

- (29) Roy, A. B.; Raychaudhury, C.; Ghosh, J. J.; Roy, S. K.; Basak, S. C. In *Quantitative Approaches to Drug Design*; Dearden, J. C., Ed.; Elsevier: Amsterdam, 1983; p 75.
- (30) Raychaudhury, C.; Roy, S. K.; Ghosh, J. J.; Roy, A. B.; Basak, S. C. *J. Comput. Chem.* **1989**, 5, 581.
- (31) Klopman, G.; Raychaudhury, C.; Henderson, R. V. *Math. Comput. Modelling* **1988**, 11, 635.
- (32) Klopman, G.; Raychaudhury, C. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 12.
- (33) Balaban, A. T.; Balaban, T. S. *J. Math. Chem.* **1991**, 8, 383.
- (34) Balaban, A. T.; Mekenyan, O.; Bonchev, D. *J. Comput. Chem.* **1985**, 6, 538.
- (35) Bertz, S. H. *Discrete Appl. Math.* **1988**, 19, 65.
- (36) Balaban, A. T. *Graph Theory and Topology in Chemistry. Study Phys. Theor. Chem.* **1987**, 51, 159.
- (37) Lovasz, L.; Pelikan, J. *Period. Math. Hung.* **1973**, 3, 175.
- (38) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 27.
- (39) Randić, M. *J. Chem. Inf. Comput. Sci.* **1975**, 15, 105.
- (40) Balaban, A. T.; Ciubotariu, D.; Medeleanu, M. *J. Chem. Inf. Comput. Sci.* **1991**, 31, 517.
- (41) Diudea, M.; Minailuc, O.; Balaban, A. T. *J. Comput. Chem.* **1991**, 12, 527.
- (42) Balaban, A. T.; Diudea, M., unpublished results.
- (43) Balaban, A. T.; Catana, C. *J. Comput. Chem.*, in press.

Comparative Study of Molecular Descriptors Derived from the Distance Matrix

ZLATKO MIHALIĆ

Faculty of Science and Mathematics, The University of Zagreb, 41000 Zagreb, The Republic of Croatia, Yugoslavia

SONJA NIKOLIĆ and NENAD TRINAJSTIĆ*

The Rugjer Bošković Institute, 41001 Zagreb, The Republic of Croatia, Yugoslavia

Received August 16, 1991

A comparative study of 10 distance indices derived from the distance matrix either in the graph-theoretical (topological) form or in the geometric (topographic) form is carried out. They are partitioned into five topological indices and five topographic indices. The adjective topological or topographic indicates which matrix each family of distance indices has originated from. All five topological indices have been known in the literature, while four out of five topographic indices are introduced in this work. Only the 3-D Wiener number has been proposed earlier. Most of distance indices are found to be intercorrelated, i.e., they contain similar structural information. Only 2-J and 3-J indices did not intercorrelate with any other distance index but themselves. The three most accurate structure-property models for predicting boiling points of alkanes are based on the connectivity index χ , its variant χ' ($= N\chi$), and on the topological distance index 2-TI. It is unclear at present why the 3-D distance indices have produced inferior structure-boiling point models in comparison with models based on the 2-D distance indices and connectivity indices.

INTRODUCTION

The distance matrix appears to be a convenient source for deriving molecular descriptors.¹⁻⁴ This matrix can be given in two forms,⁵ i.e., as the graph-theoretical (topological) distance matrix^{6-8a} and the geometric (topographic) distance matrix.⁹⁻¹¹ Molecular descriptors that can be derived from the topological distance matrix belong to the class of topological indices,¹² while those that can be obtained from the topographic distance matrix are topographic indices.¹³ Topological indices and topographic indices represent a subgroup of molecular descriptors,¹⁴ i.e., they are used to characterize the constitution and the configuration of a molecule by a single number. In order to simplify the presentation, topological and topographic distance-matrix-related indices will be called by a common term, distance indices.

There are a number of distance indices available in the literature.^{15a} Most of them will be discussed here. We will not, however, consider the information-theoretic distance indices.¹⁶⁻¹⁸ The current interest in distance indices as well as in other molecular descriptors is stimulated by their use in the nonempirical¹⁹ quantitative structure-property relationships (QSPR)² and quantitative structure-activity relationships (QSAR).²⁰

The present work is motivated by an aim to compare distance indices in two ways. First, we will try to answer the question as to the extent the distance indices are intercor-

related. In other words, we will investigate to what degree they contain the same type of constitutional and geometric information. Second, we will examine how the distance indices perform in a given structure-property correlation. For the latter purpose, the boiling points of the first 150 alkanes are selected. Each distance index considered will be used to build a QSPR model for boiling points, and the quality of each model will be judged on the basis of its statistical characteristics. In this study will also be included the connectivity index²¹ and the most-used topological index in QSPR and QSAR to date.²² Hence, the QSPR model based on the connectivity index (or on one of its variants) will be used as a standard against which the QSPR models based on distance indices will be measured.

Throughout the article we will use the graph-theoretical language.⁶ Chemical structures will be represented by hydrogen-depleted graphs in the standard manner.^{8b,15b}

The structure of this article will be as follows. The second section will contain the definitions of the distance indices based on the graph-theoretical distance matrix. In the third section, the definitions of the distance indices based on the geometric distance matrix will be presented. The fourth section will include the definition of the connectivity index and one of its variants. The intercorrelation of distance indices will be discussed in the fifth section. In the sixth section will be given all QSPR models that were considered, their statistical

characteristics, and comparison with the performance of the connectivity indices. We will end the article with concluding remarks.

INDICES DERIVED FROM GRAPH-THEORETICAL DISTANCE MATRIX

The distance matrix $2-D = 2-D(G)$ of a labeled connected graph G is a real symmetric $N \times N$ matrix whose elements $(2-D)_{ij}$ are defined as:^{5-8a}

$$(2-D)_{ij} = \begin{cases} l_{ij} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where l_{ij} is the length of the shortest path, i.e., the minimum number of edges, between the vertices i and j in G . The length of the shortest path l_{ij} is also called²³ the distance between the vertices i and j in G , hence the name "distance matrix". The name "the matrix of lengths" has also been used.²⁴ The digit 2 in $2-D$ signifies that this form of the distance matrix is related to the 2-dimensional ($2-D$) realization of a graph. Consequently, all indices derived from the $2-D$ distance matrix will contain the label "2" in their symbols.

The construction of the distance matrix for large graphs is a nontrivial task, and it is necessary to use the computer for its completion. There are a number of algorithms and computer programs available in the literature for the formation of the distance matrix.²⁵⁻³² According to Tratch et al.,³³ one of the fastest algorithms for the construction of the distance matrix appears to be the algorithm developed by Müller et al.²⁹ This algorithm will be used in this work for large alkanes. In the case of the small alkanes, the distance matrix is generated using the powers of the corresponding adjacency matrix.²⁶

Among the topological indices that can be derived from the graph-theoretical distance matrix, the most popular is the Wiener number.³⁴ Here besides the Wiener number³⁵ we will briefly review the Balaban index,^{36,37} the Schultz index,³⁸ and the determinant of the adjacency-plus-distance matrix.³⁹

(a) Wiener Number. The Wiener number was introduced in 1947 by Wiener as the path number in his pioneering investigation of the relationship between the structures and properties of saturated acyclic hydrocarbons.^{35,40} The path number was defined as the number of bonds between all pairs of atoms in an acyclic molecule.

It was Hosoya who has first pointed out that there is a simple connection between the Wiener number and the distance matrix:¹² The Wiener number $2-W = 2-W(G)$ of a graph G is defined as the half-sum of the elements in the distance matrix:

$$2-W = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (2-D)_{ij} \quad (2)$$

The numeral 2 in $2-W$ indicates that this is the 2-dimensional ($2-D$) Wiener number.^{5,11}

(b) Balaban Index. Balaban^{36,37,42} has introduced a topological index formally similar to the connectivity index of Randić.²¹ This index was named the Balaban index² and can be described as the average distance sum connectivity. The Balaban index $2-J = 2-J(G)$ of a graph G is defined as:³⁶

$$2-J = \frac{M}{\mu + 1} \sum_{i,j} (d_i d_j)^{-1/2} \quad (3)$$

where M is the number of edges in G , μ is the cyclomatic number of G , and d_i ($i = 1, 2, \dots, N$) is the distance sum. The cyclomatic number $\mu = \mu(G)$ of G is given by⁴³

$$\mu = M - N + 1 \quad (4)$$

where N is the number of vertices in G . Therefore, the cyclomatic number is equal to the minimum number of edges

necessary to be removed from G in order to convert it into the related acyclic graph.

The distance sum d_i for a vertex i of G represents a sum of all entries in the corresponding row (or column) of the distance matrix $2-D$ ³⁶

$$d_i = \sum_{j=1}^N (2-D)_{ij} \quad (5)$$

The distance sum was also separately used as a topological index, under a name the distance sum index.⁴⁴ The distance sum index appears to be a convenient measure of the centrality of a particular site in G . Obviously the Wiener number can also be expressed in terms of the distance sums:

$$2-W = \frac{1}{2} \sum_{i=1}^N d_i \quad (6)$$

(c) Schultz Index. Schultz³⁸ has introduced an index for characterizing alkanes by an integer which he denoted as MTI, an acronym for the molecular topological index. This index is also named the Schultz index.^{15a} The Schultz index is defined in the following way:⁴⁶

$$2-MTI = \sum_{i=1}^N e_i \quad (7)$$

where e_i ($i = 1, 2, \dots, N$) are the elements of the row ($1 \times N$) matrix:

$$v(A + 2-D) = [e_1, e_2, \dots, e_N] \quad (8)$$

In eq 8 A is the adjacency ($N \times N$) matrix, $2-D$ the distance ($N \times N$) matrix, and v is the valency ($1 \times N$) matrix of a tree.

A variant of the Schultz index is

$$2-(MTI)' = \sum_{i=1}^N e'_i \quad (9)$$

where e'_i ($i = 1, 2, \dots, N$) are the elements of the row ($1 \times N$) matrix:

$$v(2-D) = [e'_1, e'_2, \dots, e'_N] \quad (10)$$

(d) Determinant of Adjacency-Plus-Distance Matrix. The determinant of the adjacency-plus-distance matrix has been introduced as a topological index, denoted by $2-TI$, for numerical characterization of alkanes.³⁹ This index is defined as follows:⁴⁷

$$2-TI = \det|A + 2-D| \quad (11)$$

where A and $2-D$ are the adjacency matrix and the distance matrix, respectively.

Topological distance indices are not particularly discriminating descriptors.^{41,45-47} In Figure 1 we give the smallest pairs of trees with the same, respectively, Wiener number, Balaban index, Schultz index, modified Schultz index, and determinant of the adjacency-plus-distance matrix.

The $2-W$ numbers, $2-J$, $2-MTI$, $2-(MTI)'$, and $2-TI$ indices for the first 150 alkanes are given in Table I.

INDICES DERIVED FROM GEOMETRIC DISTANCE MATRIX

In this section we will describe five topographic distance indices. Four out of these five indices will be introduced for the first time here.

The geometric distance matrix $3-D = 3-D(G)$ of a molecule G is a real symmetric $N \times N$ matrix defined in the same way as the $2-D$ matrix [see eq 1]. However, the matrix elements $(3-D)_{ij}$ now represent the shortest Cartesian distances l_{ij} (in some arbitrary units of length) between sites i and j in G . But, for many molecules their geometries are unknown. Therefore, in order to construct the topographic distance matrix, the geometry of a molecule must be approximated in some way.

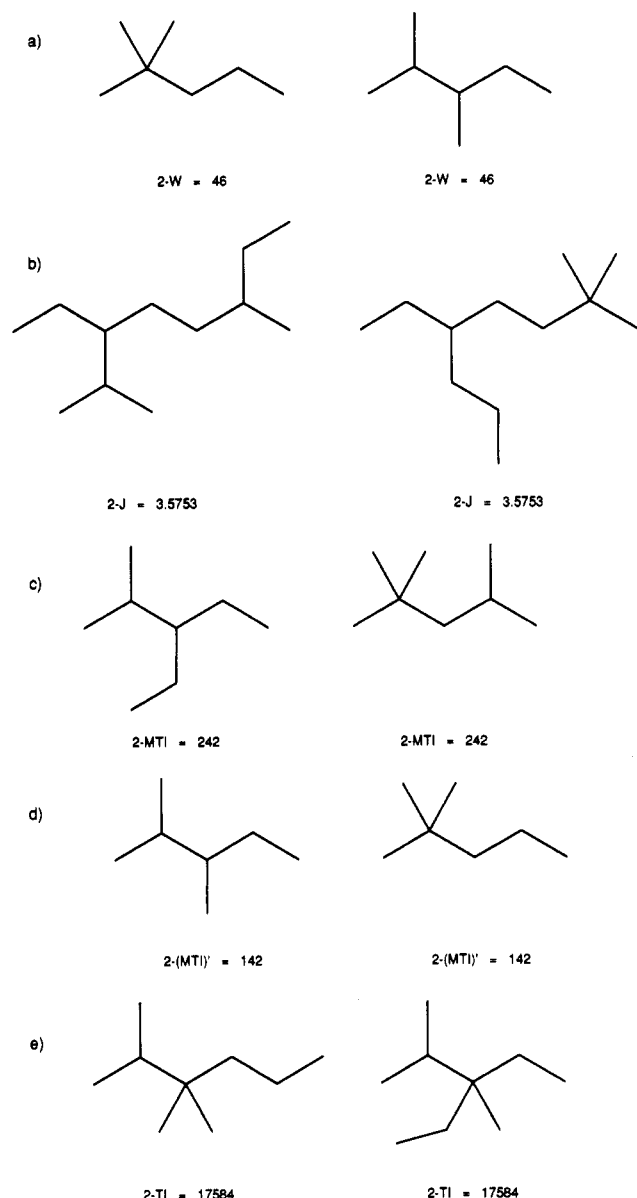


Figure 1. (a) A pair of the heptane trees with the same Wiener number. (b) A pair of the dodecane trees with the same Balaban index. (c) A pair of the octane trees with the same Schultz index. (d) A pair of the heptane trees with the same value of the modified Schultz index. (e) A pair of the nonane trees with the same 2-TI index.

In the present work this was done by means of the molecular mechanics (MM) method.⁴⁹ For each alkane molecule, we selected as an input the most extended conformation with the minimum number of gauche-butane (*g*) interactions, and, if possible, without the "forbidden" pentane (*g⁺g⁻*) interactions (see Figure 2). Such trial structures were refined to a (local) minimum by use of a standard MM program. In the case of more complicated alkanes, we continued with the procedure for determining the optimum conformation developed by Saunders.⁵⁰

The geometry of the most stable alkane conformation obtained in the above way was used to set up the corresponding geometric distance matrix. The elements of 3-D matrix, i.e., the distances between atoms in a given alkane ($3-D_{ij}$), were computed from the Cartesian coordinates, which are the part of the standard output of the MM programs:

$$(3-D)_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2} \quad (12)$$

The geometric distance matrix constructed in such a way is used as a generator of topographic distance indices. Several

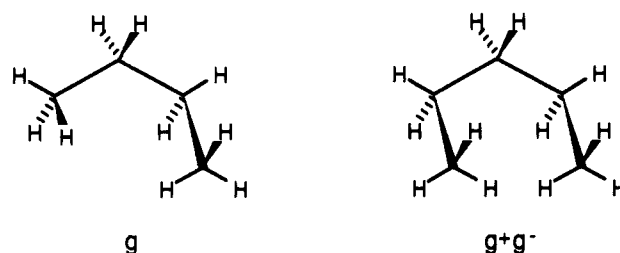


Figure 2. Diagrams of gauche-butane (*g*) and pentane (*g⁺g⁻*) interactions.

of these will be reviewed below. It should be pointed out that in most cases they do not formally differ from the corresponding topological distance indices. Their actual difference is related to the form of the distance matrix (and in a few cases also to the form of the adjacency matrix) that was used to generate them.

(a) 3-D Wiener Number. The 3-D Wiener number, denoted by 3-W, is defined as the half-sum of the elements of the geometric-distance matrix:¹¹

$$3-W = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (3-D)_{ij} \quad (13)$$

The 3-W number was computed for the whole CH skeleton of an alkane in contrast to the 2-W number which was generated only for graphs representing the carbon skeletons of alkanes. In Table I, we report the 3-W numbers for the first 150 alkanes in their optimal (minimum energy) conformations. It is worthwhile to point out that the 3-W numbers differ for different conformations of the same molecule. In this way the 3-D Wiener number may be used to characterize the conformations of a molecule by a single number.^{5,51,52} It has been also observed that the 3-W number is the largest for the most extended conformation and the smallest for the most compact conformation. Therefore, the 3-W number decreases with increasing spheroidicity of a molecule.^{52,53} The 3-W number is also a very discriminating index, and so far we have not been able to detect any case of its degeneracy for alkanes.

(b) 3-D Balaban Index. The 3-D Balaban index 3-J is formally defined in the same way as its 2-D counterpart. However, the distance sums are now given in terms of the elements of the geometric distance matrix:

$$d_i = \sum_{j=1}^N (3-D)_{ij} \quad (14)$$

The 3-W number can also be expressed in terms of the geometric distance sums:

$$3-W = \frac{1}{2} \sum_{i=1}^N d_i \quad (15)$$

(c) 3-D Schultz Index. The definition of the 3-D Schultz index, denoted by 3-MTI, also formally resembles the 2-D Schultz index (see eq 7). However, the quantities e_i ($i = 1, 2, \dots, N$) are elements of the row ($1 \times N$) matrix containing the 3-D adjacency matrix and 3-D distance matrix:

$$v[3-A + 3-D] = [e_1, e_2, \dots, e_N] \quad (16)$$

The 3-D adjacency matrix is defined as⁵⁴

$$3-A = \begin{cases} l_{ij} & \text{if atom } i \text{ is bonded to atom } j \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

where l_{ij} is the geometric distance (in arbitrary units of length) between the bonded atoms i and j in a molecule.

A variant of the 3-D Schultz index is 3-(MTI)' index in which the quantities e'_i ($i = 1, 2, \dots, N$) are the elements of the following row ($1 \times N$) matrix:

$$v(3-D) = [e_1', e_2', \dots, e_{N'}] \quad (18)$$

(d) **Determinant of 3-D Adjacency-Plus-Distance Matrix.** This index is defined as

$$3-TI = \det[3-A + 3-D] \quad (19)$$

where 3-A and 3-D are the 3-D forms of the adjacency matrix and the distance matrix, respectively.

The 3-W numbers, 3-J, 3-MTI, 3-(MTI)', and 3-TI indices for the first 150 alkanes are also given in Table I.

CONNECTIVITY INDEX

The connectivity index $\chi = \chi(G)$ of a graph G is defined as an additivity quantity:²¹

$$\chi = \sum_{i,j} (a_i a_j)^{-1/2} \quad (20)$$

where a_i and a_j are valencies of the vertices i and j making up the edge $i-j$. The summation in eq 20 is over all edges in G .

In the case of alkanes (and other saturated hydrocarbons) eq 20 can be given in a closed form. In alkane trees, only four types of vertices with respect to their valencies appear, i.e., vertices with the valency 1, 2, 3, and 4. These give rise to 10 types of edges with the following weights: $(1,1) = 1$, $(1,2) = 0.7071$, $(1,3) = 0.5773$, $(1,4) = (2,2) = 0.5$, $(2,3) = 0.4082$, $(2,4) = 0.3536$, $(3,3) = 0.3333$, $(3,4) = 0.2887$ and $(4,4) = 0.25$. If the number of each bond-type is denoted by b_{ij} ($i = 1, \dots, 4$; $j = i, \dots, 4$) and if the bond weights from above are used, then eq 21 becomes:

$$\chi = b_{11} + 0.7071b_{12} + 0.5773b_{13} + 0.5(b_{14} + b_{22}) + 0.4082b_{23} + 0.3536b_{24} + 0.3333b_{33} + 0.2887b_{34} + 0.25b_{44} \quad (21)$$

This formula reveals that the connectivity indices of alkanes are fully determined by their bond-types.

A variant of the connectivity index $\chi' = \chi'(G)$ of G is defined as

$$\chi' = N\chi \quad (22)$$

where N is the number of vertices in G .

Both the original connectivity index χ and its variant χ' for the first 150 alkanes are also given in Table I.

INTERCORRELATION OF DISTANCE INDICES

In this section we analyze the extent to which the distance indices are linearly intercorrelated. The intercorrelation of all distance indices considered in this report for the set of alkanes with up to 10 carbon atoms is studied. For this purpose alkanes are convenient systems because they represent structurally rather simple chemical structures. Their only complicated structural feature is the skeletal branching.

We have investigated the linear relationship between the pairs of distance indices, $(DI)_i$ and $(DI)_j$.

$$(DI)_i = A(DI)_j + B \quad (23)$$

The results are displayed as the intercorrelation matrix with the correlation coefficient $r \geq 0.90$ as the criterion for the intercorrelated pairs of distance indices.⁵⁵ Strongly intercorrelated pairs of the distance indices are those with $r \geq 0.98$. In Table II we give the intercorrelation matrix reflecting the pairwise linear correlation for 10 distance indices. Since this matrix is symmetric, we give only its lower triangle.

From Table II we learn several interesting points:

(1) The 2-D and 3-D counterparts are always intercorrelated.

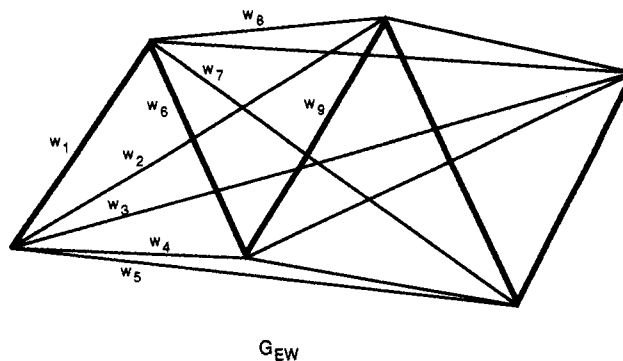


Figure 3. Edge-weighted graph G_{EW} corresponding to the anti,anti,anti- (aaa) conformation of n -hexane. w_i 's stand for the edge-weights which are equal to the geometric distances between the carbon atoms in (aaa) conformation of n -hexane. Because of the symmetry, there are only nine different weights (different geometric distances): w_1 (1.534 Å), w_2 (3.914 Å), w_3 (6.420 Å), w_4 (2.544 Å), w_5 (5.092 Å), w_6 (1.537 Å), w_7 (3.918 Å), w_8 (2.548 Å), and w_9 (1.537 Å).

(2) The majority of 2-D and 3-D counterparts are strongly intercorrelated. Only the pair 2-TI/3-TI is not.

(3) The 2-J and 3-J indices are only intercorrelated between themselves.

(4) The 2-W, 2-MTI, 2-(MTI)', and 3-(MTI)' indices intercorrelate with six other distance indices.

(5) The 2-W, 2-MTI, 2-(MTI)', and 3-W strongly intercorrelate with five other distance indices.

(6) The above items indicate that most of the distance indices contain the same type of structural information.

The strong intercorrelation between some 2-D and 3-D distance indices may be attributed to a fact that the 3-D structure of a molecule can be depicted by the edge-weighted complete graph.⁵² For example, the edge-weighted graph corresponding to (aaa) conformation of n -hexane is shown in Figure 3. Since a subgraph of G_{EW} is a graph G for which topological distance indices are computed, it is not surprising that the invariants for both G and G_{EW} are intercorrelated.

It is interesting to note that the connectivity indices χ and χ' are intercorrelated with several distance indices [2-W, 2-MTI, 2-(MTI)'] but are not strongly intercorrelated with any of them. They are strongly intercorrelated only among themselves ($r = 0.98$).

COMPARISON OF QSPR MODELS FOR BOILING POINTS BASED ON DISTANCE INDICES

In this section we report a comparison of performances of the 10 distance indices and two connectivity indices in the structure-property (property = boiling point) correlations for the first 150 alkanes. The basis of the comparison is the statistical characteristics of each correlation considered. The following correlations are tested:

$$bp = A + BI + CI^2 \quad (24)$$

$$bp = AI^B + CI^3 + D \quad (25)$$

$$bp = A(I + D)^B + C \quad (26)$$

$$bp = A + B \ln I + C (\ln I)^2 \quad (27)$$

$$bp = A + BI^{1/3} + CI^{2/3} + DI \quad (28)$$

where bp = alkane boiling point, I = distance or connectivity index, and A, B, C, \dots are statistical parameters to be determined by a least-squares regression. In Table III we list the best model obtained for each index with its statistical characteristics.

The data in Table III reveal several interesting points. They may be summarized in the following way:

Table I. Distance Indices, Connectivity Indices, and Boiling Points of the First 150 Alkanes

alkane	2-W	2-J	2-MTI	2-(MTI)'	2-TI	3-W	3-J	3-MTI	3-(MTI)'	3-TI	x	x'	bp
methane	0	0.0000	0	0	0	0.0000	0.0000	0.000	0.000	0	0.0000	0.0000	-164
ethane	1	1.0000	4	2	-4	1.5316	0.6529	6.126	3.063	-9	1.0000	2.0000	-88.6
propane	4	1.6330	16	10	16	5.6072	1.1315	23.487	14.283	48	1.4142	4.2426	-42.1
2-methylpropane	9	2.3238	36	24	-48	12.1957	1.6322	52.060	33.614	-181	1.7321	6.9282	-11.7
butane	10	1.9747	38	28	-60	13.6026	1.4302	53.787	38.434	-229	1.9142	7.6569	-0.5
2,2-dimethylpropane	16	3.0237	64	44	128	21.2640	2.1373	91.843	61.021	606	2.0000	10.0000	9.5
2-methylbutane	18	2.5395	68	52	176	23.3716	1.8793	93.919	69.298	780	2.2701	11.3503	27.8
pentane	20	2.1906	74	60	224	26.6940	1.6233	102.120	80.617	1094	2.4142	12.0711	36.1
2,2-dimethylbutane	28	3.1685	106	82	-464	35.6399	2.3516	145.387	108.348	-2474	2.5607	15.3640	49.7
2,3-dimethylbutane	29	2.9935	108	86	-512	36.9450	2.2483	146.892	112.944	-2506	2.6427	15.8564	58
2-methylpentane	32	2.6272	118	98	-656	41.0357	1.9960	158.546	127.770	-3734	2.7701	16.6203	60.3
3-methylpentane	31	2.7542	114	94	-644	39.7255	2.0769	154.011	123.201	-3311	2.8081	16.8484	63.3
hexane	35	2.3391	128	110	-817	46.2159	1.7580	173.507	145.854	-5211	2.9142	17.4853	69
2,2,3-trimethylbutane	42	3.5412	156	126	1344	52.4900	2.6778	210.259	163.834	7879	2.9434	20.6036	80.9
2,2,3,3-tetramethylbutane	46	3.1545	170	142	1728	57.8282	2.4101	226.289	183.091	11837	3.0607	21.4246	79.2
3,3-dimethylpentane	44	3.3604	162	134	1680	55.3024	2.5411	217.468	174.188	9726	3.1213	21.8492	86.1
2,3-dimethylpentane	46	3.1442	168	142	1872	57.8069	2.4070	223.216	183.055	10372	3.1807	22.2652	89.8
2,4-dimethylpentane	48	2.9532	176	150	1920	59.0680	2.3250	228.741	188.693	12425	3.1259	21.8813	80.5
2-methylhexane	52	2.6783	190	166	2372	66.1215	2.0657	250.798	213.872	17789	3.2701	22.8904	90
3-methylhexane	50	2.8318	182	158	2348	63.8276	2.1655	241.762	204.797	15779	3.3081	23.1564	92
3-ethylpentane	48	2.9923	174	150	2352	57.7615	2.3776	222.307	185.329	13050	3.3461	23.4225	93.5
heptane	56	2.4475	204	182	2840	73.3841	1.8566	273.051	239.247	23790	3.4142	23.8995	98.4
2,2,3,3-tetramethylbutane	58	4.0204	214	176	-3520	71.7166	3.0613	286.335	227.311	-22193	3.2500	26.0000	106.5
2,2,3-trimethylpentane	63	3.6233	230	196	-4912	77.8595	2.7955	302.718	250.075	-32661	3.4814	27.8510	110
2,3,3-trimethylpentane	62	3.7083	226	192	-4864	76.6911	2.8476	298.580	245.889	-30513	3.5040	28.0323	114.7
2,2,4-trimethylpentane	66	3.3889	242	208	-5056	79.9208	2.6820	312.088	259.575	-37020	3.4165	27.3320	99.2
2,2-dimethylhexane	71	3.1118	260	228	-6220	88.4539	2.4212	339.388	290.039	-56337	3.5607	28.4853	106.8
3,3-dimethylhexane	67	3.3734	244	212	-6124	80.6789	2.6743	311.300	261.885	-46578	3.6213	28.9706	112
3-ethyl-3-methylpentane	64	3.5832	232	200	-6076	77.0522	2.8167	298.297	248.800	-37111	3.6820	29.4558	118.2
2,3,4-trimethylpentane	65	3.4642	236	204	-5440	79.1004	2.7273	305.140	255.675	-34206	3.5534	28.4273	113.4
2,3-dimethylhexane	70	3.1708	254	224	-6768	87.3042	2.4640	331.447	285.125	-49387	3.6807	29.4459	115.6
3-ethyl-2-methylpentane	67	3.3549	242	212	-6832	80.3171	2.6757	307.476	261.137	-38915	3.7187	29.7500	115.6
3,4-dimethylhexane	68	3.2925	246	216	-6700	85.2243	2.5394	323.106	276.722	-42751	3.7187	29.7500	117.7
2,4-dimethylhexane	71	3.0988	258	228	-6816	85.2192	2.4949	326.491	208.254	-52767	3.6639	29.3112	109.4
2,5-dimethylhexane	74	2.9278	270	240	-6864	91.0595	2.3192	347.304	301.106	-60045	3.6259	29.0072	109
2-methylheptane	79	2.7158	288	260	-8092	100.2054	2.1106	376.198	333.122	-80985	3.7701	30.1604	117.6
3-methylheptane	76	2.8621	276	248	-8127	96.6143	2.2104	362.286	319.169	-71928	3.8081	30.4645	118
4-methylheptane	75	2.9196	272	244	-8160	95.4970	2.2491	357.541	314.418	-71970	3.8081	30.4645	117.7
3-ethylhexane	72	3.0744	260	232	-8400	86.8459	2.4628	328.350	285.216	-61816	3.8461	30.7685	118.5
octane	84	2.5301	306	280	-9487	109.5100	1.9319	405.789	365.836	-103283	3.9142	31.3137	125.7
2,2,3,3-tetramethylpentane	82	4.1447	298	256	12736	100.3735	3.2038	390.850	325.537	84771	3.8107	34.2959	140.27
2,2,3,4-tetramethylpentane	86	3.8776	312	272	14272	103.7233	3.0618	401.633	339.505	93348	3.8541	34.6865	133
2,2,3-trimethylhexane	92	3.5887	334	296	17680	112.9713	2.8090	432.062	373.251	154869	3.9814	35.8324	131.7
2,2-dimethyl-3-ethylpentane	88	3.7929	318	280	17920	106.3589	2.9906	407.692	348.830	127919	4.0194	36.1745	133.83
3,3,4-trimethylhexane	88	3.8024	318	280	17408	108.4634	2.9527	414.174	355.262	125478	4.0420	36.3784	140.5
2,3,4-tetramethylpentane	84	4.0137	304	264	14080	101.7331	3.1454	393.433	331.314	91644	3.8868	34.9808	141.5
2,3,3-trimethylhexane	90	3.7021	326	288	17584	110.7663	2.8825	423.340	364.484	145312	4.0040	36.0363	137.7
2,3-dimethyl-3-ethylpentane	86	3.9192	310	272	17584	102.9279	3.1032	395.486	336.560	110853	4.0647	36.5823	141.6
2,2,4-tetramethylpentane	88	3.7464	322	280	13312	105.0390	2.9943	411.181	346.190	108388	3.7071	33.3640	122.7
2,2,4-trimethylhexane	94	3.4673	342	304	17872	114.2161	2.7514	438.229	379.511	155267	3.9545	35.5906	126.5
2,4,4-trimethylhexane	92	3.5768	334	296	17776	111.9749	2.8245	429.386	370.621	141646	3.9772	35.7945	126.5
2,2,5-trimethylhexane	98	3.2807	358	320	17968	118.7742	2.6222	455.931	397.301	183275	3.9165	35.2485	124
2,2-dimethylheptane	104	3.0730	380	344	21008	129.4402	2.4125	490.697	435.198	256243	4.0607	36.5459	132.7
3,3-dimethylheptane	98	3.3301	356	320	21148	122.4114	2.5878	463.386	407.788	211377	4.1213	37.0919	137.3

4,4-dimethylheptane	96	3.4311	348	312	21280	120.1814	2.6561	453.980	398.374	211483	4.1213	37.0919	135.2
3-ethyl-3-methylhexane	92	3.6174	332	296	21700	110.5576	2.8788	420.436	364.776	175695	4.1820	37.6378	140.6
3,3-diethylpentane	88	3.8247	316	280	21952	104.3489	3.0551	398.291	342.552	134254	4.2426	38.1838	146.2
2,3,4-trimethylhexane	92	3.5758	332	286	19312	112.1666	2.8931	426.221	370.529	140570	4.0914	36.8228	139
2,4-dimethyl-3-ethylpentane	90	3.6776	324	298	19840	107.3463	2.9491	409.087	353.429	135290	4.0914	36.8228	136.73
2,3,5-trimethylhexane	96	3.3766	348	312	19584	115.8538	2.7036	441.631	386.033	163469	4.0366	36.3292	131.3
2,3-dimethylheptane	102	3.1553	370	336	23020	126.8899	2.4743	477.509	425.035	224691	4.1807	37.6267	140.5
3-ethyl-2-methylhexane	96	3.4101	346	312	24208	114.7661	2.7462	433.832	381.335	184017	4.2187	37.9687	138
3,4-dimethylheptane	98	3.3248	354	320	23188	122.4579	2.5865	459.894	407.348	195243	4.2187	37.9687	140.1
3-ethyl-4-methylhexane	94	3.4995	338	304	23968	112.5570	2.8079	425.165	372.603	154213	4.2567	38.3107	140.4
2,4-dimethylheptane	102	3.1513	370	336	23248	125.9825	2.4908	474.542	422.147	238348	4.1639	37.4751	133.5
4-ethyl-2-methylhexane	98	3.3074	354	320	24192	115.1150	2.7202	463.310	383.900	202872	4.2019	37.8172	133.8
3,5-dimethylheptane	100	3.2230	362	328	23256	123.5877	2.5415	465.598	413.164	210954	4.2019	37.8172	136
2,5-dimethylheptane	104	3.0608	378	344	23116	128.8376	2.4174	485.513	433.111	231462	4.1639	37.4751	136
2,6-dimethylheptane	108	2.9147	394	360	22880	133.0554	2.3198	502.621	450.272	274985	4.1259	37.1331	135.2
3-methyloctane	114	2.7467	416	368	26588	144.3925	2.1435	539.736	490.509	346786	4.2701	38.4305	142.8
4-methyloctane	110	2.8766	400	368	26784	139.6943	2.2343	521.024	471.756	311480	4.3081	38.7725	143.3
3-ethylheptane	108	2.9548	392	360	27148	137.2124	2.2903	511.196	461.922	312485	4.3081	38.7725	142.4
4-ethylheptane	104	3.0922	376	344	27832	125.5802	2.4920	471.281	421.996	278705	4.3461	39.1146	143
nonane	120	3.1753	368	336	28800	123.5549	2.5486	462.587	413.298	281586	4.3461	39.1146	141.2
2,2,3,3,4-pentamethylpentane	102	2.5951	438	408	31008	155.8198	1.9911	576.809	530.708	434901	4.4142	39.7279	150.77
2,2,3,3-tetramethylhexane	108	4.4038	390	342	-36804	129.6768	3.4662	502.151	427.265	-245019	4.1934	41.9338	166.05
3-ethyl-2,2,3-trimethylpentane	115	4.1018	416	370	-45840	139.8518	3.2152	536.076	464.595	-403201	4.3107	43.1066	158
3,3,4,4-tetramethylhexane	110	4.3283	396	350	-46032	130.8465	3.4341	503.807	432.184	-309136	4.3713	43.7132	168
2,2,3,4,4-pentamethylpentane	111	4.2818	400	354	-45136	135.6255	3.3322	519.160	447.544	-323386	4.3713	43.7132	170.5
2,2,3,4-tetramethylhexane	111	4.2311	402	354	-37440	132.2213	3.3693	513.035	438.354	-268592	4.1547	41.5470	159.29
2,2,3,4-tetramethylhexane	118	3.9418	426	382	-50448	142.7287	3.1209	544.939	476.560	-382278	4.3921	43.9206	154.9
3-ethyl-2,2,4-trimethylpentane	115	4.0729	414	370	-52032	134.8086	3.2966	517.427	449.076	-329769	4.3921	43.9206	155.3
2,3,4,4-tetramethylhexane	116	4.0341	418	374	-50176	140.5377	3.1864	535.538	467.228	-390161	4.4147	44.1472	162.2
2,2,3,5-tetramethylhexane	123	3.7348	446	402	-51072	146.9542	3.0050	562.466	494.378	-483390	4.3372	43.3722	148.4
2,2,3-trimethylheptane	130	3.5184	472	430	-59572	159.4923	2.7786	604.078	539.114	-703327	4.4814	44.8138	158
2,2-dimethyl-3-ethylhexane	122	3.8089	440	398	-63232	146.3419	3.0445	554.645	489.624	-602650	4.5194	45.1939	159
3,3,4-trimethylheptane	123	3.7784	444	402	-60112	151.1819	2.9631	571.503	506.432	-572198	4.5420	45.4204	164
3,3-dimethyl-4-ethylhexane	118	3.9711	424	382	-62272	142.7150	3.1427	539.582	474.448	-464072	4.5800	45.8005	165
2,3,3,4-tetramethylhexane	115	4.0893	414	370	-49984	139.0903	3.2272	530.241	461.882	-378917	4.4248	44.2476	164.59
3,4,4-trimethylheptane	122	3.8232	440	398	-60244	150.0670	2.9929	566.953	501.877	-573097	4.5420	45.4204	164
3,4-dimethyl-3-ethylhexane	117	4.0205	420	378	-61684	139.5135	3.2134	529.055	463.895	-436779	4.6027	46.0270	170
3-ethyl-2,3,4-trimethylpentane	112	4.2290	402	358	-50880	132.6754	3.3794	507.703	439.262	-310401	4.4474	44.4741	169.44
2,3,3,5-tetramethylhexane	120	3.8656	434	390	-50880	144.1466	3.0831	551.224	483.041	-438099	4.3599	43.5988	153
2,3,3-trimethylheptane	127	3.6334	460	418	-59712	156.0058	2.8574	590.493	525.483	-660202	4.5040	45.0404	160.1
2,3-dimethyl-3-ethylhexane	119	3.9436	428	386	-62304	141.4733	3.1700	536.056	470.982	-572964	4.5647	45.6470	169
3,3-diethyl-2-methylpentane	114	4.1535	408	366	-63504	133.9394	3.3466	509.650	444.447	-457173	4.6254	46.2536	174
2,2,4,4-tetramethylhexane	119	3.8876	432	386	-46608	143.0693	3.0979	550.658	479.403	-417948	4.2678	42.6777	153.3
2,2,4,5-tetramethylhexane	124	3.6842	450	406	-51264	148.1290	2.9696	567.462	499.362	-474762	4.3272	43.2719	148.2
2,2,4-trimethylheptane	131	3.4695	476	434	-60352	159.9161	2.7594	606.626	541.749	-505161	4.4545	44.5451	147.7
2,2-dimethyl-4-ethylhexane	126	3.6308	456	414	-63168	146.3455	3.0021	558.948	494.035	-549316	4.4925	44.9251	147
3,3,5-trimethylheptane	126	3.6419	456	414	-60516	154.0206	2.8804	583.820	518.847	-537110	4.5152	45.1517	155.68
2,4,4-trimethylheptane	127	3.6256	460	418	-60624	155.3489	2.8668	588.422	523.493	-644415	4.4772	44.7716	153
2,4-dimethyl-4-ethylhexane	122	3.8026	440	398	-62496	146.1313	3.0486	554.072	489.409	-584426	4.5378	45.3782	158
2,2,5,5-tetramethylhexane	127	3.5630	464	418	-46992	151.4077	2.8856	583.586	512.526	-550101	4.2071	42.0711	137.46
2,2,5-trimethylheptane	134	3.3555	488	446	-59956	163.0443	2.6835	619.066	554.244	-724578	4.4545	44.5451	148
2,5,5-trimethylheptane	131	3.4647	476	434	-60096	159.6441	2.7585	605.805	540.925	-672050	4.4772	44.7716	152.8
2,2,6-trimethylheptane	139	3.2055	508	466	-59152	168.0859	2.5814	640.143	575.369	-839983	4.4165	44.1650	148.2
2,2-dimethyloctane	146	3.0438	534	494	-68404	181.6059	2.4017	684.855	623.203	-1088045	4.5607	45.6066	155
3,3-dimethyloctane	138	3.2770	502	462	-69188	172.4259	2.5640	648.189	586.440	-911370	4.6213	46.2132	161.2
4,4-dimethyloctane	134	3.4175	486	446	-70644	167.5844	2.6639	628.955	567.194	-918618	4.6213	46.2132	157.5
3-ethyl-3-methylheptane	129	3.5755	466	426	-71841	154.7828	2.8728	584.012	522.200	-792312	4.6820	46.8198	163.8

Table I (Continued)

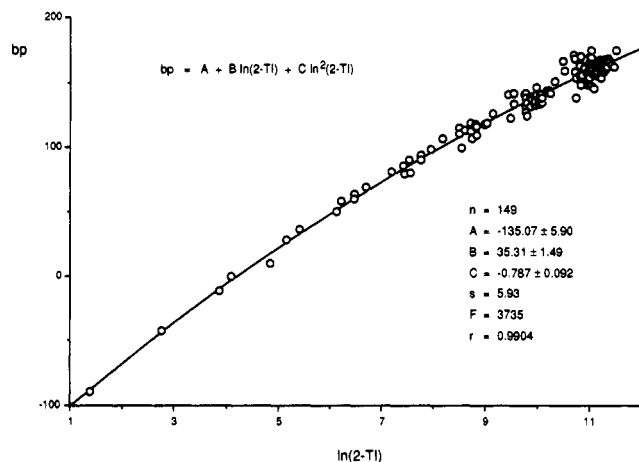
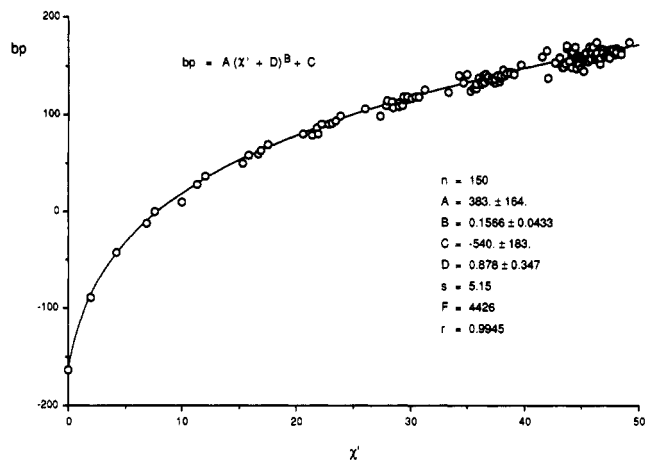
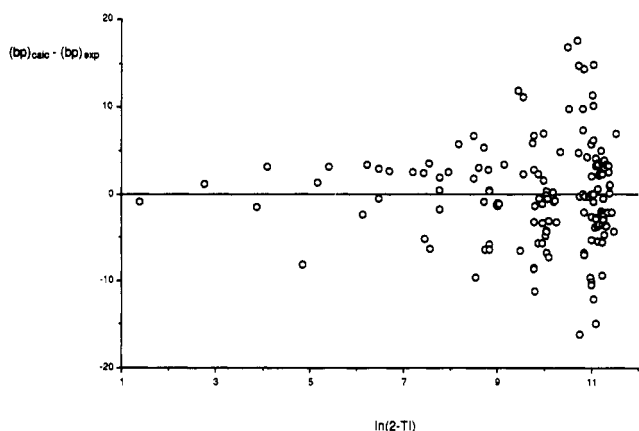
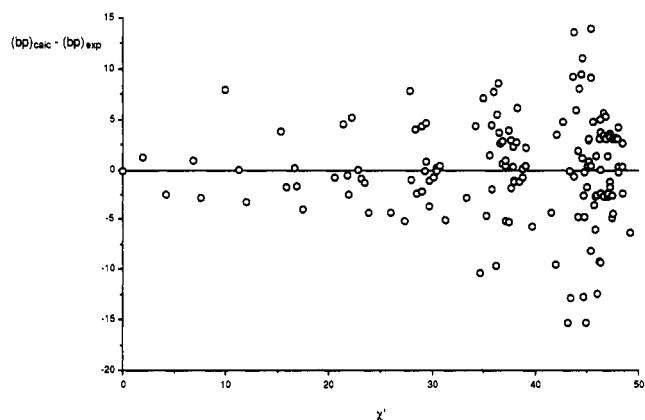
alkane	2-W	2-J	2-MTI	2-(MTI)'	2-TI	3-W	3-J	3-MTI	3-(MTI)'	3-TI	x	x'	bp
4-ethyl-4-methylheptane	126	3.6903	454	414	-74400	151.6211	2.9519	570.676	508.855	-797664	4.6820	46.8198	167
3,3-diethylhexane	121	3.8748	434	394	-76881	142.7049	3.1350	538.473	476.574	-624544	4.7426	47.4264	166.3
2,3,4,5-tetramethylhexane	121	3.8140	436	394	-55488	146.3699	3.0337	555.098	489.945	-414169	4.4641	44.6410	161
2,3,4-trimethylheptane	128	3.5833	462	422	-65680	156.3653	2.8404	589.124	527.271	-640419	4.5914	45.9142	163
2,3-dimethyl-4-ethylhexane	123	3.7561	442	402	-68544	144.1340	3.0769	544.966	483.097	-500141	4.6294	46.2943	164
2,4-dimethyl-3-ethylhexane	122	3.7979	438	398	-69072	144.7299	3.0692	546.693	484.704	-51546	4.6294	46.2943	164
3,4,5-trimethylheptane	125	3.6854	450	410	-65892	152.6252	2.9177	574.614	512.723	-566498	4.6294	46.2943	164
2,4-dimethyl-3-isopropylpentane	117	3.9835	420	378	-57600	138.2555	3.2178	525.648	460.501	-142124	4.4641	44.6410	157.04
3-isopropyl-2-methylhexane	124	3.7280	446	406	-69760	146.5882	3.0275	553.579	491.714	-685631	4.5914	45.9142	163
2,3,5-trimethylheptane	131	3.4617	474	434	-65760	158.8292	2.7717	599.858	538.067	-649495	4.5746	45.7459	157
2,5-dimethyl-3-ethylhexane	127	3.6033	458	418	-69504	147.9838	2.9770	560.816	499.039	-602963	4.5746	45.7459	157
2,4,5-trimethylheptane	130	3.5027	470	430	-65952	158.0092	2.7943	596.327	534.503	-641461	4.5746	45.7459	157
2,3,6-trimethylheptane	136	3.3014	494	454	-65008	164.6680	2.6516	623.015	561.251	-703771	4.5366	45.3658	155.7
2,3-dimethyloctane	143	3.1296	520	482	-74768	177.7434	2.4681	666.239	607.637	-956916	4.6807	46.8074	164.31
3-ethyl-2-methylheptane	134	3.3978	484	446	-78820	160.2040	2.7543	601.991	543.342	-826637	4.7187	47.1874	166
3,4-dimethyloctane	137	3.3088	496	458	-76193	171.1259	2.5895	639.631	580.933	-840267	4.7187	47.1874	166
4-isopropylheptane	131	3.4999	472	434	-82560	156.5777	2.8395	586.286	527.656	-980525	4.7187	47.1874	166
4-ethyl-3-methylheptane	129	3.5637	464	426	-81840	154.7479	2.8705	580.065	521.344	-703822	4.7567	47.5675	167
4,5-dimethyloctane	135	3.3759	488	450	-77145	168.6814	2.6376	630.116	571.410	-851030	4.7187	47.1874	162.1
3-ethyl-4-methylheptane	130	3.5299	468	430	-79408	155.3723	2.8524	583.238	524.515	-691606	4.7567	47.5675	167
3,4-diethylhexane	125	3.6982	448	410	-84084	145.6455	3.0420	548.141	489.402	-513085	4.7948	47.9475	162
2,4,6-trimethylheptane	135	3.3374	490	450	-65728	162.2136	2.6981	614.431	552.762	-765427	4.5197	45.1974	144.8
2,4-dimethyloctane	142	3.1600	516	478	-76084	175.7526	2.5019	659.018	600.470	-1017080	4.6639	46.6390	153
4-ethyl-2-methylheptane	134	3.3908	484	446	-81600	158.7189	2.7797	596.454	537.889	-921099	4.7019	47.0191	160
3,5-dimethyloctane	138	3.2686	500	462	-76644	171.1255	2.5804	640.556	581.967	-914550	4.7019	47.0191	160
2,5-dimethyloctane	143	3.4123	480	442	-79296	156.9668	2.8033	590.686	532.085	-782248	4.7399	47.3991	158.3
5-ethyl-2-methylheptane	138	3.2555	500	462	-77228	163.6745	2.6690	616.606	558.037	-869690	4.7019	47.0191	159.7
3,6-dimethyloctane	141	3.1682	512	474	-75545	175.2616	2.5000	656.352	597.755	-895137	4.7019	47.0191	160
2,6-dimethyloctane	146	3.0333	532	494	-74628	180.5990	2.4082	678.268	619.727	-1045946	4.6639	46.6390	158.54
2,7-dimethyloctane	151	2.9095	552	514	-73984	186.8390	2.3141	702.725	644.225	-1143173	4.6259	46.2590	159.87
2-methylnonane	158	2.7732	578	542	-86084	200.1000	2.1692	746.609	691.233	-1450356	4.7701	47.7006	167
3-methylnonane	153	2.8862	558	522	-86369	194.1107	2.2500	722.939	667.521	-1308981	4.8081	48.0806	167.8
4-methylnonane	150	2.9680	546	510	-87524	190.3809	2.3100	707.997	652.573	-1313400	4.8081	48.0806	165.7
3-ethyloctane	145	3.0869	526	490	-88361	175.8497	2.4920	657.231	601.795	-1192607	4.8461	48.4607	166
5-methyloctane	149	2.9984	542	506	-88464	189.1468	2.3321	703.137	647.712	-1316877	4.8081	48.0806	165.1
4-ethyloctane	141	3.2055	510	474	-92400	171.2425	2.5796	638.725	583.284	-1211227	4.8461	48.4607	163.64
4-propylheptane	138	3.2951	498	462	-9600	162.5269	2.7226	605.472	550.031	-1196023	4.8461	48.4607	162
decane	165	2.6476	604	570	-100100	213.6347	2.0387	791.227	738.976	-1812162	4.9142	49.1421	174.12

Table II. Intercorrelation Matrix of Distance Indices for Alkanes with up to 10 Carbon Atoms

	2-W	2-J	2-MTI	2-(MTI)'	2-TI	3-W	3-J	3-MTI	3-(MTI)'	3-TI
2-W	1.00									
2-J	0.47	1.00								
2-MTI	1.00	0.46	1.00							
2-(MTI)'	1.00	0.43	1.00	1.00						
2-TI	0.90	0.31	0.90	0.90	1.00					
3-W	1.00	0.43	1.00	1.00	0.89	1.00				
3-J	0.52	0.99	0.52	0.49	0.37	0.49	1.00			
3-MTI	0.99	0.49	0.99	0.99	0.91	0.98	0.55	1.00		
3-(MTI)'	0.99	0.46	0.98	0.98	0.92	0.98	0.52	1.00	1.00	
3-TI	0.84	0.10	0.85	0.86	0.93	0.85	0.15	0.84	0.85	1.00

Table III. Statistical Characteristics of Structure-Boiling Point Models with 16 Distance and 2 Connectivity Indices

index	model, eq	A	B	C	D	s	r	F
(a) Three-Parameter Models								
2-J	24	-205.8 ± 24.5	162.1 ± 16.2	-17.77 ± 2.81		28.89	0.8094	140
2-TI	27	-135.07 ± 5.90	35.31 ± 1.49	-0.787 ± 0.092		5.93	0.9904	3735
3-J	24	-197.1 ± 21.4	201.2 ± 18.2	28.08 ± 4.01		26.57	0.8416	179
3-TI	27	-165.91 ± 9.10	33.93 ± 1.85	-0.724 ± 0.092		8.10	0.9819	1967
χ	24	-172.93 ± 4.93	104.14 ± 3.02	-6.899 ± 0.459		6.36	0.9916	4318
(b) Four-Parameter Models								
2-W	25	80.08 ± 6.44	0.2980 ± 0.0138	-(8.17 ± 1.34) 10 ⁻⁶	-164.57 ± 7.59	7.93	0.9870	1836
2-MTI	25	51.63 ± 5.30	0.3080 ± 0.0146	-(1.924 ± 0.291) 10 ⁻⁷	-164.50 ± 7.80	8.17	0.9862	1729
2-(MTI)'	25	63.18 ± 6.01	0.2775 ± 0.0133	-(2.148 ± 0.344) 10 ⁻⁷	-163.73 ± 7.99	8.40	0.9854	1630
3-W	25	109.7 ± 34.1	0.2048 ± 0.0291	-(1.727 ± 0.465) 10 ⁻⁹	-347.9 ± 50.0	7.77	0.9875	1194
3-MTI	28	-159.55 ± 8.08	30.42 ± 5.55	3.82 ± 1.20	-0.3493 ± 0.0761	8.54	0.9849	1576
3-(MTI)'	28	-157.11 ± 8.15	40.55 ± 5.78	2.04 ± 1.27	-0.2768 ± 0.0833	8.90	0.9836	1448
χ'	26	383. ± 163.	0.1566 ± 0.0431	-540. ± 183.	0.878 ± 0.345	5.15	0.9945	4426

**Figure 4.** Plot of (bp)_{exp} vs 2-TI for the first 150 alkanes.**Figure 6.** Plot of (bp)_{exp} vs χ' for the first 150 alkanes.**Figure 5.** Plot of (bp)_{calc} - (bp)_{exp} vs 2-TI.**Figure 7.** Plot of (bp)_{calc} - (bp)_{exp} vs χ'.

(1) 2-D and 3-D counterparts prefer the same mathematical model in all cases studied.

(2) The most accurate model with a distance index appears to be a three-parameter logarithmic correlation with the 2-TI index:

$$\text{bp} = -135.07(\pm 5.90) + 35.31(\pm 1.49) (\ln 2\text{-TI}) - 0.787(\pm 0.092) (\ln 2\text{-TI})^2 \quad (29)$$

$$n = 149 \quad r = 0.9904 \quad s = 5.93 \quad F = 3735$$

A plot of bp vs 2-TI is shown in Figure 4, and Figure 5 is plot of residuals (bp)_{calc} - (bp)_{exp} vs 2-TI.

(3) The 3-D distance indices are less successful in modeling boiling points than their 2-D counterparts, except in the case of the 2-J/3-J pair, where 3-J gives a somewhat better model of the two. However, both QSPR models with 2-J and 3-J, respectively, are rather

poor models. Actually they appear to be the worst models among those analyzed here for predicting the boiling points of alkanes. The reason for the relative failure of the 3-D distance indices is unclear at this point. However, we are presently computing 3-D distance indices for a mixture of conformers of a given alkane. These computations are time consuming, but the preliminary results are encouraging.⁵³ They indicate that perhaps an average topographic distance index of a molecule, which can be obtained by summing up the contributions of each individual conformer in the conformational equilibrium, should be used in the QSPR modeling.

(4) All models considered with distance indices are inferior to the quadratic model (eq 24) with χ except the logarithmic model (eq 27) with 2-TI, which is comparable to it. The explicit form of the quadratic model with χ is as follows:

$$\text{bp} = -172.93(\pm 4.93) + 104.14(\pm 3.02)\chi - 6.899(\pm 0.459)\chi^2 \quad (30)$$

$$n = 150 \quad r = 0.9916 \quad s = 6.36 \quad F = 4318$$

(5) The most accurate four-parameter model obtained is the adjusted Walker-type relationship (eq 26) with χ' . Its explicit form is given by:

$$\text{bp} = 383(\pm 164)[\chi' + 0.878(\pm 0.347)]0.1566(\pm 0.0433) - 540(\pm 183) \quad (31)$$

$$n = 150 \quad r = 0.9945 \quad s = 5.15 \quad F = 4318$$

A plot of bp vs χ' is shown in Figure 6. A plot of residuals $(\text{bp})_{\text{calc}} - (\text{bp})_{\text{exp}}$ vs χ' is given in Figure 7. This result that the connectivity index or its variants provide the most accurate models is a rather general observation whenever the connectivity index is compared to topological or topographic indices.^{22,52,53,59-62} The reasons for this unexpected (almost "unreasonable") success of the χ index (and its variants) are many. Randić, the originator of the connectivity index, emphasized that this index is a solution to the bond-ordering problem for physical properties.^{19,21} Seybold et al.⁶⁰ attributed the success of the connectivity index to its ability to represent the shape of a molecule which in turn influences the packing ability of a molecule. Therefore, the connectivity index is expected to be especially effective in modeling the shape-dependent properties of molecules. However, since some of the distance indices, particularly 3-D distance indices, also possess the above properties, there might be certain additional not-yet-detected factors responsible for the effectiveness of the connectivity index in representing properties of various classes of molecules.

CONCLUDING REMARKS

In this work we carried out a comparative study of 10 molecular descriptors derived from the distance matrix in the graph-theoretical form and in the geometric form. This study belongs to our systematic efforts to discover the best possible single number molecular descriptor. Among the studied indices there were five topological distance indices, all known in the literature, and five topographic distance indices, all but the 3-D Wiener number introduced in this work. The design of the novel 3-D distance indices was rather simple. The 3-D distance indices were set to be formally identical to the 2-D indices, but the graph-theoretical distance matrix was replaced by the geometric distance matrix. In a few cases the graph-theoretical adjacency matrix was also replaced by the geometric adjacency matrix. The 2-D distance indices are not particularly discriminating descriptors, while in the case of 3-D distance indices degeneracy, i.e., two or more nonisomorphic structure possessing the same value of the index, has not been detected.

The intercorrelation analysis on the first 150 alkanes revealed that all 2-D/3-D counterparts are intercorrelated. The meaning of this result is that the 2-D/3-D pairs of indices contain similar structural information. Many of the studied indices are intercorrelated, but the indices 2-J and 3-J are only intercorrelated among themselves.

A comparison of performances of the 10 distance indices and two connectivity indices in the structure-boiling point correlations for the alkanes with up to 10 carbon atoms revealed that the most accurate QSPR models are with the connectivity indices χ and χ' . The most reliable QSPR model with the distance indices is unexpectedly one with the 2-TI descriptor. The 3-D distance indices produced less accurate models than their 2-D counterparts.

Since a great deal of information about properties (activities) of a family of molecules can be obtained from their structures, it is expected that the structure-property-activity correlations will become a standard tool of an experimentalist in chemistry and related disciplines such as biochemistry and pharmaceutical chemistry. For this reason, research on the development and application of molecular descriptors, especially those which can appear as a single number, is warranted.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science, Technology and Informatics of the Republic of Croatia via Grants 1-07-159 and 1-07-185. We thank the referees for their comments which were generally very helpful.

REFERENCES AND NOTES

- (1) Sabljčić, A.; Trinajstić, N. *Acta Pharm. Jugosl.* **1981**, *31*, 189.
- (2) Trinajstić, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1983; Vol. II, Chapter 4.
- (3) Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. *Top. Curr. Chem.* **1983**, *114*, 21.
- (4) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. *Russ. Chem. Rev.* **1988**, *57*, 191.
- (5) Mihalčić, Z.; Veljan, D.; Amiđ, D.; Nikolić, S.; Plavšić, D.; Trinajstić, N. *J. Math. Chem.*, in press.
- (6) Harary, F. *Graph Theory*, 2nd ed.; Addison-Wesley: Reading, MA, 1971.
- (7) Rouvray, D. H. *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic: London, 1976; p 175.
- (8) (a) Trinajstić, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1983; Vol. I, Chapter 4. (b) *Ibid.* Chapter 3.
- (9) (a) Crippen, G. M. *J. Comput. Phys.* **1977**, *24*, 96. (b) Crippen, G. M. *J. Math. Chem.* **1991**, *6*, 307.
- (10) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Wiley: New York, 1988.
- (11) (a) Bogdanov, B.; Nikolić, S.; Trinajstić, N. *J. Math. Chem.* **1989**, *3*, 291. (b) *Ibid.* *J. Math. Chem.* **1990**, *5*, 305.
- (12) Hosoya, H. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332.
- (13) Randić, M. *MATH/CHEM/COMP 1987*; Lacher, R. C., Ed.; Elsevier: Amsterdam, 1988; p 101.
- (14) Randić, M. *J. Math. Chem.* **1990**, *4*, 157.
- (15) (a) Trinajstić, N. *Chemical Graph Theory*; 2nd ed.; CRC: Boca Raton, FL, 1992; Chapter 10. (b) *Ibid.* Chapter 3.
- (16) Bonchev, D.; Trinajstić, N. *J. Chem. Phys.* **1977**, *67*, 4517.
- (17) Bonchev, D.; Trinajstić, N. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1982**, *16*, 463.
- (18) Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures*; Research Studies Press: Chichester, 1983.
- (19) Trinajstić, N.; Randić, M.; Klein, D. J. *Acta Pharm. Jugosl.* **1986**, *36*, 267.
- (20) Trinajstić, N.; Nikolić, S.; Carter, S. *Kem. Ind.* **1989**, *38*, 469.
- (21) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- (22) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Wiley: New York, 1986.
- (23) Harary, F. *Applied Combinatorial Mathematics*; Beckenbach, E. F., Ed.; Wiley: New York, 1964; p 185.
- (24) Kruskal, J. K., Jr. *Proc. Am. Math. Soc.* **1965**, *7*, 48.
- (25) Floyd, R. W. *Commun. ACM* **1962**, *5*, 345.
- (26) Roberts, F. S. *Discrete Mathematical Model*; Prentice-Hall: Englewood Cliffs, NJ, 1976; p 58.
- (27) Bersohn, M. *J. Comput. Chem.* **1983**, *4*, 110.
- (28) Deo, N.; Pang, C. *Networks* **1984**, *14*, 275.
- (29) Müller, W.; Szymanski, K.; Knop, J. V.; Trinajstić, N. *J. Comput. Chem.* **1987**, *8*, 170.
- (30) Mohar, B.; Pisanski, T. *J. Math. Chem.* **1988**, *2*, 267.
- (31) Senn, P. *Comput. Chem.* **1988**, *12*, 219.
- (32) Balasubramanian, K. *J. Comput. Chem.* **1990**, *11*, 829.
- (33) Tratch, S. S.; Stankevich, M. I.; Zefirov, N. S. *J. Comput. Chem.* **1990**, *11*, 899.
- (34) Rouvray, D. H. *Sci. Am.* **1986**, *254*, 40.
- (35) Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
- (36) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
- (37) Balaban, A. T. *Pure Appl. Chem.* **1983**, *55*, 199.
- (38) Schultz, H. P. *J. Am. Chem. Inf. Comput. Sci.* **1989**, *29*, 227.
- (39) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 27.
- (40) (a) Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 2636. (b) Wiener, H. *J. Chem. Phys.* **1947**, *15*, 766. (c) Wiener, H. *J. Phys. Chem.* **1948**, *52*, 425. (d) Wiener, H. *J. Phys. Chem.* **1952**, *56*, 328.
- (41) Bonchev, D.; Trinajstić, N. *J. Chem. Phys.* **1977**, *67*, 4517.
- (42) Balaban, A. T.; Ivanciuc, O. *MATH/CHEM/COMP 1988*; Graovac, A., Ed.; Elsevier: Amsterdam, 1989; p 193.
- (43) Wilson, R. J. *Introduction to Graph Theory*; Oliver & Boyd: Edinburgh, 1972; p 46.

- (44) Seybold, P. G. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1983**, 10, 95 and 103.
 (45) Balaban, A. T.; Quintas, L. V. *Math. Chem. (Mülheim/Ruhr)* **1983**, 14, 213.
 (46) Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 160.
 (47) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1991**, 31, 83.
 (48) Graham, R. L.; Pollak, H. O. *Bell Syst. Tech. J.* **1971**, 50, 2495.
 (49) Burkert, K.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, 1982.
 (50) Saunders, M. J. *Am. Chem. Soc.* **1987**, 109, 3150.
 (51) Balasubramanian, K. *Chem. Phys. Lett.* **1990**, 169, 224.
 (52) Nikolić, S.; Trinajstić, N.; Mihalić, Z.; Carter, S. *Chem. Phys. Lett.* **1991**, 179, 21.
 (53) Mihalić, Z.; Trinajstić, N. *J. Mol. Struct. (Theochem)* **1991**, 232, 65.
 (54) Nikolić, S.; Trinajstić, N.; Mihalić, Z. *J. Math. Chem.*, submitted.
 (55) Balaban, A. T.; Motoc, I. *Math. Chem. (Mülheim/Ruhr)* **1979**, 5, 197.
 (56) Motoc, I.; Balaban, A. T. *Rev. Roum. Chim.* **1981**, 26, 593.
 (57) Motoc, I.; Balaban, A. T.; Mekenyan, O.; Bonchev, D., *Math. Chem. (Mülheim/Ruhr)* **1982**, 13, 369.
 (58) Kovačević, K.; Plavšić, D.; Trinajstić, N.; Horvat, D. *MATH/CHEM/COMP 1988*; Graovac, A., Ed.; Elsevier: Amsterdam, 1989; p 213.
 (59) Razinger, M.; Chrétien, J. R.; Dubois, J. E. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 23.
 (60) Seybold, P. G.; May, M.; Bagal, U. A. *J. Chem. Educ.* **1987**, 64, 575.
 (61) Needham, D. E.; Wei, I.-C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, 110, 4186.
 (62) Hansen, P. J.; Jurs, P. C. *J. Chem. Educ.* **1988**, 65, 574.

Optimum Graph-Theoretical Models for Enthalpic Properties of Alkanes

MANUEL GARBALENA and WILLIAM C. HERNDON*

Department of Chemistry, University of Texas at El Paso, El Paso, Texas 79968

Received August 27, 1991

Simple additive schemes based on graph-theoretical properties of molecular graphs are evaluated for correlating thermodynamic properties of alkanes. The properties considered in this work are $\Delta H_f^\circ(g)$, $\Delta H_f^\circ(l)$, $\Delta H(\text{atomization})$, heats of vaporization, and heat capacities of vaporization. The original sources for thermochemical data were checked since some of the past attempts to correlate alkane properties, e.g., $\Delta H_f^\circ(g)$, have made use of calculated or interpolated values rather than restricting the analysis to experimental data. The evaluation is conducted using stepwise multilinear regression methodology. The statistical qualities of the final correlative equations are excellent (in all cases but one, squared multiple correlation coefficients > 0.995). In each case, the most successful model equation involves the use of labeled paths of length zero, one, and two as the independent variables. Therefore, each parameter has a simple molecular structure interpretation and can be obtained from the molecular graph by inspection.

1. INTRODUCTION

In general, there are causal relationships between the molecular structures of organic compounds and their chemical, physical, and biological properties.¹⁻¹⁴ Many of these relationships have been investigated using graph-theoretical descriptors of molecular structures. This is particularly true for the alkanes, which constitute the prototypical organic compounds presupposed to have graph-like character. In fact, explicitly or implicitly, nearly all additive procedures that have been proposed to systematize thermodynamic properties of alkanes use graph-theoretical topological indices or counts of particular graph features as the independent parameters to correlate the dependent enthalpic variables.^{3,15-36}

In this paper we attempt to ascertain the optimum graph-theoretical models to correlate thermodynamic enthalpic properties of alkanes. The properties treated are $\Delta H_f^\circ(g)$, $\Delta H_f^\circ(l)$, $\Delta H(\text{atomization})$, heats of vaporization [$\Delta H(\text{vapor})$], and heat capacities of vaporization [$C_p(\text{vapor})$]. The format of the report is as follows. Definitions and an outline of our procedures are described in section 2. Correlations of gas- and liquid-phase heats of formation and heats of atomization are discussed in section 3, while heats of vaporization and heat capacities of vaporization are treated and discussed in section 4. Finally, some summarizing remarks are presented in the last section.

2. PROCEDURES

In addition to various graph-theoretical molecular descriptors, the database consists of a list of abbreviations for alkanes (based on systematic nomenclature^{17,26}) and experimental values for $\Delta H_f^\circ(g)$, $\Delta H_f^\circ(l)$, $\Delta H(\text{atomization})$, $\Delta H(\text{vapor})$, and $C_p(\text{vapor})$. The compilations of thermodynamic values

due to Cox and Pilcher³⁸ and to Pedley, Naylor, and Kirby³⁹ served as the primary sources for ΔH_f° data. The original references were consulted if a cited value was not clearly identified as derived from an experiment. In some previous $\Delta H_f^\circ/\text{structure}$ studies,^{2,30} a large number of the quoted $\Delta H_f^\circ(g)$ are actually calculated values, typically estimated using an experimentally determined $\Delta H_f^\circ(l)$ and an interpolated or theoretical value for $\Delta H(\text{vapor})$.

The global molecular descriptors for each molecule consisted of the total number of carbon atoms, the Wiener numbers,^{8,16} and the Kier/Hall χ indices,² order 1-6. Descriptors based on the hydrogen-suppressed graph included counts of the vertices sorted by degrees and counts of all unlabeled paths of length 1-10. The paths were also sorted and counted based on the degrees of the vertices included in the path, which corresponds to enumeration of labeled paths in the molecular graph. Finally, the number of *gauche*-*n*-butane interactions, steric parameters for locked 1,2-interactions and 1,3-steric interactions were included according to the criteria of Kalb et al.²⁶ Of course, these latter parameters are not strictly graph-theoretical since they are derived from an examination of the (putative) most stable conformation of each individual alkane.

Many of the parameters in the data set have a dual interpretation; for example, the vertex degrees of the hydrogen-suppressed graph are linearly related to the numbers of C-H bonds according to types (primary, secondary, and tertiary). Anticipating the results of the present study, we would like to draw the readers attention to the labeled paths of length zero, one, and two; P_0 , P_1 , and P_2 , respectively. The labeled P_0 's coincide with the types of carbon atoms or single-carbon groups (CH_3 , CH_2 , CH , C). The labeled P_1 's naturally represent specific types of carbon-carbon bonds, and labeled P_2 's