Journal of Chromatography, 217 (1981) 231–237
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

CHROM. 13,830

CLASSIFICATION OF HUMAN CANCER CELLS BY MEANS OF CAPILLARY GAS CHROMATOGRAPHY AND PATTERN RECOGNITION ANALYSIS

E. JELLUM* and I. BJØRNSON
Institute of Clinical Biochemistry, Rikshospitalet, University of Oslo, Oslo 1 (Norway)
R. NESBAKKEN
Department of Neurosurgery. Ullevål Hospital, Oslo (Norway)
and
E. JOHANSSON and S. WOLD
Department of Chemistry, Umeå University, Umeå (Sweden)

SUMMARY

The metabolic profiles of brain biopsies obtained at surgery were recorded using capillary gas chromatography (GC). About 160 peaks were seen, of which 105 were used for data analysis. Three classes of brain tissue were examined: normal cerebral cortex, pituitary tumours and "brain" tumours. Pattern recognition analyses of the GC profiles using the SIMCA multivariate programme clearly resolved normal brain tissue from the tumours. Subclassification of the different tumours was more difficult, probably because the number of samples in each tumour class was too small. High-resolution two-dimensional electrophoresis separated the brain biopsies into several hundred different proteins. The combined use of the latter technique and capillary GC–mass spectrometry and pattern recognition analyses gives the possibility of the classification of diseased cells based solely on differences in their biochemical compositions.

INTRODUCTION

Several human diseases result in characteristic changes in the biochemical composition of the cells and the body fluids. Separation techniques such as liquid chromatography and gas chromatography (GC) may be used to detect such changes, *e.g.*, the accumulation of organic acids in blood and urine[1]. The recording of a metabolic profile is particularly suitable for the diagnosis and study of inborn errors of metabolism, as these usually result in the accumulation in the body of large and easily recognizable amounts of one or a few specific metabolites[1].

Most other human diseases, on the other hand, do not give rise to major alterations in the metabolite pattern. Although only minor changes may occur, it is likely that the profiles may still carry diagnostic information. The problem is to

0021-9673/81/0000–0000/$02.50

retrieve this information from the large amount of quantitative and qualitative data contained, *e.g.*, in a complex capillary gas chromatogram. One solution to this problem involves the use of computer and pattern recognition methods. Such procedures have been used to analyse the data from GC profiles of bacteria[2], pyrolysis GC profiles of moulds[3], GC profiles of volatiles[4] and field ionization[5] and pyrolysis[6] mass spectrometry profiles, etc.

This paper concerns the possibility of using GC and pattern recognition methods to study and classify human cancer cells. Different brain tumours were selected and their metabolic profiles (105 reccurring GC peaks) were determined followed by data analysis by the SIMCA pattern recognition method[7,8]. Although both chemical and analytical problems were encountered, the results demonstrate the feasibility of differentiating between various brain tissues by means of capillary GC and pattern recognition.

EXPERIMENTAL

Brain samples

Brain samples (tumour and neighbouring normal tissue) were removed surgically. Parts of the biopsies were submitted to routine neuropathological examination, and parts were immediately submerged in liquid nitrogen and subsequently stored at −70°C before GC analyses. Three categories of brain biopsies were included in the present investigation: normal cerebral cortex (class 1, $n_1 = 6$), "brain tumour" (class 2, $n_2 = 3$) and pituitary tumour (class 3, $n_3 = 7$). Class 2 consisted of a meningeoma and metastatic tumours from carcinomas. Class 3 can be sub-divided into adenomas with production of growth hormone, or prolactin or both hormones.

Gas chromatography

The frozen brain biopsies (about 20–30 mg) were subjected to methanolysis (refluxing with anhydrous methanol saturated with hydrogen chloride gas) overnight. The solvent was then removed in a stream of nitrogen and the residue was trimethylsilylated for 30 min at 80°C with bis(trimethylsilyl)trifluoroacetamide (BSTFA). The methyltrimethylsilyl derivatives were separated in a 50-m glass capillary GC column coated with SE-30. The temperature was maintained at 50°C for 3 min after injection and then increased at 4°C/min to 240°C. The gas chromatograph was a Carlo Erba Fractovap 2101 equipped with a standard flame-ionization detector.

Retention times and the pattern of neighbouring peaks were used to "identify" each GC peak, which was given a number. Quantitative information (data vectors) was obtained by measuring the peak heights. The baseline was chosen arbitrarily, but was the same in all instances. As only relative differences were sought, the choice of baseline did not affect the results.

SIMCA pattern recognition analysis

The number of samples was limited (sixteen biopsies), and the running of many duplicate analyses of each sample was deliberately avoided in order to test the computer system using unfavourable conditions. The pattern recognition method should therefore be able to utilize the chromatographic information in multivariate data

analysis regardless of the ratio between the number of parallels and the number of samples. This criterion is fulfilled by the SIMCA method[7,8].

The SIMCA procedure involves the use of eigenvector projection. Each data vector from the chromatogram (quantitative information about the 105 GC peaks present in all sixteen samples) is considered as a point in the 105-dimensional space obtained by giving each variable one orthogonal coordinate axis. When the point swarm of the sixteen-sample data set is projected down on the plane conserving most of the variance, the eigenvector projection corresponding to the first two eigenvalues of the data covariance matrix is obtained.

The SIMCA analysis also includes the use of "training sets", in this instance sixteen brain biopsies divided in three "classes" (according to neuropathological diagnosis). It is noteworthy that each training set requires only a limited number of samples, preferably five or more. Principal components models are calculated for all classes, and may range in complexity from one point to an n-dimensional hyperplane. The complexity of each class model is estimated from the data using the technique of cross-validation[9].

In the next phase of data analysis, each object data vector is related to each of the class models obtained by a multilinear regression, and the so-called residual standard deviation (RSD) is obtained and used to find distances between different class models. In this way it is possible to examine the class to which a metabolic profile from a sample (brain biopsy) belongs.

RESULTS

The gas chromatograms of human brain biopsies pre-treated as described show complex patterns (Fig. 1). No attempts have been made to identify systematically all peaks at this stage, but it is clear that many of the major components are fatty acids, carbohydrates, amino acids and cholesterol (which is the last eluted compound). It should be realized that as the samples were subjected both to methanolysis and trimethylsilylation, nearly all metabolites which are stable and volatilizable may appear in the chromatograms. The results express some differences in the profiles of the various brain samples. Fig. 1, for example, shows the chromatographic patterns of normal cerebral cortex (top), pituitary tumour (middle) and of a meningeoma (bottom). A manual interpretation of these different metabolic profiles is not easy, however. When the pattern recognition method was used to extract information from the patterns, the results shown in Fig. 2 were obtained. Fig. 2 illustrates the eigenvector projection corresponding to the first two eigenvalues of the sixteen samples. The normal brain tissue samples are clearly separated from the tumours. It is important to note that this projection is made without the use of the information about the class assignment of the samples. The resulting separation (Fig. 2) is therefore a strong indication of real differences in the metabolite pattern of the normal and the tumour tissues.

Attempts were made to separate the tumour subclasses, but this proved to be difficult at present (Fig. 3). This may partly be due to strong subgroupings within each tumour class. When such subgroupings are not recognized, severe overlapping can occur even when the subgroups are well resolved (Fig. 3). In this study the tumour samples are too few to allow the detection of subgroups even if one knows that there

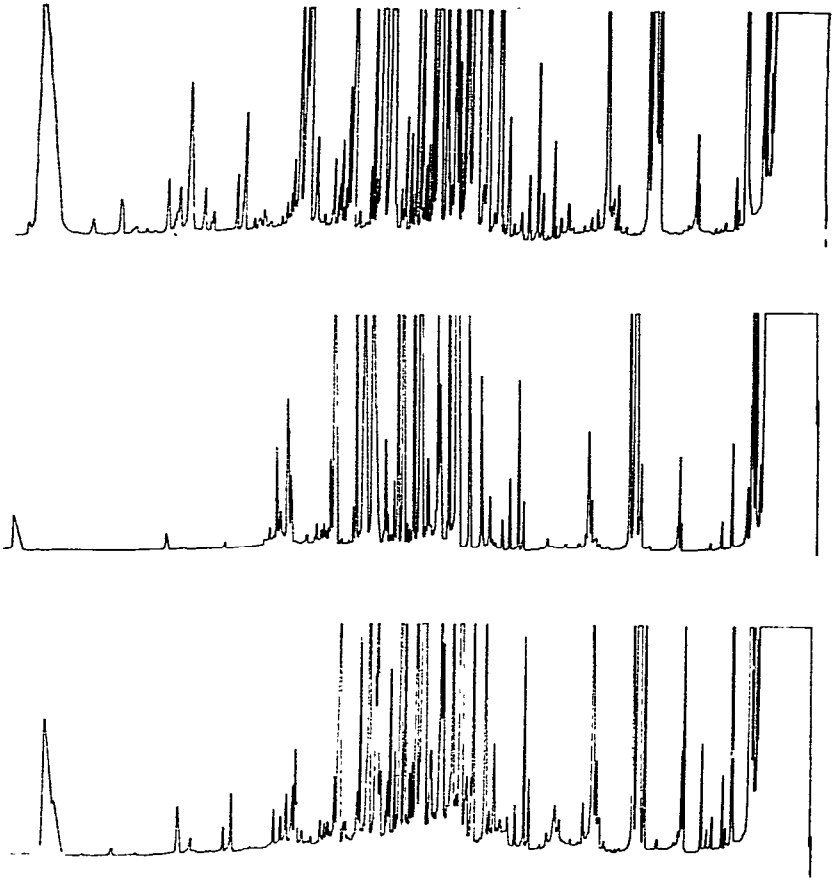

Fig. 1. Metabolic profile of human brain biopsies. Top, normal cerebral cortex; middle, pituitary tumour; bottom, meningeoma. The tissue (20–30 mg) was refluxed with anhydrous methanol–HCl overnight. After removal of the solvent, the residue was trimethylsilylated (BSTFA, 80°C, 30 minutes). The derivatives were separated in a 50-m wall-coated glass capillary column. After 3 min at 50°C, the temperature was programmed at 4°C/min to 240°C, using a Carlo Erba Fractovap 2101 gas chromatograph with a flame-ionization detector.

are indeed several types of brain tissue tumours represented in the data set. Only with an increased number of samples can we start to study this problem.

DISCUSSION

Microscopic examination of tumour cells stained in various ways often reveals morphological characteristics which the experienced pathologist uses for the classification of the diseased cells. It seems obvious that as there are visual differences in various normal and malignant cells, there must also be differences in their biochemical composition. In this work we recorded the metabolic profiles of various human brain biopsies using capillary GC. When the data were analysed by the SIMCA pattern recognition program[7,8], differentiation between normal brain tissue and

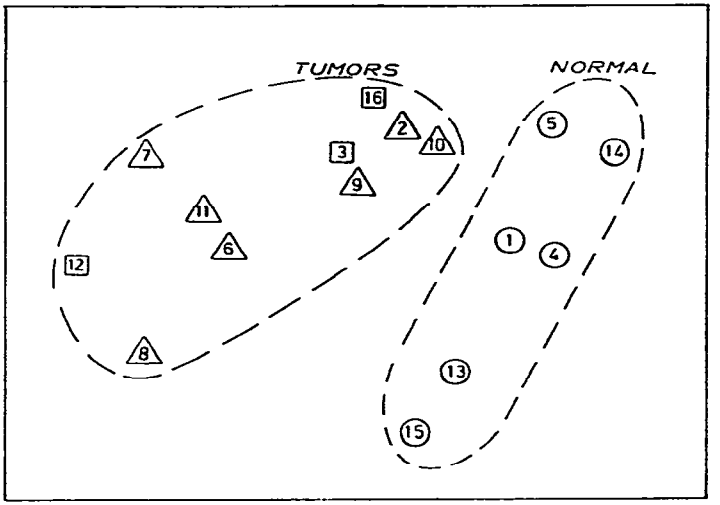

Fig. 2. Eigenvector projection corresponding to the first two eigenvalues of 16 samples. ○, Normal cerebral cortex; □, brain tumour (class 2); △, pituitary tumours (class 3).

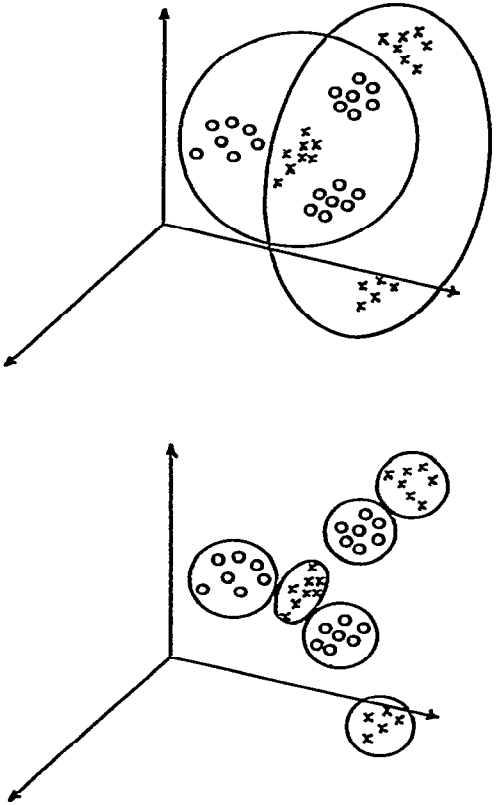

Fig. 3. Three-dimensional space with two classes (○, ×) which contain sub-groups. Top, models constructed without information about sub-classes; bottom, models constructed with knowledge of the differential diagnosis.

tumours was readily obtained. Subclassification of the various tumours, however, was more uncertain.

There may be several reasons for this. First, at this stage of the project we deliberately avoided the use of mass spectrometry and the identification of the GC peaks was based on retention times only. The confidence in the qualitative data is therefore limited, and misinterpretation of the identity of some GC peaks probably occurred. Secondly, there is also a degree of uncertainty in the quantitative aspects, *e.g.*, many of the GC peaks are inhomogeneous. In the next stage of our work we plan to incorporate mass spectrometry in order to overcome some of these problems. It is

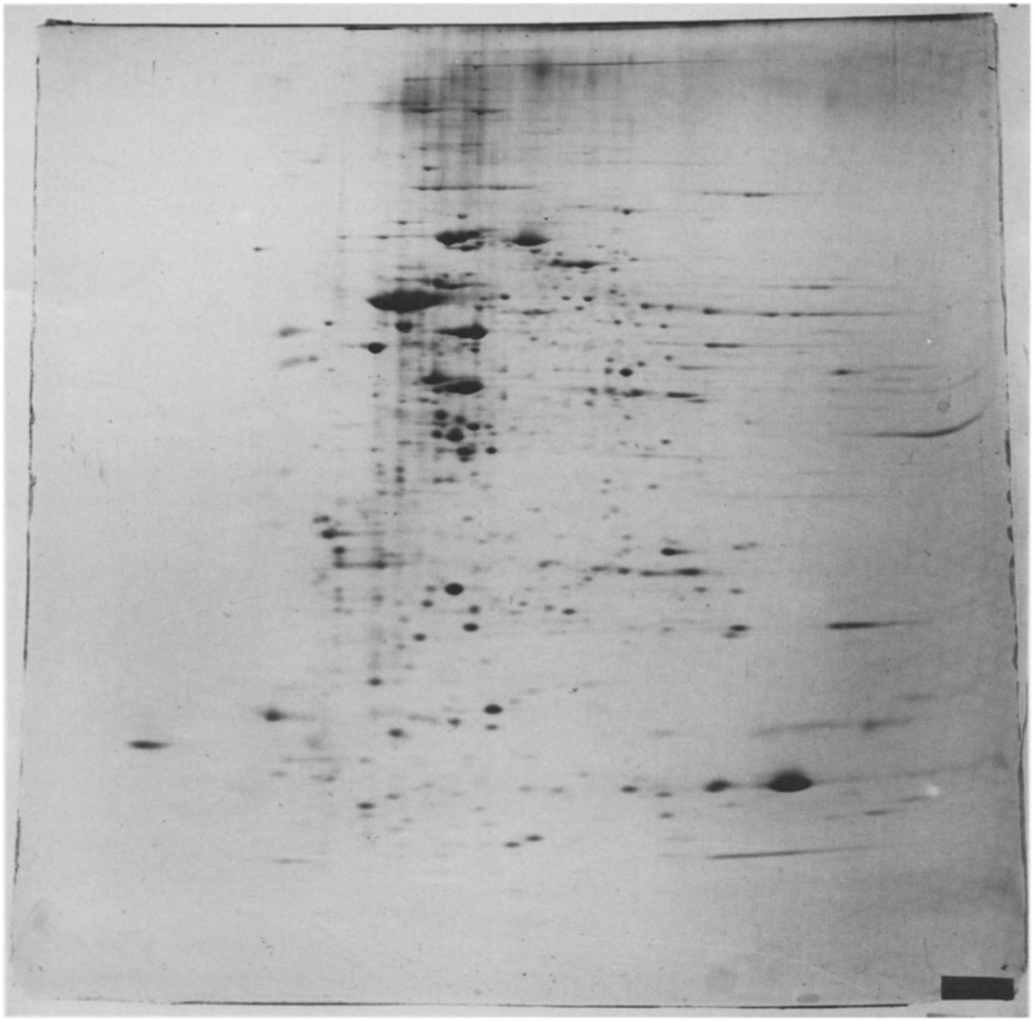

Fig. 4. Protein pattern of a human pituitary tumour. The proteins were separated by high-resolution two-dimensional electrophoresis using the ISO-DALT method described by Anderson and Anderson[13]. Isoelectric focusing was used for the first dimension (horizontal) and SDS electrophoresis in gradient polyacrylamide gel was used in the second dimension (vertical). Staining of the protein spots was effected with Coomassie Blue.

particularly tempting to implement the system described by Sweeley *et al.*[10], which provides automatic identification and quantitative analysis of each peak in a complex chromatographic profile.

Thirdly, the number of samples (tumours with a specific diagnosis) in each training set was small (probably too small). Collection of more biopsies in connection with neurosurgery is currently taking place.

Finally, the pattern recognition programs in general work better with an increasing number of variables. The chromatograms of the brain samples contained about 160–170 peaks, including the smallest ones and the many "shoulders". However, only the 105 best resolved peaks that occurred with every sample were used in the calculations. In biological practice 100–200 variables are standard[11]. As the SIMCA program was designed to handle over 1500 variables, it would be desirable also to include information other than the GC data. The other method of choice for separation might be high-resolution two-dimensional electrophoresis[12,13]. This method is currently used in one of our laboratories[14] and an example of the application of this technique to the analysis of a brain biopsy (the same tumour as in Fig. 3, middle) is shown in Fig. 4. A few milligrams of the tissue were separated into several hundred proteins (gene products). By combining capillary GC with mass spectrometry and two-dimensional electrophoresis we are at present able to separate a few milligrams of tissue into nearly 1000 constituents. It seems highly probable that when these data are subjected to pattern recognition analyses, the possibility will be open for an objective differential diagnosis of diseased cells based solely on differences in their biochemical compositions.

ACKNOWLEDGEMENTS

The support of the Norwegian Research Council for Science and the Humanities (NAVF), the Norwegian Cancer Society and NATO (Research grant 266.80) is gratefully acknowledged.

REFERENCES

1 E. Jellum, *J. Chromatogr.*, 143 (1977) 427.
2 E. Jantzen, K. Bryn, T. Bergan and K. Bøvre, *Acta Path. Microbiol. Scand., Sect. B*, 83 (1975) 569.
3 G. Blomqvist, E. Johansson, B. Søderstrøm and S. Wold, *J. Chromatogr.*, 173 (1979) 19.
4 A. B. Robinson, D. Partridge, M. Turner, R. Teranishi and L. Pauling, *J. Chromatogr.*, 85 (1973) 19
5 R. Abbott, M. Anbar, H. Faden, J. McReynolds, W. Rieth, M. Scanton, L. Verkh and B. Wolff, *Clin. Chem.*, 26 (1980) 1443.
6 H. L. C. Meuzelaar, P. G. Kistemaker and M. A. Posthumus, *Biomed. Mass Spectrom.*, 1 (1974) 312.
7 S. Wold and M. Sjøstrøm, in B. R. Kowalski (Editor), *Chemometrics, Theory and Application*, Symposium Series No. 52, American Chemical Society, Washington, DC, 1977, p. 243.
8 C. Albano, W. Dunn, III, U. Edlund, E. Johansson, B. Nordén, M. Sjøstrøm and S. Wold, *Anal. Chim. Acta*, 103 (1978) 429.
9 S. Wold, *Technometrics*, 20 (1978) 429.
10 C. C. Sweeley, N. D. Young, J. F. Holland and S. C. Gates, *J. Chromatogr.*, 99 (1974) 507.
11 P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, Freeman, San Francisco, 1973.
12 P. H. O'Farrel, *J. Biol. Chem.*, 250 (1975) 4007.
13 N. L. Anderson and N. G. Anderson, *Anal. Biochem.*, 85 (1978) 341.
14 A. K. Torsud, H. F. Haugen and E. Jellum, in B. J. Radola (Editor), *Electrophoresis '79*, Walter de Gruyter, Berlin, New York, 1980, p. 425.