

Modification of the Wiener Index. 2

Feng Yang,* Zhen-Dong Wang, and Yun-Ping Huang

Department of Environmental and Chemical Engineering, Wuhan Institute of Science and Technology,
Wuhan, 430073, P.R. China

Pei-Jiang Zhou

College of Resources and Environment Science, Wuhan University, Wuhan, 430072, P.R. China

Received March 18, 2003

A novel topological index based on the Wiener Index is proposed as $W^* = 1/2 \sum_{i,j=1}^n S_{ij}^*$, the element S_{ij}^* of the distance matrix is defined either as $S_{ij}^* = \sqrt{E_i E_j / R_{ij}}$ (atoms i and j are adjacent) or as $S_{ij}^* = \sqrt[j-i+1]{E_i \times \cdots \times E_j / R_{ij}}$ (atoms i and j are not adjacent), where E_i and E_j represent the total energy of the valence electrons of vertexes i and j , respectively, R_{ij} is the sum of the distance between the vertexes i and j in a molecular graph. The distance and the energy of the vertexes in a molecule are taken into account in this definition. Hence the application of the index W^* to multiple bond and heteroatom-containing organic systems and inorganic systems is possible. Good correlation coefficients are achieved not only in the standard formation enthalpy of methyl halides, halogen–silicon, but also in the retention index of gas chromatography of the hydrocarbons.

1. INTRODUCTION

The quantitative structure–activity relationships (QSAR) and quantitative structure–property relationships (QSPR) techniques are that the chemical molecular structures possess relations with their properties/activities. An inspection of the published QSPR/QSAR models shows that molecular graph descriptors and molecular topological indices are used with success in modeling various properties. It also demonstrates that they are valuable descriptors of chemical structure.^{1–7} Together with the molecular connectivity indices,^{1,2} the Wiener index W is one of the most frequently used graph descriptors in QSPR/QSAR models.^{8–12} It was defined in 1947 for saturated acyclic hydrocarbons as the sum of the bonds between all pairs of non-hydrogen atoms in a molecule.⁸ Subsequently, the Hosoya index,¹³ the Balaban index,¹⁴ the hyper-Wiener Index,^{15–16} the Detour Index,¹⁷ the Hyper-Detour Index,¹⁸ the Harary Index¹⁹ and the Pasareti index²⁰ were proposed. They were all based on the distance-related matrices, such as the detour Δ ,²¹ the detour-distance $\Delta\text{-D}$,²¹ and the distance-valence Dval .²² These distance-related matrices were made up of the shortest distances from a vertex i to all the other $(n-1)$ vertexes in the molecular graphs, but the molecular graphs include more structure elements than both distances and vertexes. Therefore, those who define topological indices should take into consideration atomic and molecular structure. In our work,^{23,24} a novel topological index has been defined by the electronegativity and the number of bond electrons and has been applied to correlate the standard formative enthalpies of halide methyl, the standard formative enthalpies of SiX_n ($X = \text{F, Cl, Br, I}$, $n = 1-4$), the standard formative enthalpies of saturated hydrocarbons, and the retention indices of gas chromatography in unsaturated hydrocarbons. The electronegativity and the bond length were used to define the elements of the distance matrix in a molecular graph, and a novel distance

topological index was proposed.²⁵ Good correlation coefficients were achieved in the prediction of the retention indices of gas chromatography of the hydrocarbons and the standard formation enthalpy of methyl halides, halogen–silicon, and inorganic compounds containing transition metals.

We know that two well-known concepts in the quantum chemistry, highest occupied molecular orbital energy (E_{HOMO}) and lowest unoccupied molecular orbital energy (E_{LUMO}), play an important role in QSAR/QSPR studies. That the descriptors, E_{HOMO} and E_{LUMO} , are widely used in QSPR/QSAR models by the physical, organic, analytical, medicinal chemistry, biochemistry, toxicology, and environmental sciences.^{26–31} It shows that energy is an important descriptor in QSAR/QSPR studies and an important property of the molecules. Therefore, energy should be considered in the molecular structure while defining topological indices. In this paper, we consider the energies of atoms in the molecular structure, redefine the elements of the distance matrix, propose a novel distance topological index, and extend the applicable scope of the distance matrix index.

2. DEFINITION

The distance matrix S is composed of the shortest distances from one atom to the others in a molecular graph. If a molecule contains n atoms, there is a symmetric matrix S :

$$S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix} \quad (1)$$

The Wiener number W^8 is

$$W = \frac{1}{2} \sum_{i,j=1}^n S_{ij} \quad (2)$$

* Corresponding author phone: +86-27-87801395-626; fax: +86-27-87800795; e-mail: fengy@wist.edu.cn.

In Wiener's study, W only considered the length of the edge but not the properties and the interaction of the vertexes. Energy is one of the main properties of atoms and has multiple relations with the properties of chemical molecules. Furthermore, the interaction between atoms by which the atomic energies are reassigned among them involves not only their distance but also their energies. The measurement elements of the foregoing distance-related matrices only considered their distances but not their energies. So the energy and the distance between atoms are used to reveal the properties and the interaction of the vertexes in a molecule.

Energy of Valence Electrons. The average energy of valence electrons concerns not only the principal quantum number n_i but also the azimuthal quantum number l_i of the atom. The orbital number of a subshell is $2l_i + 1$ while n_i is the same but l_i is different, and the larger the orbital number of the subshell electronic shell, the higher the electronic energy level of the subshell and the stronger the bonding ability. The energy level of the subshell s , in which there is one orbit, concerns only the principal quantum number n_i of the atom; therefore, the energy e_i of an electron in this subshell is defined as follows:

$$e_i = n_i + \sqrt{(2l_i + 1) - 1} = n_i + \sqrt{2l_i} \quad (3)$$

For the energy levels s , p , d , and f , the values of l_i are 0, 1, 2, and 3, and the values of e_i are respectively n_i , $n_i + 1.4$, $n_i + 2$, and $n_i + 2.4$.

The average energy \bar{e}_i of valence electrons can easily be calculated by using formula 4

$$\bar{e}_i = \frac{\sum e_i \cdot c_i}{\sum c_i} \quad (4)$$

where $e_i = n_i + \sqrt{2l_i}$, and c_i is the electron number of the subshell.

Therefore, the total energy of the valence electrons E_i can be defined as follows:

$$E_i = \bar{e}_i \cdot z_i \quad (5)$$

Here z_i is the number of valence electrons.

Elements of Distance Matrix. In our early article,²⁵ the idea of using Pauling's electronegativity in computing a distance topological index was already adopted to define the distance matrix S . Here, E_i is a substitute for Pauling's electronegativity. S_{ij}^* is redefined as follows:

$$S_{ij}^* = \begin{cases} \frac{\sqrt{E_i E_j}}{R_{ij}} & i, j \text{ are adjacent} \\ \frac{j-i+1 \sqrt{E_i E_k \cdots E_j}}{R_{ij}} & i, j \text{ are not adjacent} \\ 0 & i = j \end{cases} \quad (6)$$

In formula 6, E_i and E_j refer to the total energy values of the valence electrons of vertexes i and j , respectively. R_{ij} is the sum of the distance between vertexes i and j .

Chart 1. A Molecular Graph of Five Linked Atoms

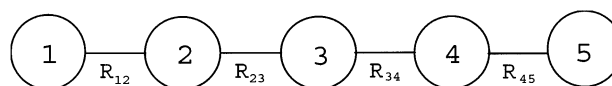


Table 1. \bar{e}_i , E_i , and d_i of Some Atoms in Methyl Halides

atoms	C	H	F	Cl	Br	I
\bar{e}_i	2.7	1	3	4	5	6
E_i	10.8	1	21	28	35	42
d_i	1	1	1	9	27	45

In Randic-Kier's work,¹⁻² δ_i^ν of vertex i was defined

$$\delta_i^\nu = \frac{z_i}{Z - z_i - 1} \quad (7)$$

In this formula, Z is the atomic number of atom i , and z_i is the number of valence electrons. We noticed that $Z - z_i$ means the number of inner-shell electrons and can reveal the size of atom i . It shows that the number of inner-shell electrons can reveal the distance between the vertexes. To keep the same form as Randic-Kier's, let $d_i = |Z - z_i - 1|$. Then, if vertexes i and j are adjacent, R_{ij} can be obtained by the formula $R_{ij} = d_i + d_j$; if vertexes i and j are not adjacent, R_{ij} is the sum of all the distance of adjacent atoms from vertexes i to j in the molecular graph.

In Chart 1, where there are five linked atoms in a molecule, the values of S_{ij}^* were calculated as follows:

$$S_{12}^* = \frac{\sqrt{E_1 E_2}}{R_{12}}, S_{23}^* = \frac{\sqrt{E_2 E_3}}{R_{23}}, S_{34}^* = \frac{\sqrt{E_3 E_4}}{R_{34}}, S_{45}^* = \frac{\sqrt{E_4 E_5}}{R_{45}}$$

$$S_{13}^* = \frac{\sqrt[3]{E_1 E_2 E_3}}{R_{12} + R_{23}}, \dots, S_{35}^* = \frac{\sqrt[3]{E_3 E_4 E_5}}{R_{34} + R_{45}}, \dots; S_{15}^* = \frac{\sqrt[5]{E_1 E_2 E_3 E_4 E_5}}{R_{12} + R_{23} + R_{34} + R_{45}}.$$

The definition of S_{ij}^* infers that the longer the distance between vertexes i to j , the smaller S_{ij}^* . The interaction between vertexes i and j becomes weaker indeed, when the distance between vertexes i and j grows.

A Novel Distance Index W^* . According to formulas 2 and 6, a novel topological index W^* takes the place of the old Wiener index W (formula 8).

$$W^* = \frac{1}{2} \sum_{i,j=1}^n S_{ij}^* \quad (8)$$

3. RESULTS AND DISCUSSION

For Methyl Halides. For such small molecules as methyl halides, hydrogen atoms have a great influence on their properties, so the effect of hydrogen atoms cannot be neglected. Hydrogen atoms must be treated as vertexes in a molecular graph. For some atoms of methyl halides, such as C, H, F, Cl, Br, and I, \bar{e}_i , E_i and d_i are listed in Table 1. For some bonds of methyl halides, such as C-H, C-F, C-Cl, C-Br, C-I, F-C-H et al., R_{ij} and S_{ij}^* are listed in Table 2.

Table 2. R_{ij} and S_{ij}^* of Some Bonds in Methyl Halides

bonds	R_{ij}	S_{ij}^*	bonds	R_{ij}	S_{ij}^*
C–F	2	7.5299	Cl–C–Cl	20	1.0191
C–Cl	10	1.7390	Cl–C–Br	38	0.5778
C–Br	28	0.6944	Cl–C–I	56	0.4166
C–I	46	0.4630	Cl–C–H	12	0.5593
C–H	2	1.6432	Br–C–Br	56	0.4223
F–C–F	4	4.2062	Br–C–I	74	0.3396
F–C–Cl	12	1.5432	Br–C–H	30	0.2410
F–C–Br	30	0.6649	I–C–I	92	0.2903
F–C–I	48	0.4416	I–C–H	48	0.1601
F–C–H	4	1.5246	H–C–H	4	0.5526

Now we take an example to illustrate how to get W^* . For the CHFCIBr molecule, the central atom C is labeled as “1”, and the others are labeled as “2”, “3”, “4”, and “5” at random. Their bond lengths with the central atom C are successively R_1 , R_2 , R_3 , and R_4 , respectively. W^* of the CHFCIBr molecule can be expressed as follows:

$$W^* = \frac{1}{2} \sum_{i,j=1}^5 S_{ij}^*$$

$$= \frac{\sqrt{E_1 E_2}}{R_1} + \frac{\sqrt{E_1 E_3}}{R_2} + \frac{\sqrt{E_1 E_4}}{R_3} + \frac{\sqrt{E_1 E_5}}{R_4} + \frac{\sqrt[3]{E_1 E_2 E_3}}{R_1 + R_2} + \frac{\sqrt[3]{E_1 E_2 E_4}}{R_1 + R_3} + \frac{\sqrt[3]{E_1 E_2 E_5}}{R_1 + R_4} + \frac{\sqrt[3]{E_1 E_3 E_4}}{R_2 + R_3} + \frac{\sqrt[3]{E_1 E_3 E_5}}{R_2 + R_4} + \frac{\sqrt[3]{E_1 E_4 E_5}}{R_3 + R_4}$$

$$= 1.6432 + 7.5299 + 1.7390 + 0.6944 + 1.5246 + 0.5593 + 0.2410 + 1.5432 + 0.6649 + 0.5778 = 16.7173$$

Listed in Table 3 are the values of W^* and the standard formation enthalpy $\Delta_f H_m^\Phi$ ($\text{kJ}\cdot\text{mol}^{-1}$)³² of the corresponding molecules, where W^* is calculated using formulas 6 and 8, and $\Delta_f H_m^\Phi$ is from ref 32. The linear correlation (LR) result is

$$\Delta_f H_m^\Phi = 539.7319 - 192.4361 \sqrt{W^*} \quad (9)$$

$$r = 0.9908, s = 36.9247, F = 2084.28, N = 41$$

Here r is the correlation coefficient, s is the standard error, F means the Fischer value, and N is the regression number of molecules.

The calculated (calc) $\Delta_f H_m^\Phi$ values by formula 9 and residuals (res) are given in Table 3.

The above-mentioned result indicates that there is a good correlativity between W^* and $\Delta_f H_m^\Phi$ for methyl halides, hence W^* can be applied to heteroatom-containing organic systems.

For Inorganic Compounds of Halogen–Silicon. Listed in Table 4 are the values of W^* and the standard formation enthalpy $\Delta_f H_m^\Phi$ ($\text{kJ}\cdot\text{mol}^{-1}$)³³ of the corresponding molecules. W^* is calculated by formulas 6 and 8, and $\Delta_f H_m^\Phi$

Table 3. W^* , $\Delta_f H_m^\Phi$ ($\text{kJ}\cdot\text{mol}^{-1}$)³² (Observed, Calculated), and Residuals Values of Methyl Halides

no.	compd	W^*	obsd ^a	calcd ^b	res ^c	pres ^d
1	CH ₂ F ₂	29.2034	−452.9	−500.2	47.3	50.0
2	CCl ₂ F ₂	29.9359	−493.3	−513.2	19.9	21.0
3	CH ₂ Cl ₂	10.5733	−95.4	−86.0	−9.4	−9.7
4	CHCl ₂ F	19.3998	−284.9	−307.9	23.0	23.6
5	CB ₂ F ₂	23.7367	−429.7	−397.8	−31.9	−33.1
6	CB ₂ Cl ₂	8.6194	−29.3	−25.2	−4.1	−4.2
7	CH ₂ Br ₂	6.6141	−14.8	44.8	−59.6	−62.6
8	CHBr ₂ F	14.3206	−223.4	−188.5	−34.9	−35.8
9	CHBr ₂ Cl	7.3902	−20.9	16.6	−37.5	−39.2
10	CB ₂ ClF	15.1086	−231.8	−208.3	−23.5	−24.1
11	CHBr ₃	5.7163	16.7	79.6	−62.9	−66.6
12	CB ₃ F	12.8747	−190.0	−150.8	−39.2	−40.3
13	CB ₃ Cl	6.8225	12.6	37.1	−24.5	−25.7
14	CHI ₃	4.3834	210.9	136.8	74.1	79.4
15	CF ₄	55.3568	−933.0	−892.0	−41.0	−50.4
16	CCl ₄	13.0706	−95.8	−156.0	60.2	61.8
17	Cl ₄	3.5938	262.9	174.9	88.0	95.3
18	CBr ₄	5.3114	79.5	96.2	−16.7	−17.8
19	CHBrF ₂	26.2236	−463.6	−445.7	−17.9	−18.7
20	CHBrCl ₂	9.3499	−58.6	−48.7	−9.9	−10.3
21	CBrCl ₂ F	17.6283	−269.4	−268.2	−1.2	−1.2
22	CH ₃ Br	8.0048	−37.7	−4.7	−33.0	−34.3
23	CBrClF ₂	26.6934	−471.5	−454.5	−17.0	−17.8
24	CH ₂ BrCl	8.4508	−50.2	−19.7	−30.5	−31.7
25	CHBrClF	16.7173	−295.0	−247.1	−47.9	−49.1
26	CH ₂ BrI	6.1382	50.2	63.0	−12.8	−13.5
27	CHF ₃	41.4253	−693.3	−698.8	5.5	6.2
28	CF ₃ I	36.9961	−589.9	−630.7	40.8	44.7
29	CCl ₃ F	20.4338	−284.9	−330.2	45.3	46.6
30	CHF ₂ Cl	29.3431	−483.7	−502.7	19.0	20.1
31	CF ₃ Cl	41.5769	−707.9	−701.1	−6.8	−7.6
32	CF ₃ Br	37.8974	−648.9	−644.9	−4.0	−4.4
33	CH ₃ F	18.6911	−237.7	−292.2	54.5	56.0
34	CH ₂ I ₂	5.6957	118.4	80.5	37.9	40.1
35	CH ₂ FCI	17.2757	−264.4	−260.1	−4.3	−4.4
36	CH ₂ ClI	7.8964	12.6	−1.0	13.6	14.2
37	CH ₃ I	7.5307	13.8	11.6	2.2	2.2
38	CH ₂ FB ₂	16.2594	−252.7	−236.2	−16.5	−16.9
39	CH ₄	9.8884	−74.9	−65.4	−9.5	−9.8
40	CBrCl ₃	10.7021	−37.2	−89.8	52.6	54.2
41	CHCl ₃	11.5954	−102.9	−115.6	12.7	13.0

^a Observed $\Delta_f H_m^\Phi$ value.³² ^b $\Delta_f H_m^\Phi$ value calculated from QSPR formula 9. ^c Obsd – calc. ^d Residual for compounds during leave-one-out (LOO) cross validation.

Table 4. W^* and $\Delta_f H_m^\Phi$ ($\text{kJ}\cdot\text{mol}^{-1}$)³³ of SiH_nX_m (X = F, Cl, Br, I, H, $m + n = 1-4$)^a

compd	W^*	$\Delta_f H_m^\Phi$	compd	W^*	$\Delta_f H_m^\Phi$
SiF	1.7630	−20.92	SiH	0.3847	377.4
SiF ₂	4.4604	−589.94	SiH ₂	0.8922	271.96
SiF ₃	8.0922	−999.98	SiH ₃	1.5225	194.97
SiF ₄	12.6584	−1625.9	SiH ₄	2.2756	34.31
SiCl	1.1309	154.81	SiH ₃ F	4.3016	−376.56
SiCl ₂	2.9964	−167.78	SiH ₂ F ₂	6.7074	−790.78
SiCl ₃	5.5965	−334.72	SiHF ₃	9.4930	−1200.81
SiCl ₄	8.9312	−662.75	SiH ₃ Cl	3.4523	−135.56
SiBr	0.6322	196.65	SiH ₂ Cl ₂	4.8481	−315.06
SiBr ₂	1.7461	−46.02	SiHCl ₃	6.4630	−499.15
SiBr ₃	3.3417	−158.99	SiH ₃ Br	2.6785	−64.02
SiBr ₄	5.4190	−415.47	SiH ₂ Br ₂	3.2199	−180.75
SiI	0.6170	259.41	SiHBr ₃	3.8998	−303.34
SiI ₂	1.6019	92.05	SiH ₃ I	2.5394	−2.09
SiI ₃	2.9547	58.58	SiH ₂ I ₂	2.9341	−38.07
SiI ₄	4.6754	−110.46	SiHI ₃	3.4597	−74.48

^a Especially, for molecules of type AB_m, W^* can be simplified.

is from ref 33. The correlation result is shown in formula 10.

$$\Delta_f H_m^\Phi = 356.1807 - 150.1426 W^* \quad (10)$$

$$r = 0.9631, s = 121.8738, F = 383.87, N = 32$$

Table 5. Statistics Parameters of LR and LOO for the Two Final Models

propriety	<i>r</i>	<i>s</i>	<i>SEP</i>	<i>r</i> _{cv}	<i>s</i> _{cv}	<i>SEP</i> _{cv}
$\Delta_f H_m^\Phi$ ^a	0.9908	36.9289	36.0168	0.9897	39.0600	38.0962
$\Delta_f H_m^\Phi$ ^b	0.9631	121.8738	116.3189	0.9564	132.2674	128.0713

^a $\Delta_f H_m^\Phi$ of methyl halides. ^b $\Delta_f H_m^\Phi$ of halogen-silicon.

W^* has a good correlation with $\Delta_f H_m^\Phi$ for SiH_nX_m ($\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}, \text{H}, m + n = 1-4$) inorganic compounds, too.

Especially, for molecules of type AB_m , W^* can be simplified.

$$W^* = \frac{m}{d_1 + d_2} \left(\sqrt{E_1 E_2} + \frac{m-1}{4} \cdot \sqrt[3]{E_1 E_2 E_2} \right) \quad (11)$$

“1” means the central atom A, “2” means the atom B. For SiX_m ($\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}, \text{H}, m = 1-4$) inorganic compounds in Table 4, W^* can easily be calculated by using the formula 11.

For Hydrocarbons. In the hydrogen-suppressed graph of the organic compounds, hydrogen atoms were neglected. To differentiate between carbon-carbon bonds of $\text{C}-\text{C}$, $\text{C}=\text{C}$ and $\text{C}\equiv\text{C}$, according to ref 25, here R_{ij} also takes the relative values of the bond length. Then, there are the same correlation results for hydrocarbons between this paper and in the aforementioned article (see ref 25).

What proceeded shows that W^* has the ability to correlate the properties of hydrocarbons with multiple bonds. Especially, W^* can be simplified for saturated hydrocarbons.

$$W^* = \frac{1}{2} k \sum_{i,j=1}^n \{1/R_{ij}\} \quad (12)$$

Here k is a constant. Therefore, W^* is the index D^{34} for saturated hydrocarbons, indeed. If $1/R_{ij}$ stands for R_{ij} in the formula 12, W^* is the Wiener number W^6 . But D or W cannot be applied in above heteroatom-containing compounds.

Model Validation. An important aspect of QSAR modeling is the development of means for the validation of the model. To attempt the model validation, two approaches to validation of the models were adopted for these data sets.

(1) Cross-Validation. The cross-validation is a practical and reliable method for testing the significance of a model. Hence, a full cross-validation test using a leave-one-out (LOO) method is used to determine the validity of all models obtained. For on observation left out of the whole data set, a model was developed from the remaining of the data. The compound left out was predicted by that model. This process was carried out on all subsets. In this way, every observation was predicted once. (The predicted residuals of methyl halides (pres) are listed in Table 3.) In principle, the performance of the models (its predictive ability) can be given by the standard error of prediction (*SEP*). Also, the predictive ability of the model is also quantified in terms of the corresponding LOO cross-validated parameters, r_{cv} , s_{cv} , and SEP_{cv} .

As a quantitative evaluation of the cross-validated results, the statistical parameters (r , s , and SEP) of the final models and the corresponding LOO cross-validated parameters (r_{cv} , s_{cv} , and SEP_{cv}) of the Jack-knifed approach are also listed in Table 5. It is expected that the values of SEP_{cv} should be

slightly larger than the values of *SEP*. One can see that for the studied properties the values of r_{cv} and s_{cv} are very close to the values of r and s , and both s_{cv} and SEP_{cv} are only slightly larger than s and *SEP*. The cross-validation results demonstrate the final models to be statistically reliable. On the other hand, the values of the predicted residuals (from LOO) are randomly distributed. The distributions of $\Delta_f H_m^\Phi$ of methyl halides and halogen-silicon are from $-933.0 \text{ kJ}\cdot\text{mol}^{-1}$ to $262.9 \text{ kJ}\cdot\text{mol}^{-1}$ and from -1625.9 to $377.4 \text{ kJ}\cdot\text{mol}^{-1}$; although the standard errors are large, we considered the final models are statistically significant and successfully validated.

(2) Randomization Test. For methyl halides, a set of 10 compounds was randomly selected as an initial validation test set. These compounds were selected randomly so that none of the compounds possesses the maximum or minimum $\Delta_f H_m^\Phi$ values among the 41 observations in the total set. The number of these compounds in Table 3 is 1, 5, 9, 13, 21, 25, 29, 33, 37, and 41. A model based on $\sqrt{W^*}$ was obtained from the 31 observations remaining after removal of the validation test set. The model was then used to predict $\Delta_f H_m^\Phi$ values for the validation test set. The mean absolute error (MAE) for the test set ($N = 10$) obtained was $\text{MAE} = 30.9278 \text{ kJ}\cdot\text{mol}^{-1}$; the computed root-mean-square (RMS) for this predicted $\Delta_f H_m^\Phi$ values is $\text{RMS} = 38.6935 \text{ kJ}\cdot\text{mol}^{-1}$. And the values of MAE and RMS for the total data set were $29.0953 \text{ kJ}\cdot\text{mol}^{-1}$ and $36.4643 \text{ kJ}\cdot\text{mol}^{-1}$, respectively. Although no accurate estimate is available for the experimental error in this data set, an MAE value of $30.9278 \text{ kJ}\cdot\text{mol}^{-1}$ and a RMS value of $38.6935 \text{ kJ}\cdot\text{mol}^{-1}$ appear to be quite reasonable. This result tends to confirm the significance of the descriptors and the model based on them. For halogen-silicon, the same procedure will be performed; the similar results can confirm that the final model is statistically significant and validated as for methyl halides.

The overall indication of these validation studies is that the models appears reasonable and likely to give useful estimations of these heteroatom-containing compounds.

The above-mentioned result indicates that there is a good correlativity between W^* and $\Delta_f H_m^\Phi$ for methyl halides, hence W^* can be applied in heteroatom-containing organic systems. For the standard formation enthalpy of these compounds, the final models obtained are statistically significant and successfully validated and have some predicted ability.

CONCLUSION

In the definition of S_{ij}^* the denominator represents the distance and the numerator represents the energy; therefore, S_{ij}^* has the meanings of potential energy.

Our study proves that W^* can be applied not only in heteroatom-containing and multiple bond-containing organic compounds but also in inorganic compounds.

ACKNOWLEDGMENT

The authors wish to express their thanks to Professor Jin-Xin Dai and Dr. Hai-Liang Zhu in WIST for their help with the manuscript.

REFERENCES AND NOTES

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

- (2) Kier, L.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, U.K., 1986.
- (3) Xu, L.; Zhong, F. Method of Topological Indices in Quantitative Study of Structure Property/Activity Relationships. *Chin. J. Appl. Chem. (in Chinese)* **1997**, *14*, 26–29.
- (4) Balaban, A. T. *From Chemical Graphs to 3D Molecular Modeling. In From Chemical Topology to three-Dimensional Geometry*; Balaban, A. T., Eds.; Plenum: New York, 1997.
- (5) 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 and Breach Science Publishers: Amsterdam, 1999.
- (6) Weida, D. R.; Rerkins, R. Y.; Welsh, W. J.; shechan, D. Evaluation of Quantitative Structure–Activity Relationship Methods for Large-Scale Prediction of Chemicals Binding to Estrogen Receptor. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669–677.
- (7) Cao, C. Z.; Hua, Y. Topological Indices Based on Vertex, Distance, and Ring: On the Boiling Points of Paraffins and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 867–877.
- (8) Wiener H. Structural determination of Paraffin Boiling Points. *J. Am. Chem. Soc.*, **1947**, *69*, 17–20.
- (9) Fritsche, H.-G.; Bonchev, D.; Mekenyan, O. The Optimum Topology of Small Clusters. *Z. Phys. Chem. (Leipzig)* **1989**, *270*, 467–476.
- (10) Bonchev, D.; Mekenyan, O. A Topological Approach to Infinite Conjugated Polymers. *Z. Naturforsch.* **1980**, *35a*, 739–747.
- (11) Mrkenyan, O.; Dimitrov, S.; Bonchev, D. Graph-Theoretical Approach to the Calculation of Physico-Chemical Properties of Polymers. *Eur. Polymer J.* **1983**, *12*, 1185–1193.
- (12) Bonchev, D.; Mekenyan, O.; Kamenska, V. A Topological Approach to the Modeling of Polymer Properties. *J. Math. Chem.* **1992**, *11*, 107–132.
- (13) Hosoya, H. A Newly Proposed Quantity Characterizing the Topological Nature of structural Isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (14) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (15) Randic, M. Novel Molecular Descriptor for Structure–Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478–483.
- (16) Gordon, C.; Sandi, K.; Marko, P. Three Methods for Calculation of the Hyper-Wiener Index of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 571–574.
- (17) Lukovits, I. The Detour Index. *Croat. Chem. Acta* **1996**, *69*, 873–882.
- (18) Linert, W.; Lukovits, I. Formulas for the Hyper-Wiener and Hyper-Detour Indices of Fused Bicyclic Structures. *Comm. Math. Comput. Chem.* **1997**, *35*, 65–74.
- (19) Plavsic, D.; Nikolic, S.; Trinajstić, N.; Mihalic, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235–250.
- (20) Lukovits, I. An All-Path Version of the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 125–129.
- (21) Ivanciuc, O.; Balaban, A. T. Design of Topological Indices. Path Matrices and Derived Molecular Graph Invariants. *Comm. Math. Comput. Chem.* **1994**, *30*, 141–152.
- (22) Ivanciuc, O. Design of Topological Indices. Part II. Distance-Valence Matrices and Derived Molecular Descriptors. *Rev. Roum. Chem.* **1999**, *44*, 619–631.
- (23) Yang, F.; Wang, Z.-D.; Sun, J. Study on the enlarged Topological Index of Distance Matrix and Its Application. *Acta Chim. Sinica (in Chinese)* **2002**, *60*, 764–768.
- (24) Wang, Z.-D.; Yang, F.; Zhou, P.-J. Research on the features of some organic Pollutants with structure index W*. *J. Wuhan Univ. (Nat. Sci. Ed.) (in Chinese)* **2002**, *48*, 672–678.
- (25) Yang, F.; Wang, Z.-D.; Huang Y.-P.; Ding, X.-R.; Zhou, P.-J. Modification of Wiener Index and Its Application. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 753–756.
- (26) Dewa M. J.; Zoeibisch E. G.; Healg E. F. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *104*, 3092–3909.
- (27) Grimmer, G. *Environmental Carcinogens: Polycyclic Aromatic Hydrocarbons, Chemistry, Occurrence, Biochemistry, carcinogenicity*; CRC Press: Boca Raton, FL, 1983.
- (28) Mekapati, S. B.; Hansch, C. On the Parametrization of the Toxicity of Organic Chemicals to Tetrahymena Pyriformis. The Problem of Establishing a Uniform Activity. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 956–961.
- (29) Yan, X.-C.; Li, X.-F.; Luo, M.-Dao; Zhang G.-Y. Studies on surface Tension of Alkyl sulfate Using a Quantum chemistry Method. *J. Wuhan Univ. (Nat. Sci. Ed.) (in Chinese)* **2002**, *48*, 655–688.
- (30) Chen, J. W.; Kong, L. R.; Zhu, C. M. Correlation between photolysis rate constants of Polycyclic Aromatic Hydrocarbons and Frontier Molecular Orbital energy. *Chemosphere* **1996**, *33*, 1143–1150.
- (31) Hansch, C.; Gao, H. Comparative QSAR: Radical Reactions of benzene Derivatives in Chemistry and Biology. *Chem. Rev.* **1997**, *97*, 2995–3059.
- (32) Dean J. A. *Lang's Handbook of Chemistry*; 13th ed.; McGraw-Hill Book Company: 1985.
- (33) Walsh, R. Thermochemistry of Silicon-Containing Compounds. *J. Chem. Soc.* **1983**, *79*, 2233–2248.
- (34) Mihalic Z.; Veljanić, D.; Amić D. The Distance Matrix in Chemistry. *J. Math. Chem.* **1992**, *11*, 223–258.

CI0340512