JOURNAL
OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 97, NUMBER 23 NOVEMBER 12, 1975

On Characterization of Molecular Branching

Milan Randić

Contribution from the Department of Chemistry, Tufts University, Medford, Massachusetts 02155. *Received April 11, 1974*

Abstract: A theoretical characterization of molecular branching is considered. Members of homologous series are ordered in a sequence and a numerical index is assigned to individual structures based on a differentiation of edge types of molecular graphs. Linear and branched alkanes having eight or less carbon atoms have been considered in particular and correlations between the derived branching index and properties which critically depend on molecular size and shape are established. The proposed index is also in satisfactory agreement with the empirical index of Kováts. The approach reveals some inherent relationships between isomers which can be traced to connectivity and molecular topology. It points, in some cases, to a considerable reduction in the number of experimentally deduced constants characterizing molecular properties, provided a sacrifice in precision can be tolerated and is compensated for by the significance of the indicated inter-relations. This point is illustrated on an analysis of the empirical constants of the Antoine equation.

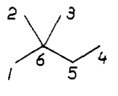
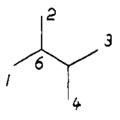
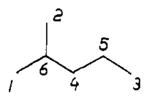
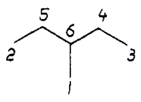
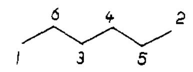
Some molecular properties depend upon molecular shape and vary regularly within a series of homologous compounds. The degree of branching of the molecular skeleton is the critical factor involved. Boiling points of hydrocarbons and the retention volumes or retention times obtained from chromatographic studies are typical of this kind of correlation. The magnitude of interaction is directly related to degree of branching, as well as molecular size. Thus the molecular topology is important in characterizing some of these experimental quantities. It would be desirable to develop a procedure which can quantitatively characterize the degree of molecular branching, providing we are able to define the intuitive concept involved. There is no unanimity of opinion as to what should be understood as molecular branching even when the discussion is confined to relatively simple systems.[1] Should 3-methylhexane, for instance, be considered more branched than 3-ethylpentane? The former has branches of unequal length and hence a longer chain, which is readily apparent from its name. The latter has branches of equal length and appears less elongated. Some authors prefer to restrict the concept and attach significance only to structures which have different number of branching vertices.[2] A less restrictive view is to associate the concept of molecular branching not exclusively with the number of branching vertices but also with their valencies and the mode of distribution.[3] Such a broader concept then permits consideration of correlations of some experimental data on such structures with a parameter characterizing the molecular branching, a concept yet to be defined.

This paper is concerned with the problem of characterizing the degree of molecular branching. The problem is approached in two stages. First, a procedure is outlined which allows ordering members of homologous series in a sequence which parallels the increase in ramification of molecular skeletons, and at least gives satisfying answers in cases beyond dispute. Next, we consider associating each individual structure with an index (parameter) which will form the same sequential ordering of the members of the series. An experimental index of branching has in fact been proposed some time ago by Kováts.[4] It is based on retention data for normal paraffins to which particular values of the index are assigned in order to provide a scale for other systems. The existence of such an index indicates the importance of topological features of the compounds involved and a search for a theoretical approach to derive these seems worthwhile. Previous theoretical characterizations of skeletal fragments have been concerned with studies of thermodynamic properties among a homologous series,[5] while more recent efforts have followed attempts to develop graph theoretical methods for discussing problems of chemistry.[6]

Ordering of Molecules According the Degree of Branching

The basis for our scheme of ordering members of a homologous series in a sequence which parallels an intuitive idea of molecular branching or degree of divergence of the molecular skeleton comes from the seemingly unrelated study of the problem of recognizing identical graphs representing molecular topology.[7] The scheme prescribes a unique numbering of atoms in a molecule based upon the requirement that the corresponding adjacency matrix associated with the graph represents the smallest number possible when interpreted as a binary number by combining the rows of the matrix into a single line. For convenience one may regard the matrix as a set of n ordered binary numbers $n_1, n_2, n_3, \ldots$ by considering the sequence of n rows instead of a single number combined from all the rows. Thus 2,2-

Table I. Numbering of Atoms for Hexane Isomers which Leads to Adjacency Matrices with the Smallest Associated Binary Labels

	$\begin{pmatrix}0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&1&0\\0&0&0&1&0&1\\1&1&1&0&1&0\end{pmatrix}$
	$\begin{pmatrix}0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&1&0\\0&0&0&0&1&0\\0&0&1&1&0&1\\1&1&0&0&1&0\end{pmatrix}$
	$\begin{pmatrix}0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&1&0\\0&0&0&0&1&1\\0&0&1&1&0&0\\1&1&0&1&0&0\end{pmatrix}$
	$\begin{pmatrix}0&0&0&0&0&1\\0&0&0&0&1&0\\0&0&0&1&0&0\\0&0&1&0&0&1\\0&1&0&0&0&1\\1&0&0&1&1&0\end{pmatrix}$
	$\begin{pmatrix}0&0&0&0&0&1\\0&0&0&0&1&0\\0&0&0&1&0&1\\0&0&1&0&1&0\\0&1&0&1&0&0\\1&0&1&0&0&0\end{pmatrix}$

dimethylbutane is represented by the sequence 1, 1, 1, 10, 101, 111010 or in decimal notation as 1, 1, 1, 2, 5, 58 (Table I). The magnitude of the binary code will depend on the assumed numbering of vertices. Although in some complex graphs the procedure to find the canonical numbering may be rather involved,[8] the unique labeling of the vertices can be found without screening many possibilities for most medium size molecular skeletons. An illustration of such a unique numbering is given for hexane isomers in Table I. In addition, the corresponding adjacency matrices are presented. The scheme not only provides the desired unique numbering of the vertices (atoms in a molecule) but also permits the ordering of different molecules in a sequence dependent on the magnitude of the associated label. For the listed hexane isomers the following sequence is obtained: 2,2-dimethylbutane, 2,3-dimethylbutane, 2-methylpentane, 3-methylpentane, *n*-hexane. The ordering follows the increase in magnitude of the unique label. Since the procedure of assigning the numbering to atoms and sequencing different structures has been based on purely mathematical considerations and constructions, it is therefore somewhat surprising to find that it also mirrors some structural features of molecular skeletons. *The above sequence does indeed parallel what may be intuitively considered as an ordering of molecules according to their branching properties.* At least the sequential order is consistent in grouping structures with similar branching qualities and gives the correct relative order for cases where there is general agreement on the degree of branching. Clearly, for example, 2,2-dimethylbutane and 2,3-dimethylbutane, on one hand, and *n*-hexane should be extremes of the sequence. We *assume* then that the sequential order based on the adjacency matrices corresponding to the smallest binary labels defines also the ordering according the molecular branching. In this way we resolve possible ambiguities in debating the degree of branching in molecules having the same number of branching vertices, or a different distribution of valencies among the branching vertices.

Table II. List of the Edge Indices of all the Ten Kinds of CC Bonds in Alkanes

Bond	Terminal valency	Contribution	(numerical)
C–C single	1–1	1	1.0000
	1–2	$1/2^{1/2}$	0.7071
	1–3	$1/3^{1/2}$	0.5774
	1–4	$1/2$	0.5000
	2–2	$1/2$	0.5000
	2–3	$1/6^{1/2}$	0.4083
	2–4	$1/(2)(2^{1/2})$	0.3536
	3–3	$1/3$	0.3333
	3–4	$1/(2)(3^{1/2})$	0.2887
	4–4	$1/4$	0.2500

An assumed concept of branching will be of little value unless it can be useful for discussing molecular properties known to depend critically on the shape of a molecule. Such are the following thermodynamic quantities: boiling points, dependence of vapor pressures on temperature, free energies, heats of solution, densities and molecular volumes, refractive indices, etc. In the next section a quantitative characterization of the branching based on a mathematical index will be compared with other theoretical indices. Finally, in the last section, the usefulness of this branching index in discussing molecular properties will be demonstrated.

Branching Index for Acyclic Structures

There are different possible approaches for defining an index associated with the concept of molecular branching. We are interested in possible approaches based on the skeleton of a molecule and even here one can approach the problem from different initial assumptions. One can consider molecular fragments and build schemes based on the decompositions implied in the earlier work of Wiener and Platt.[5] Such an approach is particularly useful when considering additivity schemes.[9] Alternatively, one can base his considerations on the characteristic polynomial and other invariants of the adjacency matrix and derive a quantity to characterize the molecular topology. Such an approach developed by Hosoya[10] provides an easily calculable quantity, but does not uniquely correspond to the individual structures and only roughly represents the topological nature of the graph, i.e., branching and cyclization.

Rather than exploring such avenues we will attempt to characterize graph branching directly from the graph. In graphs one deals with connectivities adequately characterized by adjacency relationships. To go beyond this stage requires consideration of the next nearest neighbors. A simple scheme for keeping a record of the selected neighbor relations entails differentiating individual edges according to the degrees (valencies) of the associated vertices. All edges of a graph can be classified into various (m,n)-edge types, where m and n are the degrees of the vertices in question. For graphs of hydrocarbons there are ten different edge types (Table II). For many purposes such a discrimination will suffice, but in the event it is necessary to expand the record and consider a further differentiation, this can be accomplished by using the concept of extended connectivity, which acknowledges the presence of more distant neighbors.[11] It may be mentioned here that a notion of bond type determined by the number of the adjacent CC bonds, i.e., implying the valency of a graph, was considered for conjugated molecules almost 30 years ago.[12] Particular bonds are more fully described by means of notation in terms of adjacent CC bonds but the scheme was not developed beyond the classification aspect.[13]

The next step in our approach consists of assigning special values to individual edge types so that their contributions when summed over all the bonds in a molecule and representing the molecular parameter agree, as much as possible, with the relative ordering based on the unique numbering of vertices. This leads to a set of $(n - 1)$ inequalities for each group of n isomers. The corresponding relations for butane, pentane, and hexane isomers (2, 3, and 5 isomers, respectively) give the following seven inequalities:

$$3(1,3) < 2(1,2) + (2,2)$$

$$4(1,4) < 2(1,3) + (2,3) + (1,2) < 2(1,2) + 2(2,2)$$

$$\begin{aligned}3(1,4) + (1,2) + (2,4) < 4(1,3) + (3,3) <& \\ 2(1,3) + (2,3) + (2,2) + (1,2) <& \\ 2(1,2) + (1,3) + 2(2,3) <& \ 2(1,2) + 3(2,2)\end{aligned}$$

The above seven inequalities involve seven different bond types. Isolated bond type (1,1) appears only in ethane and is of no interest for other systems. The remaining bond types (3,4) and (4,4) appear in some isomers of heptane, octane, and higher alkanes. When considering higher members of the homologous series redundancies and possible inconsistencies may arise. For instance, for the nine edge types appearing in molecular graphs of 18 isomers of octane there are 17 constraining inequalities. Of these inequalities two pairs are identical, in that they both involve the same combinations of bond types. They correspond for example to 3-methylheptane and 4-methylheptane. Clearly, it is immaterial which of the two nonequivalent (2,2)-edge types among the total number for n-heptane is converted into a (2,3) type of octane isomers. In still larger systems the number of such situations will increase even more. Therefore it seems useful to satisfy the inequalities derived from smaller members of such a family of homologous systems.

Since there are a number of ways of satisfying a set of inequalities, it would be desirable to choose one which can be interpreted in terms of graph concepts. A closer examination of the inequalities shows, however, that by assigning to (m,n) factors the reciprocal values $1/(m \cdot n)$ all the inequalities, without exception, are satisfied. The same is true if the square root of the reciprocal products of valencies is assumed. Since m,n are associated with the valencies of the respective vertices such assignments are easily interpreted. In order to avoid overlapping of indices belonging to a different set of isomers the latter alternative is chosen. Hence each vertex is assigned a value of $1/(v^{1/2})$, where v is the valency of the considered vertex. The edge index is given the value of the product of the associated vertex values: $1/(v_i \cdot v_j)^{1/2}$.

It remains to be seen how well the proposed edge indices can differentiate larger systems and characterize various molecular skeletons. In Table III a list of so derived branching indices for all alkanes up to eight carbon atoms is given; extension to higher members is straightforward. Inspection of Table III shows that the adopted scheme gives correct ordering for heptanes and octanes, although in the latter case overlapping exists. For molecules having the same number of carbon atoms the index increases as one goes from more branched to less branched structures, a trend which parallels the order defined by adjacency matrices. The largest value among isomers belongs to unbranched linear alkanes. For n-alkanes the index (the sum of all edge contributions) increases with the number of carbons, the increment being ½ per carbon atom. In his empirical approach to regularities in retention volumes observed in chromatographic works Kováts proposed an index which to a large extent reflects the branching of the molecular skeleton. Kováts used the series of normal alkanes as standards for retention values by which his empirical scale is defined. By assigning a value of $100 \cdot n$ to even n-alkanes, the index for a substance is then obtained by measuring the position of the retention peak (on a logarithmic scale) relative to the nearest defined positions of normal alkanes. The above topological branching index thus shows the same linear increase in the index which the empirical index of Kováts possesses by definition. Even the difference in the scale is conveniently simple (a factor of 200), which is gratifying, as it makes an adjust-

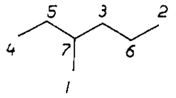
Table III. The Branching Indices for All C_2-C_8 Alkanes Expressed in Terms of Contributions from Ten Kinds of CC Bonds

Molecule	Partial contributions	Branching index
Ethane	(1,1)	1.0000
Propane	2(1,2)	1.4142
Isobutane	3(1,3)	1.7321
n-Butane	2(1,2) + (2,2)	1.9142
2,2-Dimethylpropane	4(1,4)	2.0000
2-Methylbutane	(1,2) + 2(1,3) + (2,3)	2.2701
n-Pentane	2(1,2) + 2(2,2)	2.4142
2,2-Dimethylbutane	(1,2) + 3(1,4) + (2,4)	2.5607
2,3-Dimethylbutane	4(1,3) + (3,3)	2.6425
2-Methylpentane	(1,2) + 2(1,3) + (2,2) + (2,3)	2.7702
3-Methylpentane	2(1,2) + (1,3) + 2(2,3)	2.8082
n-Hexane	2(1,2) + 3(2,2)	2.9142
2,2,3-Trimethylbutane	2(1,3) + 3(1,4) + (3,4)	2.9432
2,2-Dimethylpentane	(1,2) + 3(1,4) + (2,2) + (2,4)	3.0607
3,3-Dimethylpentane	2(1,2) + 2(1,4) + 2(2,4)	3.1213
2,4-Dimethylpentane	4(1,3) + 2(2,3)	3.1259
2,3-Dimethylpentane	(1,2) + 3(1,3) + (2,3) + (3,3)	3.1807
2-Methylhexane	(1,2) + 2(1,3) + 2(2,2) + (2,3)	3.2700
3-Methylhexane	2(1,2) + (1,3, + (2,2) + 2(2,3)	3.3081
3-Ethylpentane	3(1,2) + 3(2,3)	3.3461
n-Heptane	2(1,2) + 4(2,2)	3.4142
2,2,3,3-Tetramethylbutane	6(1,4) + (4,4)	3.2500
2,2,4-Trimethylpentane	2(1,3) + 3(1,4) + (2,3) + (2,4)	3.4165
2,2,3-Trimethylpentane	(1,2) + (1,3) + 3(1,4) + (2,3) + (3,4)	3.4814
2,3,3-Trimethylpentane	(1,2) + 2(1,3) + 2(1,4) + (2,4) + (3,4)	3.5040
2,3,4-Trimethylpentane	5(1,3) + 2(3,3)	3.5534
2,2-Dimethylhexane	(1,2) + 3(1,4) + 2(2,2) + (2,4)	3.5607
3,3-Dimethylhexane	2(1,2) + 2(1,4) + (2,2) + 2(2,4)	3.6213
2,5-Dimethylhexane	4(1,3) + (2,2) + 2(2,3)	3.6259
2,4-Dimethylhexane	(1,2) + 3(1,3) + 3(2,3)	3.6639
2,3-Dimethylhexane	(1,2) + 3(1,3) + (2,2) + (2,3) + (3,3)	3.6807
3,3-Methylethylpentane	3(1,2) + (1,4) + 3(2,4)	3.6819
2-Methyl-3-ethylpentane	2(1,2) + 2(1,3) + 2(2,3) + (3,3)	3.7188
3,4-Dimethylhexane	2(1,2) + 2(1,3) + 2(2,3) + (3,3)	3.7188
2-Methylheptane	(1,2) + 2(1,3) + 3(2,2) + (2,3)	3.7701
3-Methylheptane	2(1,2) + (1,3) + 2(2,2) + 2(2,3)	3.8081
4-Methylheptane	2(1,2) + (1,3) + 2(2,2) + 2(2,3)	3.8081
3-Ethylhexane	3(1,2) + (2,2) + 3(2,3)	3.8510
n-Octane	2(1,2) + 5(2,2)	3.9142

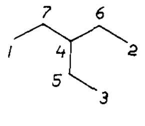
ment of the two scales simple. However, since retention times are not reproducible, an index based on such a quantity is not expected to be very reliable, and hence the topological index may provide a better standard.

The resulting index can indeed be associated with the degree of branching as suggested by extreme situations. On one hand, when a single vertex has the maximal number of adjacent vertices the branching index is at its lowest possible value for acyclic graphs having n vertices: $I_{\min} = 1/(n - 1)^{1/2}$. And on the other hand, the maximum value of the index corresponds to a linear arrangement of n vertices: $I_{\max} = (n - 3)/2 + 2^{1/2}$ $(n > 2)$.

It may be tempting to summarize the results in Table III in terms of structural requirements which determine the relative magnitudes of the branching indices, and the relative ordering based on adjacency matrices. The actual ordering is not always apparent. For instance graphs with vertices of higher valencies will generally appear earlier in the sequence. But there are systems with lesser valency which precede some systems with vertices of higher valencies (e.g., 2,3,4-trimethylpentane precedes 2,3-dimethylpentane and 3,3-dimethylpentane). In simpler molecules we can rely on intuition with some confidence in ordering isomers in a sequence of increasing (or decreasing) branching. An intuitive approach may fail to suggest the relative degree of branching, for instance, of 3-methylhexane (I) and 3-ethylpentane (II). The associated adjacency matrices provide guidance if the unique numbering of vertices is given. In fact one need not construct the adjacency matrices since the relative order *can be deduced from the labels of the vertices of the two graphs*. For the two molecules the numbering of vertices is shown below:



I



II

It can be easily found in a few trials if one starts by assigning the smallest numbers to the terminal atoms and the largest to their adjacent atoms. To establish the relative magnitudes of the associated binary labels start with the vertices labeled 1 in both graphs and compare the labels of vertices adjacent to them. If the labels of the adjacent vertices are not equal, or one has fewer adjacent vertices than the other, this should determine the magnitude of the row in the associated adjacency matrices and their relative order. If they are equal one proceeds to the next smallest labeled vertex. In the above case, both graphs show that vertices 1 and 2 have identical adjacent neighbors (7 and 6, respectively). Since the same situation arises we continue until we arrive at the vertices which have *different* adjacent connections in the two graphs. This is the case with vertex 3, which in I has neighbors 6 and 7, while in II the nearest neighbor is 5. Therefore the third row in the two adjacency matrices, the first to be different, determines the relative ordering. Isomer I is associated with the smaller binary number and by definition is considered "more branched". The branching indices of I and II are 3.308 and 3.346, respectively (Table III), in agreement with the prediction based on the unique numbering of vertices.

Comparison with Other Theoretical Indices

Besides the empirical index of Kováts, which has been found to parallel molecular branching, few theoretical indices of a similar nature have been devised. Wiener[5] related the difference in normal boiling points between a branched-chain paraffin and its straight-chain isomer to a change in two structural parameters which approximately account for a difference in a skeletal form (i.e., branching) and nonbonded contacts. The so called Wiener numbers are the sums of the number of bonds between all pairs of carbons in the molecular skeletons. Its conbinatorial origin is evident although its interpretation is less apparent (see Platt[5]). The other parameter represents a number of pairs of carbon atoms separated by three bonds.

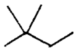
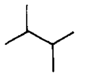
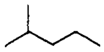
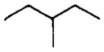
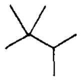
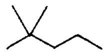
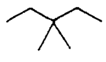
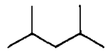
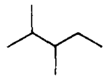
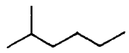
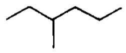
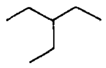
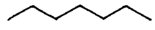
An explicit graph theoretical approach for characterizing molecular branching has been developed by Gordon and Scantlebury.[14] They define an index $B_2 = N_2/(n - 1)$, where n is the number of carbon atoms and N_2 represents the number of distinctive ways in which the considered skeleton contains *two* links (triatomic group). The discriminatory power of the procedure is limited, but the major types of skeletal ramifications are well differentiated. For instance (m,m)-disubstituted alkanes will have a different branching index from (m,n)-disubstituted alkanes, although the substituted position and the size of the substituent do not influence the magnitude of the index. Despite these limitations the B_2 indices are important theoretical quantities that are rigorously defined and when combined with other such indices involving different skeletal fragments provide a basis for a systematic study of partitioning of a system.

Hosoya[10] proposed his index $Z(p,q)$ characterizing the topological nature of structural isomers of saturated hydrocarbons, as already mentioned. It appears as an ad hoc construction, being given by the sum of the coefficients of the characteristic polynomials. However, the coefficients are related to various possible decompositions of the molecular graph, although the relationship is somewhat involved.[15] It reflects fairly satisfactorily upon the size and the mode of branching and even ring closure.

Finally, a sum of the valencies squared $\Sigma(v^2)$ of a hydrogen suppressed graph was suggested as a branching index. Σ is identical with the sum of "extended" valencies, i.e., valencies based on the so called extended connectivity[11] introduced for the needs of chemical documentation. Its capability for a discrimination of isomers is equivalent to the scheme of Gordon and Scantlebury. This, however is not sufficient in some applications; for example, a deficiency of the particular characterization of the branching was noticed in a derivation for the dependence of resonance energies on structural parameters in benzenoid systems.[16] An important disadvantage of the various schemes of Table IV is not so much that they operate with integers which somewhat limits their flexibility, nor that they provide only a limited resolution of isomers, but that the resulting parameters for molecular systems of different size overlap. This makes such schemes unsuitable for correlations which use a single scale, the requirement which accumulated relevant data seem to demand.

The branching indices based on the approach of Wiener, Gordon, and Scantlebury, Hosoya, and others represent an effort to arrive at the definition of a single parameter to describe branching effects. The considered schemes are necessarily given as integers; nevertheless the parallelism between them and the branching index, based on the differentiation of edge types introduced in this work, is satisfactory. The agreement is of considerable significance in view of the different origins of the indices. It points to an intimate relationship between bond types in a molecule and the secular equation. A particular form of such a relationship concerns the notion of the extended connectivity and the relative magnitude of the coefficients characterizing the first eigenvector of the topological matrix.[17] Ultimately this dependency has some bearing on the parallelism between MO and VB methods, the former being associated with the characteristic polynomial of a graph, while the latter description

Table IV. A Comparison between Various Theoretical Indices of Branching[a]

Molecule	B_2[14]	$Z(p,q)$[10]	Σ^2	This work
	7	9	24	2.561
	6	10	22	2.643
	5	11	20	2.770
	5	12	20	2.808
	4	13	18	2.914
	9	13	30	2.943
	8	14	28	3.061
	8	16	28	3.121
	7	15	26	3.126
	7	17	26	3.181
	6	18	24	3.270
	6	19	24	3.308
	6	20	24	3.346
	5	21	22	3.414

[a] Notice the identical differentiation ability of the schemes of Gordon and Scantlebury[14] and Gutman et al.[2] and minor inconsistancies in the Hosoya scheme. The molecules have been ordered according to the magnitude of the binary label, which is fully reflected only in the indices in the last column.

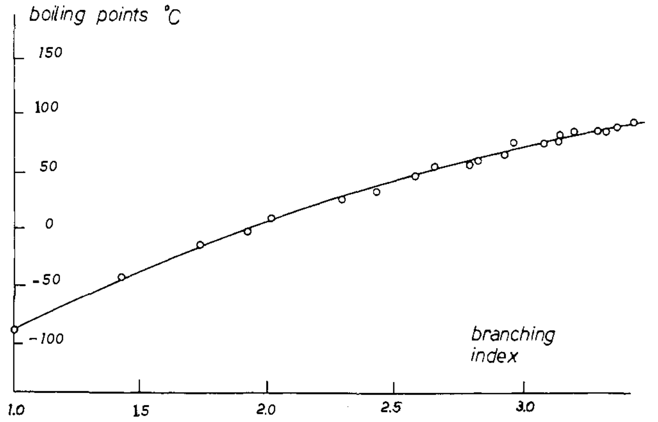


Figure 1. Boiling points of alkane isomers having from two to seven carbon atoms plotted against the topological branching index. (Experimental data are taken from ref 24.)

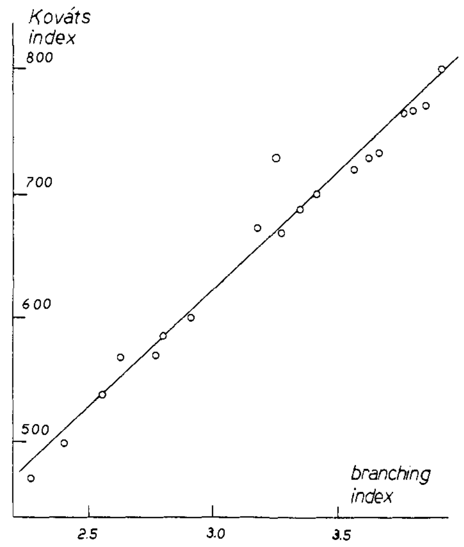


Figure 2. A correlation between the empirical branching index of Kováts and the topological branching index of this work.

is concerned with bond types and their processing.

Experimental Data and the Branching Index

Numerous attempts have been made so far to derive empirical correlations between the normal boiling points of organic substances and the structure of their molecules. The emphasis in graph theoretical methods is not necessarily on achieving an optimal correlation with the considered number of parameters, but rather one is concerned with the topological significance of the parameters involved. Such correlations should primarily be viewed as complementary to those of an empirical nature, not competitive. Clearly a *single* parameter, such as the branching index, will not provide a sufficiently flexible basis for discussing minor deviations from a correlation, even if they are systematic and attributable to some structural causes (e.g., nonbonded interactions). A scheme based on several parameters may be prefered but the former scheme remains of interest and will be instructive in indicating the degree to which more elaborated procedures owe their success to the correct account of the molecular topology.

In Figure 1 we show the correlation between the boiling points and the branching index for C_2–C_7 alkanes. The gross features of the relative magnitudes of the boiling points indeed can be related to the size and the branching mode of the molecular skeletons. In Figure 2 we plotted the empirical Kováts index, derived from chromatographic retention data,[18] against the topological branching index introduced here for several alkane isomers having five to eight carbon atoms. The scatter of the points is partly due to the limited reproducibility of the data. However, the discrepancy may also signify the importance of neglected contributions, such as nonbonded contacts, as perhaps is suggested by a larger departure from the correlation by highly branched systems.

Several other thermodynamic properties are expected to correlate with the topological nature of a compound. These include the dependence of vapor pressures on the temperature, free energies, heats of solution, densities, refractive indices, and related properties. In Figure 3 we plotted the experimental values of enthalpies of formation of alkanes in the gas phase at 298.15 K vs. the branching index.[19] An interesting diagram resulted. Normal alkanes are well represented by a linear correlation, indicating the simple dependence of $\Delta H_f°$ on the number of carbon atoms in the chain. For each C_n group of isomers there is a separate correlation line of an opposite slope. All the separate lines however have the same gradient which permits the results to be summarized in a simple bilinear expression:

$$\Delta H_f° = c_1(B) + c_2(n) + c_3$$

Approximate values for the above constants are: $c_1 = 12$ kcal/mol; $c_2 = -11$ kcal/mol; and $c_3 = -9$ kcal/mol. The

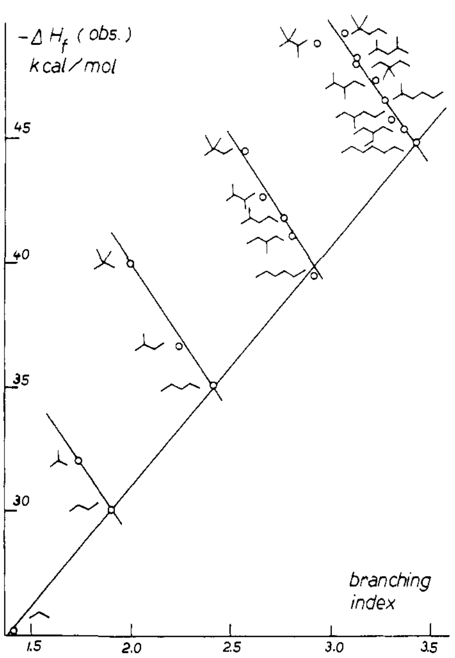


Figure 3. The experimental values of enthalpies of formation of alkane isomers having from two to seven carbon atoms plotted against the branching index. Each group of isomers has a separate correlation line approximately equally displaced and of the same slope.

significance of the above relation is that it shows the gross features of the enthalpies of branched hydrocarbons can be accounted for by a simple additivity of bond contributions. Namely, the definition of the branched index implies a superposition of bond contributions. Moreover, such a superposition is based on the simple model of differentiation of the types of bonds, i.e., the number of nearest neighbors. So in the first approximation a *single* structural parameter, the valency of a vertex of the molecular graph, contains the essential information for rationalizing the relative magnitudes involved. For a finer correlation factors beyond the molecular connectivity will be of importance. The above correlation leads to significant deviations for several branched alkanes such as 2,2,3-trimethylbutane, 2,2,4-trimethylpentane, 2,2,3,3-tetramethylbutane, etc. A statistical analysis of the existing data on the enthalpies of formation of the gaseous alkanes gives the correlation coefficient of 0.9998.[20] So apparently the major contribution to the correlation arises from a single structural parameter, the branching index, although the role of other contributions should not be dismissed.

The dependence of vapor pressures on temperature for hydrocarbons provides another correlation with the branching index. For the present discussion we adopted the Antoine equation:

$$\log P = A - B/(C + t)$$

where P is the vapor pressure, t is the temperature in °C, and A, B, and C are adjustable parameters characteristic of each compound. In Figure 4 the parameters A, B, and C for all heptane isomers are plotted on the vertical axis vs. their respective branching index values. Although the scatter of the points is occasionally larger than desirable, the overall correlation appears satisfactory. A closer look at Figure 4 reveals that in all cases of large departures from the correlation a slight revision of the corresponding A, B, and C would result in a better regression (confidence limit). A simultaneous increase of both A and B, or B and C, will have a minimal effect on the dependence of P on t, particularly when one constrains the changes in A, B, C to satisfy the following equation:

$$\Delta A - \Delta B/(C + t) + \Delta CB/(C + t)^2 = 0$$

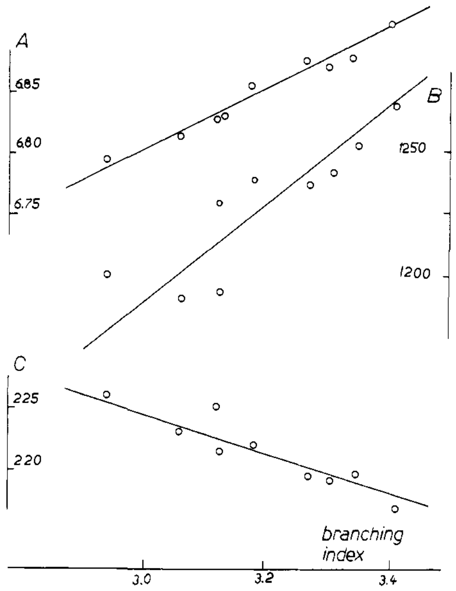


Figure 4. Correlations between empirical parameters A, B, and C based on the Antoine equation for nine isomers of C_7H_{16} and the topological branching index. (Data taken from ref 24.)

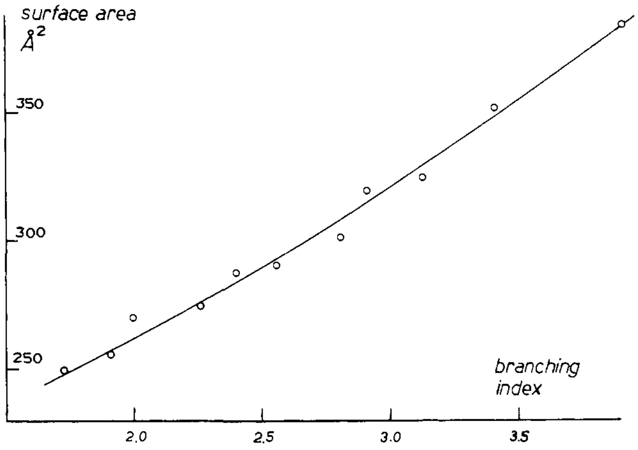


Figure 5. A correlation between theoretically calculated total surface area of selected saturated acyclic hydrocarbons based on the model of Hermann and topological branching indices.

The equation is derived from the Antoine equation by neglecting higher order contributions in ΔA, ΔB, and ΔC. An adjustment of the Antoine parameters will lead to a somewhat less accurate fit for an individual isomer, but the derived parameters for the set of isomers will be more meaningful. Detailed analysis of the parameters is outside the scope of the present work. Since the constants A, B, and C are of interest for structural correlations and are used in evaluations of other molecular properties, such as heats of vaporization and molar enthalpies of vaporization, such considerations would be of importance. Even if one is not interested in structural considerations the possibility that individual Antoine coefficients can be reduced to fewer parameters and related to the single quantity, the branching index, is in itself of considerable consequence. For instance, in the case of nine heptane isomers 27 apparently independent Antoine coefficients can be related through three linear expressions, i.e., six constants, an impressive reduction of the empirical data. Some loss in precision is not an excessive price compared with the rewarding insight into the structural interdependence of the seemingly unrelated coefficients.

Finally, the branching index will be of some interest to other theoretical approaches and may provide a useful reference. For example solubilities are related to molecular surface area.[21] Hermann developed a theoretical model in which a solute molecule is considered as a collection of spheres located at the nuclear centers, to which an envelope determined by the radius of the solvent is added.[22] The branching, cyclization, and positional isomerism are thus indirectly accounted for without the need to introduce additional terms. It is therefore of interest to see how the calculated surface area in this theoretical model depends explicitly on the branching index. The correlation based on a dozen available systems (hydrocarbons in water) is shown in Figure 5. The difference between the surface area based on the model of Hermann and derived from the correlation is less than 10 Å^2 which is less than the scatter in the correlation between the theoretical and experimental surface areas.[23] Thus the branching index can provide an alternative basis for the correlation of the solubility of hydrocarbons in water.

In concluding, we would like to emphasize the conceptual simplicity of the branching index here introduced and its importance as an expression based directly on molecular graphs. It appears that the application of graph theory which is concerned with a systematic study of connectivities may be a rewarding approach to illuminating some problems in chemistry, not so much by providing an accurate and precise scheme for predicting a particular property, but by revealing novel relationships among unsuspected quantities and by discerning the topological nature in others.

Acknowledgment. I wish to thank Dr. M. Falk (Halifax, Nova Scotia) for valuable comments and suggestions which have led to an improved presentation of the material. The correspondence with Dr. D. H. Rouvray (Johannesburg, South Africa) clarified various parts of the manuscript and is much appreciated.

References and Notes

(1) Branch and branching as technical terms are not uniquely defined, cf. J. W. Essam and M. E. Fisher, *Rev. Mod. Phys.*, **42**, 272 (1970); R. G. Busacker and T. L. Saaty, "Finite Graphs and Networks", McGraw-Hill, New York, N.Y., 1965, p 267; F. Harary, "Graph Theory", Addison-Wesley, Reading, Mass. 1972, p 35; branching vertex (atom) is however obvious.

(2) I. Gutman, B. Ruscić, N. Trinajstić, and C. F. Wilcox, Jr., *J. Chem. Phys.*, **62**, 3399 (1975).

(3) M. Randić, to be published.

(4) E. Kováts, *Z. Anal. Chem.*, **181**, 351 (1961).

(5) H. Wiener, *J. Am. Chem. Soc.*, **69**, 17 (1947); *J. Phys. Chem.*, **52**, 425 (1948); J. R. Platt, *J. Chem. Phys.*, **15**, 419 (1947); *J. Phys. Chem.*, **56**, 328 (1952).

(6) M. Gordon and J. W. Kennedy, *J. Chem. Soc., Faraday Trans. 2*, **69**, 484 (1973).

(7) M. Randić, *J. Chem. Phys.*, **60**, 3920 (1974).

(8) A. L. Mackay, *J. Chem. Phys.*, **62**, 308 (1975); M. Randić, *ibid.*, **62**, 309 (1975).

(9) E. A. Smolenskii, *Zh. Fiz. Khim.*, **38**, 1288 (1964).

(10) H. Hosoya, *Bull. Chem. Soc. Jpn.*, **44**, 2332 (1971).

(11) H. L. Morgan, *J. Chem. Doc.*, **5**, 107 (1965).

(12) B. Hartmann, *Z. Naturforsch., A*, (Feb 1947).

(13) For an application to bond length calculation, see O. Chavlet and R. Daudel, *J. Phys. Chem.*, **56**, 365 (1952).

(14) M. Gordon and G. R. Scantlebury, *Trans. Faraday Soc.*, **60**, 605 (1964).

(15) H. Sachs, *Publ. Math.* (*Debrecen*), **9**, 270 (1962); **11**, 119 (1963); L. Spialter, *J. Chem. Doc.*, **4**, 269 (1964); F. Harary, *J. Math. Phys.* (*Cambridge, Mass.*), 104 (1959); H. Hosoya, *Theor. Chim. Acta*, **25**, 215 (1972).

(16) C. F. Wilcox, Jr., *Croat. Chem. Acta*, in press.

(17) M. Randić, *J. Chem. Inf. Comput. Sci.*, **15**, 105 (1975).

(18) A recent compilation of Kovats indices for saturated hydrocarbons (used as a source of data in Figure 2) is given by V. Kubelka and M. Popl, *Chem. Listy*, **68**, 378 (1974).

(19) D. W. Scott, *J. Chem. Phys.*, **60**, 3144 (1974), or references therein.

(20) L. H. Hall, L. B. Kier, and W. J. Murray, to be published.

(21) G. Nemethy and H. A. Scheraga, *J. Chem. Phys.*, **36**, 3401 (1962).

(22) R. B. Hermann, *J. Phys. Chem.*, **76**, 2754 (1972); **75**, 363 (1971); Quantum Chemistry Exchange Program No. 225, Indiana University, Bloomington, Ind., provides the computer program for calculation of the molecular surface area.

(23) G. L. Amidon, S. H. Yalkowsy, and S. Leung, *J. Pharm. Sci.*, **63**, 1858 (1974).

(24) R. C. Wilhoit and B. J. Zwolinski, Ed., "Handbook of Vapor Pressures and Heats of Vaporization of Hydrocarbons and Related Compounds", Texas A & M University, College Station, Texas, 1971.

The Importance of Nonbonded Attraction in the Stereochemistry of the SN2′ Reaction

Ronald L. Yates,[1a] Nicolaos D. Epiotis,*[1a] and Fernando Bernardi[1b]

Contribution from the Department of Chemistry, University of Washington, Seattle, Washington 98195, and the Istituto di Chimica Organica, Università di Bologna, 40136 Bologna, Italy. *Received February 18, 1975*

Abstract: The stereochemistry of the SN2′ reaction is controlled by nonbonded attraction and electrostatic factors. An orbital symmetry analysis is used to evaluate the one-electron factors for the SN2′ reaction involving representative neutral and charged nucleophiles. The results of ab initio (STO-4G) and semiempirical SCF-MO calculations are presented to support the conclusions reached by one-electron molecular-orbital theory.

The abnormal bimolecular substitution reaction (SN2′) has been a subject of controversy ever since the first authentic example was reported by Kepner, Winstein, and Young in 1949.[2a] Other examples were added in the following years.[2b] The mechanism first proposed involved simultaneous bond breaking at the α carbon and bond making at the γ carbon atom of an allyl system. Since then, the "concertedness" of the SN2′ reaction has been questioned and alternate mechanisms have been proposed.[3,4]

The stereochemistry of the SN2′ reaction was suggested by Stork and White[5] to be consistent with a nucleophilic attack on the same side from which the leaving group departs. That is, a syn as opposed to an anti attack on the allylic system. Qualitative theoretical interpretations of the stereoselectivity of the SN2′ reaction have appeared in the literature,[6-8] but quantitative quantum mechanical calculations have never been reported in support of the qualitative theoretical arguments. In 1973, we proposed that nonbonded at-