CHROM. 8754

EMPIRICAL QUANTUM CHEMICAL APPROACH TO STRUCTURE-GAS CHROMATOGRAPHIC RETENTION INDEX RELATIONSHIPS

II. CYCLOHEXANE DERIVATIVES

G. FIRPO and M. GASSIOT

Sección Cromatografía de Gases, Depto. Quimica Analítica, Instituto Quimico de Sarriá, Barcelona (Spain)

M. MARTÍN and R. CARBÓ

Sección Química Cuántica. Depto, de Química Orgánica, Instituto Químico de Sarriá, Barcelona (Spain)

and

X. GUARDINO and J. ALBAIGÉS

Instituto Química Orgánica, Patronato Juán de la Cierva (CSIC), Barcelona (Spain)

(Received July 4th, 1975)

SUMMARY

It has been found possible to correlate chromatographic retention parameters with molecular electronic structures by means of a multilinear relationship. The empirical parameters are deduced theoretically from molecular structure calculations. In this manner, the retention index can be predicted with a small error $\left(\langle 1 \rangle \right)$.

The proposed mode1 is supported by the results of studies on 30 cydohexane derivatives.

The theoretical methodology is briefly discussed.

INTRODUCTION

A correlation between the retention data for a series of cholestanol derivatives and some of their structuraI parameters, obtained by quantum chemical empirical methods (De1 Re's method), was presented in **2** previous paperl.

It will be shown in the present work that the proposed method, MASP (multilineal adjustment of semi-empirical parameters), has a general validity, and furnishes fairly precise results for families of compounds that differ structurally from those previously studied. In this instance, retention data for a series of saturated and unsaturated alicyclic hydrocarbons of the types indicated in Fig. I have been corre i ated. Retention indices have been obtained experimentally with a known error range². This permits us to compare, in a more refiabie way, the predicted error **of** the proposed theoretical model.

Del Re's method³, as it has been applied in this study, and the process for obtaining the approximate equation by the Ieast-squares method, zre described here.

Fig. 1. Fundamental molecules considered in this work. Other molecules are derived from these, having different positions of the side-chains and different degrees of unsaturation.

Because Del Re's method does not permit the prediction of the charge density distribution and the energy due to unsaturation exhibited by the compounds studied, it was necessary to use in addition a computational procedure based on the LCAO-MO method proposed by Hückel^{4,5}.

Finally, it should be pointed out that the calculation method is comparatively easy to run, the use of a small computer (4K memory) being sufficient for its implementation.

ì

THEORETICAL

The general basis of the method assumes that the chromatographic retention of a given chemical species can be considered as a function of two characteristics directly connected with molecular structure: (a) intermolecular forces between chromatographic solute molecules and (b) intermolecular forces in the chromatographic system: solute-stationary phase.

Both characteristics will depend, in a first approach, on the electrical charge distribution and will arise as a consequence of electrostatic interactions. In order to obtain a physical model that can describe these interactions, it is necessary to achieve the following:

(1) a representation of the intrinsic nature of the molecule through some parameters that have a clear physical meaning, and which can be calculated with total independence of the experimental data;

(2) a definition of a model that is capable of describing in an adequate way the characteristic interactions of a chromatographic system;

(3) an expression of the interaction model through an equation of an empirical type, which can explain the phenomena involved to a reasonable extent.

The space distribution of the electronic charge density is peculiar to each molecule, and can be evaluated through the wave function, associated to the molecular electronic energy. Within the scope of this work, the calculation of the interaction energy and deformation of the charge clouds of the isolated systems becomes exceedingly complicated.

In order to simplify the treatment and taking into account the fact that we are trying to study the interaction of a molecular family with a stationary phase of constant composition, only those variables are considered in the constructic a of the model which are peculiar to each molecule, such as the total electronic energy (E) and the net localized charge on each atom of the molecule (Q_i) . These will therefore be the empirical parameters to be correlated with the retention index.

The model has various approximation levels, according to the technique used in the calculation of E and Q_i . In this work, as previously¹, the simplest LCAO–MO model described by Del Re⁶⁻⁸ was found to be sufficient.

Del Re's method, modified in its computational part by Carbó³, calculates the charge (O_i) distribution, as well as a parameter related to the energy of the localized electronic system. Each bond is studied separately from the remainder and the calculation involves only secular equations of dimension (2×2) .

For each atom *i*, the Coulomb integral is expressed as

 $\alpha_i = \alpha_0 + \delta_i \beta_0$

where α_0 and β_0 are reference values and δ_t is a parameter depending on the atoms to which the centre i is bonded.

The resonance integral, β_{ij} , for each $i-j$ bond is evaluated through

$$
\beta_{ij}=\varepsilon_{ij}\beta_{o}
$$

 ε_{ij} depends on the character of the bonded atoms. The problem reduces to the calculation of these parameters. The system of equations used for this purpose is

$$
\delta_i = \delta_{0,j} + \sum_j \gamma_{ij} \delta_i \quad (i = 1,n)
$$

where $\delta_{0,i}$ and γ_{ij} are empirical parameters (see Table I).

TABLE I

PARAMETERS USED IN DEL RE'S METHOD

The net charges on each atom, Q_i , are obtained from the equations:

$$
Q_{ij} = (\delta_j - \delta_i)/2\varepsilon_{ij}
$$

and

$$
Q_i = \sum_j Q_{ij}
$$

The contribution to the total energy of each bond is calculated through

 $E_{ij} = a_0 + \beta_0 [\delta_i + \varepsilon_{ij} (O_{ij} + V_1 + O_{ij}^2)]$

the energy contribution due to each atom being

$$
E_i = 2 \sum_j E_{ij}
$$

and the total energy

$$
E=\sum_{i=1}^n E_i
$$

In the above expressions, the symbol Σ refers only to the atoms bonded to the atom i.

The values of E and Q_i corresponding to each molecular structure studied have been related to the macroscopic property, the retention index (RI) , using the empirical equation

$$
RI = aE + \Sigma c_i Q_i + b \tag{1}
$$

whose coefficients are found through a least-squares procedure.

The generalization of such a method to a system of n equations with m variables $(n \gg m)$ is very easily achieved. Let

$$
Ab = c \tag{2}
$$

be the matrix equation, which represents the problem, where A is an $(m \times n)$ matrix having the values of the independent variables as elements, b is an $(n \times 1)$ vector containing the equation coefficients which are to be calculated and c is an $(m \times 1)$ vector constructed with the values of the dependent variable. It is possible to show that the solution b with the minimum quadratic error⁹ is obtained by building up the matrices

$$
D = A^T A \quad (n \times n) \tag{3}
$$

and

$$
\mathbf{d} = A^T \mathbf{c} \quad (n \times 1) \tag{4}
$$

together with the solution of the matrix equation:

$$
Db = d \tag{5}
$$

As D is a square matrix with a Gram structure with respect to the A columns, if these are linearly independent, then

Det $|D| \neq 0$

and there exists a D^{-1} , which fulfils the relationship

$$
b = D^{-1}d \tag{6}
$$

In this work, D^{-1} was calculated through Choleski's transformation¹⁰ of D, using the

fact that, if D is non-singular, then, due to the definition in eqn. 3, D is positive definite. If T is an upper triangular matrix, then

$$
D = T^T T \tag{7}
$$

The inverse T^{-1} is easily computed¹⁰, and hence

 $D^{-1} = T^{-1} (T^{-1})^T$

The general procedure can be summarized as shown in Fig. 2.

CALCULATION PARAMETERS OF MUTILINEAL EQUATION PATH

PREDICTION RETENTION INDEX PATH

Fig. 2. Scheme of the MASP method.

EXPERIMENTAL

Alicyclic hydrocarbons, upon which the present study was based and which are shown in Fig. 1, were obtained with a purity of over $98-99\%$ from Fluka (Buchs, Switzerland) (compounds Nos. 1-13, 17, 18, 24, 25) or by laboratory synthesis (Nos. 15, 16, 20–22, 26–30).

Retention indices were calculated by means of an HP 9830A desk computer² from the retention times determined with a digital integrator in an open stainlesssteel capillary column (100 ft. \times 0.01 in.) packed with Apiezon L (carrier gas helium). The gas chromatograph used was a Perkin-Elmer 990.

The operating conditions were optimized by means of a prior analysis of their influence on the reliability of the retention indices obtained². The mean absolute error was found to be 0.5–1 unit, the largest value corresponding to aromatic hydrocarbons.

The calculation of the molecular parameters E and Q_t was carried out by means of a GA 18/30 computer.

RESULTS AND DISCUSSION

Taking as variables the total electronic energy of the sigma system (E) and the localized charges in each atom of the ring $(Q_1, Q_2, ..., Q_6)$, and applying eqn. 1, the results that were obtained for the molecules studied are given in Table II, column (a).

TABLE II

STATISTICAL PARAMETERS OF MULTILINEAL ADJUSTMENTS CORRESPONDING TO THE APPLICATION OF THE DIFFERENT EQUATIONS

The individual results in columns (b) and (c) are listed in Figs. 3 and 5, respectively.

However, the individual results indicated that the error may be related to the degree of unsaturation of each molecule. The π -electron energy for the different types of unsaturation present (E_{τ}) in the molecules was therefore calculated by means of Hückel's method¹¹, the methyl group being considered through an inductive model. The coplanarity angle of the hydrogen atoms' rotational freedom in the C_{α} position was not been taken into account, as one of the main aims of this work was to achieve an easy and understandable framework.

It was taken into account that 3H or 2H in the C_{α} position produces the same inductive effect, and that 1H has no effect. The E_{τ} corresponding to each type of unsaturation is shown in Table III, and it is considered that the total electronic energy is the sum of the contributions due to the σ and π systems.

With the introduction of this factor, the total correlation obtained (Table II, column b) was considerably improved. The overall results obtained are shown in Fig. 3.

The experimental retention indices (RI) were compared with those calculated by means of eqn. 1 (CI), as well as the percentage relative error (ER), and also the percentage error relative to the CI range (EM) . The correlation obtained was excellent, from a statistical point of view, and the proposed equation is capable of explaining 99.49% of the total variance of the data (R^2) .

Similarly, the retention indices of three substances (Nos. 14, 19 and 23), which had not previously been determined experimentally were predicted.

TABLE III

 E_{π} (β UNITS) VALUES FOR THE π SYSTEMS PRESENT IN THIS CHEMICAL FAMILY

Influence of temperature

Using retention data at four or five different temperatures in the range 60-163" at intervals of20" for each molecule, it was possible to verify the hypothesis of a linear variation of retention index with temperature¹².

With isolated molecules, in all instances regression coefficients over 0.995 (minimal significance 99.95%) were obtained and the mean absolute error was always Eess than L retention index unit. The coefficients for the straight line are virtually identical for very different molecules within the same family. Functions other than linear were tried but no better results were obtained. This fact led us to amplify the proposed equation (eqn. I) to include the temperature term:

$$
RI = aE + bT + \Sigma c_t Q_t + d \tag{8}
$$

which implies a constant $\delta I/T$ value for the whole family, but with a particular ordinate for each substance.

The degree of approximation obtained with this new equation can'be observed by comparing the degree of adjustment obtained for a given group of molecules when eqn, I or 8 is used. Thus, for the sub-family of the aromatic derivatives, the results obtained can be compared by applying me or other equation, that is, by making the correlation at constant temperature (taking only E, Q_1 and Q_4 as variables) or introducing temperature as a variable (Fig. 4a).

Compound	RI	СI	ΕR	EM	Diff
Cyclohexane	696.5	694.3	0.30	0.20	2.14
Benzene	698.3	704.1	0.83	0.54	-5.82
Cyclohexene	715.4	721.5	0.85	0.58	-6.14
1,4-Cyclohexadiene	742.4	742.6	0.03	0.02	-0.22
1,3-Cyclohexadiene	699.0	6S6.3	1.80	1.19	12.62
Me hylcyclohexane	757.8	765.2	0.98	0.70	7.47
Toluene	805.9	797.2	1.07	0.81	8.65
3-Methylcyclohexene	771.0	780.6	1.25	0.91	-9.66
4-Methylcyclohexene	776.9	760.0	2.17	1.59	16.89
1-Methylcyclohexene	801.1	805.0	0.61	0.46	-4.92
Methylenecyclohexane	769.4	762.8	0.85	0.61	6.54
cis-Dimethylcyclohexane	840.1	853.9	1.64	1.30	-13.81
p-Xylene	911.9	904.7	0.78	0.67	7.19
cis-1,4-Dimethylcyclohex-2-ene		846.0			
1,4-Dimethylcyclohex-1-ene	860.8	869.7	1.03	0.84	- 8.93
1,4-Dimethylcyclohex-1(7)-ene	834.7	851.5	2.01	1.58	-16.81
Isopropylcyclohexane	961.2	948.8	1.28	1.16	12.33
Cumene.	951.1	958.9	0.82	0.74	-7.87
3-Isopropylcyclohexene		964.2			
1-Isopropylcyclohexene	949.3	943.8	0.56	0.51	5.40
4-Isopropylcyclohexene	976.3	967.1	0.93	0.86	9.17
Isopropylidencyclohexane	991.0	989.7	0.12	0.12	1.28
Iso propenylcyclohexane		939.4			
cis-p-Menthane	1027.5	1019.6	0.76	0.74	7.87
Cymene	1051.3	1051.9	0.05	0.05	-0.63
<i>n</i> -Menth-2-ene	1012.8	1011.9	0.08	0.08	0.86
p-Menth-3-ene	1003.8	1014.8	1.09	1.03	-11.01
p -Menth-1-ene	1059.0	1060.4	0.13	0.13	-1.47
p -Menth-1(7)-ene	1029.3	1017.3	1.15	1.12	11.91
p -Menth-4(8)-ene	1052.5	1060.6	0.77	0.76	-8.11
	Mean absolute error			7.62	
	Mean relative error Mean error to range			0.89	
				0.71	
		Coefficients		Signification	
	E	12.5799	100.00		
	\mathcal{Q}_1	-254.1832	99.73		
	Q,	638.0364	99.99		
	Ω,	-170.2555	85.11		
	Q.	136.0749	81.30		
	Q5	470.5475	93.18		
	Q_{6}	973.4047	99.95		
Independent term 330.3196					

Fig. 3. Computer printout giving experimental (RI) and calculated (CI) retention indices. The experimental data were obtained at 120°.

This type of calculation was also tried for the methylcyclohexane sub-family (varying in the degree of unsaturation and position of the side-chain). Although eqn. I cannot be used in this instance, as the number of variables is greater than that of the molecules (E, Q_1 , Q_2 , Q_3 , Q_4 , Q_5 , Q_6), but the mean error obtained is not greater

 $\ddot{}$

 \sim

Fig. 4. Computer printout giving experimental (RI) and calculated (CI) retention indices according to whether eqn. 1 or 8 is applied. (a) Benzene derivatives; (b) methylcyclohexane derivatives.

than 1.0 retention index unit, and the largest error given by any molecule is 2.5 index units (Fig. 4b).

It should be taken into account that the assumed experimental error lies between 0.5 and 1 index unit and the correlations presented in Fig. 4 are very similar to what is called "perfect correlation". It can be concluded, therefore, that the increase in error when using eqn. 8 is not significant.

Applying this new concept, expressed by means of eqn. 8, to the whole set of cyclohexane derivatives, a correlation is obtained with all of the significant terms,

 $\sim 10^{11}$ μ

 $\sim 10^{-11}$

 $\sim 10^7$

 \sim

G. FIRPO et al.

 $\sim 10^{-11}$

 $\ddot{}$

Fig. 5. Computer printout giving experimental (RI) and calculated (CI) retention indices when eqn. 8 is applied. For statistical data, see Table II, column (c).

whose statistical parameters appear to be very similar to those previously obtained by means of eqn. I (Table II, column b), with a slightly higher mean error (see Table II. column c).

Finally, in Fig. 5 the overall results obtained by using eqn. 8 are shown. Retention indices were predicted for three compounds, not previously studied, and retention indices were calculated at temperatures for which no experimental data are available.

REFERENCES

- 1 M. Gassiot, E. Fernández, G. Firpo, R. Carbó and M. Martín, J. Chromatogr., 108 (1975) 337.
- 2 X. Guardino, J. Albaigés, G. Firpo, R. Rodríguez and M. Gassiot, J. Chromatogr., in press.
- 3 R. Carbó, Theor. Chim. Acta, 18 (1970) 170.
- 4 E. Hückel, Z. Phys., 70 (1931) 204.
- 5 E. Hückel, Z. Phys., 76 (1932) 76.
- 6 G. Del Re, J. Chem. Soc., London, (1958) 4031.
- 7 G. Del Re and B. Pullman, J. Chim. Phys., 62 (1965) 943.
- 8 G. Del Re, in B. Pullman (Editor), Electronic Aspects of Biochemistry, Academic Press, New York, 1964, p. 221.
- 9 C. Daniel, Fitting Equations to Data, Wiley-Interscience, New York, 1971, Appendix B.
- 10 E. D. Durand, Solutions Numériques des Équations Algébriques, Tome II, Masson, Paris, 1961, pp. $74 - 77.$
- 11 A. Streitwieser, Molecular Orbital Theory for Organic Chemists, Wiley, New York, 1962.
- 12 G. D. Mitra and N. C. Saha, Chromatographie, 6 (1973) 93.