

QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIPS GENERATED WITH OPTIMIZABLE EVEN/ODD WIENER POLYNOMIAL DESCRIPTORS*

O. IVANCIUC[†], T. IVANCIUC and D. J. KLEIN

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

(Received 21 June 2000; In final form 5 September 2000)

Chemical structures of organic compounds are characterized numerically by a variety of structural descriptors, one of the earliest and most widely used being the Wiener index W , derived from the interatomic distances in a molecular graph. Extensive use of distance-based structural descriptors or topological indices has been made in QSPR and QSAR models, drug design, toxicology, virtual screening of combinatorial libraries, similarity and diversity assessment. Novel topological indices are introduced representing a partitioning of the Wiener polynomial based on counts of even and odd molecular graph distances. During the QSAR/QSPR modeling process the variables of the even and odd power functions are optimized in order to offer an improved mapping of the investigated property. These novel topological indices are tested in QSPR models for the boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, and density of alkanes. In many cases, the even/odd Wiener polynomial indices proposed here give notably improved correlations or suggest simpler QSPR models.

Keywords: Quantitative structure-property relationships; Topological indices; Graph distance descriptors; Even/odd Wiener polynomial indices

INTRODUCTION

Structural descriptors are used in molecular modeling to express in a numerical form the chemical structure. The continual development of

* Dedicated to Professor Alexandru T. Balaban on the occasion of his 70th birthday.

[†]Corresponding author.

structural descriptors and statistical models has transformed quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) into powerful and widely used methods for the prediction of physical, chemical, and biological properties. In a first approximation the chemical structure of a molecule can be represented as a molecular graph, in which vertices correspond to atoms, and edges represent covalent bonds between atoms. Using molecular graphs the chemical structure of a compound can be expressed by means of various graph matrices, polynomials, spectra, spectral moments, distance-counting sequences, paths, and walks, or yet other topological indices. A topological index (TI) is a numerical descriptor of the molecular structure based on certain topological features of the molecular graph [1–11]. Significant results have been obtained in the design of novel TIs, with improved correlational abilities [7–11]. Another direction of intense research is represented by the parametrization of vertex- and edge-weighted molecular graphs, representing organic compounds with heteroatoms and multiple bonds [8, 9]; this research direction is stimulated by the novel application of molecular graph descriptors in similarity and diversity assessment, database mining, and in the virtual screening of combinatorial libraries.

The use of modern topological indices in QSPR and QSAR begins with the Wiener index W , which was applied for establishing correlations between molecular structure and various physical properties of alkanes [12–15]. Hosoya extended the original definition to cyclic compounds with the aid of the distance matrix and gave a graph-theoretical formula for the Wiener index as “the half sum of the off-diagonal elements of a distance matrix \mathbf{D} whose element d_{ij} is a number of bonds for the shortest path between atoms i and j ” [16]. Its successful application in a large number of QSAR/QSPR models stimulated the research in the domain of structural descriptors based on graph distances [9], novel molecular matrices [10] and Wiener-type indices [11]. The relatively high degeneracy of W due to the global summation of distances can be reduced by using information theory, as shown by Bonchev and Trinajstić who used all graph distances to define indices I_D^E , \bar{I}_D^E , I_D^W , and \bar{I}_D^W [17], and by Balaban and Balaban who defined the U , V , X , and Y indices based on the vertex partitioning of graph distances [18]. Recently, the U , V , X , and Y information indices were extended for any symmetric molecular matrix derived from vertex- and edge-weighted molecular graphs [19, 20], giving the graph operators $\mathbf{U}(\mathbf{M})$, $\mathbf{V}(\mathbf{M})$, $\mathbf{X}(\mathbf{M})$, and $\mathbf{Y}(\mathbf{M})$ that were applied with success in QSPR equations [21] and in modeling the retention indices of alkylphenols in gas-liquid chromatography [22]. Using a Randić-type formula and a vertex partitioning of distances,

Balaban defined the index J [23–25], which was successfully applied in QSPR [26], QSAR [27], and virtual screening of combinatorial libraries [28]. The Balaban index J was used as a basis for the definition of the **IB** operator for vertex- and edge-weighted molecular graphs [29].

Elements of the distance matrix were used to define various molecular matrices and graph metrics. The idea to use reciprocal distances in computing degree-distance VTI indices [30] was adopted in the definition of the reciprocal distance **RD** [31–33] and distance-valency **Dval** [34] matrices. Another distance measure was defined as the resistance-distance matrix **Ω** [35]; this metric is identical with that induced by the distance matrix only for acyclic compounds, while for cyclic compounds the resistance-distance matrix offers the possibility to compute an index analogous but not identical to the Wiener index. Other recently defined distance-related matrices are the detour Δ [36], detour-distance Δ -**D** [36], distance complement **DC** [37], complementary distance **CD** [38], reciprocal complementary distance **RCD** [38], and reverse Wiener **RW** [39] matrices. The original formula of the Wiener index was extended to afford the computation of the hyper-Wiener index WW [40, 41]. Because initially WW was defined only for acyclic graphs, an extension was proposed for cyclic graphs [42] and a related molecular matrix was defined by Diudea [43, 44]. The **HyWi** graph operator, which implements a hyper-Wiener type formula for any molecular matrix [9], was used to describe the diversity and descriptor clustering for molecules selected from the National Cancer Institute AIDS database [45, 46]. Recent developments in the area of distance descriptors include the characterization of molecular cyclicity and centrality [47], Wiener-related sequences [48, 49], a multiplicative version of the Wiener index [50], and novel graph metrics [51, 52]. The mean Wiener index for alkanes, computed as an average property of isomeric series [53–55], was introduced with the aim of characterizing mixtures of organic compounds.

Hosoya used distances between pairs of graph vertices to define the Wiener polynomial, whose first derivative gives the Wiener index W [56]; several years later, an equivalent formula was proposed by a group of mathematicians [57]. The higher derivatives of the Wiener polynomial give a sequence of Wiener-type topological indices used with success in several QSPR models [58–60]. Recently we have defined a group of novel graph descriptors based on even and odd distances which are separately summed to give two distinct graph invariants that are subsequently combined into a series of new graph descriptors, that measure the molecular size and shape [61]. In this paper we propose and study two graph descriptors obtained

from the Wiener polynomial by summing into two distinct indices the terms corresponding to even and odd graph distances. These novel topological indices computed from the Wiener polynomial are tested in QSPR 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} . During the QSPR modeling process the variables of the even and odd power functions are optimized in order to generate an improved mapping of the investigated property. The statistical results of the QSPR models demonstrate that such splitting of the Wiener polynomial into even/odd distance sums is a successful one, offering topological indices with an enhanced correlational power. In several QSPR equations the optimum value of the variable suggests efficient ways for generating more simple QSPR models.

EVEN/ODD DISTANCE INDICES

Together with the connectivity indices [1, 2], the Wiener index is the most frequently used graph descriptor in QSAR/QSPR models. The Wiener index $W = W(G)$ of a molecular graph G with N vertices is [9, 29]:

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

where $\mathbf{D} = \mathbf{D}(G)$ is the distance matrix of G , and $[\mathbf{D}(G)]_{ij}$ represents the topological distance between vertices v_i and v_j . This formula, which is slightly modified from the definition proposed by Hosoya [16], can be used for vertex- and edge-weighted molecular graphs representing organic compounds with heteroatoms and multiple bonds.

In a recent study we have partitioned all graph distances into two classes, *i.e.* even and odd distances, which are separately summed to give two distinct structural descriptors, *SumE* and *SumO* [61]. The sum of even graph distances *SumE* from the molecular graph G is:

$$SumE(p, G) = \sum_{i < j}^{d_{ij} \text{ even}} [\mathbf{D}(G)]_{ij}^p \quad (2)$$

where the summation goes over all even elements in the upper triangle of the distance matrix \mathbf{D} , and p is a constant that affords an increased flexibility to this graph invariant. In a similar way we define *SumO*, the sum of odd graph distances from the molecular graph G , by collecting together all odd graph

distances:

$$SumO(q, G) = \sum_{i < j}^{d_{ij, \text{odd}}} [\mathbf{D}(G)]_{ij}^q \quad (3)$$

where the summation goes over all odd elements in the upper triangle of the distance matrix \mathbf{D} , and q is a constant that weights the odd distances. During the generation of the QSAR/QSPR models, the even and odd exponents p and q can be optimized to give better statistical indices. If $p = q = 1$, a simple relationship exists between the Wiener index W and the even/odd indices $SumE$ and $SumO$:

$$W(G) = SumE(1, G) + SumO(1, G) \quad (4)$$

The individual count of graph distances according to their length was used by Hosoya in the definition of the Wiener polynomial. Let $d(G, k)$ be the number of pairs of vertices in G that are distance k apart, and denote the largest element of \mathbf{D} by l , often called the diameter of G ; obviously, $d(G, 1)$ is equal to the number of edges. The Wiener polynomial of a graph G is defined by the following equation [56, 57]:

$$\mathbf{H}(G, z) = \sum_{k=1}^l d(G, k) z^k = \sum_{i < j} z^{d_{ij}} \quad (5)$$

From the above definition it is obvious that the procedure of separating the Wiener index W into even and odd terms can be applied also to the Wiener polynomial. The sum of Wiener polynomial terms corresponding to even graph distances defines the topological index $WiPolE$:

$$WiPolE(x, G) = \sum_{i < j}^{d_{ij, \text{even}}} x^{d_{ij}} \quad (6)$$

where the summation goes over all even graph distances d_{ij} , and the variable x is an optimizable parameter. Similarly we define $WiPolO$, the sum of all terms in the Wiener polynomial corresponding to odd graph distances:

$$WiPolO(y, G) = \sum_{i < j}^{d_{ij, \text{odd}}} y^{d_{ij}} \quad (7)$$

where the summation goes over all odd graph distances d_{ij} , and the variable y is a parameter that can be optimized for each QSAR/QSPR property and data set of chemical compounds.

MATERIALS AND METHODS

Database

The QSPR models are developed for a data set consisting of 134 alkanes between C_6 and C_{10} , as applied for the following six physical properties [62]: t_b , boiling temperature at normal pressure ($^{\circ}C$); C_p , molar heat capacity at 300 K ($J K^{-1} mol^{-1}$); $\Delta_f G_{300}^{\circ}$ (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 $25^{\circ}C$; ρ , density at $25^{\circ}C$ ($kg m^{-3}$). 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 seemingly too high when compared with the density of similar alkanes and it was not considered in the computation of the density QSPR models. There are 142 constitutional isomers for these alkanes, but some data for some of the 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.

Structural Descriptors and QSPR Models

QSPR models for all six alkane properties are obtained with multiple linear regression (MLR) equations in which three structural descriptors are tested: *MN*, the molecular number equal to the sum of the atomic number Z for all atoms in a molecule, hydrogens included, the even and odd Wiener polynomial sums $WiPolE(x)$ and $WiPolO(y)$, respectively. The x and y variables of the even and odd graph distances are optimized with a grid search algorithm in order to obtain QSPR models with highest correlation coefficient and lowest standard deviation.

In order to assess the value of QSPR models obtained with $WiPolE(x)$ and $WiPolO(y)$, we use the same molecular data base to obtain structure-property models with several widely used topological indices. This group of QSPR models is represented by biparametric equations in which one descriptor is N , the number of carbon atoms, while the second one is either a connectivity index (${}^0\chi, {}^1\chi, {}^2\chi$), the Wiener index W , or the best even/odd graph distance descriptor identified in our recent study [61]. In choosing the above set of descriptors we were guided by the following considerations: (a) the widespread use of connectivity indices in QSPR; (b) the necessary comparison with the Wiener index W , since $WiPolE(x)$ and $WiPolO(y)$ are proposed as Wiener-type indices; (c) our recent finding that even/odd graph

TABLE I Structural descriptors SD, coefficients and statistical indices for linear regression equations with the general form $P = a_0 + a_1 N + a_2 SD$, where N is the number of carbon atoms and P is a physical property of the set of 134 $C_6 - C_{10}$ alkanes

P	Eq.	a_0	a_1	a_2	SD	r	s	F
t_b (°C)	8	-73.99	34.07	-13.13	$0 \times$	0.9762	5.80	1329
	9	-78.52	14.49	20.48	$1 \times$	0.9800	5.32	1591
	10	-81.80	27.12	-7.277	$2 \times$	0.9841	4.75	2013
	11	-89.47	25.86	-0.07479	W	0.9730	6.17	1164
	12	-138.6	25.79	24.02	$W_{\circ,c}(-1,-1)$	0.9835	4.83	1939
	13	4.983	25.85	-3.458	$0 \times$	0.9873	4.07	2523
	14	3.771	21.20	4.292	$1 \times$	0.9873	4.06	2540
	15	3.226	23.68	-1.165	$2 \times$	0.9873	4.06	2533
	16	3.822	23.14	0.0009397	W	0.9870	4.11	2470
	17	-9.624	25.30	-0.06525	$W_{e \times o}(-2,0.5)$	0.9874	4.05	2545
	18	-79.66	-9.400	26.09	$1 \times$	0.8550	7.69	178
	19	-70.23	18.88	-17.45	$1 \times$	0.8218	8.45	136
	20	-69.07	9.971	2.065	$2 \times$	0.8045	8.81	120
	21	-171.3	30.09	-0.7176	W	0.9159	5.96	341
	22	-128.8	21.09	-0.1276	$W_{e+o}(1.5,2)$	0.9274	5.55	403
	23	7.323	10.21	-8.361	$0 \times$	0.9866	0.71	2399
	24	4.354	-0.1386	8.388	$1 \times$	0.9785	0.89	1477
25	3.378	4.621	-2.055	$2 \times$	0.9662	1.12	920	
26	20.87	0.5436	0.1180	W	0.9633	1.16	845	
27	-9.477	9.276	-2.719	$W_{e+o}(-1.5,-2)$	0.9796	0.87	1560	
28	1.302	-0.003034	0.01802	$0 \times$	0.9022	0.0059	284	
29	1.308	0.01559	-0.01001	$1 \times$	0.8822	0.0065	228	
30	1.308	0.01098	0.0001236	$2 \times$	0.8751	0.0067	213	
31	1.231	0.02575	-0.0005525	W	0.9490	0.0043	589	
32	1.387	-0.03370	0.03564	$SumO(-2)$	0.9663	0.0035	917	
33	523.9	-6.554	35.07	$0 \times$	0.8825	12.50	229	
34	536.2	28.36	-16.55	$1 \times$	0.8598	13.57	184	
35	535.4	21.39	-1.208	$2 \times$	0.8548	13.79	176	
36	370.5	52.29	-1.181	W	0.9459	8.62	552	
37	707.9	-76.62	77.65	$SumO(-2)$	0.9713	6.32	1084	

C_p
($J K^{-1} mol^{-1}$)

$\Delta_f G_{300}^\circ$
($kJ mol^{-1}$)

$\Delta_{vap} H_{300}$
($kJ mol^{-1}$)

n_D^{25}

ρ
($kg m^{-3}$)

distance descriptors often give better correlations than connectivity and Wiener-type indices. In Table I we present all QSPR equations used for comparison purposes, together with the corresponding statistical indices.

RESULTS AND DISCUSSION

The main objective of this investigation is to evaluate the QSPR usefulness of even/odd topological indices derived from the Wiener polynomial. Because in previous studies we found that molecular size is an important factor for QSPR modeling of the six alkane properties [61], for each property we investigate all bi- and triparametric MLR equations obtained with the three descriptors: MN , $WiPolE(x)$, and $WiPolO(y)$.

Normal Boiling Temperature

In Table II we present the coefficients, and statistical indices for the tri- and biparametric MLR equations that model the normal boiling temperature of the 134 alkanes, *i.e.*, Eqs. (38), (44), (50), and (56). The best QSPR model, with $r = 0.9917$, $s = 3.45^\circ\text{C}$, $F = 2582$, is obtained for Eq. (38) that has three descriptors; a comparison with the best results from Table I, obtained with the connectivity index ${}^2\chi$ ($r = 0.9841$, $s = 4.75^\circ\text{C}$, $F = 2013$), indicate that $WiPolE(x)$ and $WiPolO(y)$ offer a significant improvement of the correlation. The optimum value for x and y is 0.48 and 0.01, respectively, and the correlation coefficient between MN and $WiPolO(y)$ is almost equal to 1; another interesting observation concerning Eq. (38) is the opposite sign of the MN and $WiPolO(y)$ coefficients, *i.e.* negative for MN and positive for $WiPolO(y)$. Evidently MN and $WiPolO(y)$ (at $y = 0.01$) are very nearly linearly dependent (*i.e.*, very nearly scalar multiples). But the sign and sizes of the coefficients indicate that the small difference between these two (appropriately scaled) descriptors is also relevant in obtaining a good fit. This difference between MN and $WiPolO(y)$ becomes proportional to p_3 , the number of distances of length 3 in the molecular graph. Indeed Wiener recognized the importance of p_3 , and in making correlations he used p_3 together with the normalized Wiener index W/N^2 [12–15].

This observation prompted us to replace $WiPolO(y)$ with p_3 and to determine if only the count of paths of length 3 can offer the same high correlation. The results presented in Table III, Eq. (62) show that indeed the QSPR obtained with MN , $WiPolE(0.48)$, and p_3 has the same correlation coefficient and standard deviation with Eq. (38), while the Fisher test is only

TABLE II Structural descriptors, coefficients and statistical indices for linear regression equations with the general form $P = a_0 + a_1SD_1 + a_2SD_2 + a_3SD_3$, where P is a physical property of the set of 134 C_6-C_{10} alkanes

P	Eq.	a_0	a_1	SD_1	a_2	SD_2	a_3	SD_3	r	s	F	Obs.
t_b	38	33811	-3386	MN	-15.195	WiPolE(0.48)	2711803	WiPolO(0.01)	0.9917	3.45	2582	$r(SD_1, SD_3) \cong 1$
C_p	39	-10.51	3.155	MN	-333.7	WiPolE(0.05)	-0.0174	WiPolO(1.84)	0.9877	4.01	1735	
$\Delta_f G_{300}^\circ$	40	2154	-215.3	MN	2.601	WiPolE(0.99)	34312	WiPolO(0.05)	0.9688	3.69	663	$r(SD_1, SD_3) \cong 1$
$\Delta_{vap} H_{300}$	41	-3.442	0.7403	MN	-9.210	WiPolE(0.32)	-0.5114	WiPolO(0.63)	0.9814	0.84	1133	
n_D^{25}	42	2.903	-0.1564	MN	0.1004	WiPolE(0.05)	25.12	WiPolO(0.05)	0.9813	0.0027	1115	$r(SD_1, SD_3) \cong 1$
ρ	43	4052	-345.7	MN	-0.3548	WiPolE(0.96)	55490	WiPolO(0.05)	0.9877	4.17	1714	$r(SD_1, SD_3) \cong 1$
t_b	44	-13.99			0.00817	WiPolE(2.65)	26.23	WiPolO(0.52)	0.9849	4.63	2116	
C_p	45	105.7			2.502	WiPolE(1.04)	3.385	WiPolO(0.95)	0.9839	4.58	1982	
$\Delta_f G_{300}^\circ$	46	-56.38			-0.000224	WiPolE(4.83)	18.63	WiPolO(0.47)	0.9236	5.69	380	
$\Delta_{vap} H_{300}$	47	22.44			0.0830	WiPolE(1.47)	0.3998	WiPolO(1.10)	0.9630	1.17	835	
n_D^{25}	48	1.322			-0.000183	WiPolE(1.46)	0.01916	WiPolO(0.49)	0.9655	0.0036	893	
ρ	49	554.6			-0.5607	WiPolE(1.41)	39.35	WiPolO(0.49)	0.9684	6.63	980	
t_b	50	-132.5	4.256	MN	-3.389	WiPolE(0.93)			0.9806	5.25	1635	
C_p	51	-3.769	2.975	MN	-168.6	WiPolE(0.05)			0.9873	4.07	2523	
$\Delta_f G_{300}^\circ$	52	-90.20	1.673	MN	-0.000337	WiPolE(4.83)			0.8878	6.83	244	
$\Delta_{vap} H_{300}$	53	-1.381	0.6631	MN	-13.40	WiPolE(0.27)			0.9807	0.85	1645	
n_D^{25}	54	1.274	0.002041	MN	-0.000213	WiPolE(1.61)			0.9269	0.0052	396	
ρ	55	461.2	4.087	MN	-0.4917	WiPolE(1.60)			0.9232	10.21	375	
t_b	56	-44.81	1.544	MN			6.749	WiPolO(0.73)	0.9805	5.26	1628	
C_p	57	-2.939	2.909	MN			-0.000011	WiPolO(4.83)	0.9871	4.09	2497	
$\Delta_f G_{300}^\circ$	58	1947	-198.2	MN			31787	WiPolO(0.05)	0.9478	4.73	579	$r(SD_1, SD_3) \cong 1$
$\Delta_{vap} H_{300}$	59	9.291	0.3357	MN			0.0385	WiPolO(1.62)	0.9546	1.29	672	
n_D^{25}	60	2.943	-0.1605	MN			25.78	WiPolO(0.05)	0.9810	0.0027	1663	$r(SD_1, SD_3) \cong 1$
ρ	61	4069	-346.9	MN			55669	WiPolO(0.05)	0.9876	4.17	2569	$r(SD_1, SD_3) \cong 1$

TABLE III Structural descriptors, coefficients and statistical indices for linear regression equations with the general form $P = a_0 + a_1SD_1 + a_2SD_2 + a_3SD_3 + a_4SD_4 + a_5SD_5$, where P is a physical property of the set of 134 C_6-C_{10} alkanes

P	Eq.	a_0	a_1	SD_1	a_2	SD_2	a_3	SD_3	r	s	F
t_b	62	-86.02	3.219	MN	-15.20	WiPolE(0.48)	2.712	p_3	0.9917	3.45	2583
C_p	63	-3.285	2.965	MN	-189.8	WiPolE(0.05)	0.08461	p_3	0.9873	4.08	1671
$\Delta_f G_{300}^o$	64	9.054	-0.8209	MN	2.593	WiPolE(0.99)	4.281	p_3	0.9689	3.69	664
$\Delta_{vap} H_{300}$	65	-2.209	0.6798	MN	-9.008	WiPolE(0.32)	-0.09342	p_3	0.9813	0.84	1126
n_D^{25}	66	1.333	0.000609	MN	0.09897	WiPolE(0.05)	0.003136	p_3	0.9813	0.0027	1117
ρ	67	582.9	1.164	MN	-0.3697	WiPolE(0.96)	6.923	p_3	0.9877	4.17	1720
t_b	68	-71.57	2.612	MN			1.620	p_3	0.9794	5.40	1542
C_p	69	-2.726	2.913	MN			-0.07615	p_3	0.9870	4.11	2473
$\Delta_f G_{300}^o$	70	-40.05	0.4557	MN			3.968	p_3	0.9480	4.72	581
$\Delta_{vap} H_{300}$	71	-0.01811	0.5540	MN			-0.4005	p_3	0.9467	1.40	566
n_D^{25}	72	1.332	0.000638	MN			0.003217	p_3	0.9810	0.0027	1666
ρ	73	589	1.004	MN			6.946	p_3	0.9876	4.17	2577
t_b	74	-75.26	2.968	MN	-3.210	p_2	2.708	p_3	0.9906	3.67	2276
C_p	75	-3.274	2.965	MN	-0.4762	p_2	0.08515	p_3	0.9873	4.08	1671
$\Delta_f G_{300}^o$	76	-39.04	0.3589	MN	0.8746	p_2	3.672	p_3	0.9508	4.61	408
$\Delta_{vap} H_{300}$	77	-1.057	0.6540	MN	-0.9030	p_2	-0.09468	p_3	0.9811	0.84	1116
n_D^{25}	78	1.333	0.000610	MN	0.000248	p_2	0.003136	p_3	0.9813	0.0027	1117
ρ	79	588.9	1.013	MN	-0.07863	p_2	6.972	p_3	0.9876	4.18	1706
t_b	80	-90.82	3.267	MN	-1.513	p_2			0.9758	5.84	1307
C_p	81	-3.763	2.975	MN	-0.4228	p_2			0.9873	4.07	2523
$\Delta_f G_{300}^o$	82	-60.14	0.7647	MN	3.176	p_2			0.8546	7.70	177
$\Delta_{vap} H_{300}$	83	-0.5128	0.6436	MN	-0.9623	p_2			0.9804	0.85	1625
n_D^{25}	84	1.315	0.000955	MN	0.002216	p_2			0.9024	0.0059	285
ρ	85	548.9	1.782	MN	4.297	p_2			0.8825	12.50	229
t_b	86	33769	-3381	MN	-3.210	p_2	2707525	WiPolO(0.01)	0.9906	3.67	2276
C_p	87	-10.48	3.155	MN	-0.8343	p_2	-0.01737	WiPolO(1.84)	0.9877	4.01	1735
$\Delta_f G_{300}^o$	88	1799	-183.4	MN	0.8801	p_2	29401	WiPolO(0.05)	0.9507	4.62	407
$\Delta_{vap} H_{300}$	89	-2.074	0.7064	MN	-0.9262	p_2	-0.4597	WiPolO(0.63)	0.9811	0.84	1112
n_D^{25}	90	2.903	-0.1564	MN	0.000252	p_2	25.12	WiPolO(0.05)	0.9813	0.0027	1115
ρ	91	4080	-348.1	MN	-0.07057	p_2	55853	WiPolO(0.05)	0.9876	4.19	1700

one unit larger out of 2582. The even distance term is important for a good correlation, since both its omission, Eq. (68), or its substitution with p_2 , Eqs. (74) and (80), provide MLR equations with lower statistics.

Molar Heat Capacity

The results reported in Table II show that the best model for the alkane molar heat capacity is Eq. (39), computed with MN , $WiPolE(0.05)$, and $WiPolO(1.84)$. This equation has good statistical indices ($r = 0.9877$, $s = 4.01 \text{ J K}^{-1} \text{ mol}^{-1}$, $F = 1735$), slightly superior to those from Eq. (17) that was obtained with N and the product of the even and odd distances $W_{e \times o}(-2, 0.5)$. The optimum value for x is 0.05, and if the optimization continues, it has the tendency to converge toward 0, when $WiPolE$ is proportional to the path count of length 2, p_2 . This behavior suggests a more simple QSPR model, in which $WiPolE(0.05)$ is substituted with p_2 . A comparison of the C_p QSPR equations from Table III reveals that Eq. (87) obtained with MN , p_2 , and $WiPolO(1.84)$ has identical statistical indices with Eq. (39), demonstrating that the substitution of $WiPolE(0.05)$ with p_2 is an effective simplification of the QSPR model. However, by deleting the odd distance term $WiPolO(1.84)$ one obtains Eq. (81) that has only two parameters, *i.e.* MN and p_2 , but with statistics very close to those from Eqs. (39) and (87): $r = 0.9873$, $s = 4.07 \text{ J K}^{-1} \text{ mol}^{-1}$, $F = 2523$.

Standard Gibbs Energy of Formation

The use of even/odd Wiener polynomial descriptors greatly improve the QSPR model of the alkane standard Gibbs energy of formation, as revealed by a comparison of Eqs. (22) and (40). Computed with three topological indices, MN , $WiPolE(0.99)$, and $WiPolO(0.05)$, Eq. (40) has good statistical indices ($r = 0.9688$, $s = 3.69 \text{ kJ mol}^{-1}$, $F = 663$), with a significant decrease in the standard deviation in comparison with Eq. (22) computed with N and the sum of the even and odd distances $W_{e+o}(1.5, 2)$. The optimum variable for the odd distances is 0.05, suggesting again a replacement of $WiPolO(y)$ with p_3 ; indeed, Eq. (64) from Table III, computed with MN , $WiPolE(0.99)$, and p_3 , has almost identical statistical indices, demonstrating the importance of this simple graph descriptor in modeling $\Delta_f G_{300}^\circ$. Other tests intended to further simplify this QSPR models, reported in Eqs. (70), (76), (82), and (88), produced equations of lower statistical quality, indicating that the even distance term $WiPolE(0.99)$ is important in order to obtain a $\Delta_f G_{300}^\circ$ QSPR with a small standard deviation.

Vaporization Enthalpy

The QSPR model from Eq. (23) from Table I for the alkane vaporization enthalpy obtained with N and the first connectivity index ${}^0\chi$, with $r = 0.9866$, $s = 0.71 \text{ kJ mol}^{-1}$, $F = 2399$, cannot be improved with the use of the *WiPolE* and *WiPolO* indices: Eq. (53) from Table II that contains two parameters, MN and *WiPolE*(0.27) and statistical indices very close to those of Eq. (41) ($r = 0.9807$, $s = 0.85 \text{ kJ mol}^{-1}$, $F = 1645$), suggests that only a size descriptor and the even distance term is important for the $\Delta_{\text{vap}}H_{300}$ QSPR. An even simpler model, but with almost identical statistics, is obtained when *WiPolE* is replaced by p_2 in Eq. (83). Similarly with the C_p QSPR Eq. (81), this vaporization enthalpy QSPR has two simple descriptors, a size parameter MN and a branching parameter p_2 .

Refractive Index

The results reported in Table II show that Eq. (60) obtained with MN and *WiPolO*(0.05), with $r = 0.9810$, $s = 0.0027$, $F = 1663$, is a better n_D^{25} QSPR than Eq. (32) from Table I, obtained with N and the sum of odd distances *SumO*(-2). The small value of the variable for the odd distances in the Wiener polynomial prompted us to investigate several QSPR models in which *WiPolO* is replaced with p_3 ; these equations are presented in Table III. Using only two simple graph descriptors, namely MN and p_3 , Eq. (72) has statistics almost identical with those of Eq. (60); this finding is a clear indication that the refractive index of alkanes can be modeled with a size descriptor and the count of distances 3, and that several more complex Wiener-type descriptors are not able to improve significantly the statistical indices.

Density

Similarly with the refractive index QSPR, the best model for the alkane density, Eq. (61), is obtained with MN and *WiPolO*(0.05) ($r = 0.9876$, $s = 4.17 \text{ kg m}^{-3}$, $F = 2569$), showing a significant improvement over the previously reported ρ QSPR model, Eq. (37), computed with N and the sum of odd distances *SumO*(-2). The correlation coefficient between MN and *WiPolO*(0.05) is almost equal to 1, exhibiting the same tendency identified for boiling temperature, standard Gibbs energy of formation, and refractive index. In Table III we present several ρ QSPR equations obtained when *WiPolO*(0.05) is substituted with p_3 . Obviously, Eq. (73) obtained with MN

and p_3 has statistical indices as good as those of Eq. (61), demonstrating that odd graph distances larger than 3 have little importance for the QSPR model of alkane density.

CONCLUDING REMARKS

Topological indices cast the structure of organic compounds represented as molecular graphs into a numerical form, and thereby provide an important class of structural descriptors for QSPR and QSAR models. The use of molecular graph descriptors in QSAR/QSPR models may generally be used to complement other classes of structural descriptors, such as fingerprints, hydrophobicity, geometrical, electrostatic, and quantum descriptors. Three classes of topological indices are extensively used in QSAR/QSPR models, namely: local-structure counts (as in group additivity schemes); second, connectivity indices; and third, distance-related descriptors, mainly represented by the Wiener index and other Wiener-type descriptors. In this paper we have developed novel graph descriptors by separating the terms of the Wiener polynomial into even and odd graph distances. The even and odd Wiener polynomial sums $WiPolE(x)$ and $WiPolO(y)$ are used as descriptors in QSAR/QSPR models, when the even and odd variables x and y , respectively, can be optimized in order to minimize the standard deviation of the model. Structural descriptors computed with these novel even/odd distance indices 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. In many cases, the even/odd Wiener polynomial indices proposed here give notably improved correlations or suggest simpler QSPR models.

Acknowledgment

The authors acknowledge the financial support of this research by the Welch Foundation of Houston, Texas.

References

- [1] Kier, L. B. and Hall, L. H. (1976). *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, New York, p. 257.
- [2] Kier, L. B. and Hall, L. H. (1986). *Molecular Connectivity in Structure-Activity Analysis*. Research Studies Press, Letchworth, p. 262.
- [3] Balaban, A. T., Motoc, I., Bonchev, D. and Mekenyan, O. (1983). Topological indices for structure-activity correlations. *Top. Curr. Chem.*, **114**, 21 – 55.

- [4] Balaban, A. T. (1994). Local versus global (*i.e.*, atomic versus molecular) numerical modelling of molecular graphs. *J. Chem. Inf. Comput. Sci.*, **34**, 398–402.
- [5] Balaban, A. T. (1998). Topological and stereochemical molecular descriptors for databases useful in QSAR, similarity/dissimilarity and drug design. *SAR QSAR Environ. Res.*, **8**, 1–21.
- [6] Ivanciuc, O. and Balaban, A. T. (1998). 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 III, H. F. and Schreiner, P. R., Eds.). John Wiley & Sons, Chichester, pp. 1169–1190.
- [7] Ivanciuc, O. and Balaban, A. T. (1999). The graph description of chemical structures. In: *Topological Indices and Related Descriptors in QSAR and QSPR* (Devillers, J. and Balaban, A. T., Eds.). Gordon and Breach Science Publishers, The Netherlands, pp. 59–167.
- [8] Ivanciuc, O., Ivanciuc, T. and Balaban, A. T. (1999). Vertex- and edge-weighted molecular graphs and derived structural descriptors. In: *Topological Indices and Related Descriptors in QSAR and QSPR* (Devillers, J. and Balaban, A. T., Eds.). Gordon and Breach Science Publishers, The Netherlands, pp. 169–220.
- [9] Ivanciuc, O. and Ivanciuc, T. (1999). Matrices and structural descriptors computed from molecular graph distances. In: *Topological Indices and Related Descriptors in QSAR and QSPR* (Devillers, J. and Balaban, A. T., Eds.). Gordon and Breach Science Publishers, The Netherlands, pp. 221–277.
- [10] Ivanciuc, O., Ivanciuc, T. and Diudea, M. V. (1997). Molecular graph matrices and derived structural descriptors. *SAR QSAR Environ. Res.*, **7**, 63–87.
- [11] Diudea, M. V. and Gutman, I. (1998). Wiener-type topological indices. *Croat. Chem. Acta*, **71**, 21–51.
- [12] Wiener, H. (1947). Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, **69**, 17–20.
- [13] Wiener, H. (1947). Correlation of heats of isomerization and differences in heats of vaporization of isomers among the paraffin hydrocarbons. *J. Am. Chem. Soc.*, **69**, 2636–2638.
- [14] Wiener, H. (1948). Vapor pressure-temperature relationship among the branched paraffin hydrocarbons. *J. Phys. Chem.*, **52**, 425–430.
- [15] Wiener, H. (1948). Relation on the physical properties of isomeric alkanes to molecular structure: surface tension, specific dispersion and critical solution temperature in aniline. *J. Phys. Chem.*, **52**, 1082–1089.
- [16] Hosoya, H. (1971). A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Japan*, **44**, 2332–2339.
- [17] Bonchev, D. and Trinajstić, N. (1977). Information theory, distance matrix, and molecular branching. *J. Chem. Phys.*, **67**, 4517–4533.
- [18] Balaban, A. T. and Balaban, T.-S. (1991). New vertex invariants and topological indices of chemical graphs based on information on distances. *J. Math. Chem.*, **8**, 383–397.
- [19] Ivanciuc, O. and Balaban, A. T. (1999). Design of topological indices. Part 20. Molecular structure descriptors computed with information on distances operators. *Rev. Roum. Chim.*, **44**, 479–489.
- [20] Ivanciuc, O. and Balaban, A. T. (1999). Design of topological indices. Part 21. Molecular graph operators for the computation of geometric structural descriptors. *Rev. Roum. Chim.*, **44**, 539–547.
- [21] Ivanciuc, O., Ivanciuc, T., Cabrol-Bass, D. and Balaban, A. T. (2000). Evaluation in quantitative structure-property relationship models of structural descriptors derived from information-theory operators. *J. Chem. Inf. Comput. Sci.*, **40**, 631–643.
- [22] Ivanciuc, O., Ivanciuc, T., Cabrol-Bass, D. and Balaban, A. T. (2000). 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.*, **40**, 732–743.
- [23] Balaban, A. T. (1982). Highly discriminating distance-based topological index. *Chem. Phys. Lett.*, **89**, 399–404.
- [24] Balaban, A. T. (1983). Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.*, **55**, 199–206.

- [25] Balaban, A. T. (1986). Chemical graphs. Part 48. Topological index J for heteroatom-containing molecules taking into account periodicities of element properties. *MATCH (Commun. Math. Chem.)*, **21**, 115–122.
- [26] Balaban, A. T., Kier, L. B. and Joshi, N. (1992). Correlations between chemical structure and normal boiling points of acyclic ethers, peroxides, acetals, and their sulfur analogues. *J. Chem. Inf. Comput. Sci.*, **32**, 237–244.
- [27] Bonchev, D., Mountain, C. F., Seitz, W. A. and Balaban, A. T. (1993). Modeling the anticancer action of some retinoid compounds by making use of the OASIS method. *J. Med. Chem.*, **36**, 1562–1569.
- [28] Grassy, G., Calas, B., Yasri, A., Lahana, R., Woo, J., Iyer, S., Kaczorek, M., Floc'h, R. and Buelow, R. (1998). Computer-assisted rational design of immunosuppressive compounds. *Nature Biotechnol.*, **16**, 748–752.
- [29] Ivanciuc, O., Ivanciuc, T. and Balaban, A. T. (1998). 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.*, **38**, 395–401.
- [30] Ivanciuc, O. (1989). Design on topological indices. 1. Definition of a vertex topological index in the case of 4-trees. *Rev. Roum. Chim.*, **34**, 1361–1368.
- [31] Balaban, T. S., Filip, P. A. and Ivanciuc, O. (1992). Computer generation of acyclic graphs based on local vertex invariants and topological indices. Derived canonical labelling and coding of trees and alkanes. *J. Math. Chem.*, **11**, 79–105.
- [32] Plavšić, D., Nikolić, S., Trinajstić, N. and Mihalić, Z. (1993). On the Harary index for the characterization of chemical graphs. *J. Math. Chem.*, **12**, 235–250.
- [33] Ivanciuc, O., Balaban, T.-S. and Balaban, A. T. (1993). Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem.*, **12**, 309–318.
- [34] Ivanciuc, O. (1999). Design of topological indices. Part 11. Distance-valency matrices and derived molecular graph descriptors. *Rev. Roum. Chim.*, **44**, 519–528.
- [35] Klein, D. J. and Randić, M. (1993). Resistance distance. *J. Math. Chem.*, **12**, 81–95.
- [36] Ivanciuc, O. and Balaban, A. T. (1994). Design of topological indices. Part 8. Path matrices and derived molecular graph invariants. *MATCH (Commun. Math. Chem.)*, **30**, 141–152.
- [37] Randić, M. (1997). Linear combinations of path numbers as molecular descriptors. *New J. Chem.*, **21**, 945–951.
- [38] Ivanciuc, O., Ivanciuc, T. and Balaban, A. T. (2000). The complementary distance matrix, a new molecular graph metric. *A C H - Model. Chem.*, **137**, 57–82.
- [39] Balaban, A. T., Mills, D., Ivanciuc, O. and Basak, S. C. (2000). Reverse Wiener indices. *Croat. Chem. Acta*, **73**, 923–941.
- [40] Randić, M. (1993). Novel molecular descriptor for structure-property studies. *Chem. Phys. Lett.*, **211**, 478–483.
- [41] Randić, M., Guo, X., Oxley, T. and Krishnapriyan, H. (1993). Wiener matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.*, **33**, 709–716.
- [42] Klein, D. J., Lukovits, I. and Gutman, I. (1995). On the definition of the hyper-Wiener index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.*, **35**, 50–52.
- [43] Diudea, M. V. (1996). Wiener and hyper-Wiener numbers in a single matrix. *J. Chem. Inf. Comput. Sci.*, **36**, 833–836.
- [44] Ivanciuc, O., Diudea, M. V. and Khadikar, P. V. (1998). New topological matrices and their polynomials. *Ind. J. Chem.*, **37A**, 574–585.
- [45] Ivanciuc, O., Taraviras, S. L. and Cabrol-Bass, D. (2000). Quasi-orthogonal basis sets of molecular graph descriptors as a chemical diversity measure. *J. Chem. Inf. Comput. Sci.*, **40**, 126–134.
- [46] Taraviras, S. L., Ivanciuc, O. and Cabrol-Bass, D. (2000). Identification of groupings of graph theoretical molecular descriptors using a hybrid cluster analysis approach. *J. Chem. Inf. Comput. Sci.*, **40**, 1128–1146.
- [47] Bonchev, D., Balaban, A. T., Liu, X. and Klein, D. J. (1994). Molecular cyclicity and centrality of polycyclic graphs. I. Cyclicity based on resistance distances or reciprocal distances. *Int. J. Quantum Chem.*, **50**, 1–20.
- [48] Zhu, H.-Y., Klein, D. J. and Lukovits, I. (1996). Extensions of the Wiener number. *J. Chem. Inf. Comput. Sci.*, **36**, 420–428.

- [49] Klein, D. J. and Gutman, I. (1999). Wiener-number-related sequences. *J. Chem. Inf. Comput. Sci.*, **39**, 534–536.
- [50] Gutman, I., Linert, W., Lukovits, I. and Tomović, Ž. (2000). The Multiplicative Version of the Wiener Index. *J. Chem. Inf. Comput. Sci.*, **40**, 113–116.
- [51] Zhu, H.-Y. and Klein, D. J. (1996). Graph-geometric invariants for molecular structures. *J. Chem. Inf. Comput. Sci.*, **36**, 1067–1075.
- [52] Klein, D. J. (1997). Graph geometry, graph metrics, & Wiener. *MATCH (Commun. Math. Comput. Chem.)*, **35**, 7–27.
- [53] Bytautas, L. and Klein, D. J. (1998). Chemical combinatorics for alkane-isomer enumeration and more. *J. Chem. Inf. Comput. Sci.*, **38**, 1063–1078.
- [54] Dobrynin, A. A. and Gutman, I. (1999). The average Wiener index of trees and chemical trees. *J. Chem. Inf. Comput. Sci.*, **39**, 679–683.
- [55] Bytautas, L. and Klein, D. J. (2000). Mean Wiener numbers and other mean extensions for alkane trees. *J. Chem. Inf. Comput. Sci.*, **40**, 471–481.
- [56] Hosoya, H. (1988). On some counting polynomials in chemistry. *Discrete Appl. Math.*, **19**, 239–257.
- [57] Sagan, B. E., Yeh, Y.-N. and Zhang, P. (1996). The Wiener polynomial of a graph. *Int. J. Quantum Chem.*, **60**, 959–969.
- [58] Estrada, E., Ivanciuc, O., Gutman, I., Gutierrez, A. and Rodríguez, L. (1998). Extended Wiener indices. A new set of descriptors for quantitative structure-property studies. *New J. Chem.*, **22**, 819–822.
- [59] Gutman, I., Estrada, E. and Ivanciuc, O. (1999). Some properties of the Wiener polynomial of trees. *Graph Theory Notes New York*, **36**, 7–13.
- [60] Konstantinova, E. V. and Diudea, M. V. (2000). The Wiener polynomial derivatives and other topological indices in chemical research. *Croat. Chem. Acta*, **73**, 383–403.
- [61] Ivanciuc, O., Ivanciuc, T., Klein, D. J., Seitz, W. A. and Balaban, A. T., Wiener index extension by counting even/odd graph distances. *J. Chem. Inf. Comput. Sci.*, in press.
- [62] Gakh, A. A., Gakh, E. G., Sumpter, B. G. and Noid, D. W. (1994). Neural network-graph theory approach to the prediction of the physical properties of organic compounds. *J. Chem. Inf. Comput. Sci.*, **34**, 832–839.