Journal of Chromatography, 144 (1977) 1–16

CHROM. 10,181

CALCULATION OF RETENTION INDICES OF COMPOUNDS FROM THEIR STRUCTURAL FORMULAE FOR COMBINED IDENTIFICATION BY GAS CHROMATOGRAPHY–MASS SPECTROMETRY

APPLICATION TO ALIPHATIC ALCOHOLS AND SATURATED HYDROCARBONS

G. I. SPIVAKOVSKII, A. I. TISHCHENKO, I. I. ZASLAVSKII and N. S. WULFSON

Scientific Research Institute of Organic Intermediate Products and Dyes, I Bolshaya Sadowaya, Block 4, Moscow (U.S.S.R.)

(First received August 10th, 1976; revised manuscript received April 15th, 1977)

SUMMARY

A method of choosing structural groups based on the types or "forms" of bonds present for calculation of Kováts retention indices using the additivity principle is given. The form of a chemical bond and its immediate surroundings reflects the valency state of all of the atoms included and is considered to introduce an independent contribution to the retention index. In accordance with this assumption, the combined treatment of the retention data of compounds of different classes is possible in order to calculate the partial contributions of several structural groups. The values obtained by combined processing of retention data for saturated hydrocarbons and aliphatic alcohols were in satisfactory agreement with the results calculated by separate treatment of the retention data for hydrocarbons. The mean-square error of the calculated values of retention indices with respect to the experimental values is 7 retention index units. The accuracy achieved permits the use of this method in combined gas chromatographic–mass spectrometric identification.

The possibility of calculating retention indices based on the structural formulae of compounds drastically reduces the number of possible structures suggested by mass spectrometric data. The calculation method presented here is suitable for computer calculations. The accuracy of the calculation is sufficient for the required purpose.

INTRODUCTION

The combined gas chromatographic–mass spectrometric (GC–MS) technique is very useful for analysing and identifying small amounts of substances in complex mixtures of organic compounds. The last decade has seen the rapid development of useful computer techniques for unique compound identification[1–5]. A number of methods for the computer treatment of mass spectrometric data have been used

successfully for the elucidation of the structural formulae of compounds that have been preliminarily resolved by GC. These methods include several computer-matching systems[1], pattern recognition methods[2,6] and many computer programs based on the structural generation technique[3–5]. With a few exceptions[7], none of these methods utilizes GC structural information, which can be obtained simultaneously in the same experiment. The GC behaviour of the compound of interest, however, can be used successfully for choosing the correct structural formula of a compound from alternative versions obtained from mass spectrometric analysis.

Unfortunately, no generally satisfactory methods for predicting retention behaviour based on a knowledge of two-dimensional structural formulae have been derived from rigorous theories of GC. For this reason, many attempts have been made to correlate empirically retention data, especially Kováts retention indices[8], with molecular structures[9–18]. In particular, elution characteristics have been predicted based on the additive principle[9,10]:

$$\log L = \Sigma \alpha_i x_i \tag{1}$$

where L is an elution characteristic of the compound (retention time, specific retention volume, retention index), α_i is the partial contribution of the ith structural group to the total $\log L$ value and x_i is number of ith structural groups in the compound determined from its structural formula.

Recently, many studies have been devoted to the problem of establishing relationships between the retention indices of compounds and their molecular structures. Extensive reviews on this problem have been published by Schomburg and Dielmann[19], Rohrschneider[20], Takács[21] and Haken[22]. Some of the results published in these papers appear to be applicable for identification purposes. The most interesting results were found by Berezkin and co-workers[9,10], Takács and co-workers[14,15], Sanz *et al.*[16] and Castello *et al.*[17], in which the additive contribution principle of atoms, bonds or groups was utilized. Using this principle, only structural features of compounds need to be known for the calculation of their retention characteristics.

In Berezkin and co-workers' studies[9,10], partial values of α_i were determined by use of eqn. 1 for branched-chain alkanes with known retention times and structural formulae, and the values found were used for the prediction of retention times for other alkanes. Castello *et al.*[17] described a method for the identification of branched-chain alkanes by calculation of their retention indices on the basis of molecular structure and eqn. 1. Retention indices have been calculated for branched-chain alkanes heavier than nonane, formed by γ-radiolysis[23], by using α_i terms calculated from experimental values for lower alkanes on the same liquid phase[17]. Similar principles were used by Sanz *et al.*[16]. The results obtained by Takács and co-workers were discussed by Vanheertum[24].

There are some problems in the application of the additive group contribution rules for predicting retention indices from molecular structures, including the accuracy that can be achieved by this technique and the universality of the α_i terms for different classes of compounds. Both problems are caused by interactions that occur between the molecules and/or their fragments in dilute solutions.

In general, the additive principle is not completely theoretically correct for the calculation of retention indices. The retention index is a linear function of the Gibbs'

free energy of solution[25,26] and can be represented as the sum of two values, namely, enthalpic and entropic terms, as follows:

$$\frac{100}{AR}\left[\frac{1}{T}(\Delta H_{\text{sol } x} - \Delta H_{\text{sol } n}) - (\Delta S_{\text{sol } x} - \Delta S_{\text{sol } n})\right] + 100\,n \qquad (2)$$

where

$H_{\text{sol } x}$, $H_{\text{sol } n}$ = enthalpy of solution of a substance, and an *n*-alkane with *n* carbon atoms, respectively;
$S_{\text{sol } x}$, $S_{\text{sol } n}$ = entropy of solution of a substance, and an *n*-alkane with *n* carbon atoms, respectively;
A = constant.

The entropy terms depend on the total configuration of the molecule and cannot be determined in terms of additivity[27]. However, the error introduced by the presence of non-additive terms should not be considerable when the entropy terms differ only slightly for a large number of investigated compounds. This is valid, for example, for slightly branched alkanes (with side-chains no larger than ethyl), for which the mean-square deviation of the entropy term is about 0.57 (ref. 28) (Table I).

Another source of inaccuracy is the lack of universality of the group contribution terms α_i for commonly considered atomic groups. The partial energy of interaction between the solvent molecules and a given structural group depends not only on its structural features but also on the effect of surrounding atoms and atomic groups of this molecule. The interactions between a given structural group and non-bonded

TABLE I

THERMODYNAMIC CHARACTERISTICS OF BRANCHED-CHAIN ALKANES AT 100°

Stationary phase: triacontane. Data from Parcher *et al.*[28].

No.	*Compound*	$-\Delta G^\circ_{sol}$ *(cal/mole)*	$-\Delta H^\circ_{sol}$ *(cal/mole)*	$-\Delta S^\circ_{sol\ x}$	$(\Delta S^\circ_{sol\ x} - \Delta S^\circ_{sol\ n})$
1	*n*-Pentane	1814	5830	10.77	—
2	*n*-Hexane	2398	6910	12.10	—
3	*n*-Heptane	2964	8010	13.53	—
4	*n*-Octane	3527	9080	14.90	—
5	2-Methylbutane	1671	5440	10.10	0.67
6	2-Methylpentane	2225	6610	11.75	0.35
7	2-Methylhexane	2775	7590	12.90	0.63
8	2-Methylheptane	3330	8640	14.23	0.67
9	3-Methylpentane	2318	6650	11.61	0.49
10	3-Methylhexane	2837	7660	12.93	0.60
11	3-Methylheptane	3378	8680	14.21	0.69
12	4-Methylheptane	3343	8670	14.28	0.62
13	2,2-Dimethylbutane	2054	6040	10.70	1.40
14	2,3-Dimethylbutane	2229	6380	11.13	0.97
15	2,2,3-Trimethylbutane	2662	7070	11.82	1.71
16	2,2-Dimethylpentane	2554	7010	11.95	1.58
17	2,3-Dimethylpentane	2827	7530	12.60	0.93
18	3,3-Dimethylpentane	2769	7300	12.15	1.38
19	2,2,4-Trimethylpentane	2931	7670	12.70	2.20

surrounding atoms in the same molecule determine the conformational state of the solute molecules and therefore should influence the α_i value. A similar effect is caused by the available functional groups. This phenomenon prevents the use of the same α_i value for the calculation of retention indices of compounds of different classes.

Fortunately, in most instances the influence of the surrounding atoms decreases with increase in the distance of the atoms from the structural groups and becomes negligible at some finite distance.

In this paper, an attempt is made to choose structural groups by taking into consideration the above remarks. The values of α_i for saturated hydrocarbons and aliphatic alcohols were determined by a combined treatment of experimental retention indices (with squalane as the stationary phase) by the least-squares technique for 32 alcohols[29] and 65 hydrocarbons[30].

The method has the following advantages:

(1) We suggest a formalized (physically founded, however) procedure for the choice of structural groups involving forms of bonds.

(2) The introduction of such a formalized procedure allows the suggested method to be applied to different classes of compounds.

(3) The suggested procedure allows one to choose only such forms of bonds to which linear independent structural coordinates correspond, which allows us to confine ourselves to a small number of forms of bonds and a small number of experimental data for the determination of partial contributions.

(4) The accuracy of the method, despite the small number of structural groups, it is not worse than that of other methods.

METHOD

The forms of bonds in the molecules of the classes of compounds investigated were taken as structural groups each of which introduces a definite contribution to the retention index. The term "form of bond" represents the structural group formed by two bonded atoms and by atoms in its immediate surroundings[31].

We have not taken into account varieties of bonds determined by geometrical configurations, or the influence of the non-immediate surroundings of the bound atoms. If these varieties of the bonds are taken into account, it could considerably increase the number of structural groups and, consequently, the number of α_i values to be determined.

It is necessary to choose from all possible forms of bonds those structural groups which correspond to linear independent structural co-ordinates of all sets of compounds of the given classes.

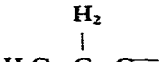
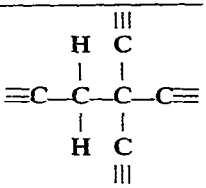
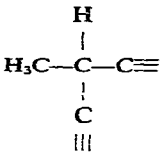
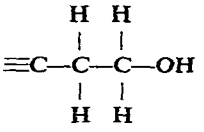
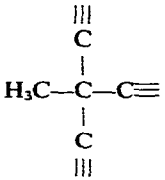
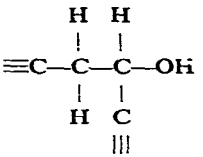
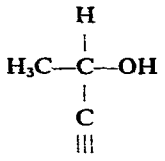
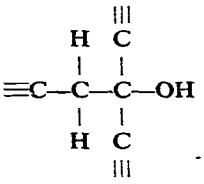
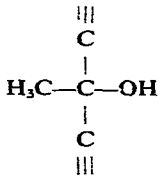
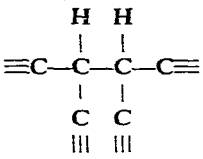
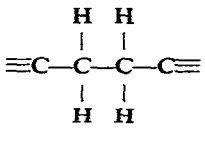
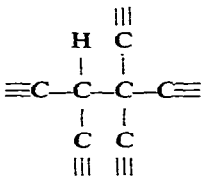
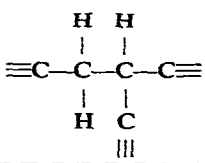
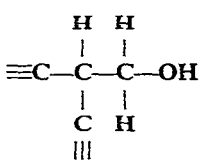
The α_i values determined in this way are assumed to be common for different classes of compounds (the non-additivity of entropy terms can be neglected). This assumption suggested a combined treatment of the experimental data for compounds of several classes for estimation of the α_i values. In this way, retention indices could be calculated and structural formulae could be elucidated by comparing the experimental value with those calculated for the proposed formulae.

Forms of bonds of saturated hydrocarbons (Nos. 1–3, 6–8, 12, 13, 17, 24–26) and aliphatic alcohols (Nos. 1–28) are given in Table II.

Only the bonds between the carbon atoms (Nos. 1–20) are linearly independent.

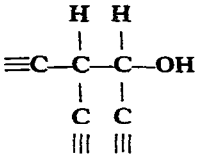
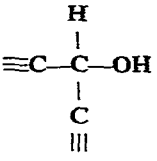
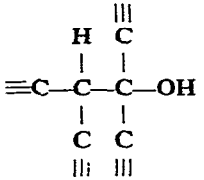
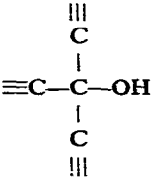
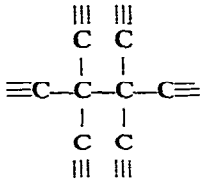
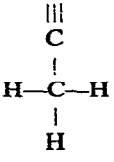
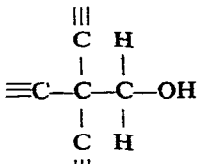
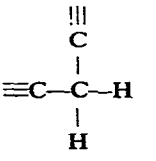
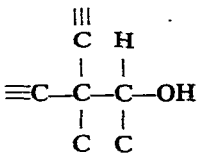
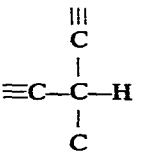
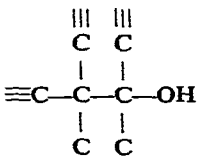
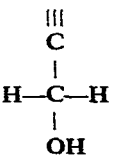
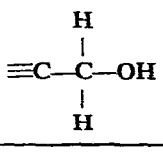
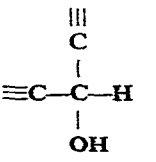
TABLE II

FORMS OF BONDS OF SATURATED HYDROCARBONS AND ALIPHATIC ALCOHOLS

No.	*Fragment*	*No.*	*Fragment*
1		8	
2		9	
3		10	
4		11	
5		12	
6		13	
7		14	

(Continued on p. 6)

TABLE II (*continued*)

No.	*Fragment*	*No.*	*Fragment*
15		22	
16		23	
17		24	
18		25	
19		26	
20		27	
21		28	

Other forms of bonds (C–H, C–OH) can be expressed linearly through C–C bonds by applying the following relationships:

$$x_{21} = x_9 + x_{14} + x_{18}$$

$$2x_{22} = x_4 + x_{10} + x_{15} + x_{19}$$

$$3x_{23} = x_5 + x_{11} + x_{16} + x_{20}$$

$$x_{24} = 3x_1 + 3x_2 + 3x_3 + 3x_4 + 3x_5$$

$$x_{25} = x_1 + 2x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11}$$

$$3x_{26} = x_2 + x_7 + 2x_{12} + x_{13} + x_{14} + x_{15} + x_{16}$$

$$x_{27} = 2x_9 + 2x_{14} + 2x_{18} \qquad (3)$$

$$x_{28} = x_{22}$$

Then it is considered that a relationship of the form of eqn. 1 is available:

$$I = \delta_0 + \Sigma\delta_i\, x_i \qquad (4)$$

where

I = retention index;
δ_i = partial contribution of the ith form of bonds to the retention index;
δ_0 = constant.

Substituting eqns. 3 into eqn. 4, we obtain

$$I = \alpha_0 + \sum_{i=1}^{20} \alpha_i\, x_i \qquad (5)$$

where

$$\alpha_0 = \delta_0$$

$$\alpha_1 = \delta_1 + (3\delta_{24} + \delta_{25})$$

$$\alpha_2 = \delta_2 + (3\delta_{24} + \tfrac{1}{3}\delta_{26})$$

$$\alpha_3 = \delta_3 + (3\delta_{24})$$

$$\alpha_4 = \delta_4 + (\tfrac{1}{2}\delta_{22} + 3\delta_{24} + \tfrac{1}{2}\delta_{28})$$

$$\alpha_5 = \delta_5 + (\tfrac{1}{3}\delta_{23} + 3\delta_{24})$$

$$\alpha_6 = \delta_6 + (2\delta_{25})$$

$$\alpha_7 = \delta_7 + (\delta_{25} + \tfrac{1}{3}\delta_{26})$$

$$\alpha_8 = \delta_8 + (\delta_{25})$$

$$\alpha_9 = \delta_9 + (\delta_{21} + \delta_{25} + 2\delta_{27})$$

$$\alpha_{10} = \delta_{10} + (\tfrac{1}{2}\delta_{22} + \delta_{25} + \tfrac{1}{2}\delta_{28})$$

$$\alpha_{11} = \delta_{11} + (\tfrac{1}{3}\delta_{23} + \delta_{25})$$

$$\alpha_{12} = \delta_{12} + (\tfrac{2}{3}\delta_{26})$$

$$\alpha_{13} = \delta_{13} + (\tfrac{1}{3}\delta_{26})$$

$$\alpha_{14} = \delta_{14} + (\delta_{21} + \tfrac{1}{3}\delta_{26} + 2\delta_{27})$$

$$\alpha_{15} = \delta_{15} + (\tfrac{1}{2}\delta_{22} + \tfrac{1}{3}\delta_{26} + \tfrac{1}{2}\delta_{28}) \qquad (6)$$

$$\alpha_{16} = \delta_{16} + (\tfrac{1}{3}\delta_{23} + \tfrac{1}{3}\delta_{26})$$

$$\alpha_{17} = \delta_{17}$$

$$\alpha_{18} = \delta_{18} + (\delta_{21} + 2\delta_{27})$$

$$\alpha_{19} = \delta_{19} + (\tfrac{1}{2}\delta_{22} + \tfrac{1}{2}\delta_{28})$$

$$\alpha_{20} = \delta_{20} + (\tfrac{1}{3}\delta_{23})$$

The values of α_i can be determined by solving eqn. 5 for compounds with known experimental retention indices. α_i are not the partial contributions of the respective forms of bonds (with the exception of $\alpha_{17} = \delta_{17}$) to the retention index, but reflect both the influence of the *i*th form of bonds and certain other bonds in accordance with eqn. 6. However, the knowledge of α_i makes it possible to calculate the retention index of a compound from its structural formula by using eqn. 5.

RESULTS AND DISCUSSION

Eqn. 5 for calculating α_i contained 97 separate equations (32 refer to aliphatic alcohols and 65 to saturated hydrocarbons) with 20 unknowns corresponding to 19 independent structural co-ordinates and the constant term α_0. The value of α_{20} could not be determined because of lack of experimental data. Retention indices for squalane at 100° were used. The values of retention indices for alcohols and hydrocarbons were taken from McReynolds[29] and Tourres[30], respectively. As the retention indices of hydrocarbons given by Tourres[30] are presented only at 50° and 70°, linear extrapolation[32] was effected and the retention indices at 100° were calculated. As was shown earlier[32–34], such an extrapolation is admissible. The error did not exceed 1 retention index unit. In this instance the error in α_i will be less than unity.

TABLE III

RESULTS OF COMBINED TREATMENT OF EXPERIMENTAL DATA FOR ALIPHATIC ALCOHOLS AND SATURATED HYDROCARBONS

Standard (root-mean-square) deviation $S = 6.88$; Correlation coefficient $R = 0.998$. Relative deviation $\lambda = \sqrt{\frac{\Sigma\Delta^2}{\Sigma I_{exp}^2}} = 0.008$. Coefficients: $\alpha_0 = 345.14$; $\alpha_1 = -9.36$; $\alpha_2 = 12.17$; $\alpha_3 = 19.79$; $\alpha_4 = 73.06$; $\alpha_5 = 55.14$; $\alpha_6 = 94.49$; $\alpha_7 = 120.66$; $\alpha_8 = 146.63$; $\alpha_9 = 214.61$; $\alpha_{10} = 174.64$; $\alpha_{11} = 175.54$; $\alpha_{12} = 174.64$; $\alpha_{13} = 216.67$; $\alpha_{14} = 232.55$; $\alpha_{15} = 215.06$; $\alpha_{16} = 235.25$; $\alpha_{17} = 279.87$; $\alpha_{18} = 236.00$; $\alpha_{19} = 238.43$.

No.	*Compound*	I_{exp}	I_{calc}	*Difference* ($\Delta = I_{exp} - I_{calc}$)
1	1-Pentanol	731	739.28	−8.28
2	2-Pentanol	670	677.97	−7.97
3	3-Pentanol	673	675.70	−2.70
4	2-Methyl-1-butanol	702	701.15	−0.85
5	2-Methyl-2-butanol	621	621.59	−0.59
6	3-Methyl-2-butanol	652	657.59	−5.59
7	2,2-Dimethyl-1-propanol	637	640.52	−3.52
8	1-Hexanol	330	833.77	−3.77
9	2-Hexanol	771	772.46	−1.46
10	3-Hexanol	771	770.19	0.81
11	2-Methyl-1-pentanol	797	795.63	1.35
12	3-Methyl-1-pentanol	805	803.78	1.22
13	4-Methyl-1-pentanol	798	799.14	−1.14
14	2-Methyl-2-pentanol	714	716.09	−2.09
15	3-Methyl-2-pentanol	772	756.73	15.27
16	4-Methyl-2-pentanol	732	737.83	−5.83
17	2-Methyl-2-pentanol	752	74[illegible].81	2.19
18	3-Methyl-3-pentanol	738	732.64	5.36
19	2-Ethyl-1-butanol	806	800.28	5.72
20	2,2-Dimethyl-1-butanol	763	758.00	5.00
21	3,3-Dimethyl-2-butanol	716	716.00	0
22	1-Heptanol	935	928.27	6.73
23	2-Heptanol	868	866.96	1.04
24	3-Heptanol	868	864.68	3.32
25	4-Heptanol	867	864.68	2.32
26	2,2-Dimethyl-1-pentanol	851	852.49	−1.49
27	2,4-Dimethyl-3-pentanol	818	823.93	−5.93
28	3-Ethyl-3-pentanol	841	843.68	−2.68
29	1-Octanol	1028	1022.76	5.24
30	2-Octanol	966	961.45	4.55
31	2-Ethyl-1-hexanol	992	989.27	2.73
32	2-Ethyl-4-methyl-1-pentanol	944	954.64	−10.64
33	2,2-Dimethylbutane	541	541.78	−0.78
34	2,3-Dimethylbutane	573	568.45	4.55
35	2-Methylpentane	570	575.27	−5.27
36	3-Methylpentane	587	579.91	7.09
37	*n*-Hexane	600	609.90	−9.9
38	2,2-Dimethylpentane	629	636.28	−7.28
39	2,4-Dimethylpentane	631	635.13	−4.13
40	2,2,3-Trimethylbutane	648	645.52	2.48
41	3,3-Dimethylpentane	666	659.26	6.74
42	2-Methylhexane	667	669.76	−2.76
43	2,3-Dimethylpentane	677	667.58	9.42

(Continued on p. 10)

TABLE III (*continued*)

No.	*Compound*	I_{exp}	I_{calc}	*Difference* ($\Delta = I_{exp} - I_{calc}$)
44	3-Methylhexane	678	674.40	3.6
45	3-Ethylpentane	689	679.04	9.96
46	*n*-Heptane	700	704.39	−4.39
47	2,2,4-Trimethylpentane	696	696.14	−0.14
48	2,2-Dimethylhexane	722	730.77	−8.77
49	2,2,3,3-Tetramethylbutane	739	743.76	−4.76
50	2,5-Dimethylhexane	730	729.62	0.38
51	2,4-Dimethylhexane	735	734.26	0.74
52	2,2,3-Trimethylpentane	746	744.65	1.35
53	2,3,4-Trimethylpentane	760	755.26	4.74
54	3,3-Dimethylhexane	750	753.75	−3.75
55	2,3,3-Trimethylpentane	771	763	8
56	2,3-Dimethylhexane	764	762.08	1.92
57	2-Methylheptane	766	764.26	1.74
58	4-Methylheptane	769	768.89	0.11
59	3,4-Dimethylhexane	777	766.71	10.29
60	3-Methylheptane	775	768.89	6.11
61	3-Ethylhexane	775	773.53	1.47
62	*n*-Octane	800	798.89	1.11
63	2,2,4,4-Tetramethylpentane	783	757.15	25.85
64	2,2,5-Trimethylhexane	780	790.63	−10.63
65	2,2,4-Trimethylhexane	796	795.27	0.73
66	2,4,4-Trimethylhexane	816	813.62	2.38
67	2,3,5-Trimethylhexane	817	821.94	−4.94
68	2,2,3,4-Tetramethylpentane	831	832.33	−1.33
69	2,2-Dimethylheptane	818	825.26	−7.26
70	2,4-Dimethylheptane	823	828.76	−5.76
71	2,2,3-Trimethylhexane	830	839.15	−9.15
72	2-Methyl-4-ethylhexane	827	833.39	−6.39
73	2,2-Dimethyl-3-ethylpentane	833	843.78	−10.78
74	2,6-Dimethylheptane	828	824.12	3.88
75	4,4-Dimethylheptane	833	833	0
76	2,5-Dimethylheptane	834	828.76	5.24
77	3,5-Dimethylheptane	836	833.39	2.61
78	3,3-Dimethylheptane	842	848.25	−6.25
79	2,4-Dimethyl-3-ethylpentane	845	854.39	−9.39
80	2,3,3-Trimethylhexane	849	857.49	−8.49
81	2-Methyl-3-ethylhexane	849	861.21	−12.21
82	2,2,3,3-Tetramethylpentane	866	861.24	4.76
83	2,3,4-Trimethylhexane	855	854.39	0.61
84	3,3,4-Trimethylhexane	864	862.13	1.87
85	3-Methyl-4-ethylhexane	863	865.84	−2.84
86	3-Methyl-3-ethylhexane	862	871.23	−9.23
87	4-Ethylheptane	860	868.02	−8.02
88	2,3-Dimethylheptane	858	856.57	1.43
89	2,3,3,4-Tetramethylpentane	872	866.74	5.26
90	3,4-Dimethylheptane	863	861.21	1.79
91	4-Methyloctane	864	863.39	0.61
92	2-Methyloctane	865	858.75	6.25
93	3-Ethylheptane	869	868.02	0.98
94	3-Methyloctane	872	863.39	8.61
95	2,3-Dimethyl-3-ethylpentane	886	880.48	5.52
96	3,3-Diethylpentane	892	894.21	−2.21
97	*n*-Nonane	900	893.38	6.62

Eqn. 5 was solved with a 220M computer using the least-squares technique. The results of calculations of the α_i values ($i = 0, 1, 2, 3, \ldots, 18, 19$) are given in Table III, together with a comparison of the calculated and experimental values of I (I_{calc} and I_{exp}). The mean-square deviation was 6.8 retention index units, which corresponds approximately to the deviation quoted by Castello *et al.*[17] for a hydrocarbon system only. The deviation obtained was slightly greater than the 5.68 units achieved by Gassiot *et al.*[18] for sterol acetate on SE-30 using an empirical quantum chemical approach.

The relative deviation (λ) was calculated using the equation

$$\lambda = \sqrt{\frac{\Sigma\Delta^2}{\Sigma I_{exp}{}^2}}$$

TABLE IV

RESULTS OF TREATMENT OF EXPERIMENTAL DATA FOR ALIPHATIC ALCOHOLS

$S = 5.44$; $R = 0.998$; $\lambda = 0.005$. $\alpha_0 = 330.93$; $\alpha_1 = 178.23$; $\alpha_2 = 132.38$; $\alpha_3 = 102.02$; $\alpha_4 = 79.00$; $\alpha_5 = 59.66$; $\alpha_6 = 95.39$; $\alpha_7 = 61.54$; $\alpha_8 = 46.10$; $\alpha_9 = 39.62$; $\alpha_{10} = -6.07$; $\alpha_{11} = -7.28$; $\alpha_{14} = -6.15$; $\alpha_{15} = -19.07$.

No.	*Compound*	I_{exp}	I_{calc}	$\Delta = I_{exp} - I_{calc}$
1	1-Pentanol	731	739.55	−8.55
2	2-Pentanol	670	677.47	−7.47
3	3-Pentanol	673	675.25	−2.25
4	2-Methyl-1-butanol	702	696.92	5.08
5	2-Methyl-2-butanol	621	621.19	−0.19
6	3-Methyl-2-butanol	652	655.62	−3.62
7	2,2-Dimethyl-1-propanol	637	637.00	0
8	1-Hexanol	830	834.94	−4.94
9	2-Hexanol	771	772.86	−1.86
10	3-Hexanol	771	770.63	0.37
11	2-Methyl-1-pentanol	797	792.31	4.69
12	3-Methyl-1-pentanol	805	804.23	0.77
13	4-Methyl-1-pentanol	798	792.23	5.77
14	2-Methyl-2-pentanol	714	716.57	−2.57
15	3-Methyl-2-pentanol	772	763.00	9.00
16	4-Methyl-2-pentanol	732	730.16	1.84
17	2-Methyl-3-pentanol	752	748.77	3.23
18	3-Methyl-3-pentanol	738	732.48	5.52
19	2-Ethyl-1-butanol	806	804.31	1.69
20	2,2-Dimethyl-1-butanol	763	759.31	3.69
21	3,3-Dimethyl-2-butanol	716	716.00	0
22	1-Heptanol	935	930.33	4.67
23	2-Heptanol	868	868.25	−0.25
24	3-Heptanol	868	866.02	1.98
25	4-Heptanol	867	866.02	0.98
26	2,2-Dimethyl-1-pentanol	851	854.69	−3.69
27	2,4-Dimethyl-3-pentanol	818	822.30	−4.30
28	3-Ethyl-3-pentanol	841	843.76	−2.76
29	1-Octanol	1028	1025.71	2.29
30	2-Octanol	966	963.64	2.36
31	2-Ethyl-1-hexanol	992	995.08	−3.08
32	2-Ethyl-4-methyl-1-pentanol	944	952.38	−8.38

and its value was 0.008. A correlation coefficient of the greatest significance was obtained (0.998).

One molecule (2,2,4,4-tetramethylpentane) gave a significant error. The extensive branching in this molecule caused considerable interactions of non-bonded atoms located further than a distance of two atoms, which are not taken into account in the method employed.

In addition to combined processing of the experimental data, separate data processing for alcohols and hydrocarbons was accomplished, and the results of the calculations are given in Tables IV and V. Good agreement was achieved for α_i values determined both by separate treatment of hydrocarbon data and by the combined treatment. Unsatisfactory agreement for α_i values for alcohols was apparently caused by the smaller number of alcohol molecules (32, compared with 65 for hydrocarbons).

The present method of choosing structural groups ensures a 1:1 correspondence between the set of molecular structures and the corresponding sets of structural co-ordinates. Therefore, no distinctions could be made between various structures with similar forms of bonds. For example, the same value of I was calculated for 3-methylheptane and 4-methylheptane. This difficulty can be avoided by the introduction of additional independent structural co-ordinates. The influence of the additional factors of unknown nature was included in the term α_0, and this term was assumed to be constant for all compounds in a given sample. This assumption is not completely correct, but Table III shows that the error introduced by this approximation is of the order of only a few retention index units.

In general, the agreement obtained between experimental and calculated values of I is not satisfactory for universal compound identification based on GC data only. The difference in I values for some compounds may be smaller than the difference in the values obtained by calculation. Therefore, the additional independent method of structural elucidations is necessary. The mass spectrometric analysis may be valuable for the achievement of a unique result.

TABLE V

RESULTS OF TREATMENT OF EXPERIMENTAL DATA FOR SATURATED HYDROCARBONS

$S = 7.12$; $R = 0.997$; $\lambda = 0.008$. $\alpha_0 = 347.54$; $\alpha_1 = -9.29$; $\alpha_2 = 12.57$; $\alpha_3 = 19.67$; $\alpha_6 = 93.6$; $\alpha_7 = 119.86$; $\alpha_8 = 145.81$; $\alpha_{12} = 172.76$; $\alpha_{13} = 215.03$; $\alpha_{17} = 278.51$.

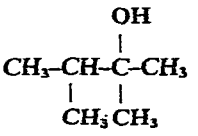
No.	*Compound*	I_{exp}	I_{calc}	$\Delta = I_{exp} - I_{calc}$
1	2,2-Dimethylbutane	541	543.08	−2.08
2	2,3-Dimethylbutane	573	570.61	2.39
3	2-Methylpentane	570	576.92	−6.92
4	3-Methylpentane	587	581.30	5.70
5	*n*-Hexane	600	609.87	−9.87
6	2,2-Dimethylpentane	629	636.71	−7.71
7	2,4-Dimethylpentane	631	637.60	−6.60
8	2,2,3-Trimethylbutane	648	646.73	1.27
9	3,3-Dimethylpentane	666	659.92	6.08
10	2-Methylhexane	667	670.55	−3.55
11	2,3-Dimethylpentane	677	668.63	8.37
12	3-Methylhexane	678	674.94	3.06

TABLE V (*continued*)

No.	*Compound*	I_{exp}	I_{calc}	$\Delta = I_{exp} - I_{calc}$
13	3-Ethylpentane	689	679.32	9.68
14	*n*-Heptane	700	703.51	−3.51
15	2,2,4-Trimethylpentane	696	697.39	−1.39
16	2,2-Dimethylhexane	722	730.34	−8.34
17	2,2,3,3-Tetramethylbutane	739	744.08	−5.08
18	2,5-Dimethylhexane	730	731.24	−1.24
19	2,4-Dimethylhexane	735	735.62	−0.62
20	2,2,3-Trimethylpentane	746	744.75	1.25
21	2,3,4-Trimethylpentane	760	755.95	4.05
22	3,3-Dimethylhexane	750	753.56	−3.56
23	2,3,3-Trimethylpentane	771	763.58	7.42
24	2,3-Dimethylhexane	764	762.26	1.74
25	2-Methylheptane	766	764.19	1.81
26	4-Methylheptane	769	768.57	−0.43
27	3,4-Dimethylhexane	777	766.64	10.36
28	3-Methylheptane	775	768.57	6.43
29	3-Ethylhexane	775	772.95	2.05
30	*n*-Octane	800	797.14	2.86
31	2,2,4,4-Tetramethylpentane	783	757.18	25.82
32	2,2,5-Trimethylhexane	780	791.02	−11.02
33	2,2,4-Trimethylhexane	796	795.41	0.59
34	2,4,4-Trimethylhexane	816	814.24	1.76
35	2,3,5-Trimethylhexane	817	822.94	−5.94
36	2,2,3,4-Tetramethylpentane	831	832.07	−1.07
37	2,2-Dimethylheptane	818	832.98	−5.98
38	2,4-Dimethylheptane	823	829.25	−6.25
39	2,2,3-Trimethylhexane	830	838.38	−8.38
40	2-Methyl-4-ethylhexane	827	833.64	−6.64
41	2,2-Dimethyl-3-ethylpentane	833	842.77	−9.77
42	2,6-Dimethylheptane	828	824.87	3.13
43	4,4-Dimethylheptane	833	833.00	0
44	2,5-Dimethylheptane	834	829.25	4.75
45	3,5-Dimethylheptane	836	833.64	2.36
46	3,3-Dimethylheptane	842	847.19	−5.19
47	2,4-Dimethyl-3-ethylpentane	845	853.96	−8.96
48	2,3,3-Trimethylhexane	849	857.21	−8.21
49	2-Methyl-3-ethylhexane	849	860.28	−11.28
50	2,2,3,3-Tetramethylpentane	866	860.92	5.08
51	2,3,4-Trimethylhexane	855	853.86	1.14
52	3,3,4-Trimethylhexane	864	861.60	2.40
53	3-Methyl-4-ethylhexane	863	864.66	−1.66
54	3-Methyl-3-ethylhexane	862	870.41	−8.41
55	4-Ethylheptane	860	866.59	−6.59
56	2,3-Dimethylheptane	856	855.89	2.11
57	2,3,3,4-Tetramethylpentane	872	867.24	4.76
58	3,4-Dimethylheptane	863	860.28	2.72
59	4-Methyloctane	864	862.20	1.80
60	2-Methyloctane	865	857.82	7.18
61	3-Ethylheptane	869	866.69	2.41
62	3-Methyloctane	872	862.20	9.80
63	2,3-Dimethyl-3-ethylpentane	886	880.43	5.57
64	3,3-Diethylpentane	892	893.62	−1.62
65	*n*-Nonane	900	890.77	9.23

Examples of the calculation of the retention indices of 2,3-dimethyl-2-butanol, 2-methyl-3-ethylpentane and 3-methyl-3-ethylpentane from their structural formulae are given below (for squalane at 100°).

(1) 2,3-Dimethyl-2-butanol:

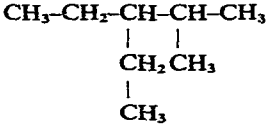


In this compound there are C–C bonds of the following forms (see Table II):

No. 2 —two bonds, $x_2 = 2$;
No. 5 —two bonds, $x_5 = 2$;
No. 16 —one bond, $x_{16} = 1$.

According to Table III, we find the values of α_0, α_2, α_5 and α_{16} and determine the retention index of 2,3-dimethyl-2-butanol as $I = \alpha_0 + \alpha_2 x_2 + \alpha_5 x_5 + \alpha_{16} x_{16} = 345.14 + 12.17 \cdot 2 + 55.14 \cdot 2 + 235.25 = 715.01$. The experimental value[29] is 715.

(2) 2-Methyl-3-ethylpentane:

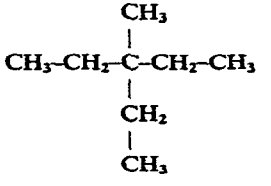


In this compound there are C–C bonds of the following forms (Table II):

No. 1 —two bonds, $x_1 = 2$,
No. 2 —two bonds, $x_2 = 2$,
No. 7 —two bonds, $x_7 = 2$,
No. 12 —one bond, $x_{12} = 1$.

Hence $I = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_7 x_7 + \alpha_{12} x_{12} = 345.14 - 9.36 \cdot 2 + 12.17 \cdot 2 + 120.66 \cdot 2 + 174.64 = 766.72$. The experimental value of the retention index at 100° was obtained by extrapolation of its values at 40°, 50° and 70°[33] and was found to be 768.

(3) 3-Methyl-3-ethylpentane:



In this compound there are C–C bonds of the following forms (Table III):

No. 1 —three bonds, $x_1 = 3$,
No. 3 —one bond, $x_3 = 1$,
No. 8 —three bonds, $x_8 = 3$.

Hence $I = \alpha_0 + \alpha_1 x_1 + \alpha_3 x_3 + \alpha_8 x_8 = 345.14 - 9.36 \cdot 3 + 19.79 + 146.63 \cdot 3 =$

776.74. The experimental value of the retention index at 100° was obtained by extrapolation of its values at 40°, 50° and 70°[33] and was found to be 784.

CONCLUSION

The suggested method is characterized by universality and a clear physical meaning and can be used with a smaller number of experimental data than other methods. The accuracy is acceptable for GLC–MS identification of compounds.

ACKNOWLEDGEMENTS

We thank Professor V. G. Berezkin for valuable comments and discussion of the results of this work.

REFERENCES

1 R. G. Ridley, in G. R. Waller (Editor), *Biochemical Application of Mass Spectrometry*, Wiley-Interscience, New York, 1972, p. 177.
2 T. L. Isenhour and P. C. Jurs, *Anal. Chem.*, 43 (1971) 20A.
3 J. Lederberg, in G. R. Waller (Editor), *Biochemical Application of Mass Spectrometry*, Wiley-Interscience, New York, 1972, p. 193.
4 A. Buchs, A. B. Delfino, A. M. Duffield, C. Djerassi, B. G. Buchmann, E. A. Feigenbaum and J. Lederberg, *Helv. Chim. Acta*, 53 (1970) 1394.
5 R. E. Garhart, D. H. Smith, H. Brown and C. Djerassi, *J. Amer. Chem. Soc.*, 97 (1975) 5755.
6 J. Stonham, I. Aleksander, M. Camp, W. T. Pike and M. A. Shaw, *Anal. Chem.*, 47 (1975) 1817.
7 H. Nau and K. Biemann, *Anal. Lett.*, 6 (1973) 1071.
8 E. Kováts, *Helv. Chim. Acta*, 41 (1958) 1915.
9 V. G. Berezkin, *Neftekhimiya*, 1 (1961) 169.
10 V. G. Berezkin and V. S. Kruglikova, *Neftekhimiya*, 2 (1962) 845.
11 G. J. Pierotti, C. H. Deal, E. L. Derr and P. E. Porter, *J. Amer. Chem. Soc.*, 78 (1956) 13 and 2989.
12 G. Schomburg and D. Henneberg, *Chromatographia*, 1 (1968) 23.
13 G. Schomburg and D. Henneberg, in C. L. A. Harbourn (Editor), *Gas Chromatography 1968*, Institute of Petroleum, London, pp. 45–54.
14 C. Káplár, C. Szita, J. M. Takács and G. Tarján, *J. Chromatogr.*, 65 (1972) 115.
15 J. M. Takács, E. Kocsi, E. Garamvölgyi, E. Eckhart, T. Lombosi, Sz. Nyiredy, Jr., I. Borbély and Gy. Krasznai, *J. Chromatogr.*, 81 (1973) 1.
16 J. Sanz, J. Calderon and M. V. Dabrio, *An. Quim.*, 71 (1975) 611.
17 G. Castello, M. Lunardelli and M. Berg, *J. Chromatogr.*, 76 (1973) 31.
18 M. Gassiot, E. Fernandez, G. Firpo, R. Carbó and M. Martin, *J. Chromatogr.*, 108 (1975) 337.
19 G. Schomburg and G. Dielmann, *J. Chromatogr. Sci.*, 11 (1973) 151.
20 L. Rohrschneider, *J. Chromatogr. Sci.*, 11 (1973) 160.
21 J. M. Takács, *J. Chromatogr. Sci.*, 11 (1973) 210.
22 J. K. Haken, *J. Chromatogr. Sci.*, 11 (1973) 144.
23 G. Castello, F. Grandi and S. Munari, *Radiat. Res.*, 45 (1971) 399 and 58 (1974) 176.
24 R. Vanheertum, *J. Chromatogr. Sci.*, 13 (1975) 150.
25 E. Kováts and P. B. Weiss, *Ber. Bunsenges. Phys. Chem.*, 69 (1965) 812.
26 B. A. Rudenko, S. S. Yufit, J. A. Krasnaya and V. F. Kucherov, *Dokl. Acad. Nauk SSSR*, 188 (1969) 156; *Izv. Akad. Nauk SSSR*, No. 10 (1970) 2199.
27 A. N. Korol, *Chromatographia*, 8 (1975) 385.
28 J. F. Parcher, P. H. Weiner, C. L. Hussey and T. N. Westlake, *J. Chem. Eng. Data*, 20 (1975) 145.
29 W. C. McReynolds, *Gas Chromatographic Retention Data*, Preston Technical Abstracts Co., Evanston, Ill., 1966.

30 D. A. Tourres, *J. Chromatogr.*, 30 (1967) 357.
31 V. M. Tatevskii, *Klassitcheskaya Teoriya Stroeniya Molecul i Kvantovaja Mechanika*, Khimiya, Moscow, 1973.
32 L. S. Ettre, *Anal. Chem.*, 36 (1964) 31A.
33 R. A. Hively and R. E. Hinton, *J. Gas Chromatogr.*, 6 (1968) 203.
34 *J. Gas Chromatogr.*, 4 (1966) 1.