Journal of Chromatography, 349 (1985) 385–393
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 657

NEW METHODS OF INTERPRETATION OF CHROMATOGRAPHIC–MASS SPECTROMETRIC DATA

B. V. IOFFE* and I. G. ZENKEVICH
Chemistry Department, Leningrad State University, 199164 Leningrad (U.S.S.R.)

SUMMARY

It is proposed to carry out the group identification of organic substances according to low-resolution mass spectra on the basis of the concept of a comparison of numbers of modulus 14. The concept of homologous groups is introduced. These groups are a total set of homologous series, the terms of which have mass numbers comparable in modulus 14. The numbers of homologous groups of molecular and fragment ions may serve as important diagnostic features for group identification according to the main peaks of the mass spectrum.

The reliability and unequivocal character of group identification increase considerably if the total intensity of the peaks of each group is taken into account and averaged spectra of ion series are used for the characterization of classes of compounds. When the spectra of ion series are compared, it is proposed to take into account standard deviations of peak intensities and to use as an additional classification feature homologous increments of retention indices calculated on the basis of the quatuordecimal system of calculus of mass numbers.

INTRODUCTION

The purpose of this paper is to consider new approaches to the interpretation of chromatographic–mass spectrometric data developed by the authors from the beginning of the 1980s and published only in Russian publications[1–7]. These methods of group identification of organic compounds according to their mass spectra are based on the concept of the comparison of natural numbers of modulus 14, or, alternatively, the application of the quatuordecimal system of calculus of mass numbers. This idea makes it possible to order and markedly simplify the procedure of decoding mass spectra by the introduction of the concept of homologous groups as new taxonomic categories of organic substances and homologous increments of additive properties, particularly retention indices. Moreover, it becomes possible to obtain simple algorithms for automatic processing of chromatographic–mass spectrometric data with the aid of a microcomputer with a limited operating memory (portable personal computers).

The entire array of experimental data used in the development of these methods for the interpretation of mass spectra covers about 7000 compounds, the

0021-9673/85/$03.30

spectra of which have been published[8–10], and approximately 1500 compounds studied by the authors over the last 12 years.

HOMOLOGOUS GROUPS OF ORGANIC COMPOUNDS, IONS AND RADICALS

The modern classification of organic compounds is based on a system of homologous series with a homologous difference CH_2, which corresponds to 14 atomic mass units. Hence, according to definition, the molecular weights of the terms of any homologous series are comparable in modulus 14 and may be represented in the general form by the equation

$$M = 14x + y \tag{1}$$

where y is a certain integer characteristic of a given series and x is the natural number determining the location of a compound in the given series. Eqn. 1 is equivalent to the presentation of the number M in a quatuordecimal system of calculus as the totality of the number of units of the lowest digit y and that of the second digit x, which may be written in the form $(x;y)$. The value of x is found as the integer quotient obtained when M is divided by 14, and y is the remainder.

The parameter y may be the same for many homologous series and it is reasonable to combine them into a single homologous group. Hence, all organic compounds may be divided into fourteen large classification units, each of which contains all the particles the mass numbers of which in the terms and symbolism of the number theory[11] form classes of modulus 14 residues:

$$M_y \equiv y(\mathrm{mod}14) \tag{2}$$

The totality of homologous series (classes of compounds) combined into a given homologous group y is formed in accordance with the mass and number of atoms composing a particle, so that compounds of different degrees of unsaturation and with different functions are in the same group. Table I gives an example of the distribution of several of the simplest classes of organic compounds in fourteen homologous groups. In the limits of each series, the parameter x determines uniquely the number of carbon atoms, n, in a particle.

According to the number theory[11], the comparisons for the same modulus may be added and multiplied just as equalities. Hence, the number of a homologous group of a compound is a simple sum of the numbers of groups of fragments composing the molecule:

$$y_M = \sum_i y_i \tag{3}$$

and for the derivatives RX the numbers of homologous groups are determined by the parameters y of the initial compound RH and the substituent X:

$$y_{RX} = y_{RH} + y_X - 1 \tag{4}$$

TABLE I

DISTRIBUTION OF THE SIMPLEST CLASSES OF ORGANIC COMPOUNDS IN HOMOLOGOUS GROUPS

Homologous group y	*Common formulae*	*Some classes of organic compounds*
0	C_nH_{2n} ($n = x$)	Alkenes, cycloalkanes
	$C_nH_{2n-2}O$ ($n = x - 1$)	Unsaturated carbonyl compounds
	$C_nH_{2n-4}O_2$ ($n = x - 2$)	Furyl alcohols and their ethers
	$C_nH_{2n}N_2$ ($n = x - 2$)	Aldazines, ketazines
	$C_nH_{2n-4}S$ ($n = x - 2$)	Thiophenes
1	$C_nH_{2n+1}N$ ($n = x - 1$)	Unsaturated amines
	$C_nH_{2n+1}N_3$ ($n = x - 3$)	Triazolines, azides
	$C_nH_{2n-1}NO$ ($n = x - 2$)	Lactams, cyanoalcohols
2	C_nH_{2n+2} ($n = x$)	Alkanes
	C_nH_{2n-12} ($n = x + 1$)	Naphthalenes, azulenes
	$C_nH_{2n}O$ ($n = x - 1$)	Aldehydes, ketones
	$C_nH_{2n+2}N_2$ ($n = x - 2$)	Diimides, alkyl hydrazones
	$C_nH_{2n+1}I$ ($n = x - 9$)	Alkyl iodides
3	$C_nH_{2n+3}N$ ($n = x - 1$)	Amines
	$C_nH_{2n+3}N_3$ ($n = x - 3$)	Hydrazidines
	$C_nH_{2n+1}NO$ ($n = x - 2$)	Amides, oximes and their ethers
4	C_nH_{2n-10} ($n = x + 1$)	Indenes, arylacetylenes, etc.
	$C_nH_{2n+2}O$ ($n = x - 1$)	Alcohols, ethers
	$C_nH_{2n}O_2$ ($n = x - 2$)	Esters, dioxanes, dioxolanes
	$C_nH_{2n+4}N_2$ ($n = x - 2$)	Diamines, hydrazines
	$C_nH_{2n+2}N_2O$ ($n = x - 3$)	Nitrosamines, hydrazides
5	$C_nH_{2n+5}N_3$ ($n = x - 3$)	Aminohydrazines
	$C_nH_{2n-9}N$ ($n = x$)	Indoles, etc.
	$C_nH_{2n+3}NO$ ($n = x - 2$)	Hydroxylamines
6	C_nH_{2n-8} ($n = x + 1$)	Styrenes, indanes, etc.
	$C_nH_{2n+2}O_2$ ($n = x - 2$)	Glycols, acetals, peroxides
	$C_nH_{2n-8}N_2$ ($n = x - 1$)	Arylpyrazolines
7	$C_nH_{2n-7}N$ ($n = x$)	Aminostyrenes
	$C_nH_{2n+1}NO_3$ ($n = x - 4$)	Alkyl nitrates
8	C_nH_{2n-6} ($n = x + 1$)	Arenes, etc.
	$C_nH_{2n-8}O$ ($n = x$)	Arylcarbonyl compounds
	$C_nH_{2n-6}N_2$ ($n = x - 1$)	Arylalkyldiimides
9	$C_nH_{2n-5}N$ ($n = x$)	Anilines, pyridines
10	C_nH_{2n-4} ($n = x + 1$)	All types of hydrocarbons with double-bond equivalent 3
	$C_nH_{2n-6}O$ ($n = x$)	Phenols, phenyl ethers
	$C_nH_{2n-8}O_2$ ($n = x - 1$)	Arylcarboxylic acids, phenyl esters
	$C_nH_{2n-4}N_2$ ($n = x - 1$)	Pyrazines
11	$C_nH_{2n-3}N$ ($n = x$)	Alkylpyrroles
12	C_nH_{2n-2} ($n = x + 1$)	Acetylenes, dienes, etc.
	$C_nH_{2n-4}O$ ($n = x$)	Furans
	$C_nH_{2n-2}N_2$ ($n = x - 1$)	Pyrazoles
13	$C_nH_{2n-1}N$ ($n = x$)	Alkyl nitriles
	$C_nH_{2n-1}N_3$ ($n = x - 2$)	Triazoles
	$C_nH_{2n-3}NO$ ($n = x - 1$)	Oxazoles, isoxazoles

These rules make it possible to reduce considerably the classification tables and to exclude from them the derivatives of polyisotopoic elements which can easily be determined at the beginning of the interpretation of the spectra according to the contour of polyisotopic peaks.

The unification into a single group of series with different extents of unsaturation, different functions and different elemental compositions greatly facilitates the group identification of the compounds under analysis, not only according to additional data obtained by other physical methods but also according to the main peaks of the mass spectra or the spectra of ion series.

IDENTIFICATION ACCORDING TO HOMOLOGOUS GROUPS OF THE MAIN PEAKS

The proposed method of group identification is based on the interpretation of numbers of homologous groups of the main peaks in the spectrum and hence rules out errors due to possible impurities in the principal substance and to the uncertainty of identification of weak peaks. It is reasonable to have a criterion for the selection of the main peaks depending on the selectivity of fragmentation. Hence, the peaks considered to be the main peaks should obey the equation

$$I_{\mathrm{rel}} \geqslant 5(N + 1) \tag{5}$$

where N is the number of peaks in the spectrum with a relative intensity of over 50%. This method of selecting the main peaks may also be used for the reduced mass spectra or for incomplete spectra registered by the chromatography of trace amounts of organic components. The minimum value of N is unity and then we have

$$I_{\mathrm{rel}}^{\mathrm{min}} = 5(1 + 1) = 10\%$$

TABLE II

PART OF CLASSIFICATION TABLE FOR GROUP IDENTIFICATION OF ORGANIC COMPOUNDS ACCORDING TO THE NUMBERS OF HOMOLOGOUS GROUPS OF THE MAIN PEAKS OF MASS SPECTRA

Numbers of homologous groups of the main peaks with the exception of the maximum peak	*Indices of classes of compounds (numbers in groups lists)*	*Numbers of homologous groups of the main peaks with the exception of the maximum peak*	*Indices of classes of compounds (numbers in group lists)*
Maximum peaks in group 0:			
1, 2	2:43	1, 3, 11, 12, 13	3:17*
1, 2, 3	3:14	1, 3, *11, 12, 13*	2:11*
Maximum peaks in group 1:			
0, 2	2:43	*0*, 2, 12, 13	4:17*
(Maximum peaks in groups 2–12):			
Maximum peaks in group 13:			
0	0:35	0, *1, 11*	0:3, 1:25, 8:35*
7, 9	6:25	*11*, 12	12:11

This implies that only peaks the intensity of which is not lower than 10% should refer to the main peaks, and if the selectivity of fragmentation is lower, the limit of selection of characteristic peaks increases correspondingly.

The totality of parameters y of the main peaks reflects the peculiarities of fragmentation of each class of organic compounds and may be used for group identification with the aid of special tables[6,7]. The structure of the last variant of these tables containing data for 450 classes of organic substances[7] is illustrated in Table II. The classes of compounds (coded with the numbers of homologous groups and the ordinal numbers in the group lists) are arranged first according to the numbers of groups of maximum peaks and then according to the sets of numbers of other main peaks (in ascending order). Series the terms of which may have maximum peaks of different groups are doubled in the corresponding sections of Table II and the numbers of these groups are in italics. If the spectrum investigated has no peak of a molecular ion, then only the series marked with an asterisk are taken into account. The terms of these series may not produce molecular ions on electron impact. For series with unique or very uncommon peculiarities of fragmentation, this approach gives an unequivocal answer. Ions of groups 4–10 prove to be the most informative. However, if fragment ions are present only in groups 1, 13 and 0 containing little information and appearing in many classes of compounds as a result of the elimination of particles $C_nH_{2n+1}^+$, $C_nH_{2n-1}^+$, $C_nH_{2n}^{+\cdot}$ and those isobaric to them, then classification becomes ambiguous and the number of alternative answers may attain 10 and even more. This peculiarity should be taken into account when mass spectrometry is used as a method of structural analysis.

For computer processing it is reasonable to represent the set of numbers of homologous groups of the main peaks y characteristic of each class of compounds by the single number $\sum_i 2^{y_i}$. This number, requiring minimum amount of on-line memory, can easily be decoded with restoration of the initial totality of y at the stage of comparison with the main peaks of the unknown compounds. A small Hewlett-Packard 2114B computer (with an 8K on-line memory) with a program in BASIC solves the problem of group identification according to the main peaks in a few tens of seconds including the printing of results in the following form:

Initial data and answer	*Comments*
DATA 4, 1, 2 DATA 179, 136, 121, 120 RUN CLASSES 11:5 READY	4 = number of the main peak in the spectrum; 1 = indication of the presence of the peak of the molecular ion; 2 = ordinal number of the maximum peak in the list; series 11:5 = aminophenols and their ethers.

The reliability and unequivocal character of this method of group identification was checked for the spectra of 100 compounds, one third of which were not contained in the initial array of data in compiling the classification tables and programs. The results of this check, given in Table III, indicate that in approximately half of the cases the answer contains less than five alternative classes of compounds. In order to obtain more definite conclusions, it is necessary to take into account the intensity of peaks and to pass to the application of the spectra ion series.

TABLE III

RESULTS OF AUTOMATIC GROUP IDENTIFICATION OF 100 CONTROL COMPOUNDS ACCORDING TO THE PARAMETERS y OF THE MAIN PEAKS OF MASS SPECTRA

Correct answers		73
Including:	Unequivocal	31
	From 2 to 5 alternative series	24
	From 6 to 10 alternative series	8
	More than 10 alternative series	10
Erroneous answers		27
Including:	Unidentified compounds in known series	14
	Erroneous assignment to known series	13

GROUP IDENTIFICATION ACCORDING TO THE SPECTRA ION SERIES

The spectra ion series are a totality of overall intensities of peaks of each of 14 homologous groups (as a percentage of ion current) calculated by the summation of the intensities of all peaks with the same numbers y:

$$I_y = \sum_x I_{(x;y)}; \qquad \sum_0^{13} I_y = 100\% \tag{6}$$

The summation of intensities of some ions with a mass difference of 14 units or a number of units divisible by 14 has long been used for the analysis of oil fractions. The first information on group identification based on the principle of ion series appeared as early as 1968[12], but the term itself "spectra ion series" was probably proposed in 1972[13]. However, the spectra ion series have not yet been widely used for group identification. Nevertheless, they can serve as an effective method of structural analysis if averaged values of intensities (within a series with the exception of its first terms) are employed and, in contrast to the procedure used by earlier workers, their standard deviations are taken into account. In the form of recording proposed by the present authors, the spectra ion series are simply expressed by a succession of 14 numbers arranged in order of numbers of groups y with subscripts indicating standard deviations of intensities:

(y =	0	1	2	3	4	5	6	7	8	9	10	11	12	13)
Alkenes:	22_8,	14_{11},	0,	0,	0,	0,	0,	0,	0,	2_1,	1_1,	9_3,	3_1,	48_{12}
Alkanes:	16_5,	56_7,	3_1,	0,	0,	0,	0,	0,	0,	0,	0,	4_1,	1_1,	20_2
Arenes:	0,	1_1,	0,	1_1,	0,	3_1,	1_1,	52_8,	24_6,	9_2,	1_1,	4_1,	0,	4_2

Hence, each homologous series is characterised by 14 pairs of numbers: the average values of the intensities of 14 ion series, I, and their standard deviations, s. At present the total spectra ion series drawn up by the present authors contains over 450 classes of organic compounds[7] and may be used for group identification both "manually" and with the aid of small personal computers.

A spectrum ion series of an unknown compound is compared successively with tabulated averaged spectra coinciding to within ±2s (with a confidence level of 95%)

or $\pm 3s$ (with a confidence level of 99%) for each of 14 values of I. Hence, the main criterion for the identification of an unknown compound in a certain homologous series is the fulfilment of the following condition for all I_y values:

$$|I_y - \bar{I}_y| \leqslant ks_y \quad (y = 0\text{–}13) \tag{7}$$

The parameter k controls the reliability of identification. At $k = 1$ it is often impossible to identify the compounds belonging to well investigated series, and at $k = 3$ the number of erroneous identifications of unknown compounds to series of another nature increases.

In order to avoid errors that might arise from the presence of small amounts of impurities in the substance under investigation, a standard deviation of 1% is ascribed to all zero values. When the answer is ambiguous (the spectrum ion series coincides with those of several classes of compounds), additional criteria for the degree of accordance between the spectra may be applied. The overall divergence between peak intensities may serve as one of these additional parameters:

$$D = \sum |I_y - \bar{I}_y| \tag{8}$$

This divergence should not exceed the sum of standard deviations of all the values of I_y:

$$D < \sum s_y \tag{9}$$

The procedure of comparison of spectra ion series can easily be formalized and programmed. An example of thesolution of thisproblem with a BASIS 108 computer according to the program[5] is shown below:

Initial data and answer	*Comments*
114, 4, 86, 10, 100, 70, 20, 59, 3, 58, 3, 57, 33, 56, 26, 55, 21, 43, 6, 42, 38, 41, 17, 40, 3, 39, 13, 30, 3, 29, 79, 28, 23, 27, 31, 26, 4, 0, 2 SIGMA 2 CYCLIC ETHERS CLASS 2:5, $s = 60$, $D = 51$ LACTONES CLASS 2:11, $s = 51$, $D = 28$	Complete mass spectrum in the sequence mass number, relative peak intensity. 0 = indication of the end of the spectrum; 2 = number of the homologous group. The parameter $k = 2$ chosen by the operator in the condition of comparison 7. The spectrum belongs to γ-caprolactone.

The results of checking the proposed method of group identification according to the spectra ion series for the same sample among 100 compounds which was used for the verification of identification according to the main peaks are shown in Table IV. It is clear that the overall probability of correct answers attains 90%. It should be particularly noted that it is possible to obtain distinct indications of the fact that the compounds under study belong to uninvestigated and unregistered homologous series. The rate of these correct negative answers was 80%.

The values of y and I_y were the main operating parameters used in this method. The parameter x, closely related to the mass number, may also be employed for the

TABLE IV

RESULTS OF GROUP IDENTIFICATION OF 100 CONTROL COMPOUNDS ACCORDING TO THE SPECTRA ION SERIES

Correct answers		90
Including:	Unequivocal	66
	From 2 to 5 alternative series	22
	From 6 to 10 alternative series	1
	More than 10 alternative series	1
Erroneous answers		10
Including:	Unidentified compounds in known series	4
	Erroneous assignment to known series	6

establishment of the overall formula if the properties of the compounds under investigation corresponding to the condition of additivity are used for their characterization.

HOMOLOGOUS INCREMENTS OF RETENTION INDICES AND THEIR APPLICATION TO GROUP IDENTIFICATION

Within a homologous series the additive properties A increase linearly by a value A_{CH_2} corresponding to the homologous difference CH_2. Hence, for the characterization of the series as a whole, it is reasonable to introduce the concept of the homologous increment of the property A, determined by the equation

TABLE V

HOMOLOGOUS INCREMENTS OF RETENTION INDICES OF SOME CLASSES OF ORGANIC COMPOUNDS WITH SIMILAR MASS SPECTRA

Homologous group	*Homologous series*	i_A *(on Carbowax 20M)*
0	Alkenes	30–85
	Cyclopentanes and cyclohexanes	145–170
	α,β-Unsaturated carbonyl compounds	465–570
	Cyclic ketones	630–665
4	Alcohols: primary	570–670
	secondary	440–520
	tertiary	330–390
	Ethers	65–155
	Esters	205–345
	Carboxylic acids	980–1120
9	Anilines	1120–1285
	Pyridines	670–750
10	Tricyclanes	330–480
	Phenols	1360–1440
	Alkyl benzoates	655–695

$$i_A = A - xA_{CH_2} \tag{10}$$

In chromatography–mass spectrometry, retention indices may be this additive property. According to definition, for these indices $A_{CH_2} = 100$ and the homologous increment of *n*-alkanes is zero. The branching of the skeleton leads to a decrease in retention indices so that in the series of isomeric alkanes the values of i_A are negative and for compounds with a high degree of branching i_A attains -100 to -130 units. Temperature has little effect on the numerical values of the homologous increments of retention indices, but they can be profoundly affected by the structural features of the compounds and the nature of the stationary phase. The structural effects are particularly pronounced on polar stationary phases (*e.g.*, Carbowax 20M), and the differences in homologous increments may be used in group identification. Table V, which is part of a more complete summary[6], illustrates the possibilities of identifying some classes of compounds that exhibit very similar mass spectra. Thus, alkanes and cycloalkanes, secondary and tertiary alcohols, anilines and pyridines can be clearly differentiated according to homologous increments of their Kováts retention indices.

CONCLUSIONS

In this brief paper it was impossible to consider all the details of the proposed new approaches to the interpretation of chromatographic–mass spectrometric data or to reproduce in detail the classification tables published in a special monograph[7].

In conclusion, we emphasize again the suitability of singling out group identification as an independent stage of the decoding of mass spectra before the final establishment of the structures of unknown compounds and making it possible to simplify and formalize the processing of the initial experimental data. It is also desirable for further information to be accumulated on the spectra ion series and homologous increments of retention parameters of insufficiently studied classes of compounds.

REFERENCES

1 B. V. Ioffe and I. G. Zenkevich, *Dokl. Akad. Nauk SSSR*, 264 (1982) 1150.
2 B. V. Ioffe and I. G. Zenkevich, *Zh. Org. Khim.*, 19 (1983) 673.
3 I. G. Zenkevich and B. V. Ioffe, *Zh. Org. Khim.*, 19 (1983) 682.
4 B. V. Ioffe and I. G. Zenkevich, *Dokl. Akad. Nauk SSSR*, 279 (1984) 637.
5 I. G. Zenkevich, V. I. Bespalov and B. V. Ioffe, *Vestn. Leningr. Univ.*, No. 4 (1985) 73.
6 B. V. Ioffe, I. G. Zenkevich, M. A. Kuznetsov and I. I. Bershtein, *Novye Fizicheskie i Fiziko-Khimicheskie Metody Issledovaniya Organicheskikh Soedinenĭ (New Physical and Physical-Chemical Methods for the Investigation of Organic Compounds)*, Leningrad University Publishing House, Leningrad, 1984, Ch. I and II.
7 I. G. Zenkevich and B. V. Ioffe, *Interpretaziya Mass-Spektrov Organicheskikh Soedinenĭ (Interpretation of Mass Spectra of Organic Compounds)*, Khimiya, Leningrad, 1986 (in press).
8 E. Stenhagen, S. Abrahamsson and F. W. McLafferty, *Atlas of Mass Spectral Data*, Interscience, New York, 1969.
9 A. Cornu and R. Massot, *Compilation of Mass Spectral Data*, Vols. 1–3, L., 1975.
10 V. A. Koptyug (Editor), *Atlas Mass Spektrov Organicheskikh Soedinenĭ (Atlas of Mass Spectra of Organic Compounds)*, Vols. 1–10, Novosibirsk, 1977–1982.
11 H. Davenport, *The Higher Arithmetic. An Introduction to the Theory of Numbers*, Harper, New York, 1963.
12 B. Petterson and R. Ryhage, *Anal. Chem.*, 39 (1967) 790.
13 D. H. Smith, *Anal. Chem.*, 44 (1972) 536.