QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs

Ovidiu Ivanciuc[†]

Department of Marine Sciences, Texas A & M University at Galveston, Fort Crockett Campus,
5007 Avenue U, Galveston, Texas 77551

Quantitative structure−property relationship (QSPR) and quantitative structure−activity relationship (QSAR) studies use statistical models to compute physical, chemical, or biological properties of a chemical substance from its molecular structure, encoded in a numerical form with the aid of various descriptors. Structural indices derived from molecular graph matrices represent an important group of descriptors used in QSPR and QSAR models; recently, their utilization was extended to molecular similarity and diversity, in database mining and virtual screening of combinatorial libraries. Initially defined from the distance matrix, the Wiener index W was the source of novel graph descriptors derived from recently proposed molecular matrices and of the Wiener graph operator. In this work we present a comparative study of several Wiener-type descriptors computed for vertex- and edge-weighted molecular graphs, corresponding to organic compounds with heteroatoms and multiple bonds. The acute toxicities toward *Tetrahymena pyriformis* of 47 nitrobenzenes are modeled with multilinear regression equations, using as structural descriptors the hydrophobicity (corrected for ionization) and various Wiener-type indices, with better results than a comparative molecular field analysis model.

INTRODUCTION

The success of the statistical models used in QSPR (quantitative structure−property relationships) and QSAR (quantitative structure−activity relationships) approaches can be explained by the possibility of estimating the properties of new chemical compounds without the need to synthesize and test them. The main assumption in QSPR and QSAR studies is that all properties (physical, chemical, and biological) of a chemical substance can be computed from its molecular structure, encoded in a numerical form with the aid of various descriptors, such as constitutional, graph theoretic and topological indices (TIs), and geometrical, electrostatic, quantum-chemical, and grid (field) descriptors. A survey of the QSPR and QSAR equations reveals that molecular graph descriptors and topological indices are used with success to model various properties[1−4] and demonstrates that they are valuable descriptors of chemical structure.[5−14] The interest in developing new graph descriptors for organic compounds revived in recent years, when topological indices found new applications in similarity and diversity assessment, in database mining, and in the virtual screening of combinatorial libraries.[15−17] The use of topological indices in chemistry begins with the Wiener index W;[18] this sensitive structural index was applied with success in modeling several physical properties of alkanes.[18−21] Although the original definition of W was not formulated with graph concepts, we present here a formal mathematical definition translated from the calculation method proposed by Wiener. Consider an acyclic graph $G = G(V,E)$ with the vertex set $V = V(G)$ and the edge set $E = E(G)$, and denote with N_i and N_j the number of vertices situated on both sides of the edge e_{ij}; vertex v_i is counted in N_i while vertex v_j is counted in N_j. For acyclic

graphs the Wiener index $W = W(G)$ of a graph G is defined with the formula

$$W(G) = \sum_{e_{ij} \in E(G)} N_i N_j \qquad (1)$$

where the summation goes over all edges from the edge set $E(G)$, $e_{ij} \in E(G)$. Hosoya gave a graph theoretic formula for the Wiener index and extended the original definition to cyclic compounds with the aid of the distance matrix **D**:[22]

$$W(G) = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{D}(G)]_{ij} \qquad (2)$$

where $[\mathbf{D}(G)]_{ij}$ is the topological distance between vertices v_i, $v_j \in V(G)$. For vertex- and edge-weighted molecular graphs representing organic compounds containing heteroatoms and multiple bonds, the weighting schemes proposed in the literature define some vertex weights that appear on the diagonal of the distance matrix.[10,16,17,23−27] Because eq 2 does not consider such cases, the current definition of the Wiener index uses a slightly modified formula that will be presented in the third section of this paper. The Wiener index is one of the most frequently used graph descriptors in QSAR/QSPR models, and its success stimulated the development of descriptors based on graph distances,[11] novel molecular matrices,[12] and Wiener-type indices.[13] Elements of the distance matrix were used to define distance−degree *VTI* indices;[28] such vector−matrix invariants can generate TIs with a low degeneracy. The idea to use reciprocal distances in computing *VTI* indices was adopted in the definition of the reciprocal distance matrix **RD**.[29−32] Several distance-related matrices were defined in recent years, such as the resistance distance Ω,[33] detour Δ,[34] detour−distance

[†] E-mail: ivanciuc@netscape.net.

WIENER DESCRIPTORS FOR WEIGHTED MOLECULAR GRAPHS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 6, 2000 **1413**

$\Delta-\mathbf{D}$,[34] distance$-$valency **Dval**,[35] distance complement **DC**,[36] complementary distance **CD**,[37] reciprocal complementary distance **RCD**,[37] distance$-$path $\mathbf{D_p}$,[38-40] reciprocal distance$-$path $\mathbf{RD_p}$,[38-40] edge Szeged $\mathbf{Sz_e}$,[41,42] path Szeged $\mathbf{Sz_p}$,[41,42] and reciprocal Szeged $\mathbf{RSz_p}$[41,42] matrices. Graph distances were used to define the Wiener polynomial,[43,44] which is the source of novel topological indices used with success in several QSPR models.[45] The original formula of the Wiener index[18] was extended to afford the computation of the hyper-Wiener index *WW*.[46] Because initially *WW* was defined only for acyclic graphs, an extension was proposed for cyclic graphs[47] and the related distance$-$path matrix $\mathbf{D_p}$ was introduced.[38-40] Recent developments in the area of distance descriptors include the information theory topological indices *U*, *V*, *X*, and *Y*[48] and their extension to other graph matrices[26,49] and three-dimensional molecular matrices,[50] Wiener-related sequences,[51] a multiplicative version of the Wiener index,[52] and the computation of the mean Wiener index for isomeric series of alkanes.[53,54]

In this paper we present a comparative study of a large number of Wiener-type indices computed from various molecular matrices, including several matrices defined in recent years. In the second section of this study we give definitions and examples for the computation of molecular matrices from vertex- and edge-weighted molecular graphs; we introduce some novel extensions of matrices previously defined only for simple (nonweighted) graphs representing alkanes and cycloalkanes. The third section is dedicated to the presentation of the most important Wiener-type descriptors; all definitions can be applied to both simple and weighted molecular graphs. In the fourth section the correlational ability of Wiener-type descriptors is compared in a QSAR study of the acute toxicities toward *Tetrahymena pyriformis* of 47 nitrobenzenes.

MOLECULAR MATRICES FOR WEIGHTED GRAPHS

Molecular graphs are nondirected chemical graphs that represent organic compounds.[1-14] In the graph representation of organic compounds, their geometrical features, such as bond lengths or bond angles, are not taken into account and the chemical bonding of atoms is regarded as being their most important characteristic; each graph vertex corresponds to an atom, while each edge represents a covalent bond of the chemical compound. An organic compound containing heteroatoms and multiple bonds can be represented as a vertex- and edge-weighted molecular graph.[10] A vertex- and edge-weighted (VEW) molecular graph $G = G(V,E, Sy,Bo,Vw,Ew,w)$ consists of a vertex set $V = V(G)$, an edge set $E = E(G)$, a set of chemical symbols for vertices $Sy = Sy(G)$, a set of topological bond orders for edges $Bo = Bo(G)$, a vertex weight set $Vw(w) = Vw(w,G)$, and an edge weight set $Ew(w) = Ew(w,G)$. The elements of the vertex and edge weight sets are computed with the weighting scheme w. Usually, hydrogen atoms are not considered in the molecular graph, and in a VEW graph the weight of a vertex corresponding to a carbon atom is 0, while the weight of an edge corresponding to a carbon$-$carbon single bond is 1. Also, the topological bond order Bo_{ij} of an edge e_{ij} takes the value 1 for single bonds, 2 for double bonds, 3 for triple bonds and 1.5 for aromatic bonds. In a VEW graph G the length of a path p_{ij} between vertices v_i and v_j, $l(p_{ij},w) = l(p_{ij},w,G)$, is equal to the sum of the edge parameters $Ew(w)_{ij}$

for all edges along the path. The topological length of a path p_{ij}, $t(p_{ij}) = t(p_{ij},G)$, in a VEW graph G is equal to the number of edges along the path. In a VEW graph, the distance $d(w)_{ij}$ between a pair of vertices v_i and v_j is equal to the length of the shortest path connecting the two vertices, $d(w)_{ij} = \min(l(p_{ij},w))$. The topological distance td_{ij} between vertices v_i and v_j from a VEW graph G is equal to the minimum topological length of the paths connecting the two vertices, $td_{ij} = \min(t(p_{ij}))$, i.e., the minimum number of bonds between vertices v_i and v_j. In simple (nonweighted) graphs, the distance d_{ij} and topological distance td_{ij} are equal, while in weighted graphs usually this is not true. Several procedures (weighting schemes) for computing vertex and edge weights in molecular graphs were proposed in the literature;[10,16,17,23-27,55,56] VEW graphs are then represented as weighted graph matrices, which are transformed into topological indices using various mathematical operations.

Weighting Schemes. In a weighting scheme w the vertex Vw and edge Ew parameters are computed from a property p_i associated with every vertex v_i from G, $v_i \in V(G)$, and the topological bond order Bo of all edges from the molecular graph. The vertex parameter $Vw(w)_i$ for the vertex v_i is[10,23,27]

$$Vw(w)_i = 1 - p_C/p_i \qquad (3)$$

and the edge parameter $Ew(w)_{ij}$ for the edge between vertices v_i and v_j is[10,23,27]

$$Ew(w)_{ij} = p_C p_C/Bo_{ij}p_ip_j \qquad (4)$$

where p_i is the atomic property of vertex v_i, p_j is the atomic property of vertex v_j, and p_C is the atomic property for carbon atom. Several weighting schemes for molecular graphs were defined by applying eqs 3 and 4 to different atomic properties: *Z*, when *p* is the atomic number *Z*;[23] *A*, when *p* is the atomic mass; *P*, when *p* is the atomic polarizability; *E*, when *p* is the atomic electronegativity; *R*, when *p* is the atomic radius.[10,16,17,27] Similar equations were used to define the *X* and *Y* weighting schemes.[24,55,56]

The *AH* weighting scheme uses the following equation to define the vertex parameter $Vw(AH)_i$ for the non-hydrogen atom i:[10,27]

$$Vw(AH)_i = 1 - A_C/(A_i + NoH_iA_H) = \\ 1 - 12.011/(A_i + 1.0079NoH_i) \quad (5)$$

The edge parameter $Ew(AH)_{ij}$ for the bond between atoms i and j is defined with the equation[10,27]

$$Ew(AH)_{ij} = A_CA_C/Bo_{ij}(A_i + NoH_iA_H)(A_j + NoH_fA_H) = \\ 12.011 \cdot 12.011/Bo_{ij}(A_i + 1.0079NoH_i)(A_j + \\ 1.0079NoH_j) \quad (6)$$

where $A_C = 12.011$ is the atomic mass for carbon, $A_H = 1.0079$ is the atomic mass for hydrogen, NoH_i is the number of hydrogen atoms bonded to the heavy atom i, and NoH_j is the number of hydrogen atoms bonded to the heavy atom j.

In the present study, the Wiener-type descriptors used in the QSAR study are computed using various molecular matrices; because from these matrices several were recently defined and their calculation for vertex- and edge-weighted molecular graphs was not presented previously, we give in this section both their definitions and examples of computation.

Adjacency Matrix. The adjacency matrix $\mathbf{A}(w) = \mathbf{A}(w,G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ real symmetric matrix whose element $[\mathbf{A}(w)]_{ij}$ is defined as[10]

$$[\mathbf{A}(w,G)]_{ij} = \begin{cases} Vw(w)_i & \text{if } i = j \\ Ew(w)_{ij} & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (7)$$

where $Vw(w)_i$ is the weight of the vertex v_i, $Ew(w)_{ij}$ is the weight of the edge e_{ij}, and w is the weighting scheme used to compute the parameters Vw and Ew.

The computation of the adjacency matrix for a weighted molecular graph is presented for phenol **1** represented by the molecular graph **2**; we have to mention that in a
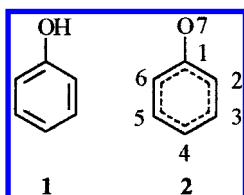
molecular graph aromatic bonds are represented with two lines, the first continuous and the second broken. Using the polarizability weighting scheme P, obtained when the property p is the atomic polarizability, one computes the following adjacency matrix for phenol **2**:

A(P,2)	1	2	3	4	5	6	7
1	0	0.667	0	0	0	0.667	2.195
2	0.667	0	0.667	0	0	0	0
3	0	0.667	0	0.667	0	0	0
4	0	0	0.667	0	0.667	0	0
5	0	0	0	0.667	0	0.667	0
6	0.667	0	0	0	0.667	0	0
7	2.195	0	0	0	0	0	−1.195

Distance Matrix. The distance matrix $\mathbf{D}(w) = \mathbf{D}(w,G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the symmetric square $N \times N$ matrix with real elements defined with the formula:[10]

$$[\mathbf{D}(w,G)]_{ij} = \begin{cases} d(w)_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (8)$$

where $d(w)_{ij}$ is the distance between vertices v_i and v_j, $Vw(w)_i$ is the weight of the vertex v_i, and w is the weighting scheme used to compute the parameters Vw and Ew. The distance matrix of phenol **2** is obtained with the vertex and edge weights from the corresponding adjacency matrix, using an efficient algorithm that transforms $\mathbf{A}$ into $\mathbf{D}$:[9,10,12]

D(P,2)	1	2	3	4	5	6	7
1	0.000	0.667	1.333	2.000	1.333	0.667	2.195
2	0.667	0.000	0.667	1.333	2.000	1.333	2.861
3	1.333	0.667	0.000	0.667	1.333	2.000	3.528
4	2.000	1.333	0.667	0.000	0.667	1.333	4.195
5	1.333	2.000	1.333	0.667	0.000	0.667	3.528
6	0.667	1.333	2.000	1.333	0.667	0.000	2.861
7	2.195	2.861	3.528	4.195	3.528	2.861	−1.195

Reciprocal Distance Matrix. The reciprocal distance matrix $\mathbf{RD}(w) = \mathbf{RD}(w,G)$ of a vertex- and edge-weighted

molecular graph G with N vertices is the square $N \times N$ symmetric matrix with real elements defined with the equation[10,24]

$$[\mathbf{RD}(w,G)]_{ij} = \begin{cases} 1/[\mathbf{D}(w,G)]_{ij} & \text{if } i \neq j \\ [\mathbf{D}(w,G)]_{ii} & \text{if } i = j \end{cases} \quad (9)$$

where $[\mathbf{D}(w)]_{ij}$ is the graph distance between vertices v_i and v_j, $[\mathbf{D}(w)]_{ii}$ is the diagonal element corresponding to vertex v_i, and w is the weighting scheme used to compute the parameters Vw and Ew. The elements of the distance matrix $\mathbf{D}(P,\mathbf{2})$ are used to compute the reciprocal distance matrix of phenol **2**:

RD(P,2)	1	2	3	4	5	6	7
1	0.000	1.500	0.750	0.500	0.750	1.500	0.456
2	1.500	0.000	1.500	0.750	0.500	0.750	0.350
3	0.750	1.500	0.000	1.500	0.750	0.500	0.283
4	0.500	0.750	1.500	0.000	1.500	0.750	0.238
5	0.750	0.500	0.750	1.500	0.000	1.500	0.283
6	1.500	0.750	0.500	0.750	1.500	0.000	0.350
7	0.456	0.350	0.283	0.238	0.283	0.350	−1.195

Distance−Valency Matrix. A generalization of the distance sum **DS** vertex invariant was proposed by introducing the distance−degree *VTI* descriptors.[28] This type of vertex invariants can be obtained by applying the vertex sum **VS** operator to the distance−valency matrices.[35] The valency of the vertex v_i, $\mathbf{val}(w)_i = \mathbf{val}(w,G)_i$, is defined as the sum of the weights $Ew(w)_{ij}$ of all edges e_{ij} incident with vertex v_i:

$$\mathbf{val}(w)_i = \sum_{e_{ij} \in E(G)} Ew(w)_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{A}(w,G)]_{ij} =$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{N} [\mathbf{A}(w,G)]_{ji} \quad (10)$$

where w is the weighting scheme used to compute the Ew parameters. The valency of the vertex v_i may be also computed as the sum of the nondiagonal elements in the row i, or column i, of the adjacency matrix $\mathbf{A}(w) = \mathbf{A}(w,G)$, of a molecular graph G with N vertices. In alkanes and cycloalkanes the degree of a vertex v_i, $\mathbf{deg}_i$, is identical with the valency of that vertex, $\mathbf{val}_i$, while for molecules containing heteroatoms and/or multiple bonds, represented as vertex- and/or edge-weighted molecular graphs, this equality is not true. The distance−valency matrix of a vertex- and edge-weighted graph G with N vertices, $\mathbf{Dval}(p,q,r,w) = \mathbf{Dval}(p,q,r,w,G)$, is a square $N \times N$ matrix, whose entries $[\mathbf{Dval}(p,q,r,w)]_{ij}$ are equal to[11,35]

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

$$(11)$$

where $Vw(w)_i$ is the weight of vertex v_i, $d(w)_{ij}$ is the graph distance between vertices v_i and v_j, and $\mathbf{val}(w)_i$ is the valency of vertex v_i, all computed with the weighting scheme w. For vertex- and edge-weighted molecular graphs the definition of the **Dval** matrix was formulated in analogy with that for the reciprocal distance matrix **RD**, in such a way that **Dval-**

$(-1,0,0,w,G) = \mathbf{RD}(w,G)$. In the particular case when $p = 1$ and $q = r = 0$, the $\mathbf{Dval}(p,q,r,w)$ matrix is identical with the distance matrix $\mathbf{D}(w)$. From the definition of the $\mathbf{Dval}$ matrix one can see that nonsymmetric matrices can be obtained if $q \neq r$; because Wiener-type indices are defined only for symmetric matrices, we will investigate only $\mathbf{Dval}$ matrices with $q = r$, although we have developed several graph descriptors that can be calculated from nonsymmetric matrices.[35] The first example of a distance−valency matrix for phenol is computed with $p = -1$, $q = 1$, and $r = 1$:

Dval(−1,1,1,*P*,2)						
	1	2	3	4	5	6	7
1	0.000	7.056	3.528	2.352	3.528	7.056	3.528
2	7.056	0.000	2.667	1.333	0.889	1.333	1.023
3	3.528	2.667	0.000	2.667	1.333	0.889	0.829
4	2.352	1.333	2.667	0.000	2.667	1.333	0.698
5	3.528	0.889	1.333	2.667	0.000	2.667	0.829
6	7.056	1.333	0.889	1.333	2.667	0.000	1.023
7	3.528	1.023	0.829	0.698	0.829	1.023	−5.753

The distance−valency matrix represents a flexible generalization of the distance and reciprocal distance matrices, which offers the possibility to generate a whole series of related topological indices. A second example of distance−valency matrix, with negative values for all three parameters, is presented:

Dval(−2,−1,−1,*P*,2)						
	1	2	3	4	5	6	7
1	0.000	0.478	0.120	0.053	0.120	0.478	0.027
2	0.478	0.000	1.266	0.316	0.141	0.316	0.042
3	0.120	1.266	0.000	1.266	0.316	0.141	0.027
4	0.053	0.316	1.266	0.000	1.266	0.316	0.019
5	0.120	0.141	0.316	1.266	0.000	1.266	0.027
6	0.478	0.316	0.141	0.316	1.266	0.000	0.042
7	0.027	0.042	0.027	0.019	0.027	0.042	−0.248

Distance−Path Matrix. Diudea[38,39] proposed a definition of the hyper-Wiener index based on the combinatorial distance−path $\mathbf{D_p}$ matrix; this novel matrix was subsequently used to compute several graph descriptors.[12,40] We present here an extension for VEW graphs of the distance−path matrix, which makes possible its computation for any organic compound. Consider the vertex- and edge-weighted graph G with N vertices and its distance matrix $\mathbf{D}(w) = \mathbf{D}(w,G)$ computed with the weighting scheme w. The distance−path matrix of the weighted graph G, $\mathbf{D_p}(w) = \mathbf{D_p}(w,G)$, is the square $N \times N$ symmetric matrix whose element $[\mathbf{D_p}(w)]_{ij}$ is defined by the formula[11]

$$[\mathbf{D_p}(w,G)]_{ij} = [\mathbf{D}(w,G)]_{ij}([\mathbf{D}(w,G)]_{ij} + 1)/2 \quad (12)$$

The elements of the distance matrix $\mathbf{D}(P,2)$ are used to compute the distance−path matrix of phenol **2**:

D_p(*P*,2)						
	1	2	3	4	5	6	7
1	0.000	0.556	1.556	3.000	1.556	0.556	3.505
2	0.556	0.000	0.556	1.556	3.000	1.556	5.524
3	1.556	0.556	0.000	0.556	1.556	3.000	7.987
4	3.000	1.556	0.556	0.000	0.556	1.556	10.894
5	1.556	3.000	1.556	0.556	0.000	0.556	7.987
6	0.556	1.556	3.000	1.556	0.556	0.000	5.524
7	3.505	5.524	7.987	10.894	7.987	5.524	0.116

Reciprocal Distance−Path Matrix. The reciprocal distance−path matrix of a VEW graph G with N vertices, $\mathbf{RD_p}(w) = \mathbf{RD_p}(w,G)$, is the square $N \times N$ symmetric matrix whose element $[\mathbf{RD_p}(w)]_{ij}$ is equal to the reciprocal of the corresponding distance−path matrix element, $1/[\mathbf{D_p}(w)]_{ij}$, for nondiagonal elements, and is equal to $[\mathbf{D_p}(w)]_{ii}$ for the diagonal elements:

$$[\mathbf{RD_p}(w,G)]_{ij} = \begin{cases} 1/[\mathbf{D_p}(w,G)]_{ij} & \text{if } i \neq j \\ [\mathbf{D_p}(w,G)]_{ii} & \text{if } i = j \end{cases} \quad (13)$$

Using the above formula, the elements of the distance−path matrix $\mathbf{D_p}(P,2)$ give the reciprocal distance−path matrix of phenol **2**:

RD_p(*P*,2)						
	1	2	3	4	5	6	7
1	0.000	1.800	0.643	0.333	0.643	1.800	0.285
2	1.800	0.000	1.800	0.643	0.333	0.643	0.181
3	0.643	1.800	0.000	1.800	0.643	0.333	0.125
4	0.333	0.643	1.800	0.000	1.800	0.643	0.092
5	0.643	0.333	0.643	1.800	0.000	1.800	0.125
6	1.800	0.643	0.333	0.643	1.800	0.000	0.181
7	0.285	0.181	0.125	0.092	0.125	0.181	0.116

Distance Complement Matrix. In the distance matrix $\mathbf{D}$ the value of the ijth element is equal to the sum of the edge weights Ew for the shortest path between the two graph vertices v_i and v_j, and the largest contribution to the numerical value of the Wiener index W arises from pairs of distant vertices. In several recently introduced molecular matrices, namely the reciprocal distance $\mathbf{RD}$,[29−32] reciprocal distance−path $\mathbf{RD_p}$,[38−40] distance−valency $\mathbf{Dval}(p,q,r)$ with negative p,[35] complementary distance $\mathbf{CD}$,[37] and the reverse Wiener $\mathbf{RW}$[57] matrices, the value of the matrix elements corresponding to pairs of vertices decreases when the distance between the vertices increases. The same property is observed for the distance complement matrix $\mathbf{DC}$,[36] defined by Randić for alkanes. Structural descriptors computed from the $\mathbf{DC}$ matrix give more weight to shorter graph distances, a property that is consistent with the commonly accepted theory that interactions are inversely proportional to the distance between interacting bodies. We present here an extension of the distance complement matrix to vertex- and edge-weighted molecular graphs, together with the definition of the reciprocal distance complement matrix $\mathbf{RDC}$. The distance complement matrix $\mathbf{DC}(w) = \mathbf{DC}(w,G)$ of a VEW graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are defined as

$$[\mathbf{DC}(w,G)]_{ij} = \begin{cases} N - [\mathbf{D}(w,G)]_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (14)$$

With this formula, the elements of the distance matrix $\mathbf{D}(P,2)$ give the distance complement matrix of phenol **2**:

DC(*P*,2)						
	1	2	3	4	5	6	7
1	0.000	6.333	5.667	5.000	5.667	6.333	4.805
2	6.333	0.000	6.333	5.667	5.000	5.667	4.139
3	5.667	6.333	0.000	6.333	5.667	5.000	3.472
4	5.000	5.667	6.333	0.000	6.333	5.667	2.805
5	5.667	5.000	5.667	6.333	0.000	6.333	3.472
6	6.333	5.667	5.000	5.667	6.333	0.000	4.139
7	4.805	4.139	3.472	2.805	3.472	4.139	−1.195

Reciprocal Distance Complement Matrix. Using a procedure similar to that employed in the computation of reciprocal matrices, one can transform the distance complement matrix into the reciprocal distance complement matrix. The reciprocal distance complement matrix $\mathbf{RDC}(w) = \mathbf{RDC}(w,G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix with real elements defined by the equation

$$[\mathbf{RDC}(w,G)]_{ij} = \begin{cases} 1/[\mathbf{DC}(w,G)]_{ij} & \text{if } i \neq j \\ [\mathbf{DC}(w,G)]_{ii} & \text{if } i = j \end{cases} \quad (15)$$

where $[\mathbf{DC}(w)]_{ij}$ is the graph distance complement between vertices v_i and v_j, $[\mathbf{DC}(w)]_{ii}$ is the diagonal element corresponding to vertex v_i, and w is the weighting scheme used to compute the vertex and edge parameters Vw and Ew. The elements of the distance complement matrix $\mathbf{DC}(P,\mathbf{2})$ are used to compute the reciprocal distance complement matrix of phenol $\mathbf{2}$:

RDC(P,2)							
	1	2	3	4	5	6	7
1	0.000	0.158	0.176	0.200	0.176	0.158	0.208
2	0.158	0.000	0.158	0.176	0.200	0.176	0.242
3	0.176	0.158	0.000	0.158	0.176	0.200	0.288
4	0.200	0.176	0.158	0.000	0.158	0.176	0.356
5	0.176	0.200	0.176	0.158	0.000	0.158	0.288
6	0.158	0.176	0.200	0.176	0.158	0.000	0.242
7	0.208	0.242	0.288	0.356	0.288	0.242	−1.195

Complementary Distance Matrix. The complementary distance matrix $\mathbf{CD}$ is another matrix in which the value of the matrix elements corresponding to pairs of vertices decreases when the distance between the vertices increases. Structural descriptors computed from $\mathbf{CD}$ were extensively tested in QSPR models for six alkane properties;[37] defined initially for alkanes and cycloalkanes, we present here the $\mathbf{CD}$ matrix definition for VEW graphs. The complementary distance matrix $\mathbf{CD}(w) = \mathbf{CD}(w,G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are defined as

$$[\mathbf{CD}(w,G)]_{ij} = \begin{cases} d(w)_{max} + d(w)_{min} - [\mathbf{D}(w,G)]_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (16)$$

where $[\mathbf{D}(w)]_{ij}$ is the ijth element of the distance matrix $\mathbf{D}(w)$ which is equal to the graph distance between vertices v_i and v_j, $d(w)_{max}$ is the maximum distance between two distinct graph vertices (the graph diameter), and $d(w)_{min}$ is the minimum distance between two distinct graph vertices (equal to 1 for alkanes and cycloalkanes):

$$d(w)_{max} = \max\{d(w)_{ij}, v_i, v_j \in V(G), v_i \neq v_j\} \quad (17)$$

$$d(w)_{min} = \min\{d(w)_{ij}, v_i, v_j \in V(G), v_i \neq v_j\} \quad (18)$$

Using formula 16, the distance matrix $\mathbf{D}(P,\mathbf{2})$ gives the complementary distance matrix of phenol $\mathbf{2}$:

CD(P,2)							
	1	2	3	4	5	6	7
1	0.000	4.195	3.528	2.861	3.528	4.195	2.667
2	4.195	0.000	4.195	3.528	2.861	3.528	2.000
3	3.528	4.195	0.000	4.195	3.528	2.861	1.333
4	2.861	3.528	4.195	0.000	4.195	3.528	0.667
5	3.528	2.861	3.528	4.195	0.000	4.195	1.333
6	4.195	3.528	2.861	3.528	4.195	0.000	2.000
7	2.667	2.000	1.333	0.667	1.333	2.000	−1.195

Reciprocal Complementary Distance Matrix. The reciprocal complementary distance matrix $\mathbf{RCD}(w) = \mathbf{RCD}(w,G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix defined by the equation

$$[\mathbf{RCD}(w,G)]_{ij} = \begin{cases} 1/[\mathbf{CD}(w,G)]_{ij} & \text{if } i \neq j \\ [\mathbf{CD}(w,G)]_{ii} & \text{if } i = j \end{cases} \quad (19)$$

The elements of the complementary distance matrix $\mathbf{CD}(P,\mathbf{2})$ are used to compute the reciprocal complementary distance matrix of phenol $\mathbf{2}$:

RCD(P,2)							
	1	2	3	4	5	6	7
1	0.000	0.238	0.283	0.350	0.283	0.238	0.375
2	0.238	0.000	0.238	0.283	0.350	0.283	0.500
3	0.283	0.238	0.000	0.238	0.283	0.350	0.750
4	0.350	0.283	0.238	0.000	0.238	0.283	1.500
5	0.283	0.350	0.283	0.238	0.000	0.238	0.750
6	0.238	0.283	0.350	0.283	0.238	0.000	0.500
7	0.375	0.500	0.750	1.500	0.750	0.500	−1.195

WIENER OPERATOR AND WIENER DESCRIPTORS

As already emphasized in the Introduction, Wiener defined W only for alkanes, and Hosoya extended the definition to cycloalkanes with the aid of the distance matrix $\mathbf{D}$. As one can see from eq 2, the formula proposed by Hosoya does not consider the diagonal elements $[\mathbf{D}]_{ii}$, thus neglecting the vertex weight contributions Vw in the case of weighted molecular graphs. The actual definition of the Wiener index, presented below, can be applied to all vertex- and edge-weighted molecular graphs; using appropriate weighting schemes, one can compute the W index for any organic compound.

Wiener Index. The Wiener index $W(w) = W(w,G)$ of a vertex- and edge-weighted graph G with N vertices is

$$W(w,G) = \sum_{i=1}^{N} \sum_{j=i}^{N} [\mathbf{D}(w,G)]_{ij} \quad (20)$$

where the distance matrix $\mathbf{D}(w)$ is computed with the weighting scheme w.

Wiener Operator. Molecular graph operators were recently proposed as an extension of topological indices and graph invariants;[35] a graph operator applies a mathematical equation to compute a whole class of related molecular graph descriptors, using different molecular matrices and various weighting schemes (sets of parameters for atoms and bonds). In this way, molecular graph operators introduce a systematization of topological indices and graph invariants by assembling together all descriptors computed with the same mathematical formula or algorithm, but with different parameters or molecular matrices. Using equations similar

WIENER DESCRIPTORS FOR WEIGHTED MOLECULAR GRAPHS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 6, 2000 **1417**

to that proposed by Hosoya, several Wiener-type indices were defined from various molecular matrices;[13] all Wiener-type indices can be conveniently defined and denoted using the Wiener operator. The Wiener operator $\mathbf{Wi(M},w) = \mathbf{Wi(M},w,G)$ of a vertex- and edge-weighted molecular graph G with N vertices is

$$\mathbf{Wi(M},w,G) = \sum_{i=1}^{N} \sum_{j=i}^{N} [\mathbf{M}(w,G)]_{ij} \qquad (21)$$

where $\mathbf{M}(w) = \mathbf{M}(w,G)$ represents the molecular matrix of G computed with the weighting scheme w. Special cases of this operator are found for certain matrices: for the distance matrix $\mathbf{D}$ the operator is identical with the Wiener index W; from the reciprocal distance matrix $\mathbf{RD}$ the operator gives the *RDSUM* index;[31] for the resistance distance matrix Ω the operator yields the Kirchhoff index Kf;[33] from the distance−path matrix $\mathbf{D_p}$ the operator gives the hyper-Wiener index WW;[38−40] the Szeged index Sz is obtained from the edge Szeged matrix $\mathbf{Sz_e}$.[11,41,42] As examples of descriptors computed with $\mathbf{Wi}$, we give here the values for the Wiener indices obtained from the molecular matrices of phenol **2**: $\mathbf{Wi(D},P,\mathbf{2}) = 35.973$, $\mathbf{Wi(RD},P,\mathbf{2}) = 15.766$, $\mathbf{Wi(Dval}(-1,1,1),P,\mathbf{2}) = 43.474$, $\mathbf{Wi(Dval}(-2,-1,-1),P,\mathbf{2}) = 7.795$, $\mathbf{Wi(D_p},P,\mathbf{2}) = 63.203$, $\mathbf{Wi(RD_p},P,\mathbf{2}) = 16.763$, $\mathbf{Wi(DC},P,\mathbf{2}) = 108.638$, $\mathbf{Wi(RDC},P,\mathbf{2}) = 3.035$, $\mathbf{Wi(CD},P,\mathbf{2}) = 63.723$, and $\mathbf{Wi(RCD},P,\mathbf{2}) = 7.360$.

Szeged Index. Using an equation similar to that proposed by Wiener, Gutman introduced a new index, the Szeged index Sz,[58,59] that can be considered another possible extension of eq 1 to cycle-containing molecular graphs (other possibilities were investigated by Hosoya[22] and by Klein and Randić).[33] The Szeged index is defined in the following way. Let e_{ij} be an edge of the molecular graph G, connecting the vertices v_i and v_j from G, v_i, $v_j \in V(G)$. Let n_i be the number of vertices v_k of the molecular graph G, having the property $td_{ki} < td_{kj}$, and let n_j be the number of vertices v_k of the molecular graph G, having the property $td_{kj} < td_{ki}$. When a vertex v_k is situated at the same distance from vertices v_i and v_j, i.e., $td_{ki} = td_{kj}$, the vertex is not counted in n_i or in n_j. For the two vertices that form the edge e_{ij}, n_i gives the number of vertices closer to vertex v_i and n_j gives the number of vertices closer to vertex v_j. The Szeged index of the molecular graph G is

$$Sz(G) = \sum_{e_{ij} \in E(G)} n_i n_j \qquad (22)$$

where the summation goes over all edges e_{ij} from the edge set $E(G)$, $e_{ij} \in E(G)$. In acyclic graphs $N_i = n_i$ and $N_j = n_j$ and the Wiener and Szeged indices coincide, while for cyclic graphs their values are usually different. The Szeged index was initially defined for simple graphs representing alkanes and cycloalkanes; its extension for weighted graphs uses vertex and edge weights that can be computed with various weighting schemes. Consider the vertex- and edge-weighted graph G with N vertices whose vertex and edge parameters $Vw(w)$ and $Ew(w)$ are computed with the weighting scheme w. The Szeged index $Sz(w) = Sz(w,G)$ of the weighted

molecular graph G is defined by the formula[11]

$$Sz(w,G) = \sum_{i=1}^{N} Vw(w)_i + \sum_{e_{ij} \in E(G)} Ew(w)_{ij} n_i n_j \qquad (23)$$

where the second summation goes over all edges e_{ij} from the edge set $E(G)$, $e_{ij} \in E(G)$, and the numbers n_i and n_j are defined by the following equations:

$$n_i = |\{v_k : v_i, v_j, v_k \in V(G), e_{ij} \in E(G), td_{ki} < td_{kj}\}| \qquad (24)$$

$$n_j = |\{v_k : v_i, v_j, v_k \in V(G), e_{ij} \in E(G), td_{kj} < td_{ki}\}| \qquad (25)$$

where td_{ij} represents the topological distance between vertices v_i and v_j, i.e., the minimum number of bonds between vertices v_i and v_j. Formula (23) conserves the equality between the Wiener W and Szeged Sz indices for acyclic weighted graphs; for cyclic graphs the two indices have different values and only accidentally or in special cases do they show identical values.

The computation of the Szeged index $Sz(P,\mathbf{2})$ for phenol **2** uses the vertex and edge weights Vw and Ew taken from the adjacency matrix $\mathbf{A}(P,\mathbf{2})$ presented before, and the numbers n_i and n_j taken from the nonsymmetric edge Szeged matrix $\mathbf{Sz_{ue}}$[11,41,42] of the nonweighted graph **2**, equivalent to the molecular graph of methylcyclohexane. Using the algorithm proposed by Diudea it is straightforward to compute the nonweighted matrix $\mathbf{Sz_{ue}}(\mathbf{2})$, and with these values for n_i, n_j, Vw, and Ew, one obtains the Szeged index $Sz(P,\mathbf{2}) = 59.973$.

$\mathbf{Sz_{ue}(2)}$							
	1	2	3	4	5	6	7
1	0	4	0	0	0	4	6
2	3	0	4	0	0	0	0
3	0	3	0	4	0	0	0
4	0	0	3	0	3	0	0
5	0	0	0	4	0	3	0
6	3	0	0	0	4	0	0
7	1	0	0	0	0	0	0

QSAR EVALUATION OF WIENER-TYPE DESCRIPTORS

Although a large number of Wiener-type descriptors were defined and used in various QSAR/QSPR models, this is the first attempt to compare their correlational ability for a QSAR data set comprising weighted molecular graphs. Previous comparative studies were limited to a smaller set of descriptors and were conducted for nonweighted molecular graphs representing alkanes and cycloalkanes. In this investigation, the acute toxicities toward *Tetrahymena pyriformis* of 47 nitrobenzenes are modeled with multilinear regression (MLR) equations, using as structural descriptors the hydrophobicity (corrected for ionization) and various Wiener-type indices.

Data. The structures of the 47 nitrobenzenes, the calculated octanol−water partition coefficient corrected for ionization at pH 7.35 (log D_{ow}), and associated acute toxicities toward the aquatic ciliate *Tetrahymena pyriformis* were taken from the literature[60,61] and are given in Table 1; the data set comprises nitrobenzene, 12 ortho-, 14 meta-, and 20 para-substituted compounds. Toxicity is expressed as log $1/IC_{50}$, the negative logarithm of the 50% growth inhibition con-

Table 1. The 47 Nitrobenzenes Used in the Structure−Toxicity Model, Calculated Octanol−Water Partition Coefficient Corrected for Ionization at pH 7.35 (log D_{ow}), Their Experimental Acute Toxicities toward the Aquatic Ciliate *Tetrahymena pyriformis* (log $1/IC_{50}$, the Negative Logarithm of the 50% Growth Inhibition Concentration, in Millimoles per Liter), Calibration Residuals, and Prediction Residuals

substituent	log D_{ow}	log $1/IC_{50}$		
		exp^a	res_{cal}^b	res_{LOO}^c
H	1.885	0.350	−0.033	−0.036
2-Br	2.468	0.863	0.019	0.021
2-C$_6$H$_5$	3.773	1.301	−0.132	−0.207
2-CH$_2$OH	0.567	−0.155	−0.049	−0.051
2-CH$_3$	2.254	0.479	−0.160	−0.168
2-CHO	1.502	0.174	−0.300	−0.314
2-Cl	2.318	0.676	−0.085	−0.091
2-CN	1.318	1.076	0.196	0.289
2-CONH$_2$	−0.010	−0.721	−0.565	−0.611
2-COOH	−3.572	−1.637	0.021	0.032
2-NH$_2$	1.776	0.077	−0.443	−0.461
2-NO$_2$	1.628	1.252	0.248	0.310
2-OH	1.483	0.770	0.287	0.302
3-Br	2.748	1.215	0.256	0.280
3-C$_6$H$_5$	3.773	1.569	0.251	0.362
3-CH$_2$OH	0.847	−0.220	−0.134	−0.142
3-CH$_3$	2.534	0.572	−0.113	−0.119
3-CHO	1.502	0.140	−0.148	−0.152
3-Cl	2.598	0.836	0.005	0.006
3-CN	1.318	0.451	−0.053	−0.058
3-COCH$_3$	1.486	0.317	−0.028	−0.030
3-CONH$_2$	0.830	−0.193	−0.268	−0.278
3-COOH	−2.042	−1.089	0.027	0.034
3-NH$_2$	1.306	0.026	−0.134	−0.139
3-NO$_2$	1.628	0.762	−0.121	−0.142
3-OCH$_3$	2.104	0.670	0.081	0.085
3-OH	1.816	0.506	0.103	0.107
4-Br	2.748	0.461	−0.290	−0.314
4-C$_2$H$_5$	3.063	0.804	0.102	0.114
4-CH$_2$Cl	2.447	1.180	0.658	0.704
4-CH$_2$CN	1.307	0.132	−0.181	−0.188
4-CH$_2$OH	0.847	0.101	0.249	0.271
4-CH$_3$	2.534	0.796	0.278	0.303
4-CH=NOH	1.868	0.678	0.078	0.084
4-CHO	1.502	0.203	0.023	0.024
4-Cl	2.598	0.559	−0.086	−0.093
4-CN	1.318	0.569	0.233	0.243
4-CONH$_2$	0.830	0.179	0.184	0.192
4-COOC$_2$H$_5$	2.545	0.710	−0.333	−0.399
4-COOCH$_3$	2.016	0.398	−0.317	−0.350
4-COOH	−2.082	−0.862	0.340	0.427
4-F	2.028	0.253	−0.204	−0.219
4-NHC$_6$H$_5$	4.011	1.886	0.208	0.283
4-NO$_2$	1.628	1.301	0.485	0.566
4-OC$_2$H$_5$	2.633	0.829	−0.021	−0.023
4-OC$_4$H$_9$	3.691	1.420	−0.135	−0.177
4-OCH$_3$	2.104	0.544	0.001	0.001

^{a} Experimental log $1/IC_{50}$ values taken from refs 60 and 61. ^{b} Calibration residuals (log $1/IC_{50,exp}$ − log $1/IC_{50,cal}$) computed with the first equation from Table 3. ^{c} Prediction (leave-one-out cross-validation) residuals (log $1/IC_{50,exp}$ − log $1/IC_{50,pre}$) computed with the descriptors of the first equation from Table 3.

centration, in millimoles per liter, of *Tetrahymena pyriformis* in a static assay with 48-h exposure time.

This data set was previously investigated with the comparative molecular field analysis (CoMFA) model,[62,63] when the steric and electrostatic field descriptors of the molecules were evaluated in 693 points from a grid large enough to include all aligned nitrobenzenes, using as probe an sp^3 carbon cation (charge +1).[60] It is interesting to note that all attempts to derive QSAR models based solely on CoMFA

steric and electrostatic descriptors failed for the complete set of 47 nitrobenzenes as well as for the subsets of ortho-, meta-, and para-substituted congeners, following that these descriptors are not sufficient in structure−toxicity models. Consequently, log D_{ow} was used in addition to both CoMFA fields in the subsequent analyses employing partial least squares (PLS) equations. The three-dimensional (3D) QSAR PLS model obtained with log D_{ow} and the CoMFA steric and electrostatic fields gave fairly good results, namely r^2 = 0.788, r_{cv}^2 = 0.761, and s = 0.314; the usual leave-one-out cross-validation technique implemented in CoMFA was used for the computation of r_{cv}. These results are close but inferior to those obtained in a multilinear regression model based on log D_{ow}, AM1 lowest unoccupied molecular orbital (LUMO) energy, and the absolute value of the change in the AM1 charge on the nitro oxygen upon substitution, with r^2 = 0.858, r_{cv}^2 = 0.826, and s = 0.255.[61]

Structural Descriptors. The structure−toxicity models are developed from a pool of 87 structural descriptors consisting of (1) molecular weight, **MW**; (2) octanol−water partition coefficient corrected for ionization at pH 7.35, log D_{ow};[60,61] (3) Kier and Hall's valence connectivity indices $^0\chi^v$, $^1\chi^v$, $^2\chi^v$, $^3\chi_p{}^v$, $^3\chi_c{}^v$;[1,2] (4) 80 Wiener-type indices. With the exception of the Szeged index Sz, which is computed with eq 23, all Wiener-type indices are computed with the Wiener operator **Wi** applied to the following matrices: distance **D**, reciprocal distance **RD**, distance−path **D$_p$**, reciprocal distance− path **RD$_p$**, distance complement **DC**, reciprocal distance complement **RDC**, complementary distance **CD**, reciprocal complementary distance **RCD**, and distance−valency **Dval**-(1,1,1), **Dval**(1,−1,−1), **Dval**(−1,1,1), **Dval**(−1,−1,−1), **Dval**(−2,0,0), **Dval**(−2,1,1), and **Dval**(−2,−1,−1) matrices. The Wiener-type indices are computed with the five weighting schemes presented in detail under Molecular Matrices for Weighted Graphs, namely P, E, R, A, and AH.[27]

QSAR Model. All studies that develop QSAR models from a large set of computed descriptors use a wide range of algorithms for selecting significant descriptors. Because the exhaustive test of all MLR equations requires too-large computational resources, we have used a heuristic method for descriptor selection. This heuristic algorithm starts from the set of structural descriptors and develops structure−toxicity QSAR 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) Biparametric regression equations are computed with all possible pairs of descriptors selected in step 1 that are not significantly correlated. Two descriptors are considered to be not significantly correlated if their intercorrelation coefficient r_{ij} is lower than a threshold, $|r_{ij}| < 0.8$. The most significant 200 pairs of molecular descriptors were used in the third step.

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

(4) The most significant 200 MLR models containing n + 1 descriptors are selected.

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

WIENER DESCRIPTORS FOR WEIGHTED MOLECULAR GRAPHS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 6, 2000 **1419**

Table 2. Structural Descriptors and Statistical Indices for the Best 15 MLR Equations with Three Independent Variables that Model the Acute Toxicity for *Tetrahymena pyriformis* of the 47 Nitrobenzenes[a]

SD$_2$	SD$_3$	r	s	F
Wi(RCD,R**)**	**Wi(Dval**(1,1,1),P**)**	0.9166	0.279	75.3
Wi(RD$_p$,E**)**	**Wi(Dval**(1,1,1),E**)**	0.9146	0.282	73.3
Wi(RD,P**)**	**Wi(RD$_p$,**E**)**	0.9130	0.285	71.8
Wi(RD$_p$,E**)**	**Wi(CD,**A**)**	0.9126	0.286	71.4
Wi(RD$_p$,E**)**	**Wi(CD,**E**)**	0.9125	0.286	71.3
$^3\chi_c{}^v$	**Wi(Dval**(−2,0,0),A**)**	0.9125	0.286	71.3
Wi(RCD,R**)**	**Wi(Dval**(1,−1,−1),A**)**	0.9122	0.286	71.0
Wi(RD$_p$,E**)**	**Wi(CD,**R**)**	0.9121	0.286	70.9
Wi(RCD,R**)**	**Wi(Dval**(−2,0,0),P**)**	0.9120	0.286	70.9
$^3\chi_c{}^v$	**Wi(RD$_p$,**A**)**	0.9120	0.286	70.9
Wi(D$_p$,P**)**	**Wi(RCD,**R**)**	0.9117	0.287	70.6
$^2\chi^v$	**Wi(RD,**A**)**	0.9113	0.288	70.2
Wi(RCD,R**)**	**Wi(Dval**(1,−1,−1),AH**)**	0.9113	0.288	70.2
Wi(RD$_p$,E**)**	**Wi(RCD,**R**)**	0.9112	0.288	70.1
Wi(RD$_p$,P**)**	**Wi(RD$_p$,**E**)**	0.9111	0.288	70.0

[a] The MLR equations have the general form log $1/IC_{50} = a_0 + a_1SD_1 + a_2SD_2 + a_3SD_3$; because for these QSAR models SD_1 is always log D_{ow}, this descriptor was not included in the table.

RESULTS

In Table 2 we present the structural descriptors and statistical indices (r, correlation coefficient; s, standard deviation; and F, Fisher test) for the best 15 MLR equations with three independent variables that model the toxicity toward *Tetrahymena pyriformis* of the 47 nitrobenzenes; the equations are arranged in the decreasing order of the F statistics. We have to mention that, in agreement with previous QSAR studies of this data set, the hydrophobicity parameter log D_{ow} is necessary to obtain good correlations and all MLR models from Table 2 contain this descriptor; consequently, it was not explicitly indicated in a separate column. 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 QSAR equations are not significant. The statistical indices of the 15 QSAR models presented in Table 2 are very similar, indicating that all these equations are of comparable statistical value, with a correlation coefficient between 0.9166 and 0.9111 and a standard deviation between 0.279 and 0.288.

The best structure−toxicity model from Table 2 is obtained with log D_{ow} and two Wiener indices computed from the reciprocal complementary distance matrix **RCD** and the distance−valency matrix **Dval**(1,1,1), namely **Wi(RCD,**R**)** and **Wi(Dval**(1,1,1),P**)**; the statistical indices of this equation ($r = 0.9166$, $s = 0.279$, $F = 75.3$) are higher than those obtained with the CoMFA field descriptors,[60] but slightly lower than those from the MLR model with quantum indices.[61] However, of the three types of descriptors, the graph indices used in the present study are the most convenient in term of computational expenses, they do not require the optimization of the three-dimensional molecular structure, and using properly defined weighting schemes, they can be computed for any organic compound.

An analysis of the presence of different molecular matrices as a source of Wiener indices in the QSAR models from Table 2 reveals several interesting tendencies: **Wi** descriptors derived from the reciprocal distance−path **RD$_p$**, distance−

valency **Dval**, and reciprocal complementary distance **RCD** matrices were selected with a greater frequency, while those computed from the complementary distance **CD**, reciprocal distance **RD**, and distance−path **D$_p$** matrices were found less important for the structure−toxicity model: **RD$_p$** nine times; **RCD** six times; **Dval** six times; **CD** three times; **RD** twice; **D$_p$** once. The Szeged index and Wiener indices computed from the distance **D**, distance complement **DC**, and reciprocal distance complement **RDC** matrices were not included in the best 15 QSAR models. Until recently, the Wiener index was mainly computed from the distance matrix, but the results from Table 2 clearly show that **Wi** descriptors computed from the novel molecular matrices are more suitable for developing relevant QSAR models. Using several atomic properties, the five weighting schemes tested in this QSAR study offer the atom and bond parameters for the computation of the Wiener indices. The QSAR models from Table 2 show that Wiener-type indices computed with atomic electronegativity parameters E were selected with a greater frequency: E nine times; R seven times; P five times; A five times; AH once.

The addition of the fourth descriptor significantly improves the QSAR model, as can be seen from the statistical indices of the best 15 MLR equations with four independent variables, presented in Table 3. Similarly with the QSAR models from Table 2, all equations with four descriptors contain the hydrophobicity parameter log D_{ow}, and therefore, we did not add it in a separate column. An inspection of the descriptors selected reveals that 13 out of the 15 structure−toxicity models are derived from the best equation with three descriptors and contain the triplet log D_{ow}, **Wi(RCD,**R**)**, and **Wi(Dval**(1,1,1),P**)**. The best equation with four descriptors has better statistics than both the CoMFA model[60] and the QSAR with quantum descriptors:[61]

$$\log 1/IC_{50} = -0.829(\pm 0.361) +$$
$$0.428(\pm 0.186)\log D_{ow} +$$
$$0.0533(\pm 0.0232)\textbf{Wi(RCD,}R\textbf{)} +$$
$$0.000444(\pm 0.000193)\textbf{Wi(Dval}(1,1,1),P\textbf{)} -$$
$$0.01226(\pm 0.00534)\textbf{Wi(Dval}(-1,1,1),A\textbf{)}$$

$$n = 47 \quad r = 0.9356 \quad s = 0.250 \quad F = 73.7$$

The predictive quality of the above QSAR model is evaluated with the leave-one-out (LOO) cross-validation procedure; the cross-validation statistics are good, very close to the statistical indices obtained in calibration, namely $r_{LOO} = 0.9180$ and $s_{LOO} = 0.27$. Again, the model that uses Wiener-type indices gives better statistics than the previous CoMFA and quantum index QSAR. However, we do not imply that graph indices can replace other descriptors and QSAR/QSPR models should be developed only with graph descriptors. In fact, each class of descriptors emphasizes particular aspects of the molecular structure, and a comprehensive exploration of the structural space must contain descriptors from all classes. The calibration and leave-one-out cross-validation residuals for the 47 nitrobenzenes were computed with the first QSAR equation from Table 3. In the CoMFA study of this data set, statistical evidence that steric and electrostatic descriptors fail to model all nitrobenzenes in a single equation was presented, and separate data sets were formed with the ortho, meta, para, and ortho plus

Table 3. Structural Descriptors and Statistical Indices for the Best 15 MLR Equations with Four Independent Variables that Model the Acute Toxicity for *Tetrahymena pyriformis* of the 47 Nitrobenzenes[a]

SD_2	SD_3	SD_4	r	s	F
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−1,1,1),*A*)	0.9356	0.250	73.7
Wi(RD$_p$,*P*)	**Wi(RD$_p$**,*E*)	**Wi(Dval**(−2,−1,−1),*P*)	0.9351	0.250	73.2
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−1,1,1),*AH*)	0.9344	0.252	72.2
Wi(RD,*P*)	**Wi(RD$_p$**,*E*)	**Wi(Dval**(−2,−1,−1),*P*)	0.9340	0.253	71.7
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−1,1,1),*R*)	0.9333	0.254	70.9
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−1,1,1),*E*)	0.9323	0.256	69.7
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(1,1,1),*A*)	0.9320	0.256	69.4
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−2,1,1),*A*)	0.9318	0.256	69.2
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(1,1,1),*AH*)	0.9314	0.257	68.8
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−2,1,1),*E*)	0.9311	0.258	68.5
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−2,1,1),*R*)	0.9306	0.259	67.8
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(Dval**(−2,1,1),*AH*)	0.9303	0.259	67.6
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(D**,*AH*)	0.9301	0.260	67.4
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(D**,*A*)	0.9299	0.260	67.1
Wi(RCD,*R*)	**Wi(Dval**(1,1,1),*P*)	**Wi(DC**,*E*)	0.9299	0.260	67.1

[a] The MLR equations have the general form $\log 1/IC_{50} = a_0 + a_1SD_1 + a_2SD_2 + a_3SD_3 + a_4SD_4$; because for these QSAR models SD_1 is always $\log D_{ow}$, this descriptor was not included in the table.

meta groups of nitrobenzenes. The residuals from Table 1 do not show systematic errors for the ortho, meta, or para groups, indicating that all 47 nitrobenzenes can be included in the same QSAR model. Because the absolute values of all calibration residuals (res$_{cal}$) and prediction (leave-one-out cross-validation) residuals (res$_{LOO}$) are smaller than 3 times the corresponding standard deviation, this structure−toxicity model does not present statistical outliers. The largest residuals are obtained for the following four substituents: 2-CONH$_2$, 2-NH$_2$, 4-CH$_2$Cl, and 4-NO$_2$. Although they are not statistical outliers, their residuals are higher than for the remaining set of nitrobenzenes. It is important to note that for the above four chemicals both the calibration and prediction residuals are high, indicating with a high probability that the errors are not coming from the QSAR equation. We can identify several causes for these large residuals: (1) there are large errors in the experimental data; (2) there are different toxicity mechanisms than for the remaining nitrobenzenes; (3) the structural descriptors do not represent in an adequate way the structure of these compounds. For three substituents we can make a comparison with isomers presenting the same substituent in another position. The computed toxicity of the 2-CONH$_2$ isomer is significantly higher than the experimental one (res$_{cal}$ = −0.565 and res$_{LOO}$ = −0.611), while for 3-CONH$_2$ (res$_{cal}$ = −0.268 and res$_{LOO}$ = −0.278) and 4-CONH$_2$ (res$_{cal}$ = −0.184 and res$_{LOO}$ = 0.192) the errors are much lower. A similar situation is encountered for 2-NH$_2$ (res$_{cal}$ = −0.443 and res$_{LOO}$ = −0.461), with lower errors for 3-NH$_2$ (res$_{cal}$ = −0.134 and res$_{LOO}$ = −0.139). The computed toxicity of the 4-NO$_2$ isomer is lower than the experimental one (res$_{cal}$ = 0.485 and res$_{LOO}$ = 0.566), but the errors for the ortho and meta isomers are small: 2-NO$_2$ with res$_{cal}$ = 0.248 and res$_{LOO}$ = 0.310; 3-NO$_2$ with res$_{cal}$ = −0.121 and res$_{LOO}$ = −0.142. For the 4-CH$_2$Cl isomer (res$_{cal}$ = 0.658 and res$_{LOO}$ = 0.704) we cannot make a similar comparison, because the ortho and meta isomers are missing; its computed toxicity is significantly lower than the experimental one, and this can indicate a higher reactivity for this compound. The above analysis demonstrates that the larger residuals observed for certain substituents are not a result of the poor descriptors, because for the same substituents situated in other positions the residuals are small.

CONCLUDING REMARKS

Structural indices derived from molecular graph matrices represent an important group of descriptors used in QSPR and QSAR models; recently, their utilization was extended to molecular similarity and diversity, in database mining and virtual screening of combinatorial libraries. Using the Wiener index *W* as a template, several Wiener-type indices were proposed in the literature; however, their use in QSPR/QSAR models was restricted to only alkanes and cycloalkanes because their definition was valid only for simple, non-weighted, molecular graphs. To make available such indices for QSPR/QSAR testing and evaluation, we have presented a comprehensive list of molecular matrices that can generate Wiener-type indices. All definitions include the case of vertex- and edge-weighted molecular graphs representing organic molecules containing heteroatoms and/or multiple bonds. Using various weighting schemes, these molecular matrices can be employed in the Wiener graph operator **Wi** to generate Wiener-type indices. We have also presented an extension of the Szeged index *Sz* to weighted graphs, together with a detailed example for its computation. The correlational ability of Wiener-type descriptors was compared in a QSAR study of the acute toxicities toward *Tetrahymena pyriformis* of 47 nitrobenzenes. Similar to previous investigations of the same data set,[60,61] we found that the octanol−water partition coefficient corrected for ionization at pH 7.35, log D_{ow}, is an essential descriptor that cannot be replaced by any combination of Wiener-type indices. The best structure−toxicity model was obtained with log D_{ow} and three Wiener-type indices; the statistical indices of this model are higher than those obtained with the CoMFA field descriptors[60] and quantum indices.[61] We have to point out that when a data set is investigated with a large number of different descriptors it is possible to find several combinations of descriptors that provide QSAR models with similar statistics; owing to the errors in the experimental data, small statistical differences between QSAR equations are not significant. The most important thing is not to select a "best" QSAR equation between statistically similar models, but to compare different approaches to the same problem, to identify their convergent and divergent conclusions; for example, if a molecule is identified as outlier in two different QSAR models, this is a

Wiener Descriptors for Weighted Molecular Graphs

J. Chem. Inf. Comput. Sci., Vol. 40, No. 6, 2000 **1421**

strong indication that there is indeed a problem with that compound, either large errors in the experimental toxicity, a greater reactivity, or a different mechanism of action. We have to emphasize that graph descriptors are not intended to replace other classes of descriptors in QSAR/QSPR models. Each class of structural descriptors reflects in a greater extent some particular aspects of the chemical structure, and a comprehensive exploration of the structural space must consider descriptors from all classes as inputs to a statistical QSAR model.

REFERENCES AND NOTES

(1) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(2) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, 1986.

(3) Diudea, M. V.; Ivanciuc, O. *Molecular Topology*; Comprex: Cluj, Romania, 1995.

(4) Balaban, A. T. From Chemical Graphs to 3D Molecular Modeling. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum: New York, 1997; pp 1−24.

(5) Balaban, A. T. Using Real Numbers as Vertex Invariants for Third Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23−28.

(6) Balaban, A. T. Local versus Global (i.e. Atomic versus Molecular) Numerical Modeling of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398−402.

(7) Ivanciuc, O.; Balaban, A. T. Graph Theory in Chemistry. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 1169−1190.

(8) Balaban, A. T.; Ivanciuc, O. Historical Development of Topological Indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach Science Publishers: Amsterdam: The Netherlands, 1999; pp 21−57.

(9) Ivanciuc, O.; Balaban, A. T. The Graph Description of Chemical Structures. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach Science Publishers: The Netherlands, 1999; pp 59−167.

(10) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Vertex- and Edge-Weighted Molecular Graphs and Derived Structural Descriptors. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach Science Publishers: The Netherlands, 1999; pp 169−220.

(11) Ivanciuc, O.; Ivanciuc, T. Matrices and Structural Descriptors Computed from Molecular Graph Distances. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach Science Publishers: The Netherlands, 1999; pp 221−277.

(12) Ivanciuc, O.; Ivanciuc, T.; Diudea, M. V. Molecular Graph Matrices and Derived Structural Descriptors. *SAR QSAR Environ. Res.* **1997**, *7*, 63−87.

(13) Diudea, M. V.; Gutman, I. Wiener-Type Topological Indices. *Croat. Chem. Acta* **1998**, *71*, 21−51.

(14) Diudea, M. V. Valencies of Property. *Croat. Chem. Acta* **1999**, *72*, 835−851.

(15) Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-Assisted Rational Design of Immunosuppressive Compounds. *Nature Biotechnol.* **1998**, *16*, 748−752.

(16) Ivanciuc, O.; Taraviras, S. L.; Cabrol-Bass, D. Quasi-Orthogonal Basis Sets of Molecular Graph Descriptors as a Chemical Diversity Measure. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 126−134.

(17) Taraviras, S. L.; Ivanciuc, O.; Cabrol-Bass, D. Identification of Groupings of Graph Theoretical Molecular Descriptors Using a Hybrid Cluster Analysis Approach. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1128−1146.

(18) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(19) Wiener, H. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636−2638.

(20) Wiener, H. Vapor Pressure−Temperature Relationship Among the Branched Paraffin Hydrocarbons. *J. Phys. Chem.* **1948**, *52*, 425−430.

(21) Wiener, H. Relation on the Physical Properties of Isomeric Alkanes to Molecular Structure: Surface Tension, Specific Dispersion and Critical Solution Temperature in Aniline. *J. Phys. Chem.* **1948**, *52*, 1082−1089.

(22) Hosoya, H. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339.

(23) Barysz, M.; Jashari, G.; Lall, R. S.; Srivastava, V. K.; Trinajstić, N. On the Distance Matrix of Molecules Containing Heteroatoms. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 222−227.

(24) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. 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.

(25) Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D.; Balaban, A. T. Comparison of Weighting Schemes for Molecular Graph Descriptors: Application in Quantitative Structure-Retention Relationship Models for Alkylphenols in Gas−Liquid Chromatography. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 732−743.

(26) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 20. Molecular Structure Descriptors Computed with Information on Distances Operators. *Rev. Roum. Chim.* **1999**, *44*, 479−489.

(27) Ivanciuc, O. Design of Topological Indices. Part 12. Parameters for Vertex- and Edge-Weighted Molecular Graphs. *Rev. Roum. Chim.* **2000**, *45*, 289−301.

(28) Ivanciuc, O. Design on Topological Indices. 1. Definition of a Vertex Topological Index in the Case of 4-Trees. *Rev. Roum. Chim.* **1989**, *34*, 1361−1368.

(29) Balaban, T. S.; Filip, P. A.; Ivanciuc, O. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and Topological Indices. Derived Canonical Labeling and Coding of Trees and Alkanes. *J. Math. Chem.* **1992**, *11*, 79−105.

(30) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235−250.

(31) Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices. *J. Math. Chem.* **1993**, *12*, 309−318.

(32) Diudea, M. V.; Ivanciuc, O.; Nikolić, S.; Trinajstić, N. Matrices of Reciprocal Distance, Polynomials and Derived Numbers. *MATCH* (*Commun. Math. Comput. Chem.*) **1997**, *35*, 41−64.

(33) Klein, D. J.; Randić, M. Resistance Distance. *J. Math. Chem.* **1993**, *12*, 81−95.

(34) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 8. Path Matrices and Derived Molecular Graph Invariants. *MATCH* (*Commun. Math. Chem.*) **1994**, *30*, 141−152.

(35) Ivanciuc, O. Design of Topological Indices. Part 11. Distance-Valency Matrices and Derived Molecular Graph Descriptors. *Rev. Roum. Chim.* **1999**, *44*, 519−528.

(36) Randić, M. Linear Combinations of Path Numbers as Molecular Descriptors. *New J. Chem.* **1997**, *21*, 945−951.

(37) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. The Complementary Distance Matrix, a New Molecular Graph Metric. *ACH−Model. Chem.* **2000**, *137*, 57−82.

(38) Diudea, M. V. Walk Numbers ᵉW$_M$: Wiener-type Numbers of Higher Rank. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 535−540.

(39) Diudea, M. V. Wiener and Hyper-Wiener Numbers in a Single Matrix. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 833−836.

(40) Ivanciuc, O.; Diudea, M. V.; Khadikar, P. V. New Topological Matrices and Their Polynomials. *Ind. J. Chem.* **1998**, *37A*, 574−585.

(41) Diudea, M. V.; Minailiuc, G.; Katona, G.; Gutman, I. Szeged Matrices and Related Numbers. *MATCH* (*Commun. Math. Comput. Chem.*) **1997**, *35*, 129−143.

(42) Diudea, M. V. Cluj Matrix Invariants. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 300−305.

(43) Hosoya, H. On Some Counting Polynomials in Chemistry. *Discrete Appl. Math.* **1988**, *19*, 239−257.

(44) Sagan, B. E.; Yeh, Y.-N.; Zhang, P. The Wiener Polynomial of a Graph. *Int. J. Quantum Chem.* **1996**, *60*, 959−969.

(45) Estrada, E.; Ivanciuc, O.; Gutman, I.; Gutierrez, A.; Rodríguez, L. Extended Wiener Indices. A New Set of Descriptors for Quantitative Structure−Property Studies. *New J. Chem.* **1998**, *22*, 819−822.

(46) Randić, M. Novel Molecular Descriptor for Structure−Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478−483.

(47) Klein, D. J.; Lukovits, I.; Gutman, I. On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50−52.

(48) Balaban, A. T.; Balaban, T.-S. New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances. *J. Math. Chem.* **1991**, *8*, 383−397.

(49) Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D.; Balaban, A. T. Evaluation in Quantitative Structure−Property Relationship Models of Structural Descriptors Derived from Information-Theory Operators. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 631−643.

(50) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Part 21. Molecular Graph Operators for the Computation of Geometric Structural Descriptors. *Rev. Roum. Chim.* **1999**, *44*, 539−547.

(51) Klein, D. J.; Gutman, I. Wiener-Number-Related Sequences. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 534−536.

(52) Gutman, I.; Linert, W.; Lukovits, I.; Tomović, Ž. The Multiplicative Version of the Wiener Index. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 113−116.

(53) Bytautas, L.; Klein, D. J. Chemical Combinatorics for Alkane-Isomer Enumeration and More. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1063−1078.

(54) Bytautas, L.; Klein, D. J. Mean Wiener Numbers and Other Mean Extensions for Alkane Trees. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 471−481.

(55) Balaban, A. T. Chemical Graphs. Part 48. Topological Index J for Heteroatom-Containing Molecules Taking Into Account Periodicities of Element Properties. *MATCH* (*Commun. Math. Chem.*) **1986**, *21*, 115−122.

(56) Balaban, A. T.; Ivanciuc, O. FORTRAN 77 Computer Program for Calculating the Topological Index J for Molecules Containing Heteroatoms. In *MATH/CHEM/COMP 1988*, Proceedings of an International Course and Conference on the Interfaces Between Mathematics, Chemistry and Computer Sciences, Dubrovnik, Yugoslavia, June 20−

25, 1988; Graovac, A., Ed.; Elsevier: Amsterdam, *Stud. Phys. Theor. Chem.* **1989**, *63*, 193-212.

(57) Balaban, A. T.; Mills, D.; Ivanciuc, O.; Basak, S. C. Reverse Wiener Indices. *Croat. Chem. Acta* **2000**, *73*, 000−000.

(58) Gutman, I. A Formula for the Wiener Number of Trees and Its Extension to Graphs Containing Cycles. *Graph Theory Notes New York* **1994**, *27*, 9−15.

(59) Gutman, I.; Dobrynin, A. A. The Szeged Index−A Success Story. *Graph Theory Notes New York* **1998**, *34*, 37−44.

(60) Schüürmann, G.; Flemmig, B.; Dearden, J. C. CoMFA Study of Acute Toxicity of Nitrobenzenes to *Tetrahymena pyriformis*. In *Quantitative Structure−Activity Relationships in Environmental Sciences−VII*; Chen, F., Schüürmann, G., Eds.; SETAC Press: Pensacola, FL, 1997; pp 315−327.

(61) Dearden, J. C.; Cronin, M. T. D.; Schultz, T. W.; Lin, D. T. QSAR Study of the Toxicity of Nitrobenzenes to *Tetrahymena pyriformis*. *Quant. Struct.-Act. Relat.* **1995**, *14*, 427−432.

(62) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(63) Kubinyi, H. Comparative Molecular Field Analysis (CoMFA). In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 448−460.