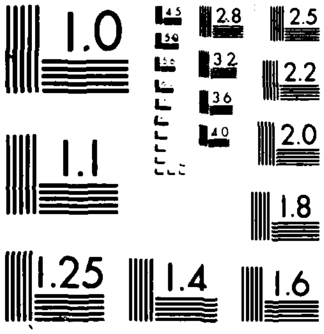
AD-A177 013 THE MODELING OF CHEMICAL PHENOMENA USING TOPOLOGICAL 1/1
INDICES(U) GEORGIA UNIV ATHENS DEPT OF CHEMISTRY
D H ROUVRAY 06 MAR 87 TR-42 N00014-84-K-0365

UNCLASSIFIED F/G 7/4 NL



CROCOPY RESOLUTION TEST CHART

(12)

AD-A177 813

…an Data Entered)

…TION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

Technical Report No. 42	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Modeling of Chemical Phenomena Using Topological Indices		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D.H. Rouvray		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Georgia Department of Chemistry Athens, GA 30602		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR051-861
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		12. REPORT DATE 3/6/87
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC
ELECTE
MAR 19 1987
S D

D

18. SUPPLEMENTARY NOTES

To be published in J. Computational Chemistry

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Graph Theory, Invariants, Topological Indices

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A class of graph invariants known today as topological indices are being increasingly realized by chemists and others to be powerful tools in the description of chemical phenomena. Topological indices generally characterize both the size and shape of chemical species; in recent years a number of such indices have been put forward which sensitively reflect the amount of branching present in molecules. Chemists are thus able to model accurately the chemical behavior of an extensvie range of chemical substances in all three thermodynamic states.

(OVER)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#20

cont'd

In discussing the manifold applications of topological indices to the description of physicochemical properties, we present a survey of the progress to date in this area, and point out some of the advantages and drawbacks of using topological indices. (Keywords:)

OFFICE OF NAVAL RESEARCH

Control N00014-84-K-0365

TECHNICAL REPORT NO. 42

THE MODELING OF CHEMICAL PHENOMENA USING TOPOLOGICAL INDICES

by

D.H. Rouvray

Prepared for publication in

J. Computational Chemistry

University of Georgia
Department of Chemistry
Athens, GA 30602

March 6, 1987

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

This document has been approved for public release
and sale; its distribution is unlimited.

The Modeling of Chemical Phenomena Using Topological Indices

Dennis H. Rouvray

Department of Chemistry, University of Georgia, Athens, Georgia 30602, U.S.A.

Abstract

A class of graph invariants known today as topological indices are being increasingly realized by chemists and others to be powerful tools in the description of chemical phenomena. Topological indices generally characterize both the size and shape of chemical species; in recent years a number of such indices have been put forward which sensitively reflect the amount of branching present in molecules. Chemists are thus able to model accurately the chemical behavior of an extensvie range of chemical substances in all three thermodynamic states. In discussing the manifold applications of topological indices to the description of physicochemical properties, we present a survey of the progress to date in this area, and point out some of the advantages and drawbacks of using topological indices.

Introduction

Topological indices are scalar numerical descriptors that are now being increasingly used by chemists and others for the characterization of molecular species. Such indices, which should be more accurately referred to as graph-theoretical indices, usually characterize both the size and shape of the species to varying extents. By size in the present context is meant the volume occupied by a molecule in 3-space; shape on the other hand expresses the distribution of the molecular volume in 3-space. Some indices reflect predominantly the size of the molecule, e.g., the carbon number index, while others, e.g. the Balaban centric index, are designed to characterize the shape. The earliest indices were generally better descriptors of size than shape whereas a number of the more recent indices provide a fairly sensitive characterization of shape. There has thus been a growing tendency to develop indices which are able to accurately express the shape of a molecule, and this has led to an increased interest in the phenomenon of branching in molecular species and its precise definition. Before discussing topological indices further, it will therefore be necessary to say something on the current state of the art in characterizing molecular branching.

We shall assume that all the molecular species of concern to us here can be represented by means of an appropriate chemical graph; readers wishing to know more on the theory of graphs are referred to introductory presentations on the subject [1,2]. For simplicity, we shall restrict our coverage mainly to hydrocarbon molecules, and these will be represented by their hydrogen-suppressed graph, i.e. only the carbon skeleton will be taken into consideration, a practice commonly adopted in this field. Branching is said to occur in a chemical graph whenever a vertex in the graph has a degree of three or greater; each vertex of this type is referred to as a *branching point*. A rough measure of the extent of branching present in a molecule is provided by the number of branching points it contains. On this basis, the three alkane molecules 2-methyl heptane, 3-methyheptane, and 4-methyheptane, for instance, all display the same extent of branching. Problems arise, however, when we wish to order such molecules according to their physicochemical or other properties. Since they have the same extent of branching, they cannot be ordered in terms of this measure;

consequently, much effort has been expended on devising more precise measures of molecular branching. To this end, use has been made of matrices, codes, sequences, Young diagrams, polynomials, and graph invariants [3]. It is on the latter means of characterizing branching that our main interest will be focused.

To determine the extent to which a given topological index reflects molecular size vis-à-vis molecular shape, it is necessary to have at our disposal effective measures of both of these parameters. The size of a molecule can be comparatively easily computed from the hard sphere van der Waals' radii of the various atoms it contains. Integration over all of the atoms yields the van der Waals' envelope for the molecule in question as a reasonably reliable indicator of size [4]. The shape of a molecule is much more difficult to assess, as meaningful shape descriptors are difficult to devise. Frequently, the shape of a molecule has been equated to the extent of branching present in the molecule. A simple measure of this was advanced by Motoc et al. [5], who proposed that the extent of branching in a molecule, B(G), be defined by the equation:

$$B(G) = \underline{n}_{BP} + \underline{n}_{SC} , \qquad (1)$$

where $\underline{n}_{BP}$ is the number of branching points, and $\underline{n}_{SC}$ is the number of side chains in the chemical graph of the molecule. It is easy to show for tree graphs that B(G) will have a minimum value of zero and a maximum value of $(\underline{n} - 2)$, where $\underline{n}$ is the number of vertices in the chemical graph. It is therefore not untenable to employ such simple measures of branching, for ultimately any definition of branching must rest on an intuitive basis [3]. Because of this circumstance, the use of sophistry in defining the concept of branching appears unlikely to lead to a more viable definition.

The Uniqueness of Topological Indices

Much effort and ingenuity have gone into the endeavor to produce topological indices that give unique characterizations of chemical species, i.e. indices that will differ in value whenever they characterize two graphs which are not isomorphic. To date, it has not been possible to construct a unique topological index, and, because the problem is very hard, the focus of attention has now shifted to the question whether such indices are theoretically possible. While waiting for a resolution of this issue, chemists have been devoting considerable time to the study of which indices are closest to being unique. For this purpose, a parameter known as the mean isomer degeneracy, $\bar{m}$, was introduced by Bonchev et al. [6]. This parameter is defined by the equation:

$$\bar{m} = N_{\text{isomers}}/t, \tag{2}$$

where N_{isomers} represents the number of chemical isomers for a given $\underline{n}$ which possess nonisomorphic graphs, and t is the corresponding number of different values assumed by the topological index. A variety of workers have determined $\bar{m}$ values for alkane tree graphs [6-8]. Examples of nonisomorphic graph pairs having identical index values are illustrated in Figure 1.

It is, of course, still quite possible to correlate molecular properties using nonunique topological indices. In fact, an index that is unique is not necessarily always the best descriptor to employ in structure-property correlations. This is especially true whenever differing structures have closely similar properties. The earliest topological index, usually known today as the carbon number index, provides an instructive example of how far one can go with an index that is very far from being unique. Although the carbon number index has been in existence for over a century [9], it is only comparatively recently that it has been recognized as a topological index. The carbon number index, $\underline{n}_C$, is defined simply as the number of vertices in the hydrogen suppressed graph of the hydrocarbon molecule. In early work the index was used to model the physicochemical properties of alkane species, such as their boiling point and refractive index [10]; in more recent times the index has been used in the study of chromatographic retention times [11], anesthetic potency [12], and carcinogenic behavior [13] among others. A plot of $\underline{n}_C$ against boiling point for normal alkane species is shown in Figure 3. From this plot, it is evident that $\underline{n}_C$ can be used to model properties such as the boiling point. However, the big disadvantage of this index is that it can be used only for normal, i.e. straight chain, species. All isomers having a given number of carbon atoms will have the same value of the $\underline{n}_C$ index. This implies that $t = 1$ in equation (2) and that the mean isomer degeneracy, $\bar{m}$, will always be equal to the number of isomers. The range of applicability of $\underline{n}_C$ is thus strictly limited.

To overcome the limitations of the carbon number index, increasingly vigorous campaigns have been mounted over the past decade to invent indices which are not only capable of effectively characterizing all the alkane isomers but which can provide a near unique representation of all graphs of chemical interest. Although over one hundred indices have been put forward to date, including a large number of information-theoretical indices and even one superindex (a summation of other indices), none has been demonstrated to be unique. Many of these indices have hardly been investigated at all since their postulation, and much work remains to be done in examining their behavior as molecular descriptors and in determining their degree of correlation with other indices [5]. Only a handful of topological indices have been widely used in chemistry for correlational studies involving the physicochemical and other properties of molecular species.

These indices we now discuss in some detail below. Before doing so, we present in Table 1 a listing of the $\bar{m}$ values [6] for the indices we consider.

The Wiener Topological Index

The first attempt to develop an index which could characterize molecular branching can be traced back to the work of Wiener [14,15] in 1947. One of the graph-theoretical parameters he put forward, originally described as the path number and nowadays referred to as the Wiener index, was intended to reflect the branching pattern in alkane species. Wiener defined his index as the sum of the chemical bonds existing between all pairs of carbon atoms in the molecule under consideration. The index was later shown by Hosoya [16] to be equivalent to one half the sum of the entries in the distance matrix of the graph of the molecule. In symbols, the Wiener index can thus be defined as follows:

$$W(G) = \frac{1}{2} \sum_{i} \sum_{j} d_{ij} , \tag{3}$$

where d_{ij} represents the ijth entry in the distance matrix of the graph G. The mathematical properties of this index have been investigated by Plesnik [17], and its applications to chemistry have been discussed by Rouvray [18].

The Wiener index has been used to model a wide range of the physicochemical properties of alkane species. Wiener himself [14,15] employed this and other indices for biparametric correlations with a number of properties, including the boiling point and various thermodynamic parameters, and obtained curves similar to that in Figure 2. The early work was followed by Stiel and Thodos [19] who used W(G) to predict critical constants; Rouvray and Crafford [20] who correlated properties such as density, viscosity and surface tension; and by Papazova et al. [21] and Bonchev et al. [22] who correlated chromatographic retention times. In all cases linear regression analysis yielded high correlation coefficients; in the latter work [22], the correlation coefficient exceeded 0.999. The index has also been used in the prediction of antibacterial activity [23]. In general, the correlations obtained are found to be good whenever molecular size rather than shape is the factor of prime importance. However, when shape is the decisive factor, the shortcomings of the Wiener index are clearly revealed. In a plot of W(G) versus boiling point for the 75 isomeric decanes, for instance, we found a very wide scatter in the points, as evidenced in Figure 3. The correlation coefficient for this particular scatter plot was only 0.0035. This finding accords with the results of Motoc and Balaban [24], who demonstrated that circa 90% of the value of the index reflects molecular size.

Some of the greatest successes with the Wiener index have been achieved for large systems. For instance, Rouvray and Pandey [25] have shown how the Wiener index can be used to gain valuable information on the mean configuration adopted by long chain alkane molecules at their boiling point. Using the concept of fractal dimensionality on such molecules, they proved that the ratios $b_1/b_{\underline{x}}$ $(1 \underline{<} \underline{x} \underline{<} \underline{n}_C)$ of the slopes of logarithmic scale plots of W(G) versus boiling point (see Figure 4) tend to the limiting value of 0.6 for normal alkanes of infinite length. As may be seen from Figure 5, this limit appears to be approached in practice. It is possible to estimate the mean configuration of alkane molecules from appropriate slope ratios in Figure [illegible]. Another use of W(G), also yielding information on long chain alkane and other polymeric species was advanced by Mekenyan et al. [26]. Here the basic idea was to normalize W(G) to give it a finite value for an infinite chain of monomeric units. Values of W(G) for some of the various monomeric units considered are listed in Table 2. Substitution of the normalized W(G) value into the corresponding regression equation afforded good estimates of properties of the infinite chain, e.g. its melting point and refractive index. A third major role for W(G) is in the modeling of solid sate phenomena. The favored vacancy positions in crystallite lattices, for instance, can be ascertained by calculating differences in W(G) for the structure in question and a reference structure. Minimization of these differences leads to the structure actually adopted. Trends in both the π-electron and LOMO energies can be modeled in this way for different lattices. The method has been applied to the study of vacancy migrations along preferred diffusion paths [27], the optimum positions in lattices of double and triple vacancies [28], the optimal positioning of defect atoms in lattices [29], and modeling of crystal growth processes [30].

The Hosoya Topological Index

After the pioneering work of Wiener, a major step forward was taken in 1971 when Hosoya proposd a new topological index [16]. The index is defined by the equation:

$$Z(G) = \sum_{\underline{k}=0}^{[\underline{n}/2]} p(G,\underline{k}) \,, \qquad (4)$$

where $p(G,\underline{k})$ is the number of ways in which $\underline{k}$ disconnected K_2 graphs can be imbedded in G as subgraphs, and $[\underline{n}/2]$ represents the maximal value assumed by the integer $\underline{k}$. From this definition, it follows that p(G,0) will be unity, p(G,1) represents the number of edges in G, and that $p(G,\underline{n}/2)$ is the number of 1-factors (Kekulé structures) in G. The Hosoya index is closely related to several other graph invariants, especially polynomials. For instance, the values adopted by the index for path graphs form members of the Fibonacci series while the values for monocyles form members of the Lucas series. Moreover, Z(G) is associated with the characteristic polynomial, P(x), of a given graph G. In the case of tree graphs, the relationship assumes the form [16]:

$$P_T(\underline{x}) = \sum_{\underline{k}=0}^{[\underline{n}/2]} (-1)^{\underline{k}} p(T,\underline{k}) \underline{x}^{\underline{n}-2\underline{k}} , \qquad (5)$$

where $P_T(x)$ is the characteristic polynomial for a tree graph, T; for cyclic graphs additional terms are added to equation (5). It may also be noted that the largest eigenvalue of $P_T(\underline{x})$ is intimately related to the extent of branching in T [31]. Numerous other mathematical relationships have been established by Hosoya and his associates [32-34].

The Hosoya index has found applications in a variety of different physical and chemical settings [35]. Like many other topological indices, the index has been used to model the physicochemical properties of hydrocarbon species, such as the boiling point [5,16]. Unlike most other indices, however, a number of empirical rules have been put forward [36] which prescribe the extent of lowering of the boiling point in various isomeric species. Thus, it was established that the lowering of the boiling point of alkanes due to monomethyl or geminal-dimethyl substitution alternates as the site of substitution moves in from the end of the main chain. Furthermore, if two substituents are far removed from each other, the lowering effect caused by both will be additive, though this will not be the case when the substituents are close. Gutman [37] has shown that Z(G) is particularly suitable for modeling alternations in boiling point in substituted alkane species. For instance, the alternation in monomethyl substituted heptanes, octanes and nonanes are well modeled by Z(G) whereas other indices, such as the Wiener index, are not capable of reflecting this behavior (see Figure 6). Z(G) has also been shown [38] to correlate in a linear fashion with the absolute entropy of alkanes species, though the correlation again becomes less reliable the greater the degree of steric overcrowding in the molecule. The interesting observation was made that log Z(G) values model well both the diminution of the rotational degree of freedom of a molecule with increasing branching and the decrease in the partition function which arises from overcrowed conformations. Additionally, the index Z(G) has been employed in the study of dimer statistics [39], in the coding and identification of chemical graphs [40], and in the prediction of the π-electron structure of unsaturated hydrocarbon species [41].

The Topological Indices of Randić

Following the Hosoya index, the next major topological index to be proposed was that of Randić [42]. This index, which is nowadays widely referred to as the molecular connectivity index, is the only index to have had two entire books [43,44] devoted to it. The index has the symbol χ and is as widely known in the biological sciences as in the physical sciences because of its very widespread application. In its original form, the index was defined in the following way:

$$\chi(G) = \sum_{\text{edges}} (p_i p_j)^{-\frac{1}{2}} , \qquad (6)$$

where the p_i and p_j represent the degrees of the adjacent pair of vertices i and j in G. The index was generalized into a series of indices [45] in which summations were made over subgraphs of G other than edges; for the types of subgraph used for this purpose, see Figure 7. In its most general form, $\chi(G)$ is defined by the equation:

$$ {}^{h}\chi_{r}(G) = \sum_{k=1}^{\sigma_h} \prod_{i=1}^{h+1} (p_i)_k^{-\frac{1}{2}} , \qquad (7)$$

where $\underline{h}$ is the number of edges in the summation subgraph used, $\underline{r}$ is the type of subgraph used, $\sigma_{\underline{h}}$ is the number of subgraphs of type $\underline{r}$ having $\underline{h}$ edges, and the index $\underline{k}$ extends over all the $\sigma_{\underline{h}}$ subgraphs.

We shall focus here only upon the applications of molecular connectivity indices to the modeling of the physicochemical properties of molecular species; readers interested in applications in the biological sphere are referred to appropriate review articles [46,47]. indices have been employed in correlations with a large number of physicochemical parameters; examples of such correlations are collated in Table 3. An illustration of the correlation of $^1\chi$ ($\underline{h}$=1) with the $\underline{n}$-octanol/water partition coefficient is presented in Figure 8 for various hydrocarbon and other species. This particular correlation is of significance because the partition coefficient is well-known [47] to correlate closely with many biophysical parameters. Moreover, the index is at least as effective a tool as experimentally determined parameters in such studies [48]. It has been demonstrated [24] that roughly 90% of the numerical value of $^1\chi$ reflects the size of a molecule. With only 10% of the value characterizing molecular shape, the generally good correlations obtained with the properties of branched species indicate the overriding importance of the size factor. In the plot of $^1\chi$ versus boiling point for the 75 isomeric decanes, shown in Figure 9, the correlation coefficient for linear regression is marginally better than that found for the Wiener index (0.3295 as against 0.0035). It is evident therefore that in general $^1\chi$ correlates only poorly with the branching structure of alkane species.

A variety of other studies have provided much corroborative evidence. Altenburg has shown [49], for example, that $^1\chi$ is closely related to the mean square radius of a molecule, $\bar{R}^2(G)$, at least in alkane species. For a given radius, $^1\chi$ grows as the sum of the degrees of the edges in G (one of the so-called Platt numbers [50]) whereas for a given Platt number a monotonic relationship exists between $^1\chi$ and the radius. Recently, Trinajstić [51] demonstrated that $\bar{R}^2(G)$ for alkanes is related to the Wiener index, W(G), by the equation:

$$\bar{R}^2(G) = W(G)/\underline{n}^2 . \tag{8}$$

Furthermore, Hall and Kier [52] found excellent correlations of the various χ indices they studied with molecular volume for various chemical species, including alcohols, ethers and ketones. It was suggested by Edward [53] that the χ indices correlate well with the properties of alkanes because the indices correlate with both the numbers and types of different carbon atoms in the molecule and the mole fraction of gauche conformations of the molecule. For normal alkanes, for instance, he derived the relationship:

$$^1\chi = 0.457\,\underline{n}_p + 0.5\,\underline{n}_s, \tag{9}$$

where $\underline{n}_p$ and $\underline{n}_s$ represent respectively the numbers of primary and secondary carbon atoms in the molecule. He concluded that the χ indices encode information in varying proportions about both the nature and number of the carbon atoms and the degree of folding of the molecule ($\underline{Zg}$) in alkane species. Kier and Hall [44] have shown that whereas $^0\chi$ and $^2\chi$ increase with the extent of branching in tree graphs $^1\chi$ decreases; the way in which higher χ terms reflect molecular branching is at present not well understood.

In order to improve upon the discriminating power of molecular connectivity indices and to reduce their mean isomer degeneracy (see Table 1), Randić put forward a modified version of his index known as the molecular identification number [54]. The newer index was based on a count of appropriately weighted paths in G, though the weighting factor used was the same as that for the χ indices.

In mathematical formalism the index may be defined as:

$$ID(G) = n + \sum_{\text{paths}} \prod_{i=1}^{z} P(e_i), \qquad (10)$$

where $P(e_i)$ represents the weighted path e_i, which may be expressed in terms of its edges as $(e_1, e_2, \ldots, e_z)$ where $z = 0$. The index is known [55] not to provide a unique characterization of molecular structure. For the alkane isomers the first single degeneracy occurs in 15-carbon-atom species; by the time 20-carbon-atom species are reached there are 88 duplicate values of the index for a total of 366,319 isomeric species [55]. Because of its very low mean isomer degeneracy, however, the index appears suited to offer a good characterization of at least the alkanes. To date, only scant use has been made of the index for correlational purposes [54,56]. Much work remains to be done using this index, for ID(G) clearly merits further detailed study.

The Topological Indices of Balaban

Two topological indices designed specifically to model the branching in alkane and other species have recently been described by Balaban. The first of these, known as the centric index [57], represents a summation of terms derived from a stepwise pruning of the chemical graph under consideration. This index is defined only for tree graphs. The terms used in the index are obtained by squaring the number of leaves, i.e. vertices of degree one, pruned away at each step in the pruning procedure. The summation of these terms yields the index:

$$C(G) = \sum_{\text{steps}} \delta_i^2, \qquad (11)$$

where δ_i represents the number of leaves pruned away at the ith step. To eliminate size effects, Balaban [57] proposed a normalization of C(G) by subtracting from it the value of the index for the path graph on n vertices. This gave a normalized index C'(G) of the form:

$$C'(G) = \tfrac{1}{2}[\sum_{\text{steps}} \delta_i^2 - 2n + \tfrac{1}{2}(1-(-1)^n)], \qquad (12)$$

where the initial factor of $\frac{1}{2}$ has been employed to ensure that the lower bound for the path graph will equal zero.

The normalized index C'(G) was chosen by Balaban and Motoc [58] to model the octane numbers of various alkane species commonly used as fuels. The efficiency of such molecules in this context is well-known to be critically dependent upon their extent of branching. In general, the more branched an alkane molecule is, the less likely it will be to self-ignite or 'knock' upon sudden compression in air in an internal combustion engine. Correlating alkane octane numbers against their C'(G) values thus provides a sensitive test of the reliability of the index in the characterization of molecular branching. For all the heptane and octane isomers, linear regression yielded a correlation coefficient of 0.945. This was the best correlation obtained from the various indices used for the purpose, indicating that C'(G) does indeed reflect the extent of branching present in these molecular species. Since the size factor is effectively eliminated by the normalization procedure, relatively low correlations with the indices Z(G) and W(G) are to be expected; the respective correlation coefficients for alkane species in the range $4 \leq n \leq 8$ are 0.07 and 0.21. Since largely size-dependent parameters, such as boiling point, do not correlate with shape-dependent parameters, such as the octane number, it is hardly surprising that indices used to reflect size will correlate poorly here.

The second topological index of Balaban is based on the distance matrix of the graph G and is known as the averaged distance sum connectivity index [59]. For tree graphs, the index may be written as:

$$J(G) = n_e \cdot \sum_{\text{edges}} (s_i s_j)^{-\frac{1}{2}}, \qquad (13)$$

where n_e is the number of edges in the tree, and s_i and s_j represent respectively the sums of the *i*th and *j*th row of the distance matrix of the tree. The index attempts to reflect both molecular size and the extent of branching present, and increases with both of these parameters. Linear correlation of J(G) with the octane number of various alkane molecules ($4 \leq n \leq 8$) afforded a correlation coefficient of 0.92, very close to that for the C(G) index [60]. Moreover, Hanson and Rouvray [61] have demonstrated that J(G) alone, and also as a product with the hydrogen deficiency number, correlates well (0.951) with the threshold soot index for a wide assortment of hydrocarbon species. Correlation of J(G) for the 75 isomeric decanes, however, yielded a correlation coefficient of only 0.0031 (see Figure 10) -- on a par with that for $^1\chi$. Since the index is one of the more discriminating currently available (see Table 1), it is certainly deserving of further study, and perhaps also extension to include paths of length greater than one, in analogy to the Randić molecular connectivity index [45].

Conclusion

Just over 40 different topological indices have been documented in the chemical literature to date, and, for the most part, these have been put forward to model the physicochemical properties of hydrocarbon molecules. In addition to these indices, some 70 information-theoretical indices [62] have been advanced for the same purpose. Many of these descriptive indices have not been examined in any detail; most have been simply advocated without any ensuing follow-up. All of the indices have, however, been collated recently and closed formulas derived [63] enabling their values to be calculated for graphs in the form of paths and monocycles. Only a handful of the topological indices presently available have thus been employed extensively in practice, the most frequently used index being the Randić molecular connectivity index [42]. Attempts to extend the usefulness of topological indices by allowing for heteroatoms, i.e. atoms other than carbon, have been made by Kier and Hall [43,44], Lall and Srivastava [64], and Barysz et al. [65], though atomic charge is difficult to treat with purely topological models. More effective ways of treating heteroatoms will probably have to be developed in the future.

The relatively small number of indices employed so far have proven themselves capable of modeling a vast array of chemical phenomena, ranging from thermodynamic properties to molecular configurations. The indices thus seem to be modeling in a reliable way factors of fundamental importance at the molecular level. We are proposing here that the two factors in question are the molecular size and shape. The former may be readily calculated from the van der Waals' envelope [4] of the molecule under consideration; the latter is much more problematical because of the current lack of general agreement on the definition of extent of branching in molecular species [3]. However, the relative importance of the shape factor can be determined by taking a simple difference, if size and shape are the only factors considered. The outcome of such investigations is that all topological indices (except the carbon number index) incorporate to varying degrees both the shape and size factors. Thus, whereas some indices emphasize predominantly the shape of a molecule, e.g. the Balaban centric index [57], others reflect mainly its size, e.g. the Wiener index [14].

Corroborative evidence exists that topological indices yield good correlations because they model the two important parameters which are determinants of physicochemical properties. In an attempt to predict the properties of 114 diverse liquid compounds, Cramer [66,67] concluded that there were only two types of intermolecular interaction which were significant in determining observable macromolecular properties. These were what he termed the bulk and the bulk-corrected cohesiveness of a molecule. In our terms, the bulk corresponds directly to what we have called the molecular size; the cohesiveness appears to be related to electrostatic interactions occurring and will certainly be dependent upon molecular shape. Cramer's observations [67] support the contention that any property which depends primarily on nonspecific and noncovalent molecular interactions will display a similar interaction mechanism. Such properties can therefore be predicted from the molecular structure alone and are not an artifact of the factor analysis. Nonspecific biological responses [46], i.e. those that are not influenced significantly by receptor shape, are also included here and may thus also be predicted by means of topological indices [47]. It is intriguing that manifold physicochemical and other properties can be modeled so effectively using no more than these comparatively simple parameters of molecular size and shape.

Acknowledgments

The author is indebted to the U.S. Office of Naval Research for partial support of this work. Thanks are also due to Ms. C. Pirie for running the scatter plots shown in Figures 3, 9, and 10.

References

1. F. Harary, Graph Theory, Addison-Wesley, Reading, Massachusetts, 1969.
2. J.W. Essam and M.E. Fisher, Revs. Mod. Phys. 42, 272 (1970).
3. D.H. Rouvray, Discrete Appl. Math., in press.
4. A. Bondi, J. Phys. Chem. 68, 441 (1964).
5. I. Motoc, A.T. Balaban, O. Mekenyan, and D. Bonchev, Math. Chem. 13, 369 (1982).
6. D. Bonchev, O. Mekenyan, and N. Trinajstić, J. Comput. Chem. 2, 127 (1981).
7. M. Barysz, J.V. Knop, S. Pejaković, and N. Trinajstić, Polish J. Chem. 59, 405 (1985).
8. M. Razinger, J.R. Chrétien, and J.E. Dubois, J. Chem. Inf. Comput. Sci. 25, 23 (1985).
9. M. Goldstein, Chem. Ber. 12, 689 (1879).
10. A.W. Francis, Ind. Eng. Chem. 33, 554 (1941).
11. J.K. Haken and I.O. Korhonen, J. Chromatogr. 357, 253 (1986).
12. P. Ménassa, T. Lussier, and C. Sandorfy, Can. J. Chem. 64, 633 (1986).
13. H. Druckrey, P. Preussmann, S. Ivanovic, and D. Schmähl, Zeitschr. Krebsforsch, 69, 103 (1967).
14. H. Wiener, J. Amer. Chem. Soc. 69, 17, 2636 (1947).
15. H. Wiener, J. Chem. Phys. 15, 766 (1947).
16. H. Hosoya, Bull. Chem. Soc. Japan 44, 2332 (1971).
17. J. Plesnik, J. Graph Theory 8, 1 (1984).
18. D.H. Rouvray, in Mathematics and Computational Concepts in Chemistry, ed. N. Trinajstić, Horwood Publishers, Chichester, U.K., 1986, p. 295.
19. L.I. Stiel and G. Thodos, Am. Inst. Chem. Eng. J. 8, 527 (1962).
20. D.H. Rouvray and B.C. Crafford, South Afr. J. Sci. 72, 74 (1976).

21. D. Papazova, M. Dimov, and D. Bonchev, J. Chromatogr. 188, 297 (1980).

22. D. Bonchev, O. Mekenyan, G. Protić, and N. Trinajstić, J. Chromatogr. 176, 149 (1979).

23. M. Medić-Sarić, D. Maysinger, and M. Movrin, Acta Pharm. Jugosl. 33, 199 (1983).

24. I. Motoc and A.T. Balaban, Rev. Roumaine Chim. 26, 593 (1981).

25. D.H. Rouvray and R.B. Pandey, J. Chem. Phys., 85, 2286 (1986).

26. O. Mekenyan, S. Dimitrov, and D. Bonchev, Eur. Polym. J. 19, 1185 (1983).

27. D. Bonchev, O. Mekenyan and H.-G. Fritsche, Phys. Stat. Sol. 55A, 181 (1979).

28. O. Mekenyan, D. Bonchev, and H.-G. Fritsche, Phys. Stat. Sol. 56A, 607 (1979).

29. O. Mekenyan, D. Bonchev, and H.-G. Fritsche, Zeit. Phys. Chem. Leipzig 265, 959 (1984).

30. D. Bonchev, O. Mekenyan, and H.-G. Fritsche, J. Cryst. Growth 49, 90 (1980).

31. D.M. Cvetković and I. Gutman, Croat. Chem. Acta 49, 115 (1977).

32. H. Hosoya, Theor. Chim. Acta 25, 215 (1972).

33. H. Hosoya and K. Hosoi, J. Chem. Phys. 64, 1065 (1976).

34. I. Gutman, T. Yamaguchi, and H. Hosoya, Bull. Chem. Soc. Japan 49, 1811 (1976).

35. H. Hosoya, in Mathematics and Computational Concepts in Chemistry, ed. N. Trinajstić, Horwood Publishers, Chichester, U.K., 1986, p. 110.

36. H. Hosoya, K. Kawasaki, and K. Mizutani, Bull. Chem. Soc. Japan 45, 3415 (1972).

37. I. Gutman, Zeit. Physik. Chem., in press.

38. H. Narumi and H. Hosoya, Bull Chem. Soc. Japan 53, 1228 (1980).

39. H. Hosoya and A. Motoyama, J. Math. Phys. 26, 157 (1985).

40. H. Hosoya, J. Chem. Docum. 12, 181 (1972).

41. H. Hosoya, K. Hosoi and I. Gutman, Theor. Chim. Acta 38, 37 (1975).

42. M. Randić, J. Amer. Chem. Soc. 97, 6609 (1975).

43. L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.

44. L.B. Kier and L.H. Hall, Molecular Connectivity in Structure-Activity Analysis, Research Studies Press, Letchworth, United Kingdom, 1986.

45. L.B Kier, W.J. Murray, M. Randić and L.H. Hall, J. Pharm. Sci. 65, 1226 (1976).

46. A. Sabljić and N. Trinajstić, Acta Pharm. Jugosl. 31, 189 (1981).

47. D.H. Rouvray, Acta Pharm. Jugosl., in press.

48. S.C. Basak, D.K. Harriss, and V.R. Magnuson, J. Pharm. Sci. 73, 429 (1984).

49. K. Altenburg, Zeit. Phys. Chem. Leipzig 261, 389 (1980).

50. J.R. Platt, J. Chem. Phys. 15, 419 (1947).

51. N. Trinajstić, Chemical Graph Theory II, C.R.C. Press, Boca Raton, Florida, 1983, p. 116.

52. L.H. Hall and L.B. Kier, Eur. J. Med. Chem. 16, 399 (1981).

53. J.T. Edward, Can. J. Chem. 60, 480 (1982).

54. M. Randić, J. Chem. Inf. Comput. Sci. 24, 164 (1984).

55. K. Szymanski, W.R. Müller, J.V. Knop, and N. Trinajstić, J. Chem. Inf. Comput. Sci. 25, 413 (1985).

56. M. Randić, Int. J. Quant. Chem., Quant. Biol. Symp. 11, 137 (1984).

57. A.T. Balaban, Theor. Chim. Acta 53, 355 (1979).

58. A.T. Balaban and I. Motoc, Math. Chem. 5, 197 (1979).

59. A.T. Balaban, Chem. Phys. Lett. 89, 399 (1982).

60. A.T. Balaban, Pure and Appl. Chem. 55, 199 (1983).

61. M.P. Hanson and D.H. Rouvray, J. Phys. Chem., in press.

62. D. Bonchev, Information-Theoretic Indices for Characterization of Chemical

Structures, Research Studies Press, Chichester, U.K., 1983.

63. D.H. Rouvray and C. Raychaudhury, J. Comput. Chem., submitted.

64. R.S. Lall and V.K. Srivastava, Math. Chem. 13, 325 (1982).

65. M. Barysz, G. Jashari, R.S. Lall, V.K. Srivastava, and N. Trinajstić, Studies Phys. Theor. Chem. 28, 222 (1983).

66. R.D. Cramer, J. Amer. Chem. Soc. 102, 1837 (1980).

67. R.D. Cramer, J. Amer. Chem. Soc. 102, 1849 (1980).

Captions for Figures and Tables

Figure 1. Examples of nonisomorphic graph pairs having the same eigenvalue spectrum, the same Wiener index, the same Randić $^1\chi$ molecular connectivity index, and the same Balaban J averaged distance sum connectivity index.

Figure 2. Plot of the carbon number index, n_C, versus boiling point for the first forty normal alkanes ($1 \le n \le 40$).

Figure 3. Scatter plot of the boiling point versus Wiener index for the 75 isomeric decane species ($C_{10}H_{22}$).

Figure 4. Plot on logarithmic scales of the Wiener index versus boiling point for the first forty normal alkanes ($1 \le n \le 40$) showing changing slope.

Figure 5. Plot of the slope ratios in Figure 6 versus an averaged carbon number for successive sets of nine points.

Figure 6. Plot of boiling point versus the Randić molecular connectivity index, $^1\chi$, the Wiener index, W, and the Hosoya index, Z, for the monomethyl substitute tridecanes. Based on ref. [38].

Figure 7. Illustration of the types of subgraph used in calculating the higher order ($h \ge 3$) Randić molecular connectivity indices, $^h\chi$.

Figure 8. Semilogarithmic plot of *n*-octanol/water partition coefficient versus the Randić molecular connectivity index, $^1\chi$, for various organic species. Based on W.J. Murray, L.H. Hall, and L.B. Kier, J. Pharm. Scis 64, 1978 (1975).

Figure 9. Scatter plot of the boiling point versus the Randić molecular connectivity index, $^1\chi$, for the 75 isomeric decane species ($C_{10}H_{22}$).

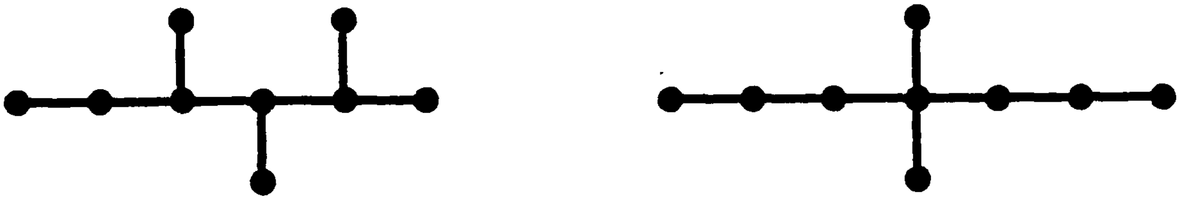
Figure 10. Scatter plot of the boiling point versus the Balaban averaged distance sum connectivity index, J, for the 75 isomeric decane species ($C_{10}H_{22}$).

Table 1. Listing of the mean isomeric degeneracy values, $\bar{m}$, for alkane tree isomers in the range $2 \leq n \leq 10$ for severn different topological indices.

Table 2. Closed formulas for the value of the Wiener index for various monomeric units which can form polymeric chains.

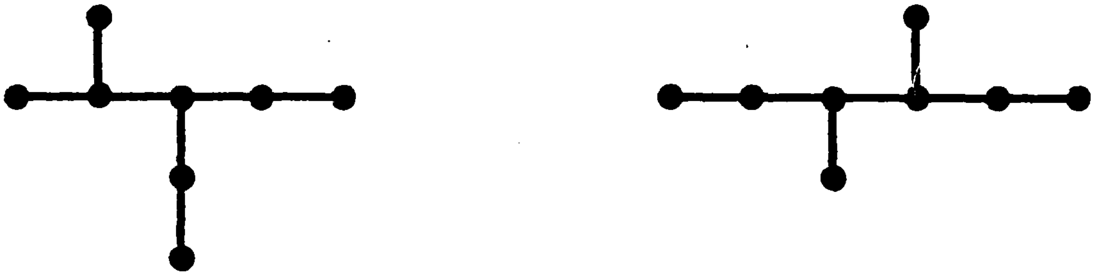
Table 3. Examples of linear regression equations obtained from correlations of various physicochemical parameters versus the Randić molecular connectivity index, $^1\chi$, and valence corrected index, $^1\chi^v$, for different organic species.

Isospectral



Same Wiener Index

Same Connectivity Index



Same Balaban Index

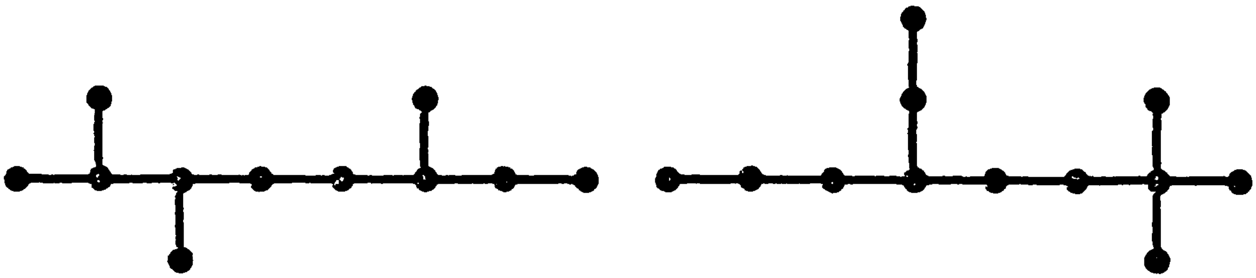


FIGURE 1

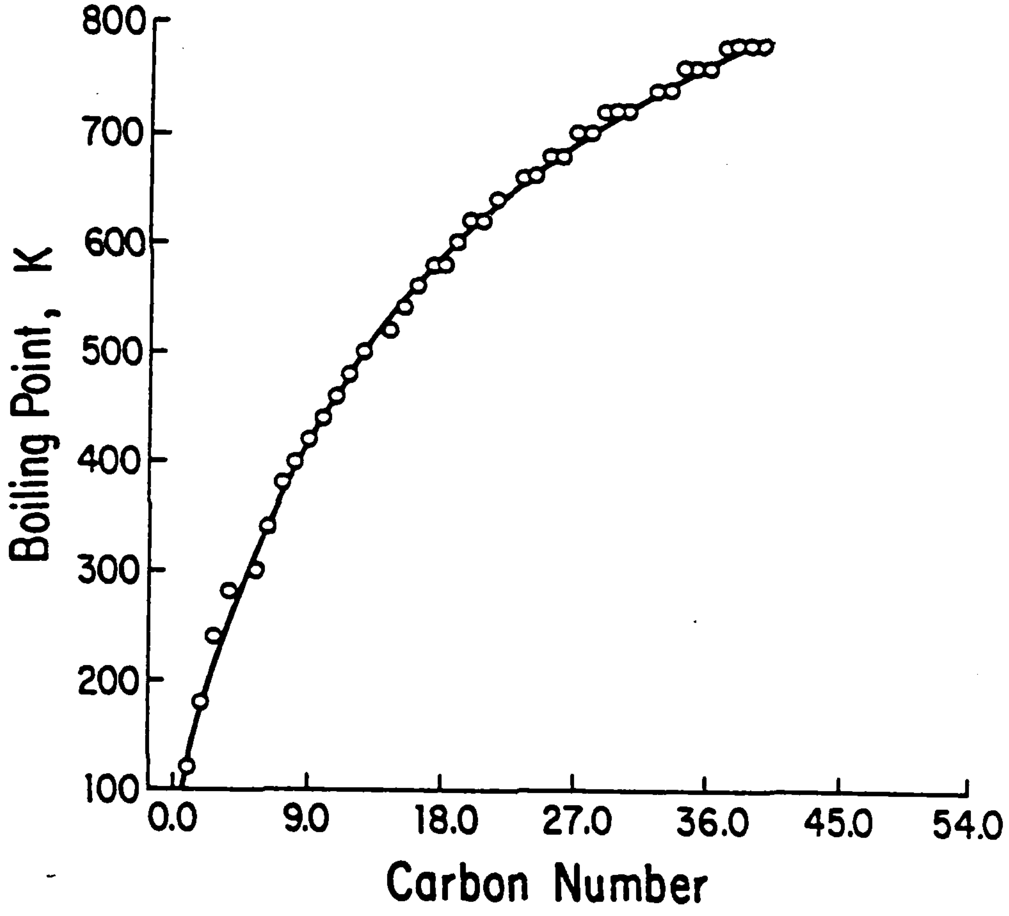

800
700
600
500
400
300
200
100
Boiling Point, K
0.0
9.0
18.0
27.0
36.0
45.0
54.0
Carbon Number

FIGURE 2

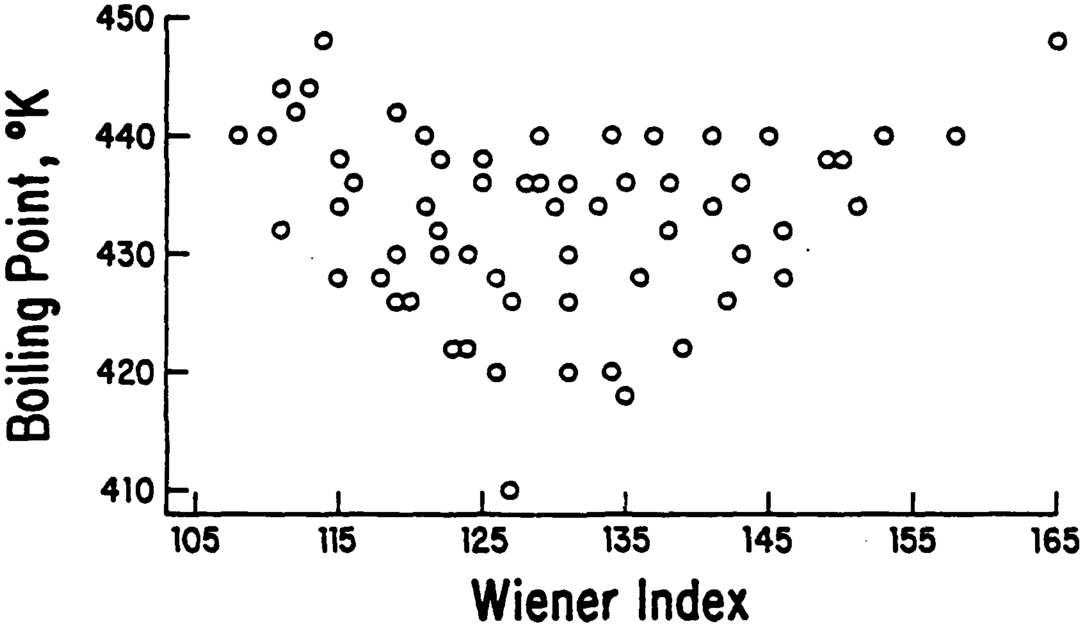

450
440
430
420
410
Boiling Point, °K
105
115
125
135
145
155
165
Wiener Index

FIGURE 3

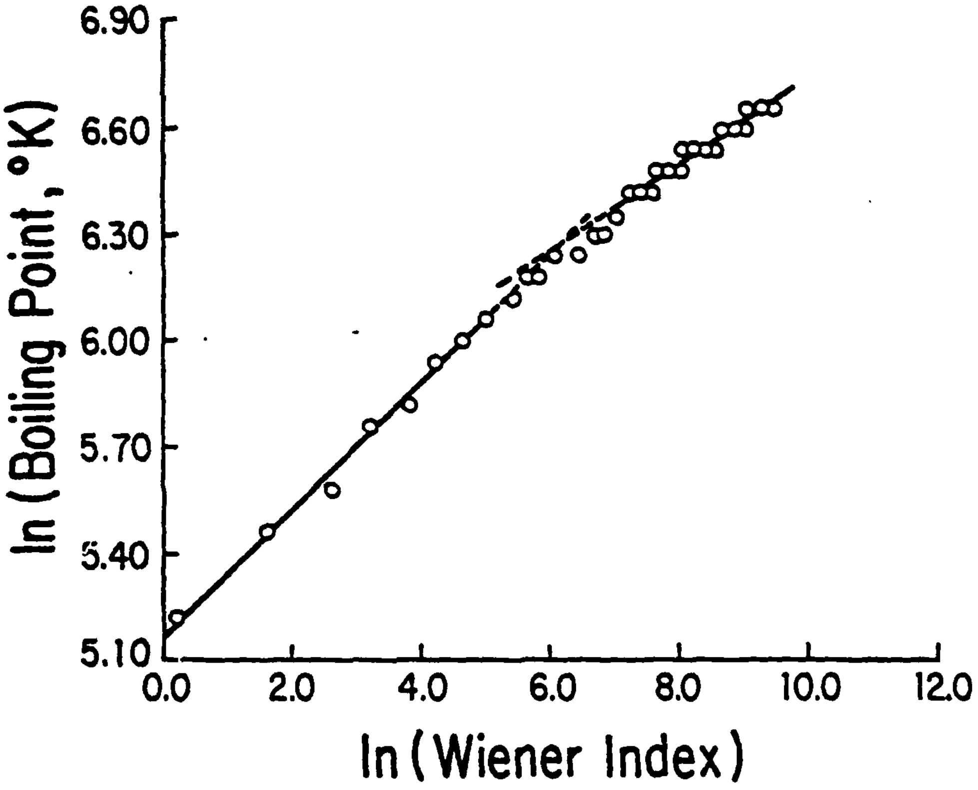

6.90
6.60
6.30
6.00
5.70
5.40
5.10
0.0
2.0
4.0
6.0
8.0
10.0
12.0
ln (Boiling Point, °K)
ln (Wiener Index)

FIGURE 4

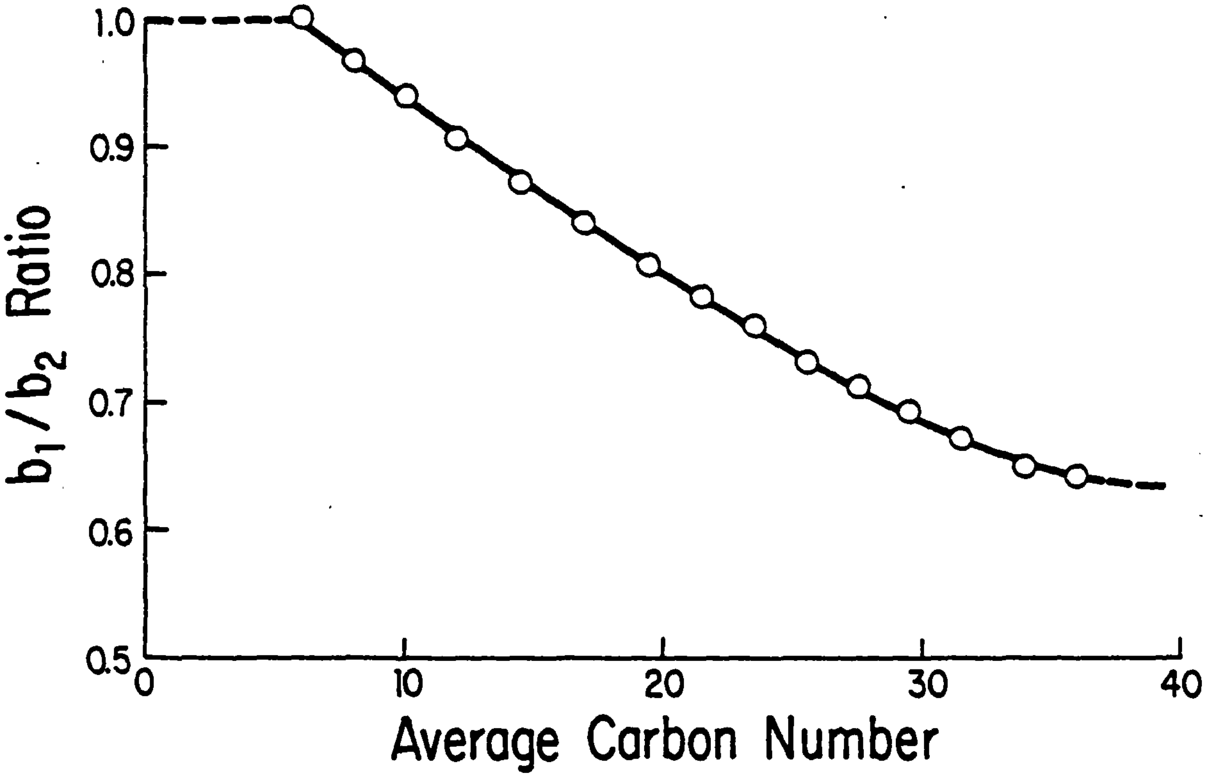


FIGURE 5

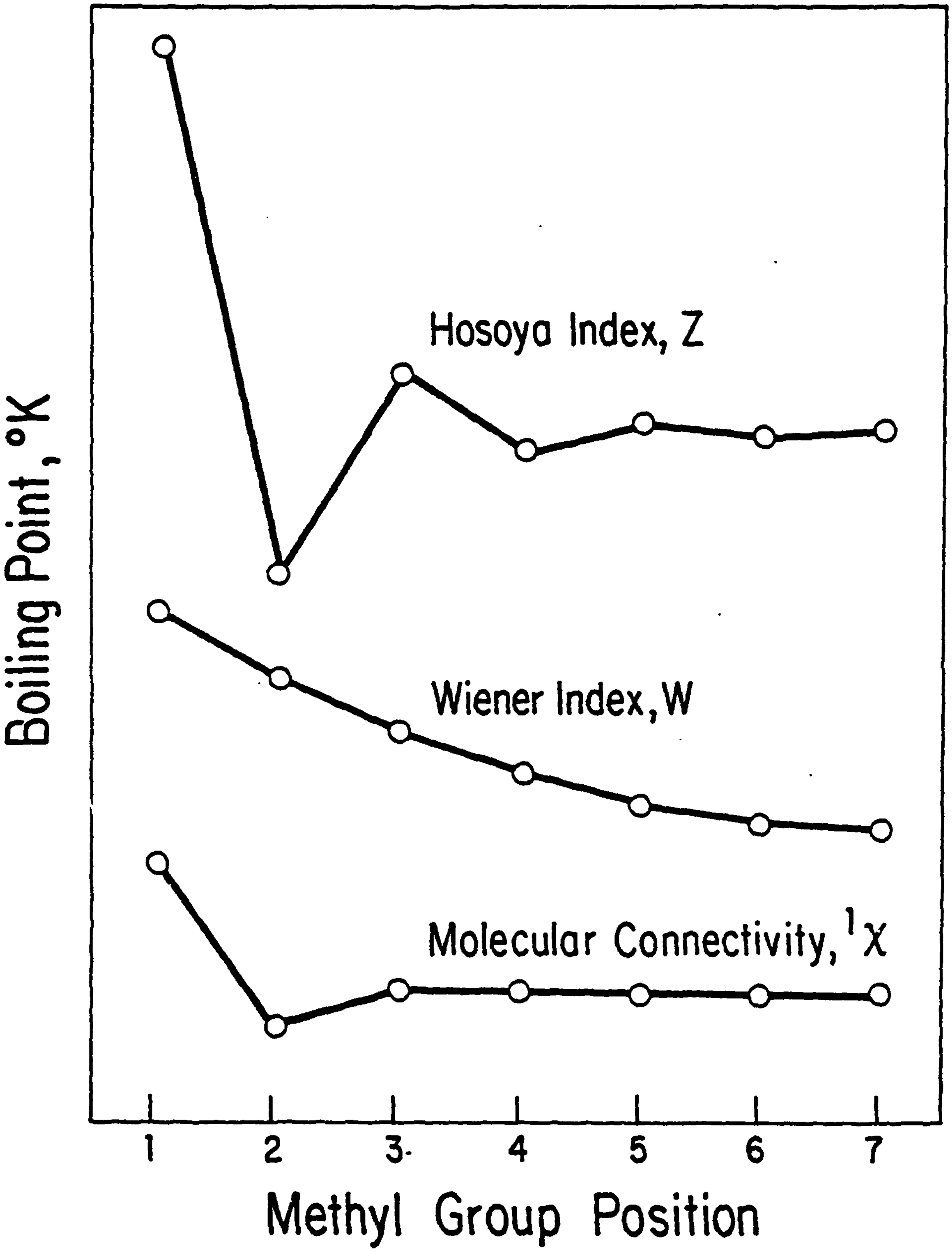

Hosoya Index, Z
Wiener Index, W
Molecular Connectivity, $^1\chi$
Boiling Point, °K
1
2
3
4
5
6
7
Methyl Group Position

FIGURE 6

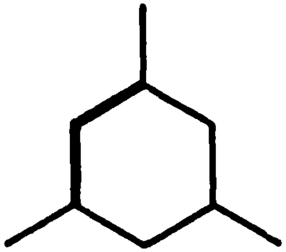
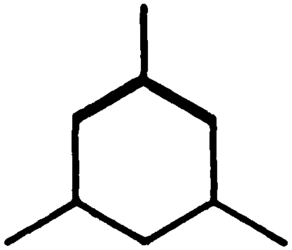
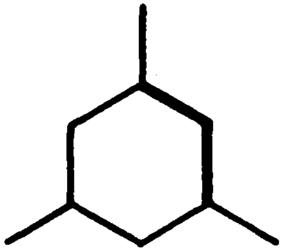
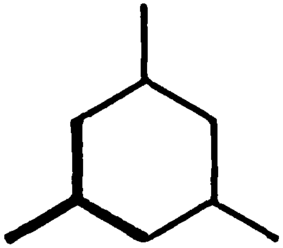
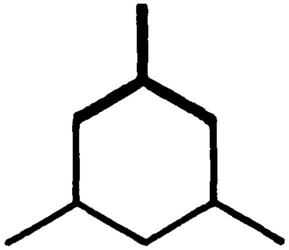
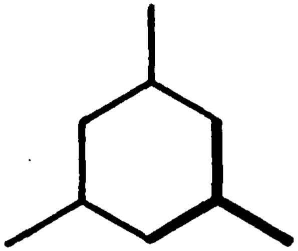
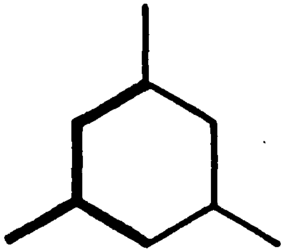
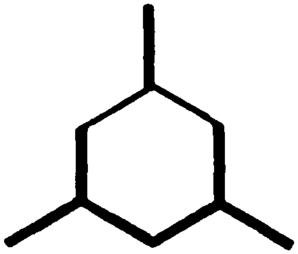
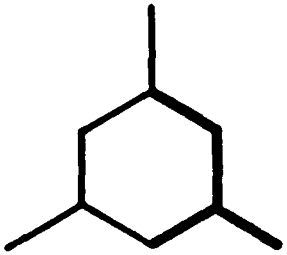
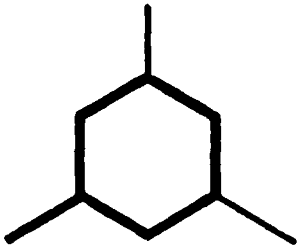
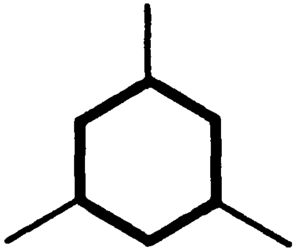
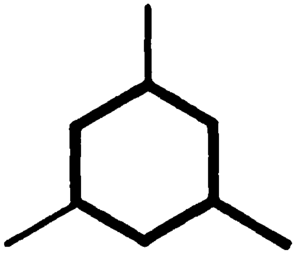
SUBGRAPH	EXAMPLES		
Path			
Cluster			
Path/Cluster			
Chain			

FIGURE 7

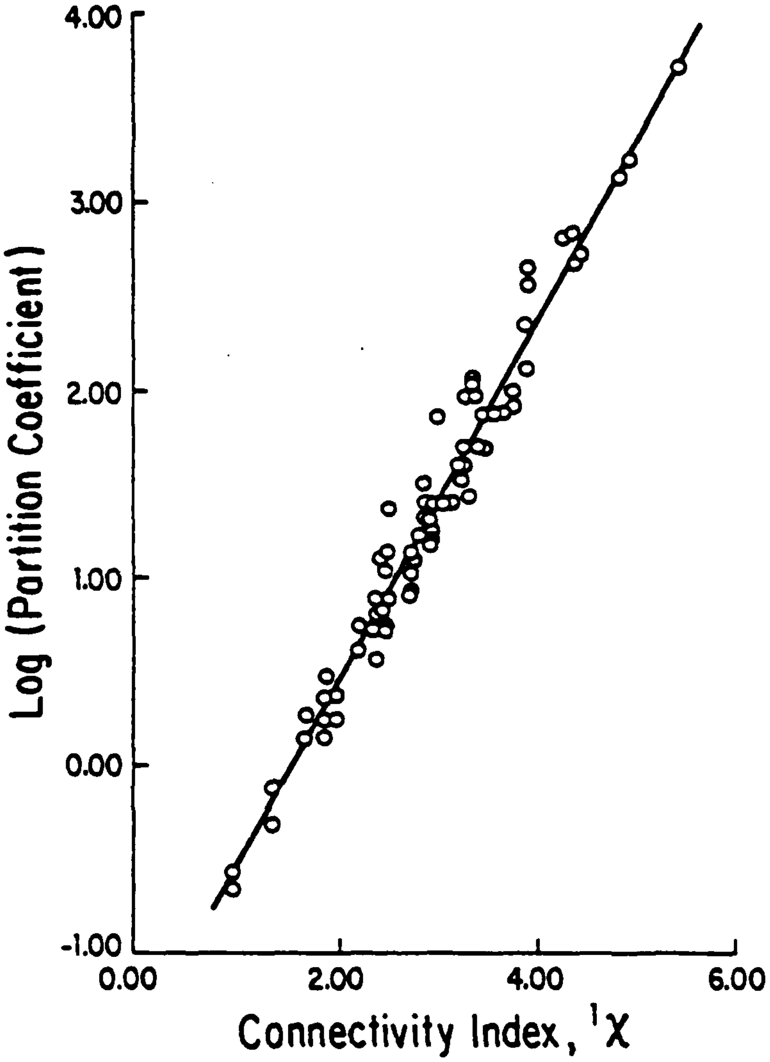
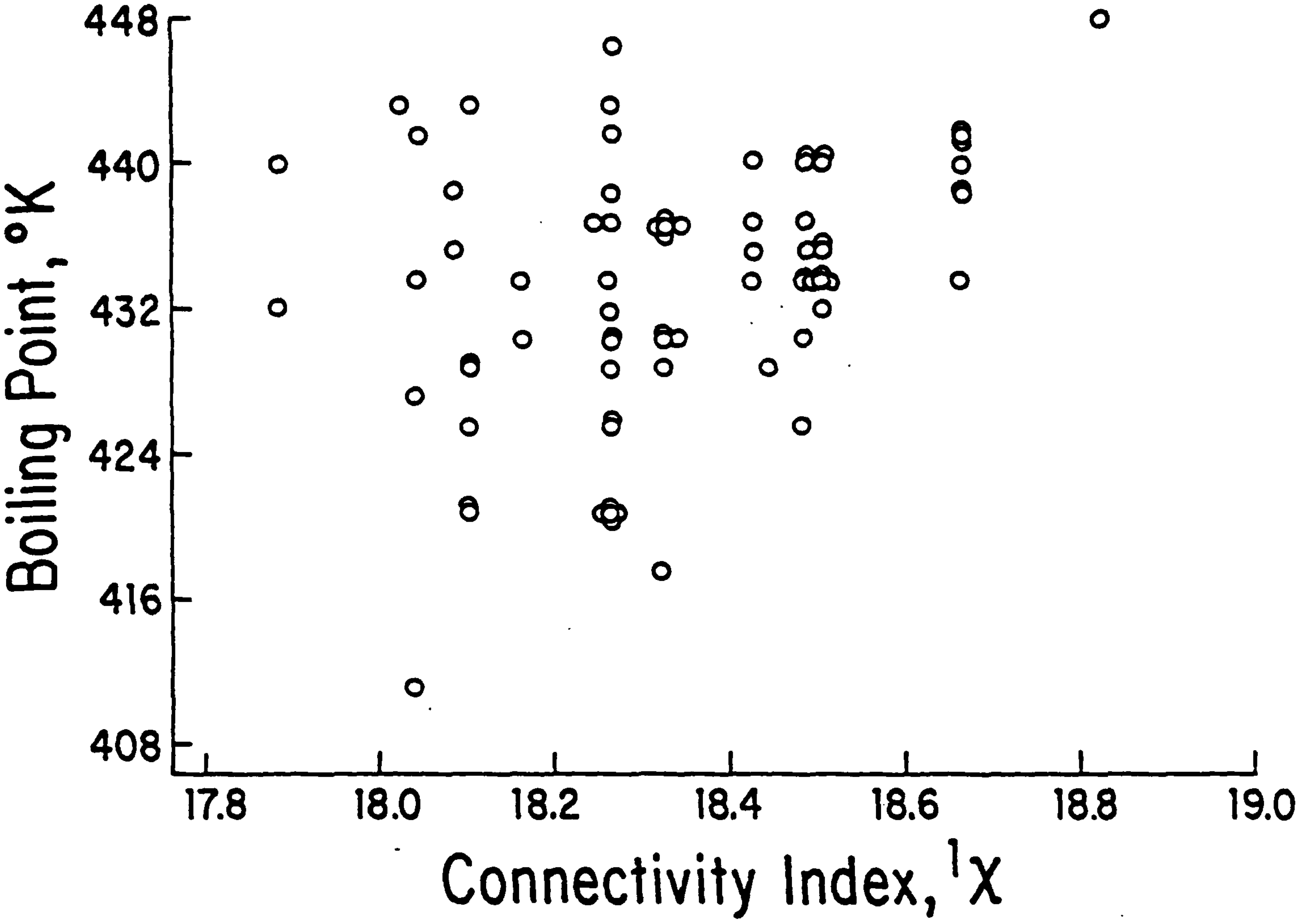
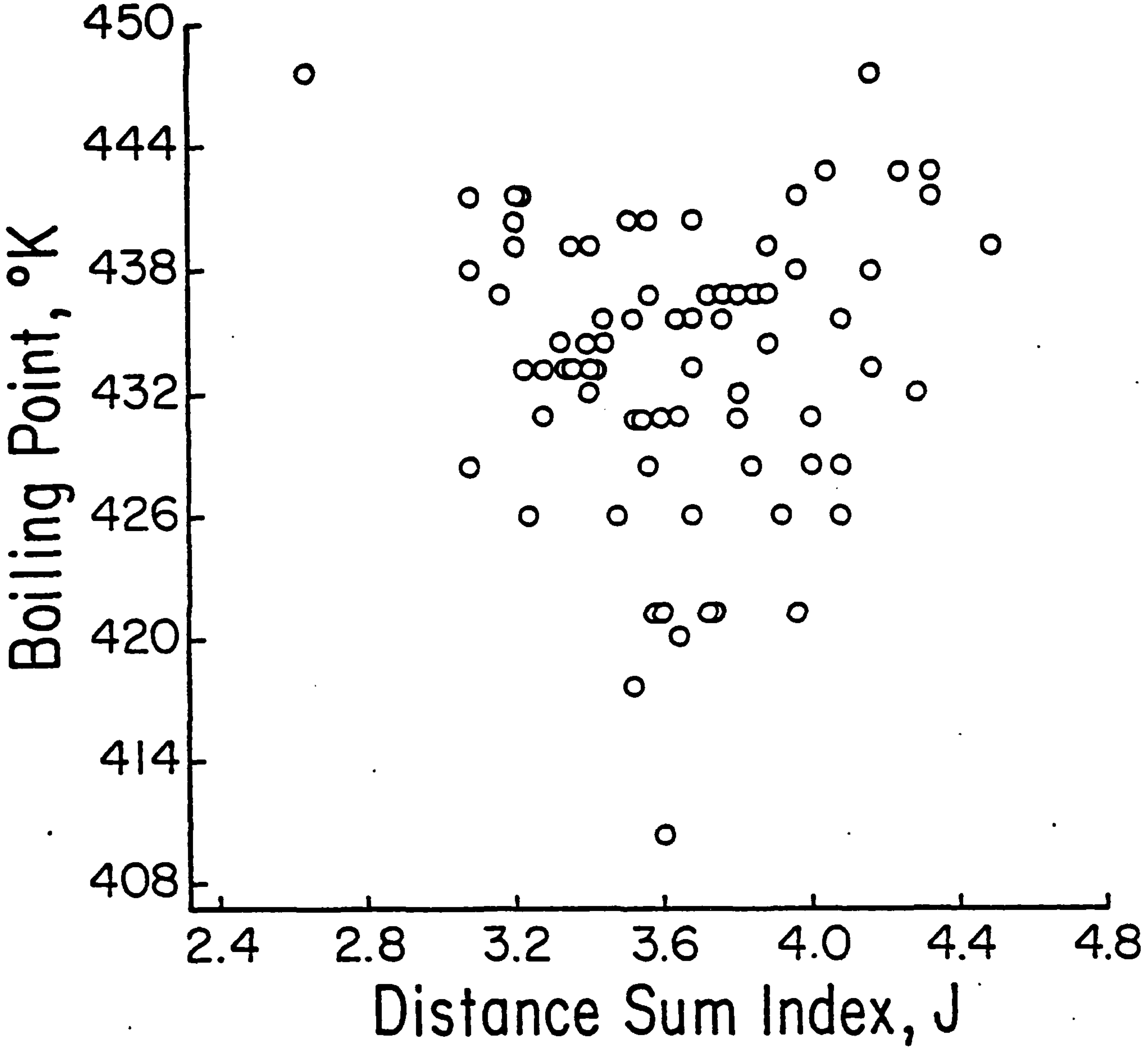

4.00
3.00
2.00
1.00
0.00
-1.00
Log (Partition Coefficient)
0.00
2.00
4.00
6.00
Connectivity Index, $^1\chi$

FIGURE 8



448
440
432
424
416
408
Boiling Point, °K
17.8
18.0
18.2
18.4
18.6
18.8
19.0
Connectivity Index, ¹X

FIGURE 9



450
444
438
432
426
420
414
408
Boiling Point, °K
2.4
2.8
3.2
3.6
4.0
4.4
4.8
Distance Sum Index, J

FIGURE 10

Table 1

Vertices in Alkane Tree	Carbon Number	Wiener	Hosoya	Randić $^1\chi$	Randić ID	Balaban C	Balaban J
2	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1
4	2	1	1	1	1	1	1
5	3	1	1	1	1	1	1
6	5	1	1	1	1	1.25	1
7	9	1.22	1	1	1	1.57	1
8	18	1.15	1.28	1.09	1	1.85	1
9	35	2.17	1.52	1.25	1	-	1
10	75	1.89	1.89	1.53	1	-	1

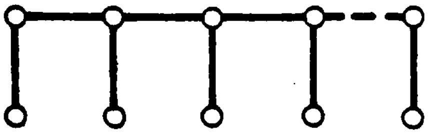
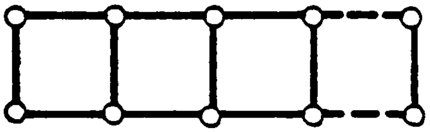
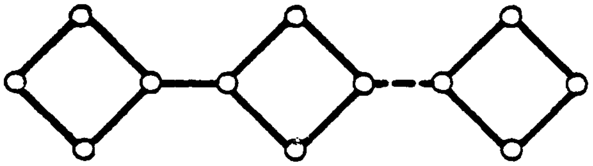
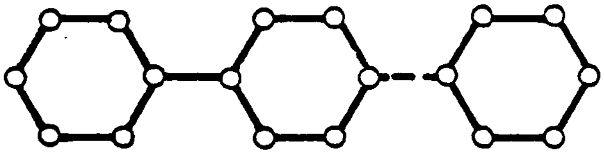
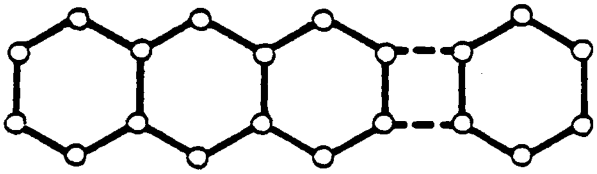
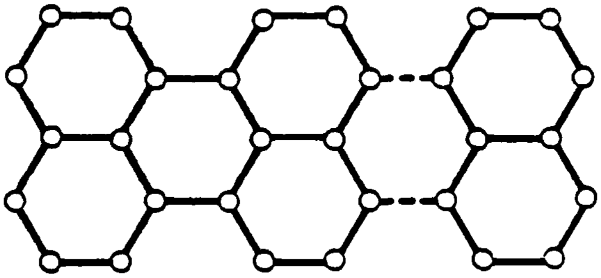
Polymeric System	**Wiener Index**
	$\frac{1}{6}(n^3-n)$
	$\frac{1}{12}(n^3+6n^2-10n)$
	$\frac{1}{2}(n^3+3n^2-4n)$
	$\frac{1}{8}n^3$
	$\frac{1}{18}(2n^3+9n)$
	$\frac{1}{12}(n^3+3n^2+2n-12)$
	$\frac{1}{30}(2n^3+235n-1080)$

TABLE 2

Table 3

Property	Correlation Equation	Data Points	Chemical Species	Correlation Coefficient
Boiling Point	$t_B = 57.85\ {}^1\chi - 97.90$	51	Alkanes	0.9851
Octanol/water Partition Coefficient	$\log p = 1.48 + 0.950\ {}^1\chi^v$	138	Various organics	0.9860
Water Solubility	$\log S = 6.702 - 2.666\ {}^1\chi$	51	Aliphatic alcohols	0.9870
Chromatographic Retention Time	$T_R = 482.12\ {}^1\chi - 559.60$	51	Arenes	0.9911
Soil Sorption	$\log K_{om} = 0.550\ {}^1\chi + 0.450$	37	Various organics	0.9730

END

4-87

DTIC