Bull.Chem.Soc.Ethiop., 1(1), 9-17 (1987).

DATA REDUCTION IN GAS CHROMATOGRAPHIC-MASS SPECTROMETRIC ANALYSIS OF COMPLEX MIXTURES

Robert J. Anderegg
Department of Chemistry, University of Maine,
Orono, Maine 04469, USA.

(Received April 23, 1987)

ABSTRACT: Interpretation of the data from a gas chromatographic - mass spectrometric (GC-MS) analysis of a complex mixture of unknown compounds can be the most time-consuming step in the investigation. Dozens or hundreds of mass spectra must be examined and interpreted. We are developing computer methods to assist in this process; allowing the computer to pre-screen the data set and retrieve only those spectra that display a spectral feature defined by the analyst. While previous methods have used the presence of a unique ion as an indicator of a compound class, we can improve the selectivity of the search by using a pattern of ions. These patterns arise from the special way a series of related compounds fragments when bombarded by energetic electrons. The goal of our methods is to reduce the number of spectra that the analyst needs to interpret; although the class-selective chromatograms generated are also useful for "fingerprinting" samples. The method is demonstrated using a fuel oil (petroleum) sample and an extract of the secretion from the scent glands of civet *(Viverra civetta)*.

INTRODUCTION

Gas chromatography coupled to mass spectrometry is perhaps the most powerful combination available for the analysis of complex mixtures of unknown compounds (1). In these experiments, the mass spectrometer is repetitively scanned over a wide mass range, to collect dozens or even hundreds of mass spectra per analysis. The most time consuming part of the analysis becomes the interpretation of the data, rather than its collection. To attempt to simplify the interpretation process, we have been developing computerized methods of selective data reduction over the past several years (2). In these methods, we train the computer to search through a large number of spectra, seeking a spectral feature of interest in our particular analysis. Rather than looking scan-by-scan through hundreds of spectra, the analyst can focus attention on those spectra that the computer has indicated will be likely to be of interest.

The simplest form of this type of data reduction is the mass

chromatogram (3). If all the compounds of interest have a common mass spectral ion, the computer can be instructed to locate all spectra in the data set displaying that ion. The resulting mass chromatogram is a plot of intensity of the ion vs spectrum, or scan, number; and it indicates where in the data set the analyst can expect to find compounds of interest. Only those spectra need further interpretation; other spectra can be ignored for the time being. The time of analysis is greatly shortened, since relatively few of the collected spectra need be fully evaluated. Other investigators have used Boolean operations to combine sets of mass chromatograms, improving somewhat the selectivity of the data reduction (4-6).

One weakness of these methods is that the relative intensity of the selected ions is not considered. We believe that the presence of ions and their correct relative abundance comprise a spectral pattern that can truly represent a compound class. Initially we studied the patterns that result from the presence of elements with unusual isotopic distributions such as chlorine and bromine. The mass spectra of these compounds display characteristic sets of ions that can be used in computerized data reduction. When the computer searches a data set retrieving spectra containing these patterns, an element-selective detector results. In this fashion, we have described programs that act as chlorine-selective (7-8) or selenium-selective (9) chromatographic detectors.

We now report that the method can be extended to other sorts of mass spectral patterns, namely those that arise from the unimolecular fragmentation of a class of related molecules. The relationship between molecular structure and mass spectral fragmentation has been well established; and forms the basis of computerized "library searches" for identification of unknown compounds. By using these patterns in computerized selective data reduction, we can improve the efficiency of data interpretation.

EXPERIMENTAL

GC-MS analysis and computer programing were conducted on a Hewlett-Packard 5985B GC-MS system. Computer programs were written in BASIC with FORTRAN subroutines. The data reduction strategy parallels that of our earlier work (7-9). A small portion of the mass spectrum of a reference compound is defined as the pattern of interest. A modified library search routine is used to calculate the similarity of the equivalent portion of an unknown mass spectrum to this reference pattern. A number from 0.0 (bad match) to 1.0 (perfect match) is assigned to the unknown spectrum, reflecting the possible presence of the pattern of interest. Because only a small portion of the spectrum is searched, the calculation is far faster than a true library search. Time for the similarity calculation is typically about 10-100 msec per spectrum, including display of the results. The process is repeated for each spectrum in the data set. Finally, a graph of the weighted similarity score (weighted for the overall intensity of the spectrum) vs spectrum number produces a "class-selective chromatogram".

Our data reduction methods were tested on two complex mixtures

analyzed in our laboratory over the past several years. A sample of fuel oil from a commerical source was diluted with methylene chloride to a concentration of 100 ppm. Two microliters were injected (splitless injection) onto a 25 m x 0.3 mm (id) fused silica capillary column, wall coated with methyl silicone stationary phase (J & W Scientific, Rancho Cordova, CA, USA). The GC oven was temperature programed from 70° to 300°C at 5°C/min. The entire GC effluent was directed into the mass spectrometer through a heated (275°C) transfer line. Electron impact mass spectra were recorded repetitively every 4 sec from m/z 40 to 500. The source temperature was 200°C and the electron energy was 70 eV.

The civet extract and fractions thereof were provided by Dr. Berhanu Abegaz of the Department of Chemistry, Addis Ababa University. They were analyzed by GC-MS in the same fashion described above. The same methyl silicone capillary GC column was used; temperature programing was from 120°C to 300°C at 8°C/min. Mass spectra were recorded every 3 sec from m/z 40 to 400.

RESULTS AND DISCUSSION

It is useful to emphasize the goal of selective data reduction. It is not intended to be used as an automated interpretation method; nor as a substitute for conventional library searching of spectral data bases. The purpose is to reduce the number of spectra that an analyst must interpret, by screening rapidly through a large data file for spectral patterns of interest. The greatest advantage is to be gained in the analysis of complex mixtures, particularly of unknown compounds. Extracts of biological fluids or tissues, or environmental matrices often provide this type of difficult analysis. To illustrate the utility of selective data reduction methods, we have chosen two examples of analyses from our laboratory.

Fuel oil is a very complex petroleum distillate fraction, composed primarily of hydrocarbons. For a number of reasons it is useful to be able to characterize these materials in terms of the saturated aliphatic hydrocarbons, olefins, and aromatic compounds present, since these fractions will have a direct bearing on fuel potential, toxicity, and environmental persistence of the oil. Many characterization schemes have been proposed based on HPLC (10), TLC (11), MS (12), or GC-MS (13). Mass chromatograms can be used to search GC-MS data sets for compound classes, for example, the m/z 57 and m/z 71 ions are present in the spectra of saturated alkanes; however, because of multiple possible formulae for these ions, the use of mass chromatograms alone lacks specificity. Gallegos has suggested that high resolution GC-MS could distinguish between various molecular fragments (14), but such an approach requires very sophisticated and expensive instrumentation. Varmuza adopts an elegant computerized chemometric approach (15) with some success.

Figure 1 shows a typical chromatogram that results from GC-MS analysis of a fuel oil sample. Hundreds of peaks are resolved; and the complete interpretation of such a data set could take many hours or days. Rather than adopt such a "brute force" approach, we prefer to use selective data reduction to simplify the task to more

manageable proportions. The mass spectra of saturated hydrocarbons show a very characteristic pattern of ions at low mass (<m/z 120). A series of ions is formed, each separated by 14 amu ($-CH_2-$) at

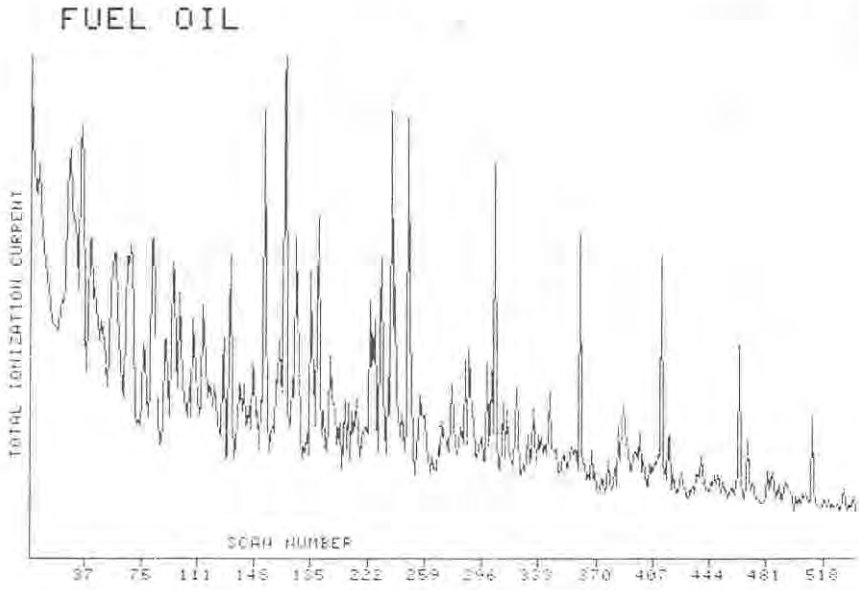

Fig. 1 The total ion chromatogram resulting from GC-MS analysis of fuel oil.

masses 43, 57, 71, 85 ... . Figure 2 shows the mass spectrum of hexadecane; the ion series at low mass is quite apparent. If one would like to quickly locate similar saturated hydrocarbons in the fuel oil data set, one might perform selective data reduction as follows: A "pattern" of ions would be defined, in this situation, the pattern of hexadecane ions between m/z 50 and 100 is used (boxed portion of Fig. 2). The computer is instructed to search through the data set (540 mass spectra) and identify other spectra that display this same pattern. The result is the class-selective chromatogram in Figure 3a. A much simpler chromatogram is obtained, relative to the total ion chromatogram in Figure 3b. The analyst can easily interpret the mass spectra of these few compounds and identify a series of saturated hydrocarbons: decane (scan 40), undecane (scan 81), dodecane (scan 133), tridecane (scan 190), tetradecane (scan 248), pentadecane (scan 305), hexadecane (scan 360), heptadecane (scan 413), octadecane (scan 463), and nonadecane (scan 511).

The smaller peaks in between these major components are due to various branched alkanes in the fuel oil. It is worth noting that in the simplification of the data set, the level of background has been greatly reduced. Also, since these methods are simply a computerized method of data presentation, the original data are in no way altered by the searching process. The full mass spectrum of

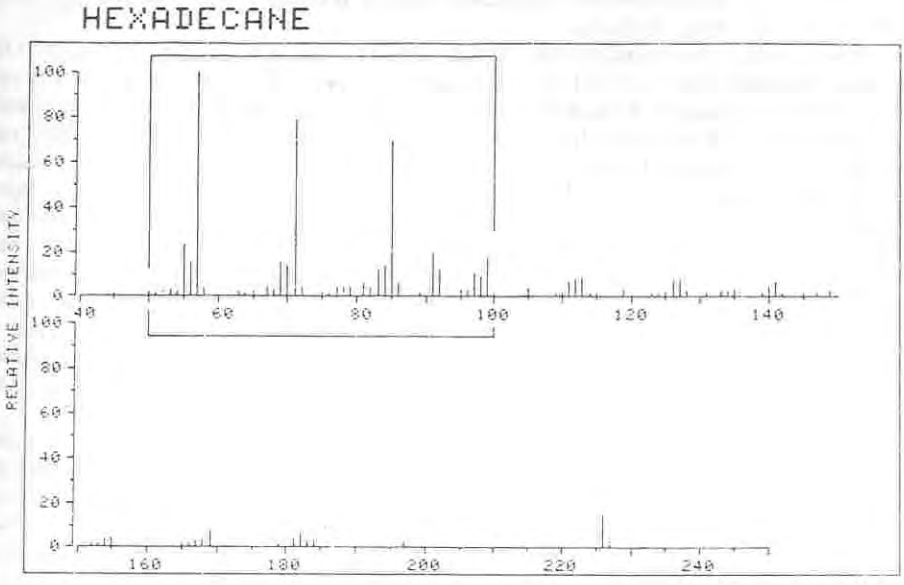

Fig. 2 Mass spectrum of hexadecane. Boxed portion indicates the pattern used for selective data reduction.

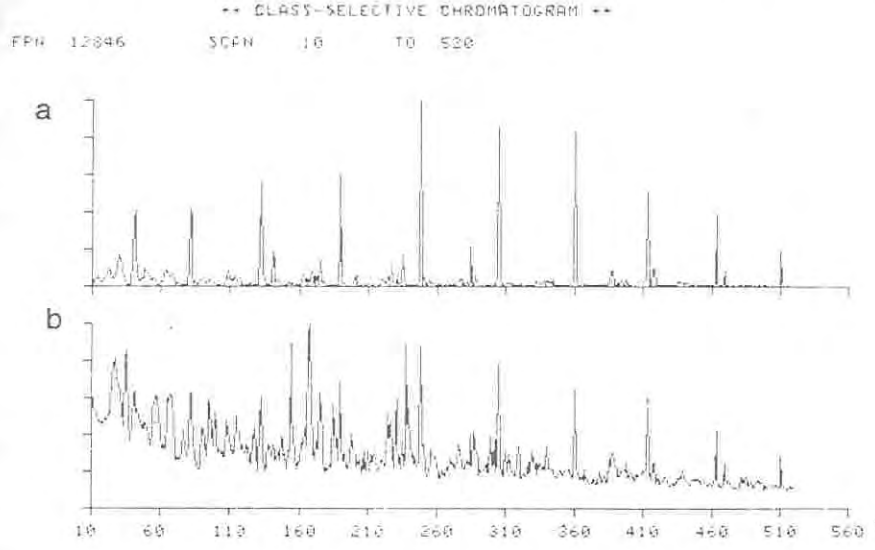

Fig. 3 a) Class selective chromatogram resulting from a search for saturated hydrocarbons and b) total ion chromatogram from the GC-MS analysis of fuel oil.

any component can be retrieved and examined at any time. The selective data reduction simply tells the analyst which spectra are worth examining. Finally, since the spectra are all stored, the data can be re-examined or searched for a different spectral pattern at any time in the future.

The mass chromatogram that would result from having the computer search for m/z 57 in the data set would also have retrieved all of the saturated alkanes. However, because many other compounds also have an ion at m/z 57 (for example, branched aromatic species with a C_4 or larger branch, alkenes, or any alcohol, acid or ketone with a hydrocarbon chain longer than four carbon atoms), many non-alkane species would also be observed. By using the masses and intensities of the other ions in the pattern, we improve the selectivity of the data reduction and locate only saturated alkanes. One can envision similar searches which might use the spectral pattern of alkenes or aromatic hydrocarbons to generate chromatograms that reflect where in the data set these compound classes occur. We are, in fact, continuing to develop such searches.

Although our primary goal is to increase the efficiency of data analysis, the class-selective chromatograms described have another important potential use. Depending on the source and previous history of an oil sample, the distribution of the various components in the sample changes. Upon weathering, for example, many of the more volatile components are lost by evaporation, leading to a relative enrichment of the heavier components. While subtle changes in distribution of components is difficult to observe in the very complex total chromatogram of the fuel oil, changes in the much simpler class-selective chromatograms ought to be relatively easy to detect. In this way, a sample can be, in some sense, characterized, or "fingerprinted" by these class-selective chromatograms, and the changes in distribution of components more readily documented.

A second application of the selective data reduction methods is in the analysis of the secretion from the scent gland of the African civet *(Viverra civetta)*. The musk from these animals is important to the perfume industry, and civet extract is an important commerical export of Ethiopia. Partial chemical characterization of the secretion was done by Van Dorp in 1973 (16). Civetone, a macrocyclic unsaturated ketone is by far the major component. The minor constitutents in the commerical extract have been less well characterized. A sample of the extract, provided by Dr. Berhanu Abegaz of the Chemistry Department of Addis Ababa University, was analyzed by GC-MS in our laboratory. Mass spectra were collected every 3 seconds during the course of the GC separation. A large number of peaks appeared (Figure 5b), with the largest component (scan 186) being civetone, as expected. The spectrum of one of the minor components (scan 136) is shown in Figure 4 and arises from the ethyl ester of tetradecanoic acid.

The pattern of ions at low mass, particularly the ions at m/z 88 and 101 are very diagnostic of fatty acid ethyl esters. To quickly search for other ethyl esters in the sample, this pattern, m/z 80 to 120 (boxed portion of Figure 4), was used for selective

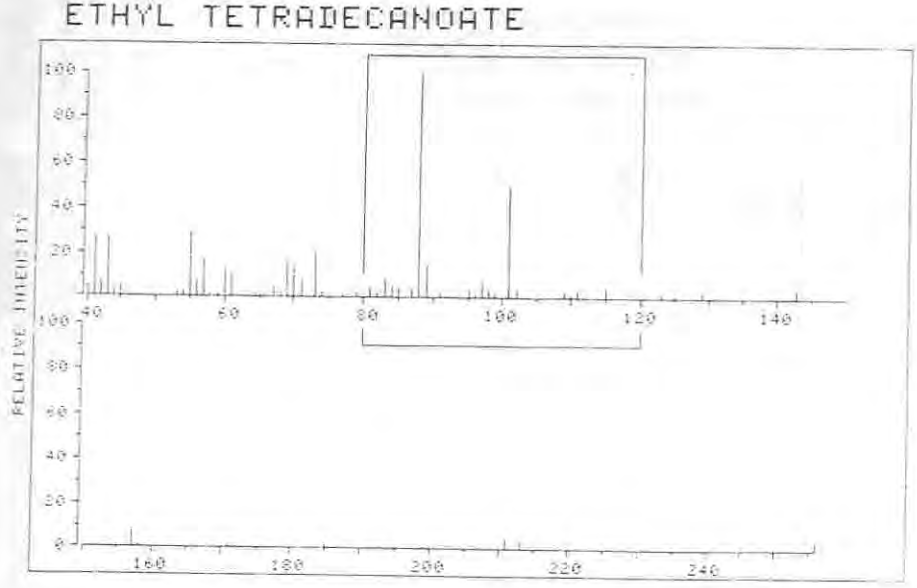

Fig. 4 Mass spectrum of ethyl tetradecanoate, from scan 136 of the data set in Fig. 5. Boxed portion indicates the pattern used for selective data reduction.

data reduction. The computer was instructed to search all spectra for the presence of this pattern. The result is the class-selective chromatogram in Figure 5a. A number of other ethyl esters are indicated: ethyl decanoate (scan 36), ethyl dodecanoate (scan 85), ethyl hexadecanoate (scan 184), and ethyl octadecanoate (scan 228). The analyst can go directly to these spectra for interpretation, momentarily ignoring all of the other spectra in the data set. In a similar fashion, data reductions using different patterns would result in other chemical classes being retrieved. The occurrence of these ethyl esters in the civet extract is somewhat surprising, but since extensive use of ethanol is made in the sample preparation, it is possible that they are artifacts of the work-up resulting from the ethylation of fatty acids.

One final point should be made regarding the data reduction methods. In the examples above, the compound class being sought was known in advance, because the mass spectrum used for a reference pattern had been previously interpreted. This need not be the case. Since what the computer is seeking is a pattern of ions, one need not know in advance what compound class that pattern represents. The results of the data reduction would indicate if more components in the mixture belonged to that same (as yet undefined) class, and where in the data set the mass spectra of those compounds could be found. The collection of spectra could then be considered and interpreted as a group.

In conclusion, the use of mass spectral patterns of ions that

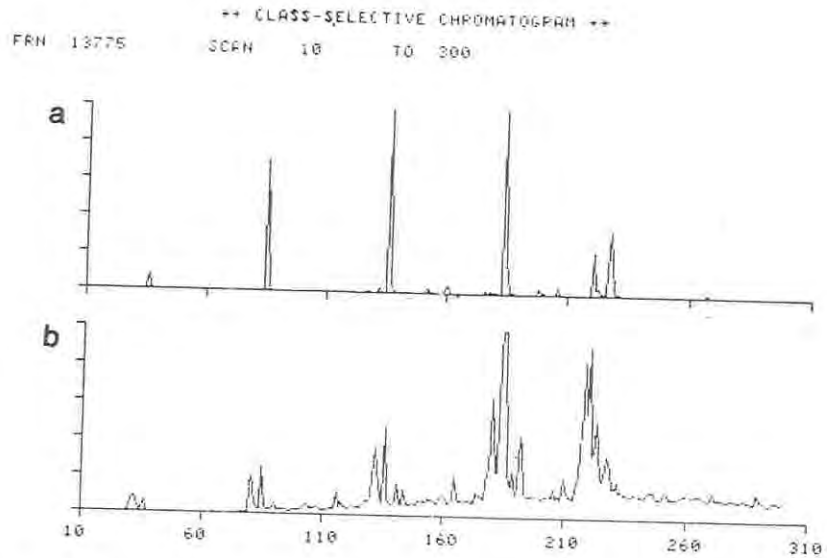

Fig. 5 a) Class-selective chromatogram resulting from a search for fatty acid ethyl esters, and b) total ion chromatogram from the GC-MS analysis of the extract of secretion from scent glands of the civet.

result from unimolecular decompositions of classes of compounds shows great potential in selective data reduction. Members of a compound class can be quickly and efficiently located and interpreted without the necessity of examining every mass spectrum in the data set. The class-selective chromatograms can also be helpful in fingerprinting complex samples for their further characterization.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the valuable assistance of A. Ghosh and S. Brackett of the University of Maine, and helpful discussions with Berhanu Abegaz, Ermias Dagne, and their colleagues at Addis Ababa University. This work was supported in part by the National Science Foundation (USA), grant number CHE-8209056.

REFERENCES

1. M.A. Grayson, J. Chromatogr. Sci., **24**, 529(1986).
2. R.J. Anderegg, Internatl. Laboratory, **16**, 62(1986).
3. R.A. Hites and K. Biemann, Anal. Chem., **42**, 855(1970).
4. D.H. Smith, Anal. Chem., **44**, 536(1972).
5. J.T. Clerc, M. Kutter, M. Reinhard, and R. Schwarzenbach, J. Chromatogr., **123**, 271(1976).

6. D.W. Kuehl, Anal. Chem., **49**, 521(1972).
7. J.L. LaBrosse and R.J. Anderegg, J. Chromatogr., **314**, 83(1984).
8. J.L. LaBrosse and R.J. Anderegg, J. Chromatogr., **314**, 93(1984)
9. R.J. Anderegg, Anal. Chim. Acta, **176**, 175(1985).
10. P.C. Hayes, Jr., and S.D. Anderson, Anal. Chem., **58**, 2384(1986).
11. T.G. Harvey, T.W. Matheson, and K.C. Pratt, Anal. Chem., **56**, 1277(1984).
12. R.M. Teeter, Mass Spectrom. Rev., **4**, 123(1985).
13. T.L. Youngless, J.T. Swansiger, D.A. Danner, and M. Greco, Anal. Chem., **57**, 1984(1985).
14. E.J. Gallegos, Anal. Chem., **56**, 701(1984).
15. H. Lohninger and K. Varmuza, Anal. Chem., **59**, 236(1987).
16. D. A. Van Dorp, R. Klok, and D.H. Nugteren, Recl. Trav. Chim. Pays-Bas, **92**, 915(1973).