Wiener−Hosoya Index − A Novel Graph Theoretical Molecular Descriptor[†]

Milan Randić*

3225 Kingman Road, Ames, Iowa 50014, and National Institute of Chemistry, Ljubljana, Slovenia

Received December 3, 2003

We describe the construction of a novel molecular descriptor, called the Wiener−Hosoya index, in view of its structural relationship to both the Wiener number W and the Hosoya topological index Z. It is shown that this index has a smaller degeneracy than many simple topological indices, including W, Z, and the connectivity index χ. In a way the index can be viewed as a particular generalization of the Wiener number.

INTRODUCTION

Among the oldest and still widely used molecular descriptors in structure−property-activity studies we find the Wiener number W,[1] the Hosoya topological index Z,[2] and the molecular connectivity index χ.[3] These three indices may also be the three most studied indices from the mathematical point of view. For example, an alternative definition of the Wiener index follows from the summation of the entries in the distance matrix above the main diagonal.[2] On the other hand, it can also be obtained from roots of the Laplacian matrix.[4] The Hosoya index Z is related to the characteristic polynomial,[2,5] and the connectivity index, as has only recently been pointed out, can be defined in terms of atomic contributions[6] and relates also to the elements of the normalized Laplace matrix.[7] Considerable interest in these three molecular descriptors is in part due to their "historical" role: The Wiener index is the first nontrivial molecular descriptor *used in combination* with paths of length three for structure−property correlations; the Hosoya Z index is the first molecular invariant *found* useful in simple regressions (based on a single descriptor); and the connectivity index is the first molecular descriptor *designed* for use in structure−property correlations. However, since then, particularly during the past decade, hundreds of various molecular descriptors have been proposed and used in structure−property-activity studies, to the point that questions can be raised not only on the justification of such designs but also on any genuine need for construction of additional new descriptors. Despite the fact that introduction of the novel molecular descriptors (often referred to those based on molecular graphs as topological indices) contributes further to yet another problem, that of selection of descriptors from a large pool, we firmly believe that new indices are not only desirable but also important for the further development of QSAR and QSPR (quantitative structure−activity relationship and quantitative structure−property relationship, respectively), just as new molecular models are not only desirable but also important for a better understanding of structure−activity relationships and structure−property relationships. A novel index may have a simpler relationship to a molecular structure and can therefore give better insight into the structure−property relationship, and also it may more clearly point out which particular structural elements are dominant for which molecular property. If such an index replaces linear combinations of two or more descriptors by a single descriptor, besides producing a better regression, it will facilitate a better interpretation of the model. An important benefit of replacing multivariate regressions with simple regressions is to avoid the pitfalls of the instabilities of the multivariate regression equations upon introduction of novel descriptors, which require orthogonalization of descriptors.[8,9] In fact, the above was the rationale for the recently reintroduced variable molecular descriptors[10−12] and the variable connectivity index in particular.[13−19] The first papers on the variable descriptors had appeared about a decade earlier.[20,21]

THE WIENER−HOSOYA INDEX

To describe the various properties of alkanes and other organic compounds, Wiener introduced two descriptors W and P to be combined in the regression analysis. He referred to W as "the path number," but Platt[22] called it the Wiener number, the name that has been exclusively used ever since. The accompanying index P, which represents the count of paths of length three, was called by Wiener the "polarity number," which is occasionally still so labeled. For W (which had only so far been defined for acyclic molecules) Wiener suggested the following interpretation: "*The path number W was calculated as the total distance between all carbon atoms. The smaller this total distance, the larger is the compactness of the molecule*." To calculate W the following algorithm was proposed:

"*The path number W is defined as the sum of the distances between any two carbon atoms in the molecule, in terms of carbon−carbon bonds. Brief method of calculation: Multiply the number of carbon atoms on one side of any bond by those on the other side; W is the sum of these two values for all bonds.*" However, as Hosoya found out, W can also be obtained by simply adding all the elements of the graph distance matrix above the main diagonal. This not only offers an alternative route to W but also allows a particular extension of W to cyclic structures!

Calculation of the Hosoya topological index, which is obtained by counting the *k* disjoint edges in a graph (for *k* = 0, 1, 2, 3, ...), is not as straightforward as it may seem, because of the combinatorial "explosion" in the case of the

[*] Corresponding author fax: (515)292-8629; e-mail: mrandic@msn.com. Professor Emeritus, Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311.

[†] This article is dedicated to W. G. A. Milne, on the occasion of his retirement after serving for 15 years as the Editor-in-Chief of the *Journal of Chemical Information and Computer Sciences.*

10.1021/ci030425f CCC: $27.50 © 2004 American Chemical Society
Published on Web 02/26/2004

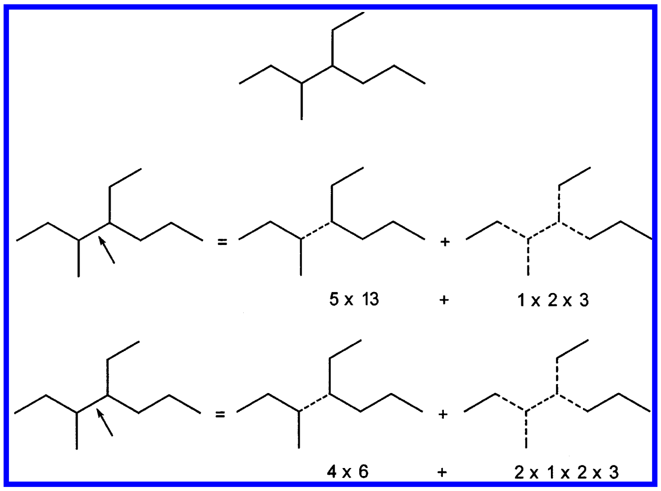

Figure 1. Construction of the Hosoya index Z (middle) and the Wiener−Hosoya index W−H (bottom).

larger structures. However, Hosoya and co-workers[23] reported an elegant composition principle, which can be expressed as the following: Z (G) = Z (G − e) + Z (G − ee).

Here (G − e) is a graph with a single edge being deleted, and (G − ee) is the same graph with deletion of the edge *e* and all incident edges to it. We will refer to this sometimes informally as operation *ee*. In Figure 1 we illustrate the composition principle on the hydrogen-depleted graph of 3-methyl-4-ethylheptane, which appeared in the most recent publication of H. Hosoya.[24] It was in fact looking at this figure that inspired this author to come forward with a novel index, to be called the Wiener−Hosoya index, for reasons that will soon become apparent. The importance of the composition principle is that it breaks a large molecule into smaller fragments, for which *Z* values may be known or are easily calculated. Observe that the first part of the illustration of the composition principle corresponds to the way Wiener suggested for the calculation of W if this is extended to all bonds. This observation suggested that we may use the Hosoya composition principle but instead of calculating Z (shown in the middle of Figure 1) we calculate W for the given components. If we now extend the same for all bonds in the structure we will obtain from the first fragment W and from the remaining parts an incremental contribution ΔW. We defined the new Wiener−Hosoya index (W−H) as W + ΔW. An alternative format of the same definition is to write the following: W−H (G) = W(G) + W(G − ee).

Figure 1 thus graphically shows the calculation of a contribution of a single bond to W−H (the bond that has been erased). Using the procedure that Wiener proposed for calculating W one obtains the following: $4 \times 6 + 2 \times 1 \times 2 \times 3 = 24 + 12 = 36$. To obtain W−H we have to do the same computations for the remaining eight CC bonds. It seems quite appropriate to call this index the "Wiener−Hosoya" index, because the first term, when summation over all bonds is made, gives the W number, and contributions of other terms arise from the algorithm that Hosoya had used in calculating Z topological index.

SOME PROPERTIES OF THE WIENER−HOSOYA INDEX

In Table 1 we have listed the Wiener−Hosoya indices for the nine heptane isomers. As we can see, just as the Wiener index decreases with the apparent increase in molecular branching, so also decreases ΔW. The combination of the two parts increases the range of the value of W−H for isomers, which diminishes the chance for degeneracy that happens when two molecules have the same values of the index. This is an important quality of the W−H indices, which in the case of W and Z as well as the connectivity index χ is evident among heptanes (W) and octanes (Z, χ). As we see from Table 1, the degeneracy between 2,2-dimethylpentane and 2,3-dimethylpentane as well as between 2,4-dimethylpentane and 3-ethylpentane has been lifted.

Table 1. Wiener−Hosoya Indices for the Nine Isomers of Heptane Partitioned into Contributions Coming from the Wiener Number and the Remaining Portion of the Composition Principle

isomer	Wiener	Σ (G-ee)	Wiener−Hosoya
C_7	56	+ 20	76
2M C_6	52	+ 13	65
3M C_6	50	+ 11	61
3E C_5	48	+ 12	60
2,2MM C_5	46	+ 6	52
2,3MM C_5	46	+ 5	51
2,4MM C_5	48	+ 6	54
33,3MM C_5	44	+ 4	48
2,3,4MMM C_4	42	+ 1	43

The calculation of ΔW is rather simple because the components obtained after the *ee* operation tend to be small and the corresponding W can be calculated fast. Terminal bonds do not contribute to the sum of (G−ee) because the operation *ee* eliminates terminal vertices and thus adds a factor of zero. Factor 1 remains as one of the factors when next to terminal bonds are eliminated.

That W−H has a visibly lower degeneracy than many simple topological indices has been established by examining all the degenerate cases of the graphs among octanes (3 pairs), nonanes (5 pairs, 5 triples, and 1 quadruple), and decane isomers (14 pairs, 2 triples, 5 quadruples). In Table 2 we listed the corresponding W and ΔW for heptanes, octanes, and nonanes, from which we can see that introducing ΔW has lifted the degeneracy in all cases except four pairs of 75 nonane isomers. We should add, however, that in Table 2 we have listed structures that have initially the same W, but W−H turned out to differentiate between them. However, it is possible that two alkane isomers having different W have the same W−H, because the increments ΔW could make different contributions to each of them resulting thus in the same W−H. The smallest such pair is found among octane isomers 4-methylheptane (W=75, DW=22, giving W−H=97) and 3-ethylhexane (W=72, DW=25, giving W−H=97). The following are the four pairs of nonane isomers that have the same W and W−H: 2,4-dimethyl-3-ethylpentane and 2,2,4-trimethylheptane (W−H=111), 3,4-dimethylheptane and 3-methyl-4-ethylhexane (W−H=123), 2-methyl-4-ethylheptane and 2,3-dimethylheptane (W−H=131), and 4-ethylheptane and 4-methyloctane.

ILLUSTRATION OF THE WIENER−HOSOYA INDEX IN CORRELATIONS

To curb the proliferation of unnecessary descriptors it has been suggested that novel molecular descriptors should satisfy some desirable requirements and that when used alone or in a combination with other descriptors, they should produce better statistics associated with the regression

Table 2. Wiener−Hosoya Indices for the Isomers of Heptane, Octane, and Nonane that Have Degenerate Wiener Number

isomer	Wiener	Σ (G-ee)	Wiener−Hosoya
Heptanes			
2,2MM C_5	46	+ 6	52
2,3MM C_5	46	+ 5	51
3E C_5	48	+ 12	60
2,4MM C_5	48	+ 6	54
Octanes			
3,3MM C_6	67	+ 12	79
2M3E C_5	67	+ 16	83
2,2MM C_6	71	+ 16	87
2,4MM C_6	71	+14	85
2,2MM3E C_4	88	+ 20	108
3,3,4MMM C_5	88	+ 13	101
Nonanes			
2,3,3,4MMMM C_5	84	+ 6	90
2,2,4,4MMMM C_5	84	+ 8	92
3,3EE C_5	84	+ 32	116
2,2,3,4MMMM C_5	86	+ 7	93
2,3 MM3E C_5	86	+ 10	96
2,3,3MMM C_6	90	+ 15	105
2,4MM3E C_5	90	+ 21	111
2,2,3MMM C_6	92	+ 17	109
2,4,4MMM C_6	92	+ 15	107
3M3E C_6	92	+ 23	115
2,3,4MMM C_6	92	+ 6	98
2,2,4MMM C_6	94	+ 17	111
3M4E C_6	94	+ 29	123
4,4MM C_7	96	+24	120
2,3,5MMM C_6	96	+ 18	114
2M3E C_6	96	+ 33	132
2,2,5MMM C_5	98	+ 21	119
2,3MM C_7	98	+ 12	110
3,4MM C_7	98	+25	123
2M4E C_6	98	+ 33	131
2,3MM C_7	102	+ 29	131
2,4MM C_7	102	+ 28	130
4E C_7	102	+ 45	147
2,2M7	104	+ 32	136
2,5MM C_7	104	+ 30	134
3E C_7	104	+ 42	146
2,6MM C_7	108	+ 34	142
4M C_8	108	+ 39	147

Table 3. Statistical Parameters of Correlations of a Selection of Physicochemical Properties of Octane Isomers C_8H_{18} with the Wiener Number W, the Hyper-Wiener Number WW, and the Wiener−Hosoya Index[a]

		W	WW	W−H
boiling point	r	0.5738	0.5699	0.6436
	s	5.34	5.35	4.99
	F	4	4	5
steric factor	r	0.9806	0.9838	0.9076
	s	0.37	0.33	0.48
	F	188	226	110
entropy	r	0.9254	0.8939	0.9331
	s	1.90	2.25	1.81
	F	39	26	44
mean radius	r	0.9055	0.9030	0.9315
	s	0.082	0.080	0.070
	F	34		49
heat of vaporization	r	0.4971	0.4653	0.5884
	s	1.146	1.169	1.068
	F	2	2	4

[a] r = correlation coefficient; s = standard error; F = Fisher number.

equations than can be obtained from the available descriptors.[25,26] To test the usefulness of the novel descriptor we will restrict attention to octane isomers and the following five properties (the sources of data are also given): the boiling points,[27] the steric factor,[28] the entropy,[29] the quadratic mean radius,[30] and the heat of vaporization.[31] We have restricted such testing to octane isomers so as to eliminate the dominant role of the molecular size on properties, which most molecular descriptors parallel, and thus produce a high regression coefficient, which however need not represent a high quality regression. The relative standard error $\Delta s/s$ would in such situations be a better measure of the quality of the regression. However, when all molecules are of the same size the regression coefficient itself gives a fair indication of the quality of a regression.

In Table 3 for the five properties of octanes C_8H_{18} listed above we have given the statistical parameters for the quadratic fitting using as molecular descriptors the Wiener number W, the hyper-Wiener number WW, and the Wiener−Hosoya index. The hyper-Wiener number WW represents a generalization of the Wiener number W and is defined as the following:[32] *The hyper-Wiener number WW is defined as the sum of the sum of distances between any two carbon atoms in the molecule, in terms of carbon−carbon bonds. Brief method of calculation: Multiply the number of carbon atoms on one side of any path by those on the other side; WW is the sum of these two values for all paths."*

First we should mention that the Wiener index, when introduced by Harry Wiener, was one of two descriptors to be used in correlations and thus has not been envisaged to be used alone as a single descriptor. The reason is that although it can be interpreted as an index of "compactness" it does not adequately represent crowding of hydrogen atoms, which is apparently well described by P (paths of length three). That is why in Table 3 we see that W does not correlate well with boiling points. The same is also true for the hyper-Wiener index and the Wiener−Hosoya index. However, there are physicochemical properties that correlate well with the Wiener number. These include the steric numbers and the entropy and also the critical pressure (not examined in this work).[33] As we can see from Table 3 in the case of entropy the novel Wiener−Hosoya index gives the best regression, associated with a standard error of only 1.81 entropy units.

In view that W was introduced as one of the two descriptors for correlations of the physicochemical properties of paraffins, it is of interest to see how indices that are closely related structurally to the Wiener number correlate in similar combinations with the polarity index P. In Table 4 we have collected such regressions for octane isomers, again applied to the five properties already selected. A close look at Table 4 shows that although the three alternative pairs of descriptors differ slightly in view of the very high intercorrelations between them (W/WW: 0.900; W−H/W: 0.982; and WW/W−H: 0.971), the statistical parameters for alternative regressions show visible differences. In particular we see that in the case of the boiling points of octane isomers the Wiener−Hosoya index represents a considerable improvement over the Wiener index, reducing the standard error from 3.39 °C to 2.35 °C. The combination (W−H, P) gives also the best regression for the mean square radius and also ΔH_f, although none of the three alternative descriptors have offered a satisfactory regression. It is only in the case of steric numbers that W is slightly better than W−H, but both of these correlations are already very good, with the coefficient of regression exceeding 0.9700. Indeed it is difficult, when

Table 4. Statistical Parameters of Correlations of a Selection of Physicochemical Properties of Octane Isomers C_8H_{18} with the Wiener Number W, the Hyper-Wiener Number WW, and the Wiener−Hosoya Index and P (Paths of Length Three) as the Second Variable[a]

		W, P	WW, P	W−H, P
boiling point	*r*	0.8543	0.8119	0.9329
	s	3.39	3.80	2.35
	F	20	15	50
steric factor	*r*	0.9767	0.9605	0.9715
	s	0.41	0.53	0.45
	F	156	89	126
entropy	*r*	0.9254	0.8939	0.9331
	s	1.90	2.25	1.81
	F	39	26	44
mean radius	*r*	0.8906	0.9031	0.9241
	s	0.088	0.083	0.074
	F	29	33	44
heat of vaporization	*r*	0.7634	0.6772	0.8554
	s	0.85	0.97	0.68
	F	10	6	20

[a] r = correlation coefficient; s = standard error; F = Fisher number.

confining regressions to isomers (and in general molecules of the same size and the same molecular mass), to get regressions with such a high correlation coefficient, and this can be seen from the study of over a dozen physicochemical properties of octanes using besides W and WW five additional distance-related molecular descriptors. As reported in ref 33 only a single regression using descriptors WW and JJ (the latter of which is a generalization of the Balaban's index J[34]) gave a regression with the correlation coefficient above 0.9700 (for the critical pressures of octanes).

The regression of the boiling points of octane isomers with (W−H, P), with the correlation coefficient $r = 0.933$, and the standard error of 2.35 °C can be compared with other two-variable regression equations reported in the literature.[35] For example, when the Wiener number and the Hosoya Z index are combined, one obtains for boiling points of octane isomers $r = 0.913$ and $s = 2.66$ °C, and when one combines Z with ^{2}Z, paths of length two derived from the Hosoya matrix (outlined in ref 35), one obtains a very slightly better regression with $r = 0.914$ and $s = 2.64$ °C, the two being the best results based on several combinations of the Wiener number and invariants derived from the modifications of the Hosoya Z index.

A comparison should be made with paths/walks shape descriptors,[36] which are in many ways remarkable descriptors that yield impressive regressions of isomeric variations of numerous physicochemical properties. These regressions are so impressive that they tempt one to refer to them as "universal descriptors," that is, descriptors that can be used for numerous molecular properties, in combination with any of the size-dependent descriptors (such as are W, Z, χ, etc.). In Table 5 we have listed the best two-variable correlations for 14 physicochemical properties of octanes when using either (p_2/w_2, p_3/w_3) or (P_2/W_2), (P_3/W_3) as descriptors. Here p_2 and p_3 are the count of atomic paths of length two and three summed over all atoms, and P_2 and P_3 are the molecular count of paths of length two and three. Similarly w_2 and w_3 are the count of atomic walks of length two and three summed over all atoms, and W_2 and W_3 are the molecular count of walks of length two and three. First observe the very high correlation coefficients for 10 out of 14 properties, being above 0.9500, and in the case of molar refraction and densities over 0.9900! The only property that has the regression coefficient less than 0.8000 is the critical volume with the correlation coefficient 0.7000. Table 5 ought to be taken as a "standard" against which new molecular descriptors should be measured! For sources of data and abbreviations see ref 37.

Table 5. Statistical Parameters of Correlations of Numerous Physicochemical Properties of Octane Isomers C_8H_{18} with the Path/Walk Shape Indices[a]

property	descriptor	*r*	*s*	*F*
boiling point	P_2/W_2, P_3/W_3	0.9340	2.33	51
heat of formation	P_2/W_2, P_3/W_3	0.9619	0.36	93
entropy	P_2/W_2, P_3/W_3	0.9541	1.57	63
mean radius	P_2/W_2, P_3/W_3	0.8547	0.100	20
heat of vaporization	P_2/W_2, P_3/W_3	0.9705	0.52	122
^{13}C NMR shift sum	P_2/W_2, P_3/W_3	0.9685	4.95	114
eccentricity	P_2/W_2, P_3/W_3	0.9879	0.005	284
steric factor	p_2/w_2, p_3/w_3	0.9781	0.39	166
density	p_2/w_2, p_3/w_3	0.9908	0.0017	350
critical temperature	p_2/w_2, p_3/w_3	0.9098	3.88	34
critical pressure	p_2/w_2, p_3/w_3	0.9868	0.209	279
molar refraction	p_2/w_2, p_3/w_3	0.9948	0.020	569
octane number	p_2/w_2, p_3/w_3	0.8050	0.11	14
critical volume	p_2/w_2, p_3/w_3	0.7031	0.011	7

[a] r = correlation coefficient; s = standard error; F = Fisher number.

The paths/walks descriptors of Table 5 are not necessarily the best pairs of descriptors for each of the properties shown but are very good and not easy to surpass. For instance, in the case of the octane numbers they are inferior to the simple correlation based on the information theoretic number I_{WD} (considering the partition of the distance matrix elements involved in construction of the Wiener number[38]) which has $r = 0.959$ and $s = 7.27$; even the very good correlation of the eccentric factor with $r = 0.9879$ is surpassed by the simple regression based on the $^2\chi$ (connectivity index of order two) with $r = 0.9920$.[37] In the case of entropy the regression based on P_2/W_2, P_3/W_3 almost ties with the simple regression based on the variation in the exponent used for calculating the connectivity index, having the same regression coefficient (on four digits) but a slightly worse standard error: $s = 1.57$ versus $r = 1.40$.

It is of interest to see how the novel Wiener−Hosoya index scores for the five properties considered in comparison with the data of Table 5. By comparing the corresponding rows of Tables 4 and 5 we find that the novel W−H index combined with P gives better results for the quadratic mean radius R^2 and is of marginally lesser quality, almost tying with the correlations for the boiling points and also with the regression for an eccentric factor, for which the combination (W, P) is almost as good as p_2/w_2, p_3/w_3.

CONCLUDING REMARKS

The novel graph theoretical descriptor outlined in this paper and based on the Wiener number augmented by a "correction" term arising from the use of the Hosoya composition principle removed partly the degeneracy of the Wiener number, but more importantly it leads to better regressions for a selection of physicochemical properties of alkanes, as demonstrated on the regressions restricted to octane isomers. Moreover, this work shows that the highly intercorrelated descriptors may nevertheless lead to visibly different regressions, thus invalidating the customary practice in filtering large pools of descriptors and discarding highly intercorrelated descriptors. In view of this, it seems that a better approach than filtering descriptors would be to first find about half a dozen descriptors that give the best regression as single variables and then select one of them and make the rest orthogonal to that descriptor. Then one correlates all the descriptors with the residual of the correlation based on the best descriptor (which is tantamount to using the first descriptor and search for the best descriptors among the orthogonalized descriptors). After finding the second descriptor all the descriptors in the pool ought to be orthogonalized to the second descriptor and the process is continued. This is, of course, a procedure based on a "greedy" algorithm, which can be supplemented by repeating the process with alternatives among the half a dozen best descriptors. Although there is no guarantee that in this way, short of an exhaustive combination of all descriptors as pairs, triples, etc., the best combination will be found, it ensures at least that the best single variables will not be overlooked, which could easily be eliminated in the filtering process based on the high intercorrelation. Alternatively one should adopt the fast algorithm of B. Lučić, N. Trinajstić et al.[39] which is quite efficient in an exhaustive search of the best combinations of descriptors, to be followed with full orthogonalization that we recommend (which may be computer intensive in view of the implied orthogonalization). Although the orthogonalization does not introduce novel structural information, it can produce results with better statistical characteristics.[40]

ACKNOWLEDGMENT

The author thanks Professor A. T. Balaban (Texas A&M at Galveston, TX) for reading the manuscript and suggesting numerous improvements in the presentation.

REFERENCES AND NOTES

(1) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(2) Hosoya, H. Topological index. A newly proposed quantity characterizing topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn* **1971**, *44*, 2332−2339.

(3) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.

(4) A rather complete bibliography on the relationship between W and the Laplacian can be found in the following: Chan, O.; Lam, T. K.; Merris, R. Wiener number as an immanant of the Laplacian of molecular graphs. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 762−765.

(5) Hosoya, H. Graphical enumeration of the coefficients of the Hückel Molecular Orbitals. *Theor. Chim. Acta* **1972**, *25*, 215−222.

(6) Gutman, I.; Araujo, O.; Rada, J. An identity for Randic's connectivity index and its applications. *ACH − Models Chem.* **2000**, *137*, 653−658.

(7) Klein, D. J.; Palacios, J. L.; Randić, M.; Trinajstić, N. Random walks and chemical graph theory. To be submitted for publication.

(8) Randić, M. Resolution of ambiguities in structure−property studies by use of orthogonal descriptors. *J. Chem. Inf. Comput. Sci*. **1991**, *31*, 311−320.

(9) Randić, M. Orthogonal molecular descriptors. *New J. Chem*. **1991**, *15*, 517−525.

(10) Randić, M.; Pompe, M. On characterization of the CC double bond in alkenes. *SAR QSAR Environ. Sci*. **1991**, *10*, 451−471.

(11) Randić, M.; Basak, S. C. Optimal molecular descriptors based on weighted path numbers. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 261−266.

(12) Randić, M.; Pompe, M. On variable molecular descriptors based on distance related matrices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 575−581.

(13) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptor for nitrogen-containing molecules. *Int. J. Quantum Chem*. **1998**, *70*, 1209−1215.

(14) Randić, M. High quality structure−property regressions. Boiling points of smaller alkanes. *New J. Chem*. **2000**, *24*, 165−171.

(15) Randić, M.; Basak, S. C. On construction of high quality structure−property-activity regressions: The boiling points of sulfides. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899−905.

(16) Randić, M.; Mills, D.; Basak, S. C. On Characterization of physical properties of amino acids. *Int. J. Quantum Chem*. **2000**, *80*, 1199−1209.

(17) Randić, M.; Basak, S. C. On use of the variable connectivity index $^1\chi^f$ in QSAR: Toxicity of aliphatic ethers. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614−618.

(18) Randić, M.; Pompe, M. The variable connectivity index $^1\chi^f$ versus the traditional molecular descriptors: A comparative study of $^1\chi^f$ against descriptors of CODESSA. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 631−638.

(19) Randić, M.; Plavšić, D.; Lerš, N. Variable connectivity index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 657−662.

(20) Randić, M. Novel graph theoretical approach to heteroatoms in QSAR. *Chemom. Intel. Lab. Syst.* **1991**, *10*, 213−227.

(21) Randić, M. On Computation of optimal parameters for multivariate analysis of structure−property relationship. *J. Comput. Chem.* **1991**, *31*, 970−980.

(22) Platt, J. R. Prediction of isomeric differences in paraffin properties. *J. Chem. Phys.* **1952**, *56*, 328−336.

(23) Hosoya, H.; Kawasaki, K.; Mizutani, K. Topological index and thermodynamic properties. I. Empirical rules on the boiling points of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3415.

(24) Hosoya, H. From How to Why. Graph-Theoretical Verification of Quantum-Mechanical Aspects of p-electron Behaviors in Conjugated Systems. *Bull. Chem. Soc. Jpn.* In print.

(25) Randić, M. Generalized Molecular descriptors. *J. Math. Chem*. **1991**, *7*, 155−168.

(26) Balaban, A. T. Topological indices and their use: A new approach for coding alkanes. *J. Mol. Struct. (THEOCHEM)* **1988**, *165*, 243−253.

(27) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc*. **1988**, *110*, 4186−4194.

(28) Randić, M. Correlation of enthalpies of octanes with orthogonal molecular descriptors. *J. Mol. Struct. (THEOCHEM)* **1991**, *233*, 45−59.

(29) Scott, D. W. Correlation of chemical thermodynamic properties of alkane hydrocarbons. *J. Chem. Phys*. **1974**, *60*, 3144−3165.

(30) Altenburg, K. Eine Bemerkung zu dem Randicschen "Molekularen Bindungs-Index (molecular connectivity index)". *Z. Phys. Chem. (Leipzig)* **1980**, *261*, 389−393.

(31) Garbalena, M.; Herndon, W. C. Optimum graph-theoretical models for enthalpic properties of alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 37−42.

(32) Randić, M. Novel molecular descriptor for structure−property studies. *Chem. Phys. Lett.* **1993**, *211*, 478−483.

(33) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener matrix invariants. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 361−367.

(34) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1992**, *89*, 399−404.

(35) Randić, M. Hosoya matrix − A source of new molecular descriptors. *Croat. Chem. Acta* **1994**, *67*, 415−429.

(36) Randić, M. Novel shape descriptors for molecular graphs. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 607−613.

(37) Randić, M. Comparative regression analysis. Regressions based on a single descriptor. *Croat. Chem. Acta* **1993**, *66*, 289−312.

(38) Bonchev, D.; Trinajstić, N. Information theory, distance matrix and molecular branching. *J. Chem. Phys*. **1977**, *67*, 4517−4533.

(39) Lučić, B.; Trinajstić, N.; Sild, S.; Karelson, M.; Katritzky, A. R. A new efficient approach for variable selection based on multiregression: Prediction of gas chromatographic retention times and response factors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610−621.

(40) Lučić, B.; Trinajstić, N.; Nikolić, S.; Juretić, D. The structure−property models can be improved using the orthogonalized descriptors. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 532−538.

CI030425F