal reverse Wiener **RRW** matrices. Structural descriptors computed with these two matrices were used to develop QSPR models for six alkane properties: normal boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density. For the generation of the QSPR models we have used a selection of the most used molecular graph descriptors: molecular weight, **MW**; connectivity indices $^0\chi$, $^1\chi$, $^2\chi$, $^3\chi_p$, $^3\chi_c$; Hosoya indices **Ho**(**M**); Wiener indices **Wi**(**M**); hyper–Wiener indices **HyWi**(**M**); minimum matrix eigenvalue **MinSp**(**M**); maximum matrix eigenvalue **MaxSp**(**M**); Ivanciuc–Balaban indices **IB**(**M**); information–theoretic operators **U**(**M**), **V**(**M**), **X**(**M**), and **Y**(**M**). The results obtained show that all six alkane properties can be modeled with multilinear regression equations with three independent variables, but each property requires a particular combination of molecular descriptors. Structural descriptors computed from the novel matrices **RW** and **RRW** are included in the best QSPR models for C_p, $\Delta_f G°_{300}$, and $\Delta_{vap}H_{300}$. Although until recently the distance matrix was the main source of graph invariants, our results demonstrate that **RW** and **RRW** matrices can generate valuable molecular descriptors. We have also demonstrated that the new generation of structural descriptors, computed with the graph operators **MaxSp**, **V**, **IB**, **X**, and **MinSp**, have better correlational abilities compared to the classical Wiener index, which used to be the main descriptor in structure–property studies. Another interesting conclusion is the high importance of subgraph–counting descriptors, represented by the Kier and Hall connectivity indices χ. Especially $^3\chi_p$, representing the weighted contribution of butane–like subgraphs, was found essential in modeling t_b, n_D^{25}, and ρ.

5 REFERENCES

[1] O. Ivanciuc and J. Devillers, Algorithms and Software for the Computation of Topological Indices and Structure–Property Models; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 779–804.

[2] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18.

[3] A. J. Stuper and P. C. Jurs, ADAPT: A Computer System for Automated Data Analysis Using Pattern Recognition Techniques, *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 99–105.

[4] O. Mekenyan, S. Karabunarliev, and D. Bonchev, The Microcomputer OASIS System for Predicting the Biological Activity of Chemical Compounds, *Computers Chem.* **1990**, *14*, 193–200.

[5] O. G. Mekenyan, S. H. Karabunarliev, J. M. Ivanov, and D. N. Dimitrov, A New Development of the OASIS Computer System for Modeling Molecular Properties. *Comput. Chem.* **1994**, *18*, 173–187.

[6] L. Tarko and O. Ivanciuc, QSAR Modeling of the Anticonvulsant Activity of Phenylacetanilides with PRECLAV (PRoperty Evaluation by CLAss Variables), *MATCH* (*Commun. Math. Comput. Chem.*) **2001**, *44*, 201–214.

[7] SciQSAR, SciVision, Inc., 200 Wheeler Road, Burlington, MA 01803, U.S.A., Phone: 1–781–272–4949, Fax: 1–781–272–6868, E–mail: scivision@delphi.com, www http://www.scivision.com.

[8] A. R. Katritzky, L. Mu, V. S. Lobanov, and M. Karelson, Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics, *J. Phys. Chem.* **1996**, *100*, 10400–10407.

[9] T. Ivanciuc and O. Ivanciuc, Quantitative Structure–Retention Relationship Study of Gas Chromatographic

Retention Indices for Halogenated Compounds, *Internet Electron. J. Mol. Des.* **2002**, *1*, 94–107, http://www.biochempress.com.
[10] R. Hiob and M. Karelson, Quantitative Relationship between Rate Constants of the Gas–Phase Homolysis of N–N, O–O and N–O Bonds and Molecular Descriptors, *Internet Electron. J. Mol. Des.* **2002**, *1*, 193–202, http://www.biochempress.com.
[11] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
[12] H. Hosoya, A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
[13] M. Randić, On Characterization of Molecular Branching, *J. Am. Chem. Soc*. **1975**, *97*, 6609–6615.
[14] A. T. Balaban, Highly Discriminating Distance–Based Topological Index, *Chem. Phys. Lett*. **1982**, *89*, 399–404.
[15] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Design of Topological Indices. Part 10. Parameters Based on Electronegativity and Covalent Radius for the Computation of Molecular Graph Descriptors for Heteroatom–Containing Molecules, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 395–401.
[16] O. Ivanciuc, QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs, *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 1412–1422.
[17] O. Ivanciuc, T. Ivanciuc, and D. Cabrol–Bass, QSAR for Dihydrofolate Reductase Inhibitors with Molecular Graph Structural Descriptors, *J. Mol. Struct.* (*Theochem*) **2002**, *582*, 39–51.
[18] I. Rios–Santamarina, R. García–Domenech, J. Cortijo, P. Santamaria, E. J. Morcillo, and J. Gálvez, Natural Compounds with Bronchodilator Activity Selected by Molecular Topology, *Internet Electron. J. Mol. Des.* **2002**, *1*, 70–79, http://www.biochempress.com.
[19] A. A. Toropov and A. P. Toropova, QSAR Modeling of Mutagenicity Based on Graphs of Atomic Orbitals, *Internet Electron. J. Mol. Des.* **2002**, *1*, 108–114, http://www.biochempress.com.
[20] D. J. G. Marino, P. J. Peruzzo, E. A. Castro, and A. A. Toropov, QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants, *Internet Electron. J. Mol. Des.* **2002**, *1*, 115–133, http://www.biochempress.com.
[21] S.–S. Liu, H.–L. Liu, Y.–Y. Shi, and L.–S. Wang, QSAR of Cyclooxygenase–2 (COX–2) Inhibition by 2,3–Diarylcyclopentenones Based on MEDV–13, *Internet Electron. J. Mol. Des.* **2002**, *1*, 310–318, http://www.biochempress.com.
[22] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Optimum Structural Descriptors Derived from the Ivanciuc–Balaban Operator, *Internet Electron. J. Mol. Des.* **2002**, *1*, 319–331, http://www.biochempress.com.
[23] R. García–Domenech, A. Catalá–Gregori, C. Calabuig, G. Antón–Fos, L. del Castillo, and J. Gálvez, Predicting Antifungal Activity: A Computational Screening Using Topological Descriptors, *Internet Electron. J. Mol. Des.* **2002**, *1*, 339–350, http://www.biochempress.com.
[24] E. Estrada, The Structural Interpretation of the Randić Index, *Internet Electron. J. Mol. Des*. **2002**, *1*, 360–366, http://www.biochempress.com.
[25] B. D. Gute, S. C. Basak, D. Mills, and D. M. Hawkins, Tailored Similarity Spaces for the Prediction of Physicochemical Properties, *Internet Electron. J. Mol. Des.* **2002**, *1*, 374–387, http://www.biochempress.com.
[26] I. Lukovits, A. Miličević, S. Nikolić, and N. Trinajstić, On Walk Counts and Complexity of General Graphs, *Internet Electron. J. Mol. Des.* **2002**, *1*, 388–400, http://www.biochempress.com.
[27] C. Cao and H. Yuan, A Modified Distance Matrix to Distinguish *Cis*/*Trans* Isomers of Cycloalkanes, *Internet Electron. J. Mol. Des.* **2002**, *1*, 401–409, http://www.biochempress.com.
[28] O. Ivanciuc and D. J. Klein, Building–Block Computation of Wiener–Type Indices for the Virtual Screening of Combinatorial Libraries, *Croat. Chem. Acta* **2002**, *75*, 577–601.
[29] O. Ivanciuc and D. J. Klein, Computing Wiener–Type Indices for Virtual Combinatorial Libraries Generated from Heteroatom–Containing Building Blocks, *J. Chem. Inf. Comput. Sci*. **2002**, *42*, 8–22.
[30] O. Ivanciuc, Building–Block Computation of the Ivanciuc–Balaban Indices for the Virtual Screening of Combinatorial Libraries, *Internet Electron. J. Mol. Des.* **2002**, *1*, 1–9, http://www.biochempress.com.
[31] O. Ivanciuc, T. Ivanciuc, and M. V. Diudea, Molecular Graph Matrices and Derived Structural Descriptors, *SAR QSAR Environ. Res*. **1997**, *7*, 63–87.
[32] O. Ivanciuc, T.–S. Balaban, and A. T. Balaban, Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices, *J. Math. Chem*. **1993**, *12*, 309–318.
[33] M. V. Diudea, Wiener and Hyper–Wiener Numbers in a Single Matrix, *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 833–836.
[34] O. Ivanciuc, Design of Topological Indices. Part 23. Structural Descriptors Derived from the Distance–Path Matrix of Vertex– and Edge–Weighted Molecular Graphs, *Rev. Roum. Chim*. **2001**, *46*, 543–552.
[35] D. J. Klein and M. Randić, Resistance Distance, *J. Math. Chem*. **1993**, *12*, 81–95.
[36] O. Ivanciuc, QSAR and QSPR Molecular Descriptors Computed from the Resistance Distance and Electrical Conductance Matrices, *ACH – Models Chem*. **2000**, *137*, 607–631.

[37] O. Ivanciuc and A. T. Balaban, Design of Topological Indices. Part 8. Path Matrices and Derived Molecular Graph Invariants, *MATCH* (*Commun. Math. Chem.*) **1994**, *30*, 141–152.
[38] M. V. Diudea, O. Minailiuc, G. Katona, and I. Gutman, Szeged Matrices and Related Numbers, *MATCH* (*Commun. Math. Comput. Chem.*) **1997**, *35*, 129–143.
[39] M. V. Diudea, B. Pârv, and M. I. Topan, Derived Szeged and Cluj Indices, *J. Serb. Chem. Soc*. **1997**, *62*, 267–276.
[40] O. Ivanciuc, Design of Topological Indices. Part 13. Structural Descriptors Computed from the Szeged Molecular Matrices, *Rev. Roum. Chim*. **2000**, *45*, 475–493.
[41] O. Ivanciuc, Design of Topological Indices. Part 15. The Szeged Index of Vertex– and Edge–Weighted Molecular Graphs, *Rev. Roum. Chim*. **2000**, *45*, 895–903.
[42] M. V. Diudea, Cluj Matrix Invariants, *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 300–305.
[43] O. Ivanciuc, Design of Topological Indices. Part 11. Distance–Valency Matrices and Derived Molecular Graph Descriptors, *Rev. Roum. Chim*. **1999**, *44*, 519–528.
[44] O. Ivanciuc, Design of Topological Indices. Part 14. Distance–Valency Matrices and Structural Descriptors for Vertex– and Edge–Weighted Molecular Graphs, *Rev. Roum. Chim*. **2000**, *45*, 587–596.
[45] M Randić, Linear Combinations of Path Numbers as Molecular Descriptors. *New J. Chem*. **1997**, *21*, 945-951.
[46] A. T. Balaban, D. Mills, O. Ivanciuc, and S. C. Basak, Reverse Wiener Indices, *Croat. Chem. Acta* **2000**, *73*, 923–941.
[47] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, The Complementary Distance Matrix, a New Molecular Graph Metric, *ACH – Models Chem*. **2000**, *137*, 57–82.
[48] O. Ivanciuc, Design of Topological Indices. Part 16. Matrix Power Operators for Molecular Graphs, *Rev. Roum. Chim*. **2000**, *45*, 1027–1044.
[49] O. Ivanciuc, Design of Topological Indices. Part 17. The Szeged Operator as a Source of New Structural Descriptors, *Rev. Roum. Chim*. **2000**, *45*, 1105–1114.
[50] O. Ivanciuc, Design of Topological Indices. Part 18. Modeling the Physical Properties of Alkanes with Molecular Graph Descriptors Derived from the Hosoya Operator, *Rev. Roum. Chim*. **2001**, *46*, 129–141.
[51] O. Ivanciuc, Design of Topological Indices. Part 19. Computation of Vertex and Molecular Graph Structural Descriptors with Operators, *Rev. Roum. Chim*. **2001**, *46*, 243–253.
[52] O. Ivanciuc, Design of Topological Indices. Part 22. Structural Descriptors Computed from Truncated Molecular Matrices, *Rev. Roum. Chim*. **2001**, *46*, 411–420.
[53] M. Randić, Novel Molecular Descriptor for Structure-Property Studies, *Chem. Phys. Lett*. **1993**, *211*, 478–483.
[54] D. J. Klein, I. Lukovits, and I. Gutman, On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures, *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 50–52.
[55] A. T. Balaban and T. –S. Balaban, New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances, *J. Math. Chem*. **1991**, *8*, 383–397.
[56] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Evaluation in Quantitative Structure–Property Relationship Models of Structural Descriptors Derived from Information–Theory Operators, *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 631–643.
[57] O. Ivanciuc and A. T. Balaban, Design of Topological Indices. Part 21. Molecular Graph Operators for the Computation of Geometric Structural Descriptors, *Rev. Roum. Chim*. **1999**, *44*, 539–547.
[58] A. A. Gakh, E. G. Gakh, B. G. Sumpter, and D. W. Noid, Neural Network–Graph Theory Approach to the Prediction of the Physical Properties of Organic Compounds, *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 832–839.
[59] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
[60] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, 1986.
[61] J. Devillers and A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, Amsterdam, 1999.
[62] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, 2000.
[63] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley–VCH, New York, 2000.
[64] L. B. Kier and L. H. Hall, *Molecular Structure Description*: *The Electrotopological State*, Academic Press, San Diego, CA, 1999.
[65] A. T. Balaban, QSAR and Computational Methods in Drug Discovery; in: *Encyclopedia of Analytical Chemistry*, Ed. R. A. Meyers, Wiley, Chichester, 2000, vol. 8, pp. 7288–7311.
[66] A. T. Balaban, A personal view about topological indices for QSAR/QSPR, in QSAR/QSPR; in: Studies by Molecular Descriptors, Ed. M. V. Diudea, Huntington, New York, 2000, pp. 9–37.
[67] M. Randić, Topological Indices; in: *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, Wiley, Chichester, 1998, pp. 3018–3032.