Journal of Chromatography, 360 (1986) 25–32
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 18 524

ACCURACY IN MODELLING OF RETENTION INDICES IN GAS CHROMATOGRAPHY

NENO P. DIMOV
Chemical Pharmaceutical Institute, 1156 Sofia (Bulgaria)

(First received January 23rd, 1985; revised manuscript received January 31st, 1986)

SUMMARY

The accuracy of 18 models for calculation of retention indices, *I*, of isoalkanes separated on squalane has been examined. These models emerged in the period from 1967 till 1985. Values are given for the dispersions, the maximum standard deviations between the experimental and calculated retention indices, the number of discrepancies greater than 1 index unit (i.u.) as a percentage of all compounds studied and the Fisher test for homogeneity of dispersions. On this basis a classification of the models as correlative or predictive is proposed. It is shown that there are no predictive models. The accurate precalculation of *I* on squalane with an accuracy of 1 i.u. is possible only for C_6 isoalkanes. Some mathematical models have an accuracy better than ± 3 i.u. for the C_6–C_8 isoalkanes.

INTRODUCTION

Controversial opinions can be found in the literature concerning the prediction of retention indices, *I*, by calculation methods. Ladon[1] postulated the so-called non-calculability principle in predicting *I* from molecular structure. On the other hand, there are a lot of mathematical models which relate *I* with solute structure and corresponding properties. These methods have been reviewed[2,3].

In the present paper the available calculation methods are examined with respect to their accuracy. A critical analysis of some of the methods can be found in the literature[4–8], but a complete comparison is lacking even in the above mentioned reviews. First, methods elaborated for the calculation of *I* for isoalkanes separated on squalane will be examined since most of the accumulated data and published methods refer to hydrocarbons separated on squalane[9].

Recently, Johansen *et al.*[10] showed that the *I* values for isoalkanes on OV-101 as the stationary phase are comparable with the I_{exp} values obtained on squalane. A similar comparison of DC-200 and SE-30 as stationary phases was reported by Ramsey *et al.*[11]. Hence, a calculation method must be valid for non-polar phases other than squalane. This was proven, at least for isoalkanes, by Dimov and Papazova[12].

0021-9673/86/$03.50

THEORETICAL

The accuracy of the methods was estimated according to statistical approaches[13]. The following additional parameters, however, are considered to be more important for estimating the identification ability of a method in gas chromatography (GC): the value of the maximum discrepancy, S_{max}, between I_{exp} (ref. 14) and I_{calc}; the number of deviations greater than 1 index unit (i.u.), expressed as a percentage of all cases studied. These parameters serve as a measure of the reliability of the methods examined. Deviations of 1 i.u. are accepted as practically attainable in interlaboratory reproducibility. The Fisher test (see Table I) gives an idea of the extrapolation possibility of the method.

Methods evaluated

Included in the examination are methods based on different ideas, but not using experimental chromatographic data[14]. The methods are arranged in the order of their publication.

In 1967 Bonastre and Grenier[15] suggested the following equation:

$$I_p = 100n + 100 \cdot \frac{\log p_n^0 - \log p_x^0}{\log p_n^0 - \log p_{n+1}^0} \tag{1}$$

The vapour pressure of the solute, p_x^0, as well as that of the normal alkanes with n and $n + 1$ carbon atoms, p_n^0 and p_{n+1}^0, at the analysis temperature can be found, *e.g.*, in refs. 16 and 17.

In 1968 Altenburg[18] proposed a linear relationship between I and the molecular mass, M, the density, d_4^{20}, and a geometric property of the hydrocarbon molecule, K_R^2. Values of K_R^2 can be calculated but they have also been tabulated[18]. The data obtained according to the following equation

$$I = A + B(M \cdot d_4^{20}) + C \cdot K_R^2 \tag{2}$$

are taken in the present comparison.

Dimov and Shopov[19] established in 1969 that the discrepancies between I_{exp} and I_p decrease considerably, especially for cycloalkanes, if the molecular volumes, V_{mol}, of the solute and n-alkanes are included in the equation for I_p:

$$\mathrm{PCI} = 100n + 100 \cdot \frac{\log (p_n^0 \cdot V_n) - \log (p_x^0 \cdot V_x)}{\log (p_n^0 \cdot V_n) - \log (p_{n+1}^0 \cdot V_{n+1})} \tag{3}$$

They named the value obtained the physicochemical index (PCI), because it includes physicochemical characteristics of the compounds. However, high discrepancies still existed between PCI and I_{exp} and in 1971 they proposed the so-called corrected PCI[20]. The new value takes into account the contributions of specified structural groups determined, however, experimentally.

Stemming from the Altenburg approach, Martinov and Vigdergauz[21] suggested in 1970 the following equation

$$I = 800 - 4.5(125.655 - t_b) + 22.5(2.0252 - K_R^2) \tag{4}$$

which gives improved results for isooctanes, as well as the equations

$$I = 900 - 4.7(150.796 - t_b) + 15(2.3657 - K_R^2) \tag{5}$$

$$I = 175.813 + 2.243(0.01t_b) - 398.695(2d_4^{20}) - 87.973K_R^2 + 1283.461(2d_4^{20})^2 + 38.774(K_R^2)^2 \tag{6}$$

for isononanes and isodecanes, where t_b is the boiling point.

In 1971, based on a thermodynamic approach, Hammers and de Ligny[22] introduced in the calculation the molecular volume of the stationary phase, $V_{st.p.}$:

$$I_{p.V} = 100n + 100 \cdot \frac{\ln \dfrac{p_n^0 \cdot V_n}{p_x^0 \cdot V_x} - \dfrac{V_n - V_x}{V_{st.p.}}}{\ln \dfrac{p_n^0 \cdot V_n}{p_{n+1}^0 \cdot V_{n+1}} - \dfrac{V_n - V_{n+1}}{V_{st.p.}}} \tag{7}$$

In the same year Robinson and Odell[23] proposed the so-called standard retention index, I_{std}, calculated according to:

$$I_{std} = 100n + 100 \cdot \frac{\log t_{b_x} - \log t_{b_n}}{\log t_{b_{n+1}} - \log t_{b_n}} \tag{8}$$

This index is proposed for characterizing the stationary phases, but especially for alkenes the coincidence with the experimental values is very good.

Castello *et al.*[24] studied thoroughly the reasons for the discrepancies observed between I_{calc} and I_{exp} and came to the conclusion that a possible way of calculating I would be via the correlation of the total number of carbon atoms and the contributions of some of the characteristic structural groups of the solute. As characteristic groups they chose, *e.g.*, quaternary, and tertiary carbon atoms. They defined eleven such structural groups and tabulated their contributions in index units.

Dubois and Chrétien[25,26] applied the DARC system for topological description of various molecular behaviours to the calculation of the I values of isoalkanes. The equations used had up to 23 parameters.

Two years later Randic[4], the author of the connectivity index, χ, discussed some of the previous papers in this field and proposed a considerably simplified topological equation for I calculation

$$I = 200(\chi - 1.4142) + 300 + (T_3)^2 \tag{9}$$

where T_3 restricts the butane chains, introduced by Altenburg[18], to those having methyl groups at both ends of the chain.

In a series of studies started in 1971, Takács *et al.*[27] elaborated the idea that I comprises three contributions

$$I = I_a + I_b + I_{i(T)} \tag{10}$$

where I_a is the atomic contribution, I_b the bond contribution, and the term $I_{i(T)}$ is associated with the interaction between the solute and the stationary phase at temperature T. The following arbitrary values to I_a and I_b are given:

$$I = 3.93(I_a + I_b) = 3.93I_m \qquad \text{and} \qquad I_a = 1.2n + 0.10 \tag{11}$$

where I_m is the so-called molecular contribution and n is the number of carbon atoms. About 122 types of C–C bonds and 178 types of C–H bonds for calculation of I_b were defined and their values in index units (i.u.) tabulated[28].

In 1976 I proposed a regression equation for the calculation of the PCI corrections[29], referred to herein as the structural number, StN

$$I = \text{PCI} + \text{StN} \qquad \text{StN} = A + \Sigma(B_i X_i) \tag{12}$$

where X is any structural element and i is the number of elements. Five such elements are defined to be most important: the total number of carbon atoms, the number of carbon atoms in the straight chain of the isoalkane, the number of methyl groups, the number of quaternary carbon atoms and the number of butane chains, n_B.

In 1971 Spivakovskii *et al.*[30] critically examined the additivity principle proposed in GC by Berezkin[31] as early as in 1961. Their studies corroborate the opinion that I is a function of the Gibbs free energy of solution, not only of the molar free heat of solution. The molar free entropy of solution depends on the general molecular configuration[32], hence, the contribution of identical structural groups linked to different neighbours cannot be the same, which corresponds to Ladon's opinion. In spite of this, the authors gave several equations for calculating the total I value. The results with the highest correlation coefficient (r = 0.998), including 19 parameters, are used in the present comparison.

In 1980 Korol and Misjuk[33] published I_{calc} values for about 50 isoalkanes obtained on a thermodynamic basis. Korol was one of the first to emphasize the important rôle of the entropy of solution in GC[34]. Korol[32,35] defined the so-called entropic selectivity, F^0, of the stationary phase. On the basis of equations for calculating the enthalpy of solution and F^0, he obtained the values of I_{calc} used in the present comparison.

The equation of Papazova *et al.*[36], published in 1980, could be related to the topological approaches:

$$I = A + B \cdot \bar{I}_{\text{D}}^{\text{E}} + \text{StN} \tag{13}$$

It could be said that this equation is a form of eqn. 12, in which PCI is replaced with the distance quantity $\bar{I}_{\text{D}}^{\text{E}}$[37].

In 1982 Saura-Calixto and García-Raso[38] postulated a linear relationship between I and the Van der Waals volume of the isoalkane, V_w. The latter is calculated according to Bondi[39]. In the present comparison the following equation

$$I_{V_w} = -30.23 + 9.3084V_w \tag{14}$$

which has a correlation coefficient, r, of 0.9978.

Recently Saura-Calixto *et al.*[40] repeated the use of the topological index $\bar{I}_{\text{D}}^{\text{E}}$ and executed a mathematical model on a pure topological basis. They reported an accuracy of 1.25% for the studied isoalkanes.

RESULTS AND DISCUSSION

The data published in the above mentioned papers as well as those calculated by me, according to the recommendations given in the methods considered, are summarized in Table I.

I shall define three groups of equations. The quantitative classification criterion will be the magnitude of the discrepancies between the calculated and the experimental value of the retention indices. If these discrepancies are statistically equal to the corresponding interlaboratory reproducibility, taking into account the compounds separated and the stationary phase used, the equations are adequate. I have taken the reproducibility of ± 1 index unit (i.u.) as the maximum deviation of I_{calc} from I_{exp} for isoalkanes separated on squalane.

The equations giving statistically equal variances (when greater than 1) for the different groups of hydrocarbons (C_6, C_7, C_8, etc.) and for which $s_{\max}$ is greater than ± 1 i.u. but less than ± 5 i.u., are classified as predictive.

The third group of equations should also have a correlation coefficient above 0.99, but the discrepancies obtained are too great and the methods are not satisfactory for identification nor for predictive purposes. Such equations have, however, practical meaning, because they give useful information about the most significant characteristics of the solute. These equations will be termed correlative.

This study shows that two equations, those of Takacs *et al.*[27] and Dimov[29], could be considered as adequate, but only for hexanes. Thus, the comparison shows that there is still no adequate mathematical model for calculation of the I values for all C_6–C_{10} isoalkanes.

The variances were tested according to the Fisher test (F and $F^{0.95}_{fi,fj}$) and the following equations could be classified as predictive:

Altenburg[18], for C_6–C_8 isoalkanes; $F = 1.91$, $F_{4,16} = 3.01$, $s_{\max} = 4.4$ i.u.

Dimov *et al.*[20], for C_6–C_8 isoalkanes; $F = 4.8$, $F_{16,4} = 5.8$, $s_{\max} = 3.1$ i.u.

Dubois and Chrétien[25], for C_6–C_7 isoalkanes; $F = 1.66$, $F_{8,4} = 6.04$, $s_{\max} = 4.9$ i.u.

Takács *et al.*[27], for C_6–C_7 isoalkanes; $F = 2.4$, $F_{8,4} = 6.04$, $s_{\max} = 3.4$ i.u.

Chrétien and Dubois[26], for C_6–C_7 isoalkanes; $F = 5.66$, $F_{8,4} = 6.04$, $s_{\max} = 4.2$ i.u.

Dimov[29], for C_6–C_9 isoalkanes; $F = 1.62$, $F_{28,8} = 3.08$, $s_{\max} = 3.9$ i.u.

Korol and Misjuk[33], for C_9–C_{10} isoalkanes; $F = 1.4$, $F_{24,23} = 1.98$, $s_{\max} = 4.8$ i.u.

This list encourages me to believe that the problem of prediction of I values with the necessary accuracy is solvable.

The data from Table I allow two other conclusions. First, different experimental values for I have been taken in the different studies. If a unified value of I for isoalkanes separated on squalane existed, the discrepancies would have been smaller. Secondly, a combination of thermodynamic and structural factors is the most promising way of constructing an adequate predictive equation. Such an equation should have two general contributions: one based on some thermodynamic properties (extensive factors) and should be the principal contributor; the second one could be a

TABLE I

VARIANCES, s^2, OF n ISOPARAFFIN RETENTION INDICES, CALCULATED ACCORDING TO DIFFERENT AUTHORS, THE MAXIMUM DISCREPANCIES, s_{max}, PERCENTAGES OF DISCREPANCIES GREATER THAN 1 i.u. (%) AND FISHER CRITERIA (F_{exp} AND F_{tabl}) AT THE 95% LEVEL

Author(s)	*Isoparaffin*	s^2	n	s_{max}	%	F_{exp}	F_{tabl}	*Ref. for* I_{exp}
Bonastre and Grenier[15] (1967)	Hexanes*	9.4	4	4.2	100	2.65	6.04	
	Heptanes*	24.4	8	6.4	88	1.52	2.59	
	Octanes*	37.2	14	11.9	86			
Altenburg[18] (1968)	Hexanes	9.0	4	4.4	75	1.8	3.63	
	Heptanes	5.0	8	3.8	63	1.06	2.59	41
	Octanes	4.7	16	4.2	75			
				mean	71.3			
Dimov and Shopov[19] (1969)	Hexanes	19.2	4	6	75	2.03	6.04	
	Heptanes	39.0	8	11	88	1.32	3.24	
	Octanes	51.6	14	13	100			
Martinov and Vigdergauz[21] (1970)	Octanes	4.6	17	5	76	1.8	1.94	42
	Nonanes	2.5	34	3	53			
				mean	60.7			
Hammers and de Ligny[22] (1971)	Hexanes	4.9	2	3	50	2.43	19.3	
	Heptanes	11.8	7	6.4	86	1.5	3.55	43
	Octanes	17.7	13	11.2	77			
Robinson and Odell[23] (1971)	Hexanes	17.2	4	8	50			
	Heptanes	37	8	13	87			
Dimov and Shopov[20] (1971)	Hexanes	0.6	4	1.2	25	1.83	6.04	
	Heptanes	1.1	8	2.1	25	2.64	3.2	
	Octanes	2.9	16	3.1	62			
				mean	46			
Castello *et al.*[24] (1973)	Hexanes**	12.8	4	5	75	1.7	3.84	
	Heptanes**	7.4	8	7	38	3.0	3.24	
	Nonanes**	20.2	33	10	88	1.1	2.08	
	Octanes**	22.5	14	9	71			
				mean	76.3			
Dubois and Chrétien[25] (1974)	Hexanes	6.0	4	4	50	1.66	6.04	
	Heptanes	10.0	8	4.9	100	1.03	2.55	
	Octanes	9.6	16	7.4	50			
				mean	64.3			
Takacs *et al.*[27] (1971)	Hexanes*	1.2	4	1	0	2.4	6.04	
	Heptanes*	2.9	8	3.4	38	2.5	3.2	
	Octanes*	7.1	16	9.9	19	4.05	2.22	
	Nonanes*	28.8	26	19.1	46			
				mean	33.4			

TABLE I (*continued*)

Author(s)	*Isoparaffin*	s^2	n	s_{max}	%	F_{exp}	F_{tabl}	*Ref. for* I_{exp}
Chrétien and Dubois[26] (1976)	Hexanes	1.2	4	2.5	50	5.44	6.04	42, 44
	Heptanes	6.8	8	4.2	62	2.06	3.22	
	Octanes	14.0	15	6.2	87	2.56	2.22	
	Nonanes	35.9	34	12.8	97	2.44	1.56	
	Decanes	87.6	72	18.4	90			
				mean	88.6			
Dimov[29] (1976)	Hexanes	0.3	4	0.8	0	4.33	6.04	
	Heptanes	1.3	8	2.5	37	1.5	2.59	
	Octanes	0.85	16	2.1	25			
	Nonanes	2.1	28	3.9	27			
				mean	29.6			
Spivakovskii *et al.*[30] (1977)	Hexanes	24.9	4	7.1	75	3.66	6.04	
	Heptanes	41.2	8	10.0	100	1.75	2.64	
	Octanes	23.6	15	10.3	73	1.58	2.23	
	Nonanes	37.5	33	12.2	85			
				mean	83.3			
Randic[4] (1978)	Hexanes*	14.6	4	7.2	75	2.7	6.04	45
	Heptanes*	39.3	8	11.9	75	3.25	3.20	
	Octanes*	127.8	16	21.9	94			
Korol and Misjuk[33] (1980)	Nonanes	3.0	23	4.8	48	1.4	2.0	42
	Decanes	4.2	24	3.8	71			
				mean	59.7			
Papazova *et al.*[36] (1980)	Hexanes	20	4	5	100	1.3	4.12	
	Heptanes	15.4	8	7.5	38			
	Octanes	32.5	15	10.6	67			
Saura-Calixto and García-Raso[38] (1982)	Hexanes*	17.7	4	8	75	1.76	6.04	
	Heptanes*	31.3	8	7.7	88	1.58	2.55	
	Octanes*	19.8	17	9.4	71			
Saura-Calixto *et al.*[40] (1985)	Hexanes	53	4	10.3	75			
	Heptanes	205	7	24.7	71			
	Octanes	106	15	31.3	100			
	Nonanes	100	14	20.6	86			

* Present calculations.

** The corresponding I_{exp} were obtained on SF-96.

sum of structural and/or geometrical (intensive) factors and should have only a levelling contribution.

REFERENCES

1 A. W. Ladon, *Chromatographia*, 4 (1971) 171.
2 A. N. Korol, *Usp. Khim.*, 51 (1982) 1225.

3 M. V. Budahegyi, E. R. Lombosi, T. S. Lombosi, S. Y. Mészáros, Sz. Nyiredy, G. Tarján, I. Timár and J. M. Takács, *J. Chromatogr.*, 271 (1983) 213.
4 M. Randić, *J. Chromatogr.*, 161 (1978) 1.
5 P. Souter, *J. Chromatogr.*, 92 (1974) 231.
6 P. Souter, *J. Chromatogr. Sci.*, 12 (1974) 418.
7 P. Souter, *J. Chromatogr. Sci.*, 12 (1974) 423.
8 R. Vanheertum, *J. Chromatogr. Sci.*, 13 (1975) 150.
9 E. Matisová, J. Krupčík, P. Čellár and J. Garaj, *J. Chromatogr.*, 303 (1983) 151.
10 N. G. Johansen, L. S. Ettre and R. L. Miller, *J. Chromatogr.*, 256 (1982) 393.
11 J. D. Ramsey, T. D. Lee, M. D. Osselton and A. C. Moffat, *J. Chromatogr.*, 184 (1980) 185.
12 N. Dimov and D. Papazova, *Chromatographia*, 12 (1979) 443.
13 S. Brandt, *Statistical and Computational Methods in Data Analysis*, North Holland, Amsterdam, 1970, p. 118.
14 E. sz. Kováts, *Helv. Chim. Acta*, 41 (1958) 1915.
15 J. Bonastre and P. Grenier, *Bull. Soc. Chim. Fr.*, (1967) 1395.
16 *Selected Values of Properties of Hydrocarbons*, A.P.I. Research Project 44, Texas University, College Station, TX, 1966.
17 A. Tatevskii (Editor), *Fisiko-khimicheskie Svoistva Individualnikh Uglevodorodov*, Izdatelstvo Lenina, Moscow, 1960.
18 K. Altenburg, in P. Struppe (Editor), *Gas Chromatography 1968*, Akad. Verlag, Berlin, 1969, p. 1.
19 N. Dimov and D. Shopov, *J. Chromatogr.*, 44 (1969) 170.
20 N. Dimov and D. Shopov, *J. Chromatogr.*, 63 (1971) 223.
21 A. A. Martinov and M. S. Vigdergauz, *Neftekhimiya*, 10 (1970) 76.
22 W. E. Hammers and C. L. de Ligny, *Recl. Trav. Chim. Pays-Bas*, 90 (1971) 175.
23 P. G. Robinson and A. L. Odell, *J. Chromatogr.*, 57 (1971) 1.
24 G. Castello, M. Lunardelli and M. Berg, *J. Chromatogr.*, 76 (1973) 31.
25 J. E. Dubois and J.-R. Chrétien, *J. Chromatogr. Sci.*, 12 (1974) 811.
26 J.-R. Chrétien and J.-É. Dubois, *J. Chromatogr.*, 126 (1976) 171.
27 J. Takács, C. Szita and G. Tarján, *J. Chromatogr.*, 56 (1971) 1.
28 T. S. Lombosi, E. R. Lombosi, I. Bernát, Z. S. Bernoit, E. C. Takács and J. M. Takács, *J. Chromatogr.*, 119 (1976) 307.
29 N. Dimov, *J. Chromatogr.*, 119 (1976) 109.
30 G. I. Spivakovskii, A. L. Tishchenko, I. I. Zaslavskii and N. S. Wulfson, *J. Chromatogr.*, 144 (1977) 1.
31 V. G. Berezkin, *Neftekhimiya*, 1 (1961) 169.
32 A. N. Korol, *Chromatographia*, 8 (1975) 385.
33 A. N. Korol and L. Misjuk, *Teor. Eksp. Khim.*, 16 (1980) 792.
34 J. F. Buteiko and A. N. Korol, *Ukr. Khim. Zh., Russ. Ed.*, 6 (1968) 542.
35 A. N. Korol, *Chromatographia*, 8 (1976) 335.
36 D. Papazova, N. Dimov and D. Bonchev, *J. Chromatogr.*, 188 (1980) 297.
37 D. Bonchev and N. Trinaistich, *J. Chem. Phys.*, 67 (1977) 4557.
38 F. Saura-Calixto and A. García-Raso, *Chromatographia*, 15 (1982) 521.
39 A. Bondi, *J. Phys. Chem.*, 68 (1964) 441.
40 F. Saura-Calixto, A. García-Raso and I. García-Raso, *J. Chromatogr.*, 322 (1985) 35.
41 M. B. Evans, *J. Gas Chromatogr.*, 4 (1966) 1.
42 A. Matukuma, in C. L. A. Harbourne and R. Stock (Editors), *Gas Chromatography 1968*, Institute of Petroleum, London, 1969, p. 55.
43 R. H. Hively and R. E. Hinton, *J. Gas Chromatogr.*, 6 (1968) 203.
44 J. A. Rijks, *Ph.D. Thesis*, Eindhoven, 1973.
45 D. A. Tourres, *J. Chromatogr.*, 30 (1967) 357.