

CHROM. 12,405

CALCULATION OF GAS CHROMATOGRAPHIC RETENTION INDICES OF ISOALKANES BASED ON A TOPOLOGICAL APPROACH

D. PAPAZOVA

Economic Combine "Neftochim", Institute of Petroleum, Burgas 8104 (Bulgaria)

N. DIMOV

Analytical Department, Chemical Pharmaceutical Research Institute, Sofia 1156 (Bulgaria)

and

D. BONCHEV

Higher School of Chemical Technology, Burgas 8010 (Bulgaria)

(First received June 6th, 1979; revised manuscript received September 10th, 1979)

SUMMARY

The combination of topological and structural molecular characteristics seems to be a promising tool for the prediction of retention indices. The Wiener number and the total and mean information on both distance equality and distance magnitude were combined with seven structural elements, and a linear equation with a regression coefficient 0.9986 was deduced. The mean deviation of retention indices calculated for 118 isoalkanes on squalane from their experimental values was less than 8 retention index units.

INTRODUCTION

The precise calculation of the gas chromatographic retention indices of hydrocarbons on a rigorous thermodynamic basis^{1,2} is not yet possible because of the lack of a detailed non-aqueous solution theory. Additivity approaches and other mathematical methods³⁻⁶ are insufficiently accurate. The handling of the retention index as a result of two contributions, the physico-chemical index (*PCI*) and a structural number (*StN*)⁷⁻⁹, allows the calculation of a theoretical retention index (I^{theor}) for hydrocarbons on different stationary phases with very good accuracy. The calculation of *PCI*, however, requires the vapour pressure of the hydrocarbon at the analysis temperature. Such data, especially for higher hydrocarbons, are not always available. Therefore, a calculation based on the chemical structure and conformation of compounds seems to be more useful.

A remarkably good correlation has been obtained between a number of physico-chemical properties of an organic compound and its topological invariants¹⁰. The Wiener number^{11,12}, the Randić molecular connectivity¹³⁻¹⁵, etc., made use of molecular topology for retention index prediction with a maximum error of 15 retention

index units (i.u.). The Wiener method of factor analysis is a promising approach, providing an insight into solute-solvent interactions^{16,17}. Related to this, this paper describes a study of the importance of different topological invariants in retention index calculations. We assumed that the equation¹⁸

$$I_t^{\text{theor}} = PCI_t + StN \quad (1)$$

could be transformed into

$$I_t^{\text{theor}} = a + bx + \sum_{i=1}^n c_i y_i \quad (2)$$

where a , b and c_i are constants, x is a suitable topological increment, y_i are suitable structural increments and t is the temperature of analysis.

TOPOLOGICAL INVARIANTS CALCULATION

The following topological invariants were chosen for examination in this study: Wiener number (W), total (I_D^E) and mean (\bar{I}_D^E) information on distance equality or diversity, and total (I_D^W) and mean (\bar{I}_D^W) information on distance magnitude. A brief outline of these quantities is sufficient here; more details can be found, for example, in refs. 11, 12 and 19-24.

A molecule of any hydrocarbon could be represented, according to the graph theory, by its molecular graph, in which carbon atoms are depicted by points and chemical bonds by edges. For example, the molecular graph of 2,3-dimethylbutane is



The corresponding topological distance matrix, which uniquely describes the molecule, is symmetrical:

$$D = \begin{array}{c|cccccc} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 2 & 3 \\ 1 & 0 & 1 & 2 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 & 1 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 4 & 3 & 2 & 1 & 0 & 3 \\ 5 & 2 & 1 & 2 & 3 & 0 \\ 6 & 3 & 2 & 1 & 2 & 3 \end{matrix} \end{array}$$

Every entry $d_{i,j}$ is defined by the number of the edges between i and j atoms. The total number, N , of $d_{i,j}$ is obviously

$$N = \frac{n(n-1)}{2} \quad (3)$$

where n is the number of carbon atoms in the molecule. The sum, W , of all the elements of the triangular submatrix,

$$W = \frac{1}{2} \sum_{i,j=1}^n d_{i,j} \quad (4)$$

is the so-called path number or Wiener number¹¹, which is a measure of the compactness of the molecule.

The matrix elements could be classed in k subsets according to their magnitude. In the above example there are three such subsets, those of distances equal to 1, 2 and 3, respectively.

The distance $d = 1$ appears 5 times in the triangular submatrix, while distances $d = 2$ and $d = 3$ appear 6 and 4 times, respectively. Denoting in general these appearance numbers by n_i , one can express in a concise form the essence of the distance matrix, writing the sequence $1^{*}12^{*}23^{*}3 \dots$ (or, in our example, $1^52^63^4$), which coincides with the Randić molecular code. Now, the Wiener number is easily calculated by the equation

$$W = (5 \times 1) + (6 \times 2) + (4 \times 3) = 29$$

Additional indices characterizing the distance matrix on the basis of information theory²⁴ are given in refs. 22 and 23.

The distance matrix elements are regarded as the elements of every structure, distributed according to a certain criterion into a number of classes. A certain probability is ascribed to each of the elements belonging to a certain class. Thus, a probability distribution is constructed to which an information measure (information content) is specified. Two criteria are used for classing the distance matrix elements: their equality and their magnitude. Correspondingly, two different information measures are defined on the topological distance matrix. The total information, I_D^E , and mean information, \bar{I}_D^E , on distance equality or diversity are given in bits, according to the equations

$$I_D^E = \frac{n(n-1)}{2} \cdot \log_2 \left[\frac{n(n-1)}{2} \right] - \sum_{i=1}^k n_i \log_2 n_i \quad (5)$$

and

$$\bar{I}_D^E = 2I_D^E/n(n-1) \quad (6)$$

The total and mean information on distance magnitude, I_D^W and \bar{I}_D^W , respectively, are given in bits according to the equations

$$I_D^W = W \log_2 W - \sum_{i=1}^k n_i d_i \log_2 d_i \quad (7)$$

and

$$\bar{I}_D^W = I_D^W/W \quad (8)$$

where d_i represents the distinct topological distances in a molecular graph and n_i their appearance number. In our example, $n_1 = 5$, $n_2 = 6$, $n_3 = 4$ and $d_1 = 1$, $d_2 = 2$, $d_3 = 3$.

RESULTS AND DISCUSSION

Experimental retention indices of 56 isoalkanes separated on squalane were taken from ref. 25. The validity of the linear equation

$$I^{\text{calc}} = a + bI_{\text{inv}} \quad (9)$$

where I_{inv} is any kind of topological invariant in use and a and b are constants, was examined. Independently of the type of I_{inv} , the differences between the calculated (I^{calc}) and the experimental retention indices (I^{exp}) vary from 1 to 59 i.u. Such discrepancies are not acceptable and I^{calc} cannot be a useful aid in gas chromatography. A combination of one of the topological indices and three of the most important structural elements²⁶ (n_R , n_B and n_0) was then studied in order to find which of the topological invariants could best replace PCI , where n_R is the number of substituents in the isoalkane, n_B is the number of butane chains according to Altenburg²⁷, for example for *n*-butane $n_B = 1$, for isobutane $n_B = 0$ and for 2,2-dimethylbutane $n_B = 3$, and n_0 is the total number of carbon atoms in the isoalkane.

The resulting regression equations and the corresponding correlation coefficients, R , are as follows:

$$I_1^{\text{calc}} = 254.96 + 0.96W - 17.43n_R + 13.18n_B + 46.07n_0 \quad (10)$$

with $R = 0.9862$;

$$I_2^{\text{calc}} = 241.48 + 1.10I_D^E - 16.95n_R + 13.28n_B + 47.78n_0 \quad (11)$$

with $R = 0.9862$;

$$I_3^{\text{calc}} = 196.79 + 0.09I_D^W - 20.19n_R + 11.92n_B + 59.68n_0 \quad (12)$$

with $R = 0.9856$;

$$I_4^{\text{calc}} = 67.99 + 58.70I_D^E - 15.03n_R + 13.98n_B + 61.59n_0 \quad (13)$$

with $R = 0.9868$;

$$I_5^{\text{calc}} = 87.13 + 5.35I_D^W - 23.00n_R + 10.73n_B + 76.22n_0 \quad (14)$$

with $R = 0.9852$.

Again, none of the topological indices under study dominates the others. It seems, however, that I_D^E is the most suitable index to be inserted in eqn. 1 instead of PCI . This further allows us to keep constant I_D^E as x in eqn. 2 and to choose

suitable structural elements. The simplest equation²⁸ found for the calculation of retention indices of isoalkanes is

$$I_i^{\text{theor}} = PCI_i + c_0 + c_1x_1 + c_2x_2 \quad (15)$$

with

$$x_1 = n_R + n_t - n_d \quad (16)$$

and

$$x_2 = n_{\text{CH}_3}n_B/n_L(n_0 - n_d) \quad (17)$$

where c_0 , c_1 and c_2 are constants; n_t is the number of carbon atoms in the substituents, counting from, and including the third atom from either end of the straight chain of the isoalkane; for example, for 3-methylpentane, 2,3-dimethylpentane and 4-methylheptane $n_t = 1$, and for 3-ethylhexane and 2-methyl-3-ethylhexane $n_t = 2$; n_d is the number of carbon atoms between substituents R in the alkane, when $R \geq 2$; for example, for 2,3-dimethylpentane $n_d = 0$ and for 2,4-dimethylpentane $n_d = 1$; n_{CH_3} is the number of methyl groups in the isoalkane; n_L is the number of carbon atoms in the straight chain of the isoalkane; and n_q is the number of quaternary carbon atoms in the isoalkane.

Replacing PCI in eqn. 15 by I_D^E , the new values of c_i are calculated from a 56-membered matrix. A considerably better coincidence was found between the experimental and re-calculated retention indices. It is, however, still unsatisfactory. We therefore examined a polynomial equation that includes all significant structural elements by a step regression analysis. The starting model was

$$I = F(I_D^E, n_0, n_R, n_L, n_d, n_{\text{CH}_3}, n_t, n_q) \quad (18)$$

The best linear model was found to be

$$I = 128.8 - 12.9I_D^E - 21.6n_R + 21.4n_B + 57.8n_0 - 12.5n_t + 16.8n_L \quad (19)$$

with $R = 0.9986$.

The mean deviation of I^{calc} from I^{exp} for all C_6 - C_8 isoalkanes is now less than 4 i.u. (Table I). The accuracy achieved is still not completely satisfactory, but the replacement of PCI by I_D^E is evidently possible. To support this concept, a calculation of the retention indices for 118 C_6 - C_{10} isoalkanes was made. The results obtained corroborate both the possibility of using I_D^E instead of PCI and the significance of the structural elements chosen. The mean difference between experimental and calculated values of the retention index is now 8 i.u., but the result is encouraging. We assume also that a further simplification of the model, for example by the use of suitable combinations of structural elements, is possible. A simple equation for retention index calculations, based solely on topological and structural characteristics of solutes, would be very useful for laboratory purposes. At this stage in our investigation, some helpful conclusions can be made:

TABLE I

DIFFERENCES (Δ) BETWEEN CALCULATED (ACCORDING TO EQN. 19) AND EXPERIMENTAL RETENTION INDICES OF C₅-C₈ ISOALKANES²⁹

| Hydrocarbon | I_{D}^{SQ} | I^{calc} | Δ |
|-------------------------|--------------|------------|----------|
| 2-Methylpentane | 569.7 | 574.6 | +4.9 |
| 3-Methylpentane | 584.2 | 587.2 | +3.0 |
| 2,3-Dimethylbutane | 567.2 | 564.8 | -2.4 |
| 2,2-Dimethylbutane | 536.8 | 540.1 | +3.3 |
| 2-Methylhexane | 666.6 | 659.8 | -6.8 |
| 3-Methylhexane | 676.2 | 675.2 | -1.0 |
| 2,4-Dimethylpentane | 629.8 | 629.9 | +0.1 |
| 3-Ethylpentane | 686.0 | 683.1 | -2.9 |
| 2,2-Dimethylpentane | 625.6 | 628.0 | +2.4 |
| 2,3-Dimethylpentane | 671.7 | 665.1 | -6.6 |
| 3,3-Dimethylpentane | 658.9 | 652.6 | -6.3 |
| 2-Methylheptane | 764.9 | 760.3 | -4.6 |
| 3-Methylheptane | 772.3 | 771.7 | -0.6 |
| 4-Methylheptane | 767.2 | 770.1 | +2.9 |
| 2,5-Dimethylhexane | 728.4 | 726.5 | -1.9 |
| 2,2-Dimethylhexane | 719.4 | 724.2 | +4.8 |
| 2,4-Dimethylhexane | 731.9 | 730.6 | -1.3 |
| 2,3-Dimethylhexane | 760.1 | 760.8 | +0.7 |
| 3,4-Dimethylhexane | 770.6 | 770.8 | +0.8 |
| 3,3-Dimethylhexane | 743.6 | 747.9 | +4.3 |
| 2-Methyl-3-ethylpentane | 761.4 | 764.9 | +3.5 |
| 2,2,4-Trimethylpentane | 689.9 | 687.2 | -2.7 |
| 2,3,4-Trimethylpentane | 762.4 | 747.8 | -14.6 |
| 3-Methyl-3-ethylpentane | 774.0 | 775.8 | +1.8 |
| 2,2,3-Trimethylpentane | 737.1 | 745.3 | +8.2 |
| 2,3,3-Trimethylpentane | 759.4 | 756.9 | -2.5 |

(1) the calculation of the retention indices of isoalkanes without vapour pressure data is possible;

(2) the topological index I_D^E expresses to an acceptable degree those physico-chemical properties of isoalkanes that are included in the term PCI (eqn. 1);

(3) the chosen structural elements proved their significance when the initial set of 56 isoalkanes was extended to 118;

(4) though I_D^E is calculated on the basis of the molecular graph, some structural characteristics of the molecules are evidently not involved with the necessary influence. Otherwise, the topological indices of molecules are likely to be insufficient for the satisfactory description of solute properties. Therefore, combination with the structural elements mentioned above is necessary. In this respect, the extended set of topological and structural characteristics might be regarded as a new approach aimed at predicting molecular solute properties which will be of interest to theoretical chemists.

ACKNOWLEDGEMENTS

This work was supported by Economic Combine „Neftochim”, Burgas. Assistance with the calculations by Ms. S. Dekalo of the Petrochemical Computer Centre is gratefully acknowledged.

REFERENCES

- 1 C. E. Figgins, T. H. Risby and P. C. Jurs, *J. Chromatogr. Sci.*, 14 (1976) 453.
- 2 A. N. Korol, *J. Chromatogr.*, 129 (1976) 125.
- 3 B. G. Berezkin, *Neftekhimiya*, 1 (1961) 169.
- 4 J. Takács, C. Szita and G. Tarján, *J. Chromatogr.*, 56 (1971) 1.
- 5 M. Randić, *J. Chromatogr.*, 161 (1978) 1.
- 6 G. Spivakovskii, A. I. Tishchenko, I. I. Zaslavskii and N. S. Wulfson, *J. Chromatogr.*, 144 (1976) 1.
- 7 N. Dimov and D. Papazova, *J. Chromatogr.*, 137 (1977) 265.
- 8 D. Papazova and N. Dimov, *J. Chromatogr.*, 137 (1977) 259.
- 9 N. Dimov and D. Papazova, *J. Chromatogr.*, 148 (1978) 11.
- 10 J. Chrétien and J. E. Dubois, *J. Chromatogr.*, 158 (1978) 43.
- 11 H. Wiener, *J. Amer. Chem. Soc.*, 69 (1947) 17; 69 (1947) 2536; *J. Phys. Chem.*, 62 (1948) 425.
- 12 J. R. Platt, *J. Chem. Phys.*, 15 (1947) 419; 56 (1962) 328.
- 13 M. Randić, *J. Amer. Chem. Soc.*, 97 (1975) 6609.
- 14 L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- 15 J. Chrétien and J. Dubois, *Anal. Chem.*, 49 (1977) 747.
- 16 P. H. Wiener and D. Howery, *Can. J. Chem.*, 50 (1972) 448.
- 17 R. B. Selzer and D. G. Howery, *J. Chromatogr.*, 115 (1975) 139.
- 18 N. Dimov, *J. Chromatogr.*, 119 (1976) 109.
- 19 F. Harary, *Graph Theory*, Addison-Wesley, Reading, Mass., 2nd ed., 1971.
- 20 A. T. (Editor), *Chemical Applications of Graph Theory*, Academic Press, New York, 1976.
- 21 D. H. Rouvray, *Inf. Commun. Match Math. Chem.*, 1 (1975) 125.
- 22 D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, 67 (1977) 4517.
- 23 D. Bonchev and N. Trinajstić, *Int. J. Quantum Chem., Symp.*, 12 (1978) 293.
- 24 L. Brillouin, *Science and Information Theory*, Academic Press, New York, 1956.
- 25 A. Matukuma, in C. L. A. Harbourn and R. Stock (Editors), *Gas Chromatography, 1968*, Institute of Petroleum, London, 1969, p. 55.
- 26 N. Dimov and D. Papazova, *Chromatographia*, 12 (1979) 720.
- 27 K. Altenburg, in H. G. Struppe (Editor), *Gas Chromatographie 1968*, Akademie Verlag, Berlin, 1968, p. 1.
- 28 N. Dimov and D. Papazova, *Chromatographia*, 12 (1979) 443.
- 29 J. Rijks and C. Cramers, *Chromatographia*, 7 (1974) 99.