Journal of Chromatography, 163 (1979) 125—133
Biomedical Applications
© Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

CHROMBIO. 323

CAPILLARY COLUMN GAS CHROMATOGRAPHIC PROFILE ANALYSIS OF VOLATILE COMPOUNDS IN SERA OF NORMAL AND VIRUS-INFECTED PATIENTS

ALBERT ZLATKIS, KWAN YOUNG LEE, COLIN F. POOLE and GUNTHER HOLZER

University of Houston, Chemistry Department, Houston, Texas 77004 (U.S.A.)

(Received November 6th, 1978)

SUMMARY

Using a transevaporator sampling technique, the volatile profiles from 70 µl of serum were obtained by capillary-column gas chromatography. The complex chromatograms were interpreted by a combination of manual and computer techniques and a two-peak ratio method devised for the classification of normal and virus-infected sera. Using the K-nearest neighbor approach 85.7% of the unknown samples were classified correctly. Some preliminary results indicate the possible use of the method for the assessment of virus susceptibility.

INTRODUCTION

Recent years have seen a great deal of growth in the development of profile analysis techniques for the recognition of disease disorders [1—20]. A complete profile of all the constituents of a biological fluid is at present impossible with available analytical techniques. The goal at which most workers aim is the development of a complete profile of a selected group of substances (such as steroids, amino acids, etc.) or of compounds with similar physical properties (for example, volatiles). A comparison is then made between the profile obtained with "normal samples" and "pathological samples" to establish any quantitative differences that might be of value for diagnostic purposes. A knowledge of the chemical constitution of abnormal peaks or peaks of abnormal concentration may suggest a possible biochemical reason for the disease and lead to new methods of treatment.

Except in a few cases in which a strong clue exists to implicate the role of causative agents to a disease disorder, most work still relies on a serendipitous finding in a screening program involving the study of many diseases. To maximize the chance of discovery as many substances as possible are included in the profile. The very nature of this approach leads to problems of an analytical nature. For the case of the analysis of volatiles, the current sampling

techniques generate very complex chromatograms which only reveal a full picture of their complexity when high-resolving capillary columns are used for the gas chromatographic separation. The amount of information subsequently generated can no longer be handled by empirical means and the use of pattern recognition techniques and computer sorting are obligatory. Although the selected sample can be characterized as "volatile", under that heading is masked the chemical complexity of the mixture which covers the complete spectrum of polarity. The physical property shared by all components is that they fit into a distinct boiling-point range. The successful use of such techniques as pattern recognition makes high demands on the reproducibility of the profile; a feature which it is all too easy to demonstrate is more often impaired by poor sampling techniques than chromatographic error [21]. The volatiles are present at trace levels in biological fluids, which consist principally of water, so that the sampling technique has to serve as a concentration device capable of reproducibly stripping a diverse range of compounds from a water matrix. For the analysis of urine [1—3, 6, 8—12, 14] sample size and availability are rarely a problem but for serum [7, 16, 20] this is not so, and only relatively small quantities can be obtained from patients.

In an effort to establish the methodology for the early diagnosis of viral diseases, a series of capillary-column chromatographic profiles were obtained using a "transevaporator" sampling technique [22, 23]. The transevaporator is capable of providing reproducible chromatographic profiles of volatile constituents from as little as 10—200 μl of biological fluids and is ideally suited to the analysis of serum samples [23]. The complex chromatographic profiles were analyzed by computer techniques to develop a means of differentiation between normal and virus-infected sera using a two-peak ratio method. A training set was developed and the K-nearest neighbor technique [24—27] used to establish the predictability of virus-infected serum identification. Gas chromatography—mass spectrometry was used to identify the most prominent serum components important to this study.

EXPERIMENTAL

Serum samples

A total of thirty-six serum samples were used from twelve male volunteers who had been exposed to either "England" or "Rhinovirus". These samples form part of a study of immune response to respiratory virus infection conducted by Drs. R. Couch, V. Knight and S. Criswell of the Influenza Research Center (Houston, Texas). Serum samples (200—800 μl) were stored in glass vials at —20° prior to analysis.

The serum samples fall into two categories. The first (Nos. 1—6) were obtained from volunteers who developed clinical symptoms of influenza after virus infection. The second group (Nos. 7—12) were obtained from volunteers who did not develop clinical symptoms after virus infection. Each volunteer provided three serum samples corresponding to a baseline sample taken prior to virus infection (Group I), a serum sample taken one day after exposure to the virus (Group II) and a final sample taken 14—21 days after virus infection (Group III). Group I represents normal serum samples, Group II virus-infected

serum samples (irrespective of whether or not clinical symptoms developed after exposure), and Group III virus-infected serum samples from which all volunteers had recovered clinically from the infection.

Adsorbents and reagents

Porasil E (80—100 mesh) and glass beads (80—100 mesh) were obtained from Analabs (North Haven, Conn., U.S.A.). Prior to use they were washed with distilled diethyl ether and conditioned for 12 h at 280° in a stream of dry helium.

2-Chloropropane (Eastman-Kodak, Rochester, N.Y., U.S.A.) was distilled from phosphorus pentoxide and stored at 5° in the dark prior to use.

Sampling procedure

The serum volatiles were collected from 70 μl of serum sample by the "trans-evaporator" technique described previously [23]. The serum sample is deposited on a Porasil E micro-column and the volatiles stripped from the sample by 2-chloropropane and transferred in the vapor phase to a glass-bead collection column. The volatiles were desorbed thermally at 280° and transferred to a stainless-steel precolumn (30 cm × 1 mm I.D. coated with SF-96) cooled in liquid nitrogen with a helium flow-rate of 7 ml min^{-1} for 10 min. Similarly, the sample is transferred to the analytical column, the first 30 cm of which are cooled in liquid nitrogen by briefly heating the precolumn to about 180° with an air heating gun while passing helium through the system at a rate of 1.5 ml min^{-1}.

Gas chromatography

A Hewlett-Packard 5830A gas chromatograph (Hewlett-Packard, Avondale, Pa., U.S.A.) equipped with flame ionization detectors and connected to a Hewlett-Packard 18850 gas chromatography terminal was used. The serum volatiles were separated on a stainless-steel (100 m × 0.25 mm I.D.) capillary column coated with Witconal LA-23 (Witco, Houston, Texas, U.S.A.) by the dynamic coating method [29]. With a helium carrier gas flow-rate of 1.5 ml min^{-1}, the column was held isothermally at 50° for 10 min and then programmed at 1.5° min^{-1} to 160° and maintained at this temperature for 80 min.

Gas chromatography—mass spectrometry (GC—MS) was performed on an LKB 9000 instrument with a single jet separator and a Perkin-Elmer 900 gas chromatograph. Analytical conditions were as above. Mass spectra were recorded at 70 eV with a scan speed of 4 sec for the mass range 15—300 a.m.u. When possible, identification was confirmed by comparison with standard compounds available in the laboratory, otherwise manual interpretation and comparison to library spectra [30, 31] were used.

Data handling and computer interpretation

The operations that constitute the interpretative procedure are summarized in Fig. 1. Visual inspection was used to identify those peaks common to all chromatograms. The peaks were normalized and this formed the data set for calculation of system variations and variation between individuals within each

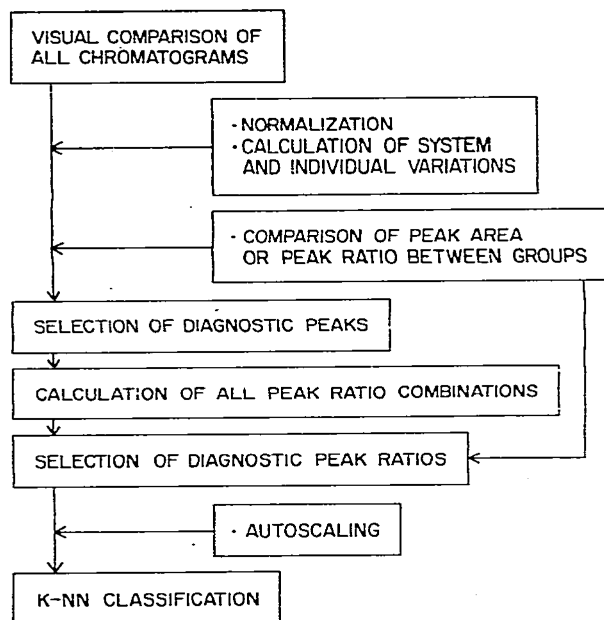

Fig. 1. Schematic diagram of data processing of chromatographic information.

group. A second data set selected on the basis of the magnitude of their difference between groups was used to calculate all possible peak ratio combinations and diagnostic ratios selected. These ratios were autoscaled [28] and used as input for K-nearest neighbor classification. All programs were written in Fortran IV and run on the University of Houston computing center's Honeywell 66/60.

RESULTS AND DISCUSSION

A total of thirty-six virus-infected and normal serum samples as well as six pooled serum samples were studied. The processing of a 70-μl serum sample including sampling by the transevaporator technique, separation by gas chromatography and data input and computation, can be completed within a 2.5-h period. A typical chromatogram obtained from a pooled virus-infected serum sample is shown in Fig. 2A. Peaks identified in the chromatogram by GC—MS are summarized in Table I. The peak numbers in Table I correspond to those marked in Fig. 2. Five of these substances (peaks 5, 8, 10, 13 and 28) have been identified previously in normal serum [20].

The profile of volatile substances in serum is complex, with more than 150 peaks in the chromatogram. To simplify data handling, a visual comparison was made of all chromatograms; peaks due to background and stripping solvent were ignored and 37 peaks which appeared consistently in all chromatograms were selected as the data base. Six replicate analyses of a pooled serum sample were used to establish the magnitude of system variation within the data base (i.e. the effect of experimental variables on reproducibility). All peaks were normalized and the percentage relative standard deviation calculated. Reproducibility depended very much on the compound itself and consequently a

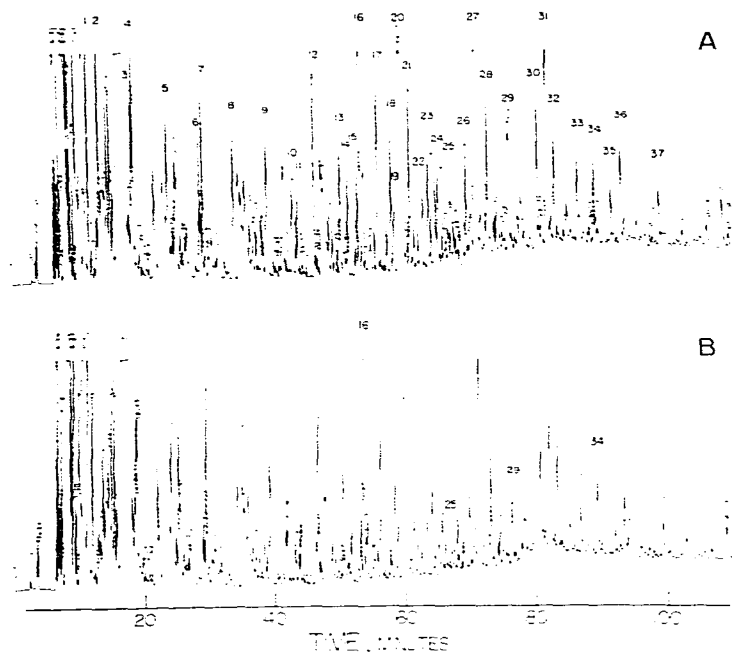

Fig. 2. Profile of organic volatiles by capillary-column gas chromatography. A, virus-infected serum; B, normal serum.

TABLE I

SUBSTANCES IDENTIFIED BY GAS CHROMATOGRAPHY—MASS SPECTROMETRY IN VIRUS-INFECTED SERUM

Components indicated by numbers in Fig. 1.

Peak No.	Compound	Peak No.	Compound
1	2-Methyl-1-hexene	19	2-Octanone
2	Dimethylcyclopentane	20	*n*-Octanal
5	2-Propanol	22	6-Methyl-5-heptene-2-one
7	2-Hexanone	23	5-Nonanone
8	*n*-Hexanal	25	2-Ethylhexanal
10	*n*-Butanol	27	2-Octenol
12	2-Heptanone	28	Benzaldehyde
13	Heptanal	29	2-Ethyl-1-hexanol
14	4-Heptanone	30	2-Decanone
15	4-Octanone	32	*o*-Tolualdehyde
16	6-Methyl-2-heptanone	33	Acetophenone
17	Cyclohexanone and 5-methyl-3-heptanone	34	Trimethyl-2-cyclohexanone

wide variation was found between individual peaks reflecting their different chemical properties. The smallest variation was 6.5% (peak 20), a median value 18.3% (peak 5) and the largest 51.1% (peak 21) relative standard deviation.

Another source of variance is the variability between individuals in any one group. This was calculated for each group in turn for the 37 peaks. All peaks

were normalized and their variance calculated. For example, the 12 serum samples of Group I (normal serum) had a lowest value of 17.1% (peak 1), a median value 32.9% (peak 32) and a highest value of 87.5% (peak 4) relative standard deviation. Results for Group III were similar but Group II showed far less variation.

A second data set was selected, based on the magnitude of the difference between the averaged normalized peak areas for the different groups. the criteria for selection was that the relative standard deviation of the selected peaks must be greater than the variation between individuals in any one group. Seven peaks were found to meet this requirement (peaks 16, 19, 20, 25, 29, 31 and 34).

The normalized peak areas in the second data set can show both negative and positive variation between groups. Under these conditions a ratio of two peaks should prove more sensitive to inter-group differences. Also, if there is an interaction between peaks in each group then peak ratios will be more reliable for classification purposes. The seven normalized peak areas were arranged in ascending order and all possible peak ratios calculated for Groups I and II. The two peak ratios 16/25 and 29/34 were found to be most suitable for sample identification.

The reproducibility of retention time and normalized peak areas for the four selected diagnostic peaks in the pooled serum sample (six replicate analyses) is shown in Table II. Retention times can be reproduced very accurately in the

TABLE II

REPRODUCIBILITY OF NORMALIZED PEAK AREA AND RETENTION TIME WITHIN A POOLED SERUM SAMPLE (6 REPLICATIONS)

Peak No.	Normalized peak area (%)			Retention time (min)		
	Mean	S.D.	C.V. (%)*	Mean	S.D.	C.V. (%)*
16	5.36	1.44	26.9	53.75	0.57	1.0
25	1.34	0.25	18.8	65.82	0.63	0.9
29	4.15	1.37	31.7	76.51	0.62	0.8
34	1.42	0.29	21.0	89.77	0.80	0.9

*Percentage relative standard deviation.

analytical system and this forms an adequate parameter for peak identification. The relative standard deviation of the normalized peak areas between individuals for the four selected peaks is given in Table III for Groups I—III. The variation in peak area for infected serum (Group II) is much less than for normal serum (Group I) and can be more correctly defined. Peak 25 in Group I shows a greater variation than the others due to the inclusion of one extraordinarily large peak in the data set.

To differentiate between normal serum (Group I) and infected serum (Group II) the two peak ratios 16/25 and 29/34 are calculated in Table IV. Visual inspection shows that generally the peak ratio 16/25 decreased upon virus infection (except samples 7, 8 and 9) and that the peak ratio 29/34 increased by virus infection (except samples 2, 3 and 5). However, the use of either peak

TABLE III

VARIATION OF NORMALIZED PEAK AREAS FOR THE SELECTED FOUR PEAKS IN ALL SERUM SAMPLES

Peak No.	Relative standard deviation (%)		
	Before infection (Group 1)	1 day after infection (Group II)	14—21 days after infection (Group III)
16	48.9	32.6	44.2
25	89.9	20.4	47.4
29	36.3	32.7	30.5
34	53.8	30.3	62.9

TABLE IV

SELECTED TWO-PEAK RATIO DATA AT DIFFERENT VIRUS-INFECTED CONDITIONS

Sample No.	Peak ratio (16/25)			Peak ratio (29/34)		
	Before infection	1 day after infection	14—21 days after infection	Before infection	1 day after infection	14—21 days after infection
Serum samples with clinical symptoms after exposure						
1	17.73	4.18	8.73	3.53	7.30	5.54
2	17.93	8.90	9.65	3.85	2.56	3.11
3	24.28	4.77	3.71	44.00	3.24	6.17
4	3.32	2.80	3.67	3.57	4.65	1.88
5	15.78	2.61	5.77	4.15	3.70	3.55
6	17.48	9.51	12.07	0.69	2.43	1.82
Serum samples with no clinical symptoms after exposure						
7	3.38	9.03	8.97	0.99	4.13	1.35
8	6.58	5.05	31.51	0.66	4.40	5.21
9	1.35	1.44	7.56	6.13	7.23	23.22
10	9.24	4.95	2.75	1.94	11.89	1.75
11	33.69	5.59	4.59	1.54	2.52	4.00
12	17.59	8.24	23.15	1.99	8.43	2.22

ratio does not in itself provide a sufficient classification between the two groups. The autoscaled data set of Groups I—III is plotted in two-peak ratio dimensions in Fig. 3. It can be seen that two separate clusters are formed for Groups I and II and thus the two-peak ratios selected are adequate for the identification of normal and infected serum.

In order to test the predictability of the proposed method the K-nearest neighbor approach was used. A randomly chosen training set of ten serum samples (5 normal and 5 infected) was used to assess the predictive accuracy of the two-peak ratio technique using the remaining 14 samples as unknowns. Predictive ability was calculated as the percentage of the 14 unknowns correctly classified. The three nearest neighbor (3-NN) computation assigned 85.7% of the samples correctly (1-NN, 71.4%; 5-NN, 78.5%). Clearly the proposed

method has excellent selectivity for the differentiation of normal and virus-infected serum.

After 14—21 days (Group III) complete recovery from infection is observed clinically. However, the two-peak ratio classification places approximately half the serum samples in the virus-infected category. The implications of this finding can only be speculated on at the moment, especially as normal samples are correctly classified by this technique.

The general usefulness of this profiling technique would be enhanced if it could be used to predict susceptibility to virus infection. The normalized two-peak ratio data for Group I (normal serum) samples are shown in Fig. 4.

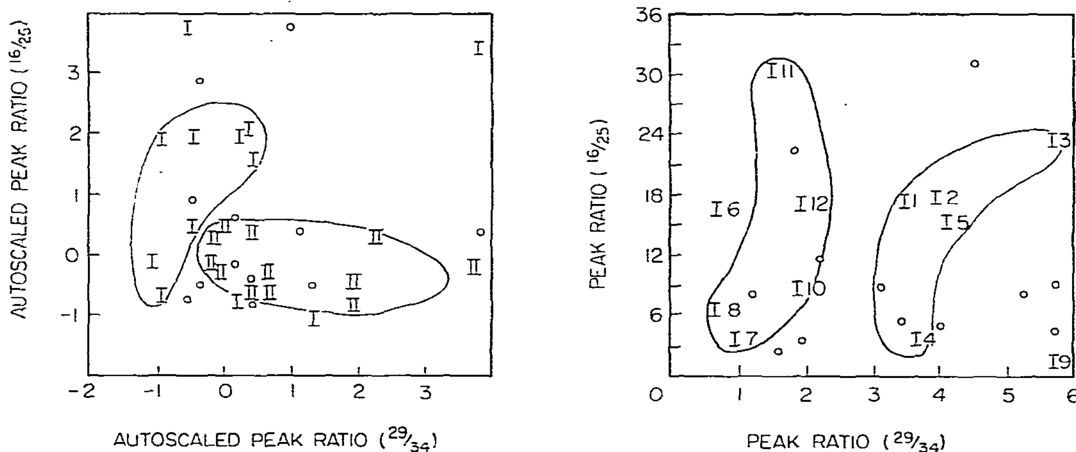

Fig. 3. Autoscaled two-peak ratio diagram. I, normal serum before infection; II, infected sera (24 h); o, recovered sera (14—21 days after infection).

Fig. 4. Two-peak ratio diagram. Samples I1—I6 are normal samples from volunteers who contracted virus infection after exposure; samples I7—I12 the same, except no clinical symptoms developed after exposure.

Samples I1—I6 were normal serum samples from volunteers who upon exposure to virus infection developed clinical symptoms of the disease, and samples I7—I12 did not develop clinical symptoms upon infection. Again two well-defined clusters (with two exceptions: I6, and I9) are formed and demonstrate the possible use of the method for the diagnosis of virus susceptibility. However, the twelve samples available are too small a data base to provide a training set and sufficient unknowns to test the predictability of the method. A much larger sampling program will be required to assess the accuracy of this method for the determination of susceptibility to virus infection.

CONCLUSIONS

The transevaporator sampling technique is shown to be useful for the volatile profile analysis of 70 μl of serum sample. A two-peak ratio method has been developed for the characterization of normal and virus-infected sera with a

percentage predictability of 85.7% of correctly classified unknowns. A similar two-peak ratio method is indicated as a possible means of assessing virus susceptibility.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the National Aeronautics and Space Administration, Life Sciences Directorate, Johnson Space Center, Houston, Texas (Contract NAS9-14629).

REFERENCES

1 A. Zlatkis, W. Bertsch, H.A. Lichtenstein, A. Tishbee, F. Shunbo, H.M. Liebich, A.M. Coscia and N. Fleischer, Anal. Chem., 45 (1973) 763.
2 A. Zlatkis, H.A. Lichtenstein and A. Tishbee, Chromatographia, 6 (1973) 67.
3 H.M. Liebich, O. Al-Babbili, A. Zlatkis and K. Kim, Clin. Chem., 21 (1975) 1294.
4 P.A.G. Malya, J.R. Wright and W.R. Nes, J. Chromatogr. Sci., 9 (1971) 700.
5 B. Krotoszynski, G. Gabriel, H. O'Neill and M.P.A. Claudio, J. Chromatogr. Sci., 15 (1977) 239.
6 R. Teranishi, T.R. Mon, A.B. Robinson, P. Gray and L. Pauling, Anal. Chem., 44 (1972) 18.
7 E. Stoner, D. Cowburn and L.C. Craig, Anal. Chem., 47 (1975) 344.
8 A.B. Robinson and L. Pauling, Clin. Chem., 20 (1974) 962.
9 H. Dirren, A.B. Robinson and L. Pauling, Clin. Chem., 21 (1975) 1970.
10 A.B. Robinson, H. Dirren, A. Sheets, J. Miguel and P.R. Lundgren, Exp. Gerontol., 11 (1976) 11.
11 M. Anbar, R.L. Dyer and M.E. Scolnick, Clin. Chem., 22 (1976) 1503.
12 R.A. Chalmers, M.J.R. Healy, A.M. Lawson, J.T. Hart and R.W.E. Watts, Clin. Chem., 22 (1976) 1292.
13 R.D. Malcolm and R. Leonards, Clin. Chem., 22 (1976) 623.
14 T. Kitagawa, B.A. Smith and E.S. Brown, Clin. Chem., 21 (1975) 735.
15 L. Eldjarn, E. Jellum and O. Stokke, Clin. Chem., 21 (1975) 63.
16 F.W. Bultitude and S.J. Newham, Clin. Chem., 21 (1975) 1329.
17 S.I. Goodman, P. Helland, O. Stokke, A. Flatmark and E. Jellum, J. Chromatogr., 142 (1977) 497.
18 E. Jellum, J. Chromatogr., 143 (1977) 427.
19 H.M. Liebich and O. Al-Babbili, J. Chromatogr., 112 (1975) 539.
20 H.M. Liebich and J. Wöll, J. Chromatogr., 142 (1977) 505.
21 M. Novotny, M.L. McConnell and M.L. Lee, J. Agr. Food Chem., 22 (1974) 765.
22 A. Zlatkis and K. Kim, J. Chromatogr., 126 (1976) 475.
23 K.Y. Lee, D. Nurok and A. Zlatkis, J. Chromatogr., 158 (1978) 377.
24 T.M. Cover and P.E. Hart, IEEE Trans. Inform. Theory, IT-13 (1967) 21.
25 D.W. Peterson, IEEE Trans. Inform. Theory, IT-14 (1970) 26.
26 B.R. Kowalski and C.F. Bender, Anal. Chem., 44 (1972) 1405.
27 J.J. Leary, J.B. Justice, S. Tsuge, S.R. Lowry and T.L. Isenhour, J. Chromatogr. Sci., 11 (1973) 201.
28 H.A. Clark and P.C. Jurs, Anal. Chem., 47 (1975) 374.
29 W. Bertsch, F. Shunbo, R.C. Chang and A. Zlatkis, Chromatographia, 7 (1974) 128.
30 E. Stenhagen, S. Abrahamsson and F.W. McLafferty (Editors), Registry of Mass Spectral Data, Wiley, New York, 1974.
31 Eight Peak Index of Mass Spectra, Mass Spectrometry Data Center, Aldermaston, 2nd ed., 1974.