391

MOLAR VOLUMES OF ALKANES AND TOPOLOGICAL INDICES

T.P. RADHAKRISHNAN and W.C. HERNDON

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

Abstract

This paper examines possible explanations for correlations that exist between graph-theoretical topological indices and observed physical properties of molecules. The general line of approach is to demonstrate that a particular graph-theoretical correlation has an analytical correspondence to an observed additive structural description. The basic approach is exemplified by showing that the correlation of partial molar volumes of normal alkanes with χ (chi) topological indices corresponds quantitatively to a previously obtained molecular structure relationship involving group additivity and average gauche conformation numbers. A recurrence relation is tested for the gauche numbers, and the results confirm a previously proposed linear dependence on molecular size.

1.   Introduction

The connectivities or topologies of physical structures can be represented by graphs [1,2]. The pictorial representation of a graph G can be described in terms of a set V of vertices (points, nodes) and a set E of edges (lines, arcs) which join pairs of vertices. A chemical molecular graph is a graph in which the vertices identify nuclear positions of atoms, and edges represent chemical bonds [3,4]. For many graph-theoretical treatments of organic chemical problems, the molecular graphs (usually undirected) are drawn after deleting all of the hydrogen atom vertices and the edges connecting those vertices to the remainder of the graph (H-suppressed graphs). The vertices and/or the edges of the molecular graph may be labeled (weighted) when heteroatoms or different types of bonds are present in the structure.

Molecular graphs are possible starting points for quantitative descriptions of molecular structures. One approach is to use indices derived from the graph structure, commonly referred to as topological indices [5–7], that essentially express the overall connectedness of a molecular graph as a numerical term. These quantities can then be tested and used as numerical descriptors of physical and chemical

properties [6–10]. Surprisingly good correlations have been reported [8–10] for a wide range of physical observables like molar volumes, refractive indices, enthalpies of formation, etc., and for a great variety of organic compounds. In general, reasons for these congruencies are only vaguely understood in terms of the common dependence of indices and properties on molecular structure.

This paper follows the methodology and suggestions by Edward [11–14] to search for molecular explanations of correlations between topological indices and observed properties of molecules. The salient aspect of the approach arises from the inference that previous molecular structure group additivity methods [15,16] provide the most useful simple structural models for structure–property relationships. One attempts, therefore, to discover how a topological index relates (analytically) to a relevant additivity model for the property under investigation. In general, we believe that this kind of understanding of topological indices will allow: (i) the validation of known results, and (ii) the systematic investigation and improvement of correlations. The specific aim of the present work is to show how a particular correlation of the molar volumes of n-alkanes with topological indices (Randić-type branching indices [17], see later) corresponds to a previous additivity relation. Edward [11,13] has considered this same problem from a generalized and more qualitative viewpoint that includes branched alkanes in the analysis.

2.   Definitions and background

The graph G can be described by the set of vertices V and the set of edges E. The set E is a subset of the product $V \times V$:

$$V \times V = \{(v_i, v_j) \mid v_i, v_j \in V\}. \tag{1}$$

If E is a subset of the set of symmetric pairs in $V \times V$, then G is an undirected graph. If $(v_i, v_j) \in E$, we say that the vertices i and j are connected. An m-path in a graph (molecular graph) is a sequence of m adjacent edges (bonds) with no vertex (atom) appearing more than once.

The adjacency matrix $[A]$ is a useful representation of the graph. The elements of this matrix, also called the connectivity matrix, are defined by

$$a_{ij} = \begin{cases} 1 & \text{when } (v_i, v_j) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Many topological indices can be directly obtained [18] from the adjacency matrix or from the m-path matrix $[(A_m)_{ij}]$, where the matrix elements are

$$(a_m)_{ij} = \begin{array}{l} 1 \quad \text{if and only if } v_j \text{ is the } m\text{th neighbor of } v_i \\ 0 \quad \text{otherwise.} \end{array} \tag{3}$$

Some of the most widely used topological indices in structure-activity analysis are those which are based on Randić's branching index [17]. For the *H*-suppressed molecular graphs of the saturated alkanes, this index χ (chi) is defined as

$$\chi = \sum (d_i d_j)^{-0.5}. \tag{4}$$

where d_i is the number of edges incident at v_i (called the degree of vertex v_i). The sum is over all edges $(v_i, v_j) \in E$. Randić originally proposed this index to rank order alkanes, according to a perceived extent of branching. Larger values of d_i reflect greater substitution on C atoms, and the inverse square root was chosen as one algorithm preserving inequalities between isomeric alkanes. χ can be expressed in the alternate formulation

$$\chi = \sum [(A^2)_{ii} (A^2)_{jj}]^{-0.5} \tag{5}$$

since $(A^2)_{ii}$, the diagonal elements of the matrix $[A]^2$, gives the degrees of the vertices i. Other, more branched alkanes have smaller χ values.

Kier and Hall [8,10] extended the concept of the χ index to incorporate other structural fragmentations of molecular frameworks. Accordingly, the following second- and third-order indices for two- and three-bond fragments are defined:

$$^2\chi = \sum (d_i d_j d_k)^{-0.5} \tag{6}$$

$$^3\chi = \sum (d_i d_j d_k d_l)^{-0.5}. \tag{7}$$

where the summations for $^2\chi$ and $^3\chi$ are over 2-paths and 3-paths, respectively. Randić's original algorithm, which included only one-bond dissections, is $^1\chi$ on this basis. In addition, cluster (c) and path/cluster (pc) indices were defined for other commonly occurring molecular fragments, as illustrated in **1 − 3**.
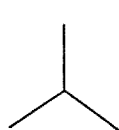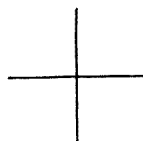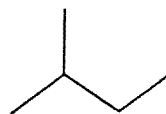
1 $(^3\chi_c)$ 2 $(^4\chi_c)$ 3 $(^4\chi_{pc})$

3. Molar volume of *n*-alkanes

Edward and coworkers [19] obtained partial molar volume experimental data in CCl_4 solution for alkanes (branched as well as normal), and found that the partial molar volumes could be calculated with good accuracy using the group additivity scheme given in eq. (8):

$$V^0 = V_c + \sum_i n_i l_i - 2.5(Z_n). \tag{8}$$

V_c is a constant $(11.61 \text{ ml mol}^{-1})$ called the covolume, the l_i are volume increments for various types of carbon atoms, i.e. $l(CH_3) = 26.85$ and $l(CH_2) = 17.36$, and the n_i are the numbers of such groups in the molecule. Z_n is the average number of gauche conformations for the alkanes given by the equation

$$Z_n = \gamma_g g, \tag{9}$$

where g is the number of gauche interactions in any conformation, and γ_g is the mole fraction of that conformation, obtained using the Pitzer steric partition function [20]. The summation is over all possible conformations. Our discussion is limited to the *n*-alkanes, and in this case the values of Z_n were obtained previously from the empirical equation

$$Z_n = 0.380 + 0.293(n - 4), \tag{10}$$

which is accurately linear in the number n of carbon atoms for the series of normal alkanes, pentane to nonane [19]. Since this equation was extrapolated to deal with molecules with as many as 32 carbon atoms, we have investigated more general algorithms (see appendix) to calculate the Z_n for *n*-alkanes and have verified the results of Edward and coworkers [19].

A graph-theoretical approach to correlate the same experimental data was developed by Hall and Kier [21]. They found an excellent correlation of the molar volumes with the first-order path, second-order path, and fourth-order cluster terms of the χ branching topological indices, described by eq. (11):

$$V^0 = 24.87(^1\chi) + 11.86(^2\chi) - 2.844(^4\chi_{pc}) + 39.79. \tag{11}$$

A relationship involving the liquid alkane densities required two additional topological parameters for equivalent accuracy [22].

4.   Results and discussion

Equation (8) is so far the best additive scheme available to calculate molar volume of alkanes. One important consideration is that it takes account of Z_n, which was a factor neglected in previous formulations, and it gives a rational molecular level picture of the molar volume. External factors such as temperature that might influence the molar volume are explicitly built into eq. (8) through the Z_n term. In eq. (11), on the other hand, the variables explicitly include only topological factors, which are present as sums of inverse square root multiplicative terms. Other factors can be only indirectly included in the coefficients of the equation. One infers that the average conformation factor is expressed in the Hall and Kier equation, but the actual condition for inclusion is not obvious. Below we show, for the case of the normal alkanes, that both eq. (8) and eq. (11) reduce to the same simple linear equation in n, thus clarifying the correspondence of the two approaches.

For n-alkanes, the only types of carbon atoms present are methyl (l_1) and methylene (l_2). Then, incorporating the linear dependence of Z_n on n given in eq. (10), and using the values for the various parameters obtained by Edward and co-workers [19], the Edward eq. (8) reduces to

$$V^0 = V_c + 2l_1 + (n-2)l_2 - 2.5(0.380 + 0.293(n-4))$$

$$= c_1(n) + c_2$$

$$= 16.63(n) + 32.57. \tag{12}$$

Now consider the Hall/Kier eq. (11) for n-alkanes with n carbon atoms. A linear chain graph with n vertices contains $(n-1)$ 1-paths and $(n-2)$ 2-paths. The $^1\chi$ and the $^2\chi$ values for an n vertex chain graph are

$$^1\chi = 2(1 \cdot 2)^{-0.5} + (n-3)(2 \cdot 2)^{-0.5}$$

$$= 0.5(n) - 0.086 \tag{13}$$

$$^2\chi = 2(1 \cdot 2 \cdot 2)^{-0.5} + (n-4)(2 \cdot 2 \cdot 2)^{-0.5}$$

$$= 0.35(n) - 0.414 \tag{14}$$

$^4\chi_{pc}$ is zero for a chain graph, since there are no path clusters. Equation (11) therefore reduces to

$$V^0 = 24.87(0.5n - 0.086) + 11.86(0.35n - 0.414) + 39.79$$

$$= c_3(n) + c_4$$

$$= 16.59(n) + 32.74. \tag{15}$$

We see that eq. (12) and eq. (15) are identical in form, with almost identical parameters. The small discrepancies in the parameters may result from the fact that Edward and coworkers did not allow complete flexibility in the determination of the volume increment independent variables during regression analysis. In fact, a regression treatment of eq. (8) and eq. (11) restricted to the n-alkanes gives slightly different parameters than those reported earlier, and leads to identical values of the final co-efficients in eq. (12) and eq. (15). Either result establishes an explicit correspondence between the molecular structure additivity rules and the topological index approach for this observed physical property, namely, the molar volume of n-alkanes. It should be mentioned that Edward's previous work [11,13] anticipated this type of result. We believe that the expositions of these equivalences are useful, since they substantiate the rather abstract approach of the connectivity index by the more physically tangible additivity model.

There is scope for improvement in the procedure we have suggested above. The linear dependence of Z_n on n has not been analytically proved. However, since there is a recurrence relation for the conformation distribution function $d(m, g)$ (see appendix), an explicit summation might lead to the desired proof. The extension of the analysis to the branched chain alkanes is also under investigation. The combinatorial problem is much more difficult, and the rational demonstration of a parallel between the toplogical approach and the additivity model is evidently correspondingly complex. Establishing such equivalences will lend useful support for the topological method, which utilizes fewer parameters compared to the traditional models (cf. eqs. (8) and (11) without sacrificing accuracy. Finally, we note that an additivity relation similar to eq. (8) and a χ index equation similar to eq. (11) exist for the heat of vaporization of alkanes [8,10–14]. The similitudes of these expressions can be worked out exactly, as in the case of molar volume that we have discussed. There are also other thermodynamic properties of molecules that may prove to be amenable to such studies, since it has been demonstrated that several of these properties depend upon path numbers and other types of similar simple topological indices to a first approximation [11,23,24].

Acknowledgement

Appendix

Edward and coworkers [19] empirically demonstrated that the average gauche conformation number Z_n for n-alkanes (pentane to nonane) is linearly dependent on n, the number of carbon atoms, and they calculated Z_n for higher alkanes from the linear relationship because the number of configurations was deemed too large to handle. Recurrence relations developed by Wurflinger [25] and Wille [26] allow a calculation of the Z_n term for any size linear alkane. The treatment by Wurflinger is recast below.

Consider the n-alkane chain of n carbon atoms. The H-suppressed graph will be a chain graph on n vertices with $(n - 1)$ edges.
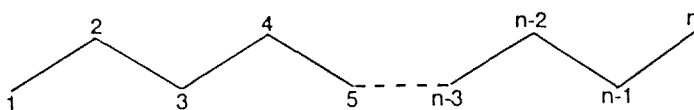
Graph of the n-alkane with n C atoms.

Different conformations of the alkane are obtained by distributing anti, gauche ($+$) and gauche ($-$) conformations on $(n - 3)$ of the bonds. The terminal C-C bonds have equal energy conformations for all rotational positions of the methyl groups. We will assign a constant value of energy ($a = 700$ cal/mol [20]) to the gauche forms with respect to the lowest energy anti conformation. Those configurations where consecutive bonds occur in gauche ($+$) and gauche ($-$) conformations will be neglected, since they invoke very high energies due to second-neighbor interactions and lead to negligible contributions to the partition function.

Now the problem of distributing the three types of conformations in $(n - 3)$ consecutive bonds with the restriction mentioned is the same as the combinatorial problem of putting yellow (anti), red ($+$), and blue ($-$) objects in $(n - 3)$ consecutive boxes (one in each box) in such a way that red and blue objects never reside in adjacent boxes. One can easily verify [25] that the number of distributions of g objects (red and/or blue) in m boxes ($m = n - 3$, $g \leqslant m$) with the above restriction is given by eq. (16):

$$d(m, g) = d(m - 1, g) + d(m - 1, g - 1) + d(m - 2, g - 1). \tag{16}$$

Using this recurrence relation and values $d(m, 0) = 1, d(m, m) = 2$, and $d(2, 1) = 4$, any $d(m, g)$ can be calculated. A computer calculation is, of course, very helpful for large values of m (see ref. [20(b)], p. 45, for a table of values for $d(m, g)$). The energy of conformations having g gauche bonds is given by (ga), and the mole fraction [20] of such conformations at temperature T will, therefore, be given by

$$\gamma_g = d(m, g)\{\exp(-ga/kT)\} \bigg/ \sum_{g=0}^{m} d(m, g)\{\exp(-ga/kT)\} \ . \tag{17}$$

Z_n, the average number of gauche forms in an n carbon atom linear alkane at temperature T, is obtained by summing over values of g from 0 to $n - 3$ ($= m$):

$$Z_n = \sum_{g=0}^{m} gd(m, g)\{\exp(-ga/kT)\} \bigg/ \sum_{g=0}^{m} d(m, g)\{\exp(-ga/kT)\} \ . \tag{18}$$

Values of Z_n obtained from eq. (18) are compared with the Z_n calculated using the Edward eq. (10) in table 1. The new values compare very well with the values

Table 1

Average number of gauche conformations for n-alkanes
(T = 298 K)

n	Z_n: eq. (18)	Z_n: eq. (10)
10	2.154	2.138
20	5.129	5.068
30	8.105	7.998
40	11.081	10.928
50	14.056	14.056
60	17.032	16.788
70	20.007	19.718
80	22.983	22.648
90	25.959	25.578
100	28.934	28.508

given by the linear eq. (10), and this result confirms the linear dependence of Z_n on n. A linear regression of the eq. (18) Z_n data versus n gives

$$Z_n = 0.368 + 0.298(n - 4). \tag{19}$$

This equation is strictly linear, correlation coefficient equal to unity, and exactly reproduces the listed eq. (18) values. The newly obtained regression parameters are in reasonable agreement with those used in the previous work by Edward and co-workers [19], as can be seen by comparison with eq. (10).

References

[1]   F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).
[2]   *Applications of Graph Theory,* ed. R.J. Wilson and L.W. Beineke (Academic Press, London, 1979).
[3]   *Chemical Applications of Graph Theory,* ed. A.T. Balaban (Academic Press, London, 1976).
[4]   N. Trinajstić, *Chemical Graph Theory,* Vols. I and II (CRC, Boca Raton, Fla, 1983).
[5]   H. Hosoya, Bull. Chem. Soc. Japan 44(1971)2332.
[6]   D.H. Rouvray and A.T. Balaban, in ref. [2], ch. 7;
      D.H. Rouvray, Sci. Amer. 254(1986)40;
      A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Topics Curr. Chem. 114(1983)21.
[7]   A. Sablijić and N. Trinajstić, Acta Pharm. Jugosl. 31(1981)189.
[8]   L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).
[9]   D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures* (Wiley, New York, 1983).
[10] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis* (Wiley, New York, 1986).
[11] J.T. Edward, Can. J. Chem. 58(1980)1897.
[12] J.T. Edward, Can. J. Chem. 59(1981)3192.
[13] J.T. Edward, Can. J. Chem. 60(1982)480.
[14] J.T. Edward, Can. J. Chem. 60(1982)2573.
[15] S.W. Benson, *Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters* (Wiley, New York, 1976).
[16] W.J. Lyman, *Handbook of Chemical Property Estimation Methods* (McGraw-Hill, New York, 1982).
[17] M. Randić, J. Amer. Chem. Soc. 97(1975)6609.
[18] M. Barysz, D. Plavsić and N. Trinajstić, MATCH 19(1986)89.
[19] J.T. Edward, P.G. Farrell and F. Shahidi, J. Phys. Chem. 82(1978)2310.
[20] K.S. Pitzer, (a) J. Chem. Phys. 8(1940)711; (b) Chem. Rev. 27(1940)39.
[21] L.B. Kier and L.H. Hall, in ref. [10], p. 31.
[22] L.B. Kier and L.H. Hall, in ref. [8], pp. 142 – 145.
[23] M. Randić and C.L. Wilkins, J. Phys. Chem. 83(1979)1525.
[24] H. Narumi and H. Hosoya, Bull. Chem. Soc. Japan 58(1985)1778.
[25] A. Wurflinger, Molecular Phys. 49(1983)413.
[26] L.T. Wille, Molecular Phys. 61(1987)409.