Topological Index and Thermodynamic Properties. 5.[†] How Can We Explain the Topological Dependency of Thermodynamic Properties of Alkanes with the Topology of Graphs?

Haruo Hosoya,* Machiko Gotoh, Miyuki Murakami, and Shigeko Ikeda

Departments of Chemistry and Information Sciences, Ochanomizu University, Bunkyo-ku, Tokyo 112, Japan

Correlations between the thermodynamic properties of alkanes and topological indices of integer values (ITI) for the corresponding graphs are studied. A set of higher order Z-indices, Z_n, for n-step graphs, $\mathbf{G}_n$, are newly defined. Fairly good correlation with the density of liquid was obtained for a single index Z_3, representing the static topology of alkane molecules. For the boiling point, which is determined by the dynamical motion of molecules, the quantity, $4Z_1 + Z_3$, was shown to have remarkably good correlation among other combinations of ITI's. Physicochemical interpretations of these correlation problems are given.

INTRODUCTION

Recently a number of topological indices have been proposed and analyzed by many authors for attacking various chemicophysical problems.[1−5] In general there seem to be two different attitudes of researchers toward the specified topological index or indices concerned. They are, so to speak, practical and conceptual, or computational and mathematical attitudes. Of course, the former one prevails these days, under the name of QSAR or QSPR, to pursue as precise and reliable prediction of a compound or a group of compounds as possible with desired properties and application. The present paper is a contribution from the standpoint of the opposite side.[6−8]

In this respect only integer topological indices (ITI's) with a simple and straightforward definition are concerned for clarifying the chemicomathematical secret of the topological dependency of typical thermodynamic properties of the most common alkanes.[1,9−12] Those indices are derived from either the distance matrix, $\boldsymbol{D}$, or nonadjacent number, $p(\mathrm{G},k)$.[2] From the former category Wiener's w and p (path number)[1] are selected. Although the distance polynomial has a powerful discriminating power and conceals a number of interesting graph-theoretical properties,[13,14] it lacks direct relevance to QSAR problems. Thus it is excluded from this study. Here the higher order Z-indices are introduced as an extension of the original Z-index.[15] Although still one cannot completely explain all the details concerning the secret of this problem yet, a novel interpretation to the relation between the mechanism of the two typically different topological behaviors of alkane properties, i.e., boiling point and density, could be obtained by using the set of the higher order Z-indices.

DEFINITIONS

For a nondirected connected graph $\mathbf{G}$ with N vertexes the distance matrix, $\boldsymbol{D}$, is defined as an $N \times N$ symmetrical matrix with elements, d_{ij}, as the smallest number of steps between vertexes, i and j. We will be concerned only with those simple graphs without any loop or multiple edge. The half sum of the off-diagonal elements of $\boldsymbol{D}$ gives w,[1,2] while the number of pairs of vertexes with $d_{ij} = 3$ is called the path number, p.[1]

The nonadjacent number, $p(\mathrm{G},k)$, is the number of ways for choosing k disjoint edges from $\mathbf{G}$, with $p(\mathrm{G},0)$ being defined as unity.[2] The topological index, Z, is the sum of $p(\mathrm{G},k)$s for $\mathbf{G}$. Let us define an n-step graph, or nth graph, $\mathbf{G}_n$, for $\mathbf{G}$ as the graph representing the shortest n-step relation among the vertexes in $\mathbf{G}$. Namely, in the adjacency matrix $\boldsymbol{A}_n$ for $\mathbf{G}_n$ only those matrix elements are unity for the pair of vertexes with $d_{ij} = n$ in $\boldsymbol{D}$ for $\mathbf{G}$, and all others are null. The definition of the higher order topological index Z_n for $\mathbf{G}_n$ follows that of the original Z, or Z_1. For such a higher order graph $\mathbf{G}_n$ that is disconnected the Z_n value is defined as the product of the Z values of all the components (see Figure 1). However, for a vacant graph $\mathbf{G}_n$ the Z_n value is set to be zero in this case.

As exemplified in Figure 1 for the two pentane isomers the set of $\mathbf{G}_n$'s and the corresponding Z_n's can easily and straightforwardly be obtained, when several simple recursion formulas are available.[2,16] On the other hand, although extensive studies for enumerating the value of w for special series of graphs have been recently performed, no simple and general recursion relation for w and p has ever been reported.[17,18] For heptane isomers the relevant ITI values are listed in Table 1. Note that the nine heptane isomers are consecutively numbered with Z_1, whose discriminability is lacking in w.

Merrifield and Simmons defined another topological index σ for the adjacency relation among the vertexes in $\mathbf{G}$ instead of edges.[19] Namely, $q(\mathrm{G},k)$ is defined as the number of ways for choosing k disjoint *vertexes* from $\mathbf{G}$ and also σ as the sum of $q(\mathrm{G},k)$'s in $\mathbf{G}$. As evident from Table 1, Z and σ exhibit a reverse dependency on molecular structure especially for alkanes and benzenoid hydrocarbons.[19] Actually as the correlation coefficients between Z and σ for lower

---

n	1	2	3	4
G_n				
Z_n	8	$6 = 3 \times 2$	$4 = 2 \times 2$	2
G_n				ϕ
Z_n	7	$8 = 4 \times 2$	3	– a

Figure 1. Examples of the *n*-step graphs and higher-order topological indices. For footnote a, in this case Z_n for a vacant graph is set to be null.

**Table 1.** List of Topological Indices for the Graphs of Heptane Isomers

G	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	w	p	S
	13	40	13				42	6	44
	14	28	8	4			46	4	43
	15	24	9	7			48	4	41
	16	30	14	2			44	6	40
	17	24	17	3			46	6	38
	18	18	12	6	3		52	4	37
	19	18	16	6	2		50	5	36
	20	16	18	4			48	6	35
	21	15	12	8	4	2	56	4	34

members of alkanes ($N < 12$) are larger than 0.99, no new information will come out from the analysis of σ in the QSAR study for alkanes.[20] Thus no further discussion on σ will be given in this paper.

HIGHER ORDER *Z*-INDICES

Here it is worth introducing several characteristic features of the higher order *Z*-indices. The sets of $\{Z_n\}$ for heptane and octane isomers are diagrammatically given in Figures 2 and 3, respectively. It is interesting to see that the overall shape of the set of patterns of $\{Z_n\}$, if properly drawn as in Figures 2 and 3, seems to reflect the plane-projected shape of the graph fairly well. Namely, the Z_n-pattern of a long molecule has a long structure with monotonic tailing, while a ball-shaped molecule has a short and fat pattern. Figure 4 illustrates diagrammatically the effect on the Z_n-pattern by the change of the mode of branching for the special series of nonane isomers. Similar trends can be observed for cyclic graphs, but they are not so simple.

For alkane graphs up to decanes, no two graphs have the same set of $\{Z_n\}$, while for cyclic graphs even with five and six vertexes, respectively, two and thirteen pairs or triplets of graphs are degenerate with respect to $\{Z_n\}$. A few examples are given in Figure 5. Thus one can conclude that
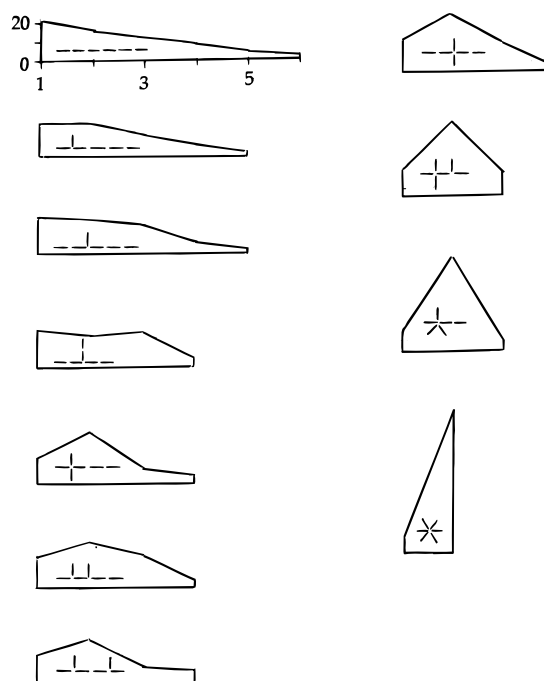
Figure 2. Diagrammatic expressions for the set of $\{Z_n\}$ for the isomeric graphs of heptanes. The last two graphs, which do not have corresponding alkane molecules, are given for graph-theoretical interest.
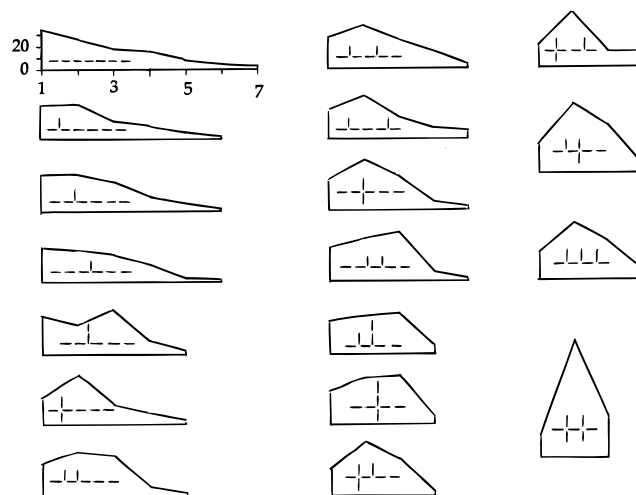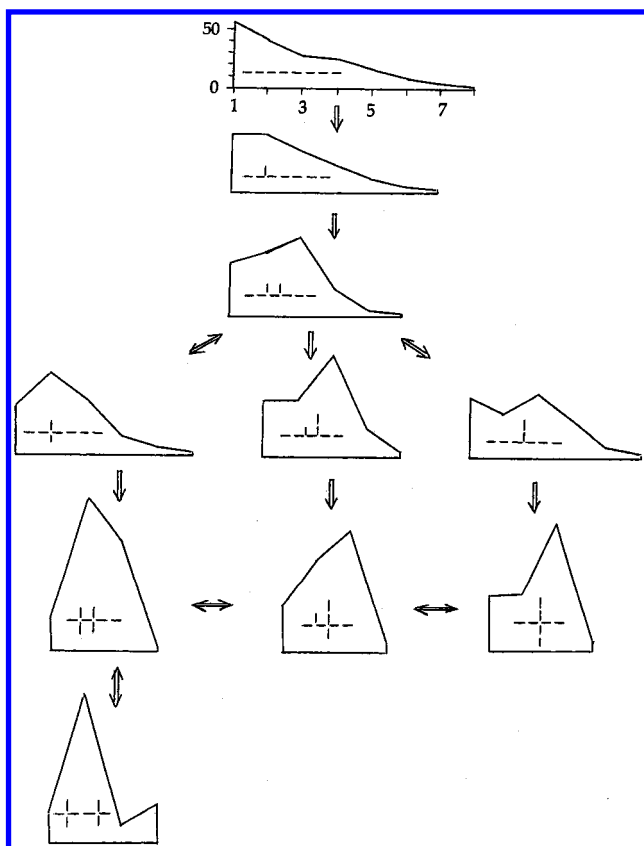

Figure 3. Diagrammatic expressions for the set of $\{Z_n\}$ for the isomeric graphs of octanes.

although the set of $\{Z_n\}$ has no efficient discriminating ability for graphs, it may have global shape discriminability of graphs, which will be disclosed in the later part of this paper.

CORRELATION AMONG ITI'S

Before going into a detailed discussion on QSAR of alkanes, one has to know how the relevant ITI's correlate with each other. In Table 2 are given the correlation coefficients, r's, among the several ITI's involved. It has been known that as well as the topological dependency of various thermodynamic properties correlation among various topological quantities in general also changes with the size of molecules.[11,12]

Except for the correlation between Z_1 and Z_2, the r value for a special pair of ITI's changes rather dramatically

Figure 4. Diagram showing the effect of branching on the Z_n-pattern of nonane isomers.

Figure 5. Examples of pairs and triplets of graphs with the identical set of $\{Z_n\}$ for graphs with five and six vertexes.

Table 2. Correlation Coefficients for the Several Pairs of Integer Topological Indices

	N	Z_2	Z_3	w	p
Z_1	7	0.902	0.507	0.796	0.089
	8	0.513	0.063	0.761	0.257
	9	0.892	0.032	0.666	0.138
	10	0.885	0.095	0.640	0.257

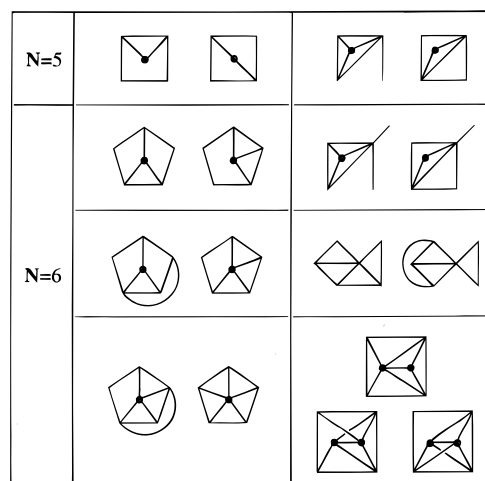|	N	Z_3	w
p	7	0.767	0.648
	8	0.917	0.787
	9	0.935	0.775
	10	0.919	0.845

between $N = 7$ and 8. For lower members of alkanes, specificity of individual molecules is sometimes so strong and the number of isomers is so small that it is rather difficult to draw a general conclusion or tendency only from those limited data. Thus, to get clear and general understanding of the topological dependency of thermodynamic properties of alkanes, only the r values for molecules larger than octane will be considered here.

Now it is easy to read the following tendency from Table 2. (i) The degree of correlation with Z_1 is in the following order: $Z_2 > w \gg p > Z_3$. (ii) The degree of correlation with p is in the following order: $Z_3 > w \gg Z_1$.

Since combined use of a pair of almost independent indices or descriptors is known to have potentially the widest application to QSAR study in general,[12,21,22] the most promising candidate for attaining the best result for this problem is expected for the pair of Z_1 and Z_3 rather than the pair of w or Z_1 with p, which was employed in our former analysis.[12]

THERMODYNAMIC PROPERTIES OF ALKANES

Let us summarize the hitherto known facts obtained from the QSAR studies of thermodynamic properties of alkanes.[23] The research group of the present authors has shown that except for the melting points almost all the thermodynamic properties of alkanes are classified into several different types depending on the relative weights of the two orthogonal parameters, namely, dynamical Z relating to the rotational degree of freedom and static p governing the bulkiness of molecules.[11,12] Boiling point, absolute entropy, and heat of vaporization are the typical members of A-type properties,

which are roughly dependent on Z, while density of liquid, refractive index, and the reciprocal of molar volume form B-type properties and are closely related to the mysterious p. Critical temperature, volume, and density belong to the W-type, in which both the factors, Z and p, are competing. Melting point is one of the most typical X-type properties, which are neither dependent on Z or p. Since these X-type properties are strongly dependent on the crystal structure and stability of each substance, it is dangerous to develop a general theory using graph-theoretical concepts. These tendencies are summarized in Table 3.

According to our dogma as stated in the Introduction let us consider only the A- and B-type properties in this paper. Although boiling point has a mixed character of A- and B-types, here it is assumed to be a representative of the A-type, while the density of a liquid is chosen to be a typical B-type property.

Since Z_2 is rather strongly correlated with Z_1 (see Table 2), it was excluded from this analysis. We have calculated the correlation coefficients of many different combinations of Z_1 and Z_3 and also of w and p with boiling point and liquid density. The results are summarized in Table 4, where large r values to be noted are printed with boldface.

The correlation between the density of liquid with either p or Z_3 is remarkably strong, as demonstrated in Figure 6 for 18 octane isomers. Although the r value for the p number is generally larger than that of Z_3, the difference is very small. The reason for the similar behaviors of these two indices can be inferred from their definitions. It is still an open question why the alkane molecules with a large number of

THERMODYNAMIC PROPERTIES OF ALKANES

J. Chem. Inf. Comput. Sci., Vol. 39, No. 2, 1999 **195**

Table 3. Size Dependency of Correlation Types of Thermodynamic Properties of Alkanes[a]

| symbol	property	6	8	10	12
			N (no. of C atoms)		
ΔS	absolute entropy	A			
bp	boiling point	A			W $\rightarrow$
ΔH_f	heat of formation	A	— W — B		$\rightarrow$
d	density of liquid				
n_D	refractive index				
V_m	molar volume	B			
R_m	molar refraction				
p_c	critical pressure				
T_c	critical temp				
d_c	critical density	W — B —		?	$\rightarrow$
V_c	critical volume				
G_f	free energy of formation	X —B			$\rightarrow$
mp	melting point	X			

[a] A-type: dynamic property. B-type: static property. W-type: A + B. X-type: no topological correlation.

Table 4. Correlation Coefficients of Several Topological Indices with Boiling Point and Density of Liquid of Alkanes.

		7	8	9	10
				N	
	Z_1	**0.966**	0.889	0.748	0.549
	Z_3	0.586	0.324	0.458	0.477
	w	**0.732**	0.546	0.279	0.071
bp	p	0.032	0.077	0.333	0.404
	$4Z_1 + Z_3$	**0.966**	**0.915**	**0.850**	**0.742**
	$w + 3p$	**0.992**	**0.963**	**0.750**	0.449
	$w + 4p$	**0.939**	**0.975**	0.877	0.632
	$w + 5p$	0.835	**0.884**	**0.913**	0.787
	$w + 6p$	0.725	0.762	**0.880**	**0.862**
	Z_1	0.205	0.114	0.105	0.237
d	Z_3	**0.818**	**0.943**	**0.949**	**0.883**
	w	0.383	0.589	0.742	0.790
	p	**0.912**	**0.971**	**0.958**	**0.936**
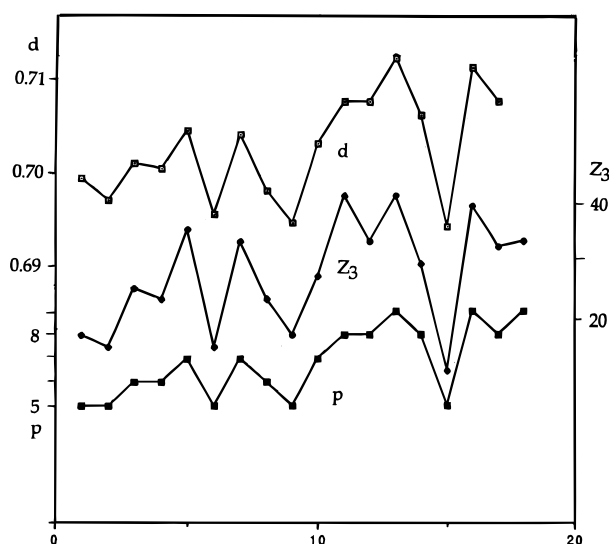
Figure 6. Correlation of density of liquid of octane isomers with two topological indices, *p* and Z_3.

3-step pairs of carbon atoms can have dense packing in forming the liquid state. A possible mechanism is that the more 3-step pairs of CC atoms alkane molecules have the easier they would take on a conformation similar to that of the diamond lattice, giving high-density packing. It is interesting to observe that the *r* value of *w* for liquid density increases with the size of molecules. Further, we can show
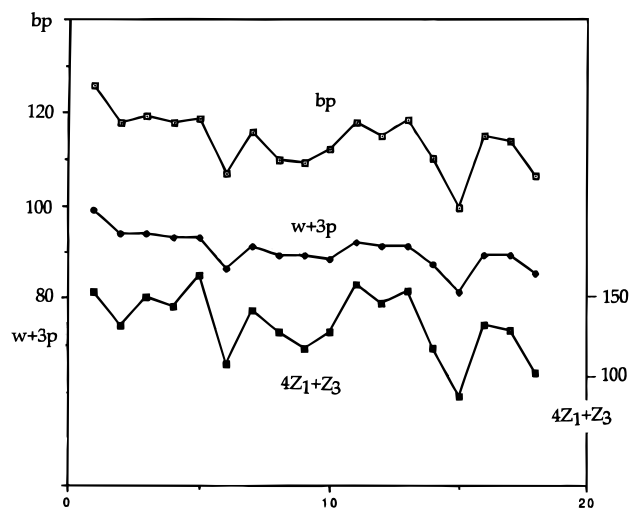
Figure 7. Correlation of boiling point of octane isomers with two pairs of topological indices, $w + 3p$ and $4Z_1 + Z_3$.

that for larger alkane graphs *w* can be correlated well with the linear combination of Z_1 and Z_3 with a fairly high *r* value. It means that the index *w* has essentially a double-face character in the topological sense and, thus, is not suitable for the analysis from the physicochemical standpoint.

With respect to the boiling point no single ITI is well correlated. Although Z_1 has the highest *r* value among them, its character becomes double-faced as *N* increases. However, both the pairs, Z_1 and Z_3 and *w* and *p*, can have remarkably good correlation with boiling point by a suitable ratio of combination. See Figure 7 for their good correlation with the boiling points of 18 isomers of octane. The combination of $4Z_1 + Z_3$ gives the highest *r* value for all the isomer groups, while the weight of *p* is increasing with *N*. This tendency can be attributed to the double-face character of *w*.

In some sense it is remarkable that the combined index of $w + np$ shows fairly good correlation with boiling point irrespective of their individually poor correlation. However, again it is very difficult to draw clear physicochemical interpretation to this observation. There might be a possibility that boiling of an especially large alkane is a kind of synergetic phenomenon refusing to be interpreted by a microscopic mechanism involving pseudorotation around each CC bond of individual alkane molecules. This is acutually the case with the melting point, for which it is almost impossible to find a simple QSAR relation using only indices derived from the topology of individual molecules.

Now the existence of a correlation between the boiling points of large alkanes and an index of the Z-family supports the following interpretation. Parallelisms among Z_1, heat of vaporization ΔH_v, and entropy *S* in the same A-type are explained from the fact that the branching of alkanes decreases the Z_1 value and also the degree of freedom of rotation which lowers *S*. Further, through the formula

$$S = \int_0^T \frac{C_p}{T} \, dT$$

and Trouton's law, i.e., boiling point is proportional to heat of vaporization, these properties have a similar dependency on molecular branching.[11]

CONCLUSION

We have seen that the pair of Z-indices, Z_1 and Z_3, are combined to give global understanding of the two typically different topological properties of alkanes, boiling point and density of liquid. The former is determined by the dynamical behavior of individual molecules. On the other hand, the latter is a static property of molecular topology and governed by the stacking of molecules depending on the number of 3-step pairs of carbon atoms, which might function to attain a dense packing of alkane molecules. For clarifying this problem more deeply another direction of study might be helpful, such as a molecular dynamics simulation calculation of gigantic size. Anyway, for global understanding of this problem graph-theoretical analysis will have a key role.

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 the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339.

(3) Rouvray, D. H. The Topological Matrix in Quantum Chemistry. In *Chemical Application of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; pp 175−221.

(4) Trinajistić, N. *Chemical Graph Theory*, 1st ed.; CRC Press: Boca Raton, FL, 1983.

(5) Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. Topological indices for structure−activity correlations. *Top. Curr. Chem.* **1983**, *114*, 21−25.

(6) Hosoya, H. On some counting polynomials in chemistry. *Discrete Appl. Math.* **1988**, *19*, 239−257.

(7) Hosoya, H. Introduction to Graph Theory. In *Graph Theoretical Approaches to Chemical Reactivity*; Bonchev, D., Mekenyan, O., Eds.; Kluwer Academic Publ.: Dordrecht, The Netherlands, 1994; pp 1−36.

(8) Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*, Springer: Berlin, 1986.

(9) Hosoya, H.; Kawasaki, K.; Mizutani, K. Topological index and thermodynamic properties. I. Empirical rules on the boiling point of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3415−3421.

(10) Narumi, H.; Hosoya, H. Topological index and thermodynamic properties. II. analysis of the topological factors on the absolute entropy of acyclic saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1980**, *53*, 1228−1237.

(11) Narumi, H.; Hosoya, H. Topological index and thermodynamic properties. III. classification of various topological aspects of properties of acyclic saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1985**, *58*, 1778−1786.

(12) Gao, Y.-D.; Hosoya, H. Topological index and thermodynamic properties. IV. Size dependency of the structure−activity correlation of alkanes. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 3093−3102.

(13) Hosoya, H.; Murakami, M.; Gotoh, M. Distance polynomial and characterization of a graph. *Natl. Sci. Rept. Ochanomizu Univ.* **1973**, *24*, 27−34.

(14) Hosoya, H.; Nagashima, U.; Hyugaji, S. Topological twin graphs. Smallest pair of isospectral polyhedral graphs with eight vertexes. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 428−431.

(15) This concept was originally born in the author's group more than 20 years ago, but for some reason it was allowed to be buried deeply under a pile of files.

(16) Hosoya, H. Topological index and Fibonacci numbers with relation to chemistry. *Fibonacci Q.* **1973**, *11*, 255−266.

(17) Polansky, O. E.; Randić, M.; Hosoya, H. Transfer matrix approach to the Wiener number of catacondensed benzenoids. *MATCH (Math. Chem.)* **1989**, *24*, 3−28.

(18) Gutman, I. Wiener numbers of benzenoid hydrocarbons: Two theorems. *Chem. Phys. Lett.* **1987**, *136*, 134−136.

(19) Merrifield, R. E.; Simmons, H. E. *Theor. Chim. Acta* **1980**, *55*, 55−75.

(20) Gutman, I.; Hosoya, H.; Uraković, G.; Ristić, L. Two variants of the topological index and the relations between them. *Bull. Chem. Soc. Jpn.* **1992**, *65*, 14−18.

(21) Randić, M. Orthosimilarity. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1092−1097.

(22) Randić, M. On Characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.

(23) Data were taken from American Petroleum Institute Research Project 44 at the National Bureau of Standards, Selected Values of Physical and Thermodynamic Properties of Hydrocarbons, Texas A&M University, 1945.