SOME STUDIES ON MASS SPECTROMETRY.

A thesis presented for the degree of

DOCTOR OF PHILOSOPHY

to

The University of Glasgow

by

IAN JARDINE.

The Chemistry Department.

October, 1973.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

To my parents and brothers.

C O N T E N T S.

ACKNOWLEDGEMENTS.

The author wishes to express his sincere gratitude to his supervisor, Dr. R.I. Reed, for his ideas, advice and encouragement during this work.

He also thanks his colleagues, Dr. M.E.S.F. Silva, Dr. J.McK. Halket, Miss E. Jardim, Dr. F.G. Jiminez, Mr. A. McLean and Mr. J. Hardy for helpful discussions and support. He is also grateful to Dr. S.-O. Lawesson for samples and Mrs. C. Kennedy for typing the manuscript.

The provision of a maintenance grant by the Institute of Petroleum is gratefully acknowledged.

Summary of Ph.D. Thesis by Ian Jardine

entitled

"Some Studies On Mass Spectrometry."

Chapter One is a short historical introduction to the subject of Mass Spectrometry. Current research in this field is briefly reviewed to place the work of this thesis in perspective.

Chapter Two is the main body of the work and is entitled "An Application of Pattern Recognition to Mass Spectral Data." The particular problemn being considered, the detection and identification of highly toxic atmospheric pollutants, is discussed and the applicability of pattern recognition techniques is explored. The general ideas and methods of pattern recognition are mentioned. The particular unsupervised method of pattern recognition, "Cluster Analysis", is described in greater detail and an outline of the mathematics involved in employing the method is given.

Four separate studies on cluster analysis of mass spectral data are described and discussed. The first study is on ninety compounds of various types containing only carbon, hydrogen and sulphur. They are taken from the classes non-cyclic thioethers, cyclic thioethers, thiophenes and thiols. The second study is on sixty compounds which are recognised as organic atmospheric pollutants. The third study is on twenty-two different pyrazines. The forth study is on the compounds found as volatile metabolites in normal subjects (forty-two compounds) and subjects with diabetes mellitus (sixty compounds).

These studies show that the pattern recognition approach to the analysis of large quantities of mass spectral data is of considerable potential, especially for data reduction and compound identification.

Chapter Three contains the mass spectra of thirty-five various benzo - 2,1,3 - thiadiazoles, many of which have been shown to have considerable herbicidal and fungicidal activity. The mass spectra of all the compounds are discussed.

An additional publication is included after Chapter Three. This is "A Coincidence Time-of-Flight Mass Spectrometer," by A. Harvey, I. Jardine, M-de-L.F.Monteiro and R.I. Reed, from "Dynamic Mass Spectrometry," Vol. 3, edited by D. Price, Pub. Heyden and Son Ltd., London 1972.

CHAPTER 1

INTRODUCTION

It is virtually impossible to peruse a current scientific journal or text-book without coming across an article devoted to, or at least containing some reference to, a mass spectrometer. However, it is not surprising that the study of one of the basic aspects of matter, its mass, prompts wide and intensive study for its own sake and finds a myriad of uses and applications in all branches of science.

Goldstein[1] in 1886 first observed positively charged electrical particles and Wien[2] in 1898 demonstrated the basic principle of mass spectrometry, the separation and registration of atomic masses by deflecting positive rays in electric and magnetic fields. J.J. Thomson[3] (1912) similarly demonstrated the existence of two neon isotopes.

Instruments of more sophisticated design were built by Dempster[4] (1918) and Aston[5] (1919). Aston's mass spectograph was particularly useful for accurate mass measurements as ions could be velocity focused and collection was made on a photographic plate placed in the focal plane. Velocity focusing is the focusing of a beam of ions, homogeneous in mass, moving in the same initial direction, but at different speeds. Dempster's instrument allowed more accurate measurement of abundances while being less suitable for mass measurement. The instrument had the property of direction focusing. That is, charged particles of a given mass and energy diverging from a slit in a magnetic field were focused after deflection through 180° and were collected, one species at a time in a Faraday Cup.[6]

Resolution[7] was about 100, where resolution or resolving power is defined as the largest mass at which two adjacent peaks of equal height, differing by one unit of mass, exhibit a valley not greater than a certain percentage (usually 10%) of the peak height. When a photographic plate is the ion detector the term mass spectrograph is used, and when detection is by electrical means, mass spectrometer is common.

Barber[8] and Stephens[9] showed that any wedged shaped magnetic field could achieve direction focusing and a variety of sector angles are employed[10] in modern mass spectrometers. The design of modern single focusing mass spectrometers, usually using electron multiplier detection,[11] allow good sensitivity[7] with resolution up to about 10,000.

Instruments which incorporate the two principles of velocity and direction focusing are known as double focusing instruments. The theory of such instruments was developed by Herzog.[12] Modern double focusing instruments are of two main designs.

The Mattauch - Herzog[13] geometry employs consecutive electrostatic and magnetic fields which deflect the ions in opposite directions. The ions come to focus in the same plane, where they are collected usually on a photographic plate.[14] In the Nier - Johnson[15] design the consecutive electrostatic and magnetic fields deflect the ion beam in the same direction and the ions are focused at the intersection of the velocity focusing curve with the direction focusing curve. Each mass over charge ratio is focused, usually on an electrical detector, for a related value of the magnetic field. At the expense of sensitivity, these instruments can give very high mass resolution when very narrow slits are used and second order image aberrations are corrected.[16] Resolution of 1,000,000 has been obtained.[17]

Various methods may be employed to produce ions. The most usual method is electron bombardment using sources based on the efficient Nier design.[18] There is some [illegible] focusing within this ion gun and hence it can be used with single focusing (direction focusing) instruments.

Other methods of ion production which are finding increasing application are photon impact,[19] field ionization,[20] high voltage spark,[21] and chemical ionization.[22]

Many other types of instrument which use different principles of mass analysis, have been developed. The time-of-flight mass spectrometer[23] effects mass separation of a pulse of ions of various masses and same initial kinetic energy by the differences in flight times of ions of different mass when they are accelerated along a field-free tube. Radio-frequency techniques have been used to achieve mass separation in instruments of the Bennett[24] and Redhead[25] types. The quadrupole mass spectrometer[26] employs a radio-frequency field in conjunction with a static field.

These instruments are generally of lower resolving power and mass range than those employing magnetic deflection, but, as they have no magnet, they are generally less bulky.

Instruments using combinations of radio-frequency fields with magnetic fields, such as the omegatron,[27] the mass synchrometer,[28] and the ion cyclotron resonance (I.C.R.) mass spectrometer[29] are also used. The I.C.R. instrument is currently the most useful, especially for ion-molecule reaction studies.

Mass spectrometers only became reliable as chemical analytical instruments after the electronic advances of the Second World War. The petroleum industry was the first to recognise the potential of the method and it was employed extensively in the study of hydrocarbon mixtures.[30]

Only gases were studied until the advent of heated inlet systems,[31] whereby liquid samples of low vapour pressure could be introduced to the ionization chamber via a leak. However, it was not until the development of methods[32] for the direct insertion of sample into the ionization chamber that involatile compounds, thermally unstable compounds and solids could be analysed. Samples can now be placed on a probe and inserted directly into the ionization chamber without breaking the mass spectrometer vacuum by using vacuum lock techniques. By this method only a small quantity ($\sim 10^{-9}$g) of material is required for analysis. This sensitivity is one of the major advantages of mass spectrometric methods of chemical analysis.

The analysis of mixtures of compounds has received much attention. Methods are being developed[33] to permit the identification of the components of a mixture when it is not possible or convenient to separate them. The combination of gas-liquid chromatography (g.l.c.), for separation of components of a mixture, and a fast scanning mass spectrometer, for analysis of the individual components, has been widely applied[34] with great success. The development of molecular separators[35] for the removal of carrier gas from the g.l.c. effluent has been an integral part of this success. Currently, development is underway to couple liquid chromatography columns to mass spectrometers.[36] The liquid column has the advantage over the g.l.c. column of not requiring volatile compounds or derivatives for resolving mixtures.

A vast number of compounds have now been examined by mass spectrometry.[37] Generally, when the spectrum of a pure unknown compound is obtained it can be compared with files of the mass spectra of known compounds.[38] Computers are frequently used,[39] either on or off line, for this type of processing.

When the spectrum does not compare satisfactorily with the references, or no references are available, a more detailed study of the spectrum is necessary.

Generally, an empirical and well proven approach, which uses the same basic principles as organic solution chemistry i.e. resonance; inductive and steric effects; stabilities of carbonium ions; etc., is used. The formation of abundant ions in the mass spectrum was suggested by McLafferty[40] to depend on the relative stabilities of the ion and neutral fragment formed by decomposition of an ion, on the strengths of the bonds of the decomposing ion, and on the possibility of fragmentation through a transition state. One of the major assumptions of this approach was that of charge localization at favoured positions in the molecule.

Djerassi[41] and his school have used the charge localization concept extensively to rationalize mass spectral fragmentations. Biemann[42] proposed a set of empirical rules, without the charge localization principle, summarizing known fragmentations, and with Mandelbaum[43] he has shown that charge is not irreversibly localized at any particular site in the molecule.

A theoretical approach to the reactions occurring in the mass spectrometer has also been advanced to rationalize mass spectral data. The quasi-equilibrium theory (Q.E.T.), initially developed by Rosenstock et al[44] in 1952 and subsequently improved,[45] assumes that

i) the electron impact ionization of a molecule is a Franck - Condon Transition,

ii) the time of residence of the molecular ion in the source is sufficient to allow any excess electronic energy to be randomly distributed over the molecular ion,

iii) because of the energies (10 - 80 eV) normally used in electron impact sources, the ions formed are distributed over a large number of electronically excited states,

iv) ions with sufficient energy will decompose unimolecularly and, if the internal energy is sufficient to allow n decomposition pathways, these will usually occur as competing unimolecular decompositions.

The rate constant for each of these pathways can be calculated provided it is possible to determine a detailed knowledge of the energetic states of the ion.

Fair agreement with experimental results has been obtained with the Q.E.T. with some small molecules[46] using a number of approximations. Recently quantitative spectral features have been explained using Q.E.T.[47]

The effectiveness of mass spectrometric analysis was greatly enhanced when Beynon[48] demonstrated that differences in the nuclear packing fraction[49] of the elements make it possible to determine the elemental constitution of an ion. High resolving power of a double focusing mass spectrometer is required. Some considerable labour may be involved in obtaining such data for each ion. However, Biemann[50] introduced a system whereby a computer was used for the calculation of the masses of all the ions in the spectrum. The data presentation was in the form of an "element map" - separate columns being used for different heteroatom content. Burlingame[51] and McLafferty[52] use other representations of the high resolution data.

The interpretation of high resolution spectra[53] again requires the consideration of fragmentation patterns. Computers have been extensively used to ease this interpretation,[54] research into which is currently being tackled in three ways.

Firstly, the school of Isenhour[55] has employed a computerized learning machine approach[56] to this problem. Learning involves applying feedback to a decision process to minimize the probability of undesired responses. The process requires a training set of mass spectra and from this set pattern classifiers are empirically developed correctly to determine molecular formulas and structures. General techniques of pattern recognition are being developed[57] to try and provide a generalized approach to the data analysis problem of mass spectrometry. Chapter 2 of this work is a study of a pattern recognition approach applied to the interpretation of mass spectral data.

Secondly, a very sophisticated approach has been used by Lederberg and coworkers.[58] Their Dendral project aims at emulating in a computer programme the inductive behaviour of the scientist interpreting mass spectral data. Given the data of the mass spectrum of an unknown compound, a workable number of plausible solutions are induced. The programme then deduces the mass spectrum predicted by the theory of mass spectrometry for each of the candidates and selects the structure whose predicted spectrum most closely matches the data. The programming for this approach is complex and the computer facility used is large but results can be obtained more quickly and more efficiently than a human interpreter. High resolution data of complex molecules have also been elucidated in this approach.[59]

Thirdly, McLafferty[60] uses an intermediate approach. This is a self-training interpretive and retrieval system for unknown mass spectra. The system utilizes directly data of all available reference spectra, but does not require prior spectra or structure correlations of these data either by computer or human. However, classes of spectral data which are known to have high structural significance, such as characteristic ions, series of ions, and masses of neutrals lost, are defined for the computer.

The study of the physics of the reactions occurring in the mass spectrometer is of paramount importance in the understanding of the subject. Heats of formation of organic ions, appearance potentials, ion decomposition rates, scrambling and rearrangement reactions, substituent effects, ion - molecule reactions, metastable ions etc., all receive considerable attention.[61]

In particular, metastable ions have been a source of much information to the organic mass spectroscopist.[62] They have become increasingly important because they provide a means of determining reaction pathways and thermodynamic information about mass spectrometric processes. McLafferty[63] and coworkers have used metastable ions extensively to deduce ion structures, while Beynon[64] and coworkers have obtained and used metastable data for a variety of kinetic energy studies.

Systems have been devised[65] which permit the detection of metastable peaks with considerable sensitivity and independent of normal ions. In the technique of Barber[65(d)] the Ion Kinetic Energy Spectrum (I.K.E.S.) obtained displays all the metastable peaks. To identify the metastable transitions the I.K.E.S. must be mass analysed which necessitates the setting of the electrostatic and magnetic fields at values related to both the mass of the parent and the fragment ions. Computer control and processing of scanning is usually necessary.[66]

Recently the study of metastable decompositions after the magnetic sector, that is, after mass analysis, has been made either using scintillation collection[67] or using an electrostatic sector for energy analysis.[68]

Ion decompositions and structures are currently being studied by the use of isotopic labelling.[69]

Useful isotopes are deuterium, carbon - 13, nitrogen - 15, and oxygen - 18. Care must be taken when using isotopes, however, since scrambling of atoms and isotope effects, especially with deuterium,[70] occur.

Ion structures are also studied using ion - molecule reactions in medium and high pressure sources[71] and in I.C.R. spectrometers.[42]

As well as studies of ion - molecule reactions, the high pressure source technique, chemical ionization, is becoming very popular in organic chemical analysis since much simpler mass spectra are commonly produced and identification of the examined compound is made easier.[73]

From this brief introduction it can be seen that the field of mass spectrometry encompasses many disciplines. Many others,[74] such as inorganic applications, organometallic compound analysis, chemical physics applications, including reaction kinetics, surface and solid state studies, and negative ion studies, have not been mentioned although they are no less important.

REFERENCES.

1. E. Goldstein, Berl. Ber., 1886, 39, 691.

2. W. Wien, Ann. Physik, 1898, 65, 446.

3. (a) J.J. Thomson, Phil. Mag., 1911, 21, 225;
(b) J.J. Thomson, "Rays of Positive Electricity," Longmans, Green and Co., London 1913.

4. A.J. Dempster, Phys. Rev., 1918, 11, 316.

5. F.W. Aston, Phil. Mag., 1919, 38, 707.

6. J.H. Beynon, "Mass Spectrometry and Its Applications to Organic Chemistry," Elsevier Publishing Company, Amsterdam, 1960, p. 200.

7. J. Roboz, "Introduction to Mass Spectrometry Instrumentation and Techniques," John Wiley and Sons, 1968, p. 11.

8. N.F. Barber, Proc. Leeds Phil. Lit. Soc., Sci. Sect., 1933, 2, 427.

9. (a) W.E. Stephens and A.L. Hughes, Phys. Rev., 1934, 45, 123;
(b) W.E. Stephens, Phys. Rev., 1934, 45, 513.

10. Reference 6, p. 9.

11. (a) J.S. Allen, Phys. Rev., 1939, 55, 336;
(b) L.G. Smith, Rev. Sci. Instr., 1951, 22, 166.

12. R.F.K. Herzog, Z. Physik, 1934, 89, 447.

13. J. Mattauch and R.F.K. Herzog, Z. Physik, 1934, 89, 786.

14. D.M. Desiderio, Jr., in "Mass spectrometry, Techniques and Applications," ed. G.W.A. Milne, Wiley - Interscience, 1971, p. 11.

15. E.G. Johnson and A.O. Nier, Phys. Rev., 1953, 91, 10.

16. H. Matsuda, S. Fukumoto, Y. Kuroda, M. Nojiri, Z. Naturforsch., 1966, 21a, 25.

17. H. Matsuda, S. Fukumoto, T. Matsuo, M. Nojiri, Z. Naturforsch., 1966, 21a, 1304.

18. A.O. Nier, Rev. Sci. Instr., 1940, 11, 212 and 1947, 18, 398.

19. (a) F.P. Lossing and I. Tanaka, J. Chem. Phys., 1956, 25, 1031;
(b) H.D. Beckey and F.J. Comes, in "Topics in Organic Mass Spectrometry," ed. A.L. Burlingame, Wiley - Interscience, 1970, p. 1.

20. (a) M.G. Inghram and R. Gomer, Z. Naturforsch., 1955, 10a, 863;
(b) Reference 19(b).

21. (a) A.J. Dempster, Proc. Am. Phil. Soc., 1935, 75, 755;
(b) A. Cornu, in "Advances in Mass Spectrometry," Vol. 4, ed. E. Kendrick, The Institute of Petroleum, London, 1968, p. 737.

22. (a) M.S.B. Munson and F.H. Field, J. Amer. Chem. Soc., 1965, 87, 3289;
(b) J.M. Wilson, in "Mass Spectrometry, Vol. 1, Specialist Periodical Reports, The Chemical Society," 1971, p. 2.

23. (a) A.E. Cameron and D.F. Eggers, Rev. Sci. Instr., 1948, 19, 605;
(b) W.C. Wiley and I.H. McLaren, ibid., 1955, 26, 1150.

24. W.H. Bennett, J. Appl. Phys., 1950, 21, 143.

25. P.A. Redhead, Canad. J. Phys., 1952, 30, 1.

26. (a) W. Paul and H. Steinwedel, Z. Naturforsch., 1953, 8a, 448;
(b) W. Paul and H. Raether, Z. Physik, 1955, 140, 262.

27. J.A. Hipple, H. Sommer and H.A. Thomas, Phys. Rev., 1949, 76, 1877; ibid., 1951, 82, 697.

28. (a) L.G. Smith, Rev. Sci. Instr., 1951, 22, 115;
(b) L.G. Smith and C.C. Damm, ibid., 1956, 27, 638.

29. J.D. Baldeschwieler, Science, 1963, 159, 263.

30. (a) H. Hoover and H.W. Washburn, Amer. Inst. Min. Met. Engineer, 1940, Tech. Publ. No. 1205; Calif. Oil World, 1941, 34, 21.
(b) H.W. Washburn, H.F. Wiley, S.M. Rock and C.E. Berry, Ind. Eng. Chem. Analyt., 1945, 17, 74;
(c) A.K. Brewer and V.H. Dibeler, J. Res. Nat. Bur. Stand., 1945, 35, 125.

31. (a) M.J. O'Neal and T.P. Wier, Analyt. Chem., 1951, 23, 830.
(b) V.J. Caldecourt, Analyt. Chem., 1955, 27, 1670.

32. (a) P. de Mayo and R.I. Reed, Chem. and Ind., 1956, 1481;
(b) J.H. Beynon, R.A. Saunders and A.E. Williams, Appl. Spectroscopy, 1963, 17, 63.

33. L. Fraser Monteiro and R.I. Reed, Internat. J. Mass Spectrometry Ion Phys., 1969, 3, 406.

34. "Biomedical Applications of Mass Spectrometry," ed. G.R. Waller, Wiley - Interscience, 1972.

35. (a) R. Ryhage and S. Wikström, Reference 14, p. 91;
(b) J.T. Watson, Reference 34, p. 29.

36. (a) M.A. Baldwin and F.W. McLafferty, Org. Mass Spectrom., 1973, 7, 1111;
(b) R.E. Lovins, S. Ellis, G. Tolbert and C. McKinney, 6th International Mass Spectrometry Conference, Edinburgh, 1973.

37. For example, "Eight Peak Index of Mass Spectra," from Mass Spectrometry Data Centre, Aldermaston, contains essential mass spectral data of >17,000 organic compounds.

38. R.A. Hites and K. Biemann, Reference 21(b), p. 37.

39. (a) F.W. McLafferty, in "Advances in Mass Spectrometry," Vol. 5, ed. A. Quayle, The Institute of Petroleum, 1971, p. 589;
(b) S.D. Ward, Reference 22(b), p. 253;
(c) R.G. Ridley, Reference 34, p. 177.

40. F.W. McLafferty, in "Mass Spectrometry of Organic Ions," ed. F.W. McLafferty, Academic Press, New York, 1962, Chapter 7.

41. H. Budzikiewicz, C. Djerassi and D.H. Williams, "Mass Spectrometry of Organic Compounds," Holden - Day, Inc., San Francisco, 1967, p. 9.

42. K. Biemann, "Mass Spectrometry," McGraw - Hill, New York, 1962.

43. A. Mandelbaum and K. Biemann, J. Amer. Chem. Soc., 1968, 90, 2975.

44. H.M. Rosenstock, M.B. Wallenstein, A.L. Wahrhaftig and H. Eyring, Proc. Nat. Acad. Sci. U.S.A., 1952, 38, 667.

45. H.M. Rosenstock, Reference 21(b), p. 523.

46. H.M. Rosenstock and M. Krauss, in "Advances in Mass Spectrometry," Vol. 2, ed. R.M. Elliot, Pergamon Press, 1963, p. 251.

47. (a) D.H. Williams, Reference 39(a), p. 569;
(b) I. Howe, Reference 22(b), p. 31.

48. J.H. Beynon, in "Advances in Mass Spectrometry," Vol. 1, ed. J.D. Waldron, Pergamon Press, London, 1959, p. 328.

49. F.W. Aston, "Mass Spectra and Isotopes," Edward Arnold and Co., London, 1942, p. 80.

50. K. Biemann, P. Bommer, D.M. Desiderio and W.J. McMurray, in "Advances in Mass Spectrometry," Vol. 3, ed. W.L. Mead, The Institute of Petroleum, London, 1966, p. 639.

51. A.L. Burlingame and D.H. Smith, Tetrahedron, 1968, 24, 5749.

52. R. Venkataraghavan and F.W. McLafferty, Analyt. Chem., 1967, 39, 278.

53. K. Biemann, Reference 19(b), p. 185.

54. (a) S.D. Ward, Reference 22(b), p. 253;
(b) Reference 39(c).

55. (a) P.C. Jurs, B.R. Kowalski and T.L. Isenhour, Analyt. Chem., 1969, 41, 21;
(b) P.C. Jurs, B.R. Kowalski, T.L. Isenhour and C.N. Reilley, ibid., 1969, 41, 690; ibid., 1969, 41, 695;
(c) T.L Isenhour and P.C. Jurs, ibid., 1971, 43, 20A;
(d) P.C. Jurs, ibid., 1971, 43, 22;
(e) W.L. Felty and P.C. Jurs, ibid., 1973, 45, 885;
(f) J. Schechter and P.C. Jurs, Applied Spect., 1973, 27, 30.

56. N.J. Nilsson, "Learning Machines," McGraw - Hill, New York, 1965.

57. (a) B.R. Kowalski and C.F. Bender, J. Amer. Chem. Soc., 1972, 94, 5632; Analyt. Chem., 1973, 45, 590;
(b) B.R. Kowalski, C.F. Bender and H.D. Shepherd, Analyt. Chem., 1973, 45, 617.

58. J. Lederberg, Reference 34, p. 193 and references therein.

59. D.H. Smith, B.G. Buchanan, R.S. Engelmore, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg and Carl Djerassi, J. Amer. Chem. Soc., 1972, 94, 5962.

60. F.W. McLafferty, R. Venkataraghaven, K.-S. Kwok and G. Psyna, Reference 36(b).

61. (a) I. Howe, Reference 22(b), p. 31;
(b) A.G. Harrison and C.W. Tsang, Reference 34, p. 135;
(c) M.M. Bursey and M.K. Hoffman, Reference 14, p. 373.

62. J.H. Beynon and R.M. Caprioli, Reference 34, P. 157.

63. F.W. McLafferty, T. Wachs and W.T. Pike, Reference 21(b), p. 153.

64. J.H. Beynon, Analyt. Chem., 1970, 42, 97A.

65. (a) K.R. Jennings, J. Chem. Phys., 1965, 43, 4176;
(b) J.H. Futtrell, K.R. Ryan and L.W. Sieck, J. Chem. Phys., 1965, 43, 1832;
(c) N.R. Daly, A. McCormick and R.E. Powell, Rev. Sci. Instr., 1968, 39, 1163;
(d) M. Barber, R.M. Elliot, ASTM E-14 Meeting on Mass Spectrometry, Montreal, 1964.

66. T.W. Shannon, T.E. Mead, C.G. Warner and F.W. McLafferty, Analyt. Chem., 1967, 39, 1748.

67. N.R. Daly, A. McCormick and R.E. Powell, Reference 39(a), p. 300.

68. (a) K.H. Maurer, C. Brunée, G. Kappus, K. Habfast, V. Schröder and P. Schulze, 19th Ann. Conf. on Mass Spectrometry and Allied Topics, Atlanta, 1971;
(b) Miss M.E.S.F. Silva and R.I. Reed, Reference 36(b).

69. (a) M.F. Grostic and K.L. Rinehart, Jr., Reference 14, p. 217;
(b) I. Howe, Reference 22(b), p. 53.

70. A.N. Yeo and D.H. Williams, J. Chem. Soc. (B), 1968, 26666.

71. P.F. Knewstubb, in "Mass Spectrometry and Ion Molecule Reactions," Cambridge University Press, 1969.

72. (a) G. Eadon, J. Dieckmann and C. Djerassi, J. Amer. Chem. Soc., 1969, 91, 3986;
(b) G.C. Good, R.M. O'Malley, K.R. Jennings, Nature, 1970, 227, 1093.

73. J.M. Wilson, Reference 22(b), p. 2.

74. See, for example, the categories listed by The Mass Spectrometry Bulletin, Mass Spectrometry Data Centre, Aldermaston.

CHAPTER 2

AN APPLICATION OF PATTERN RECOGNITION TO MASS SPECTRAL DATA

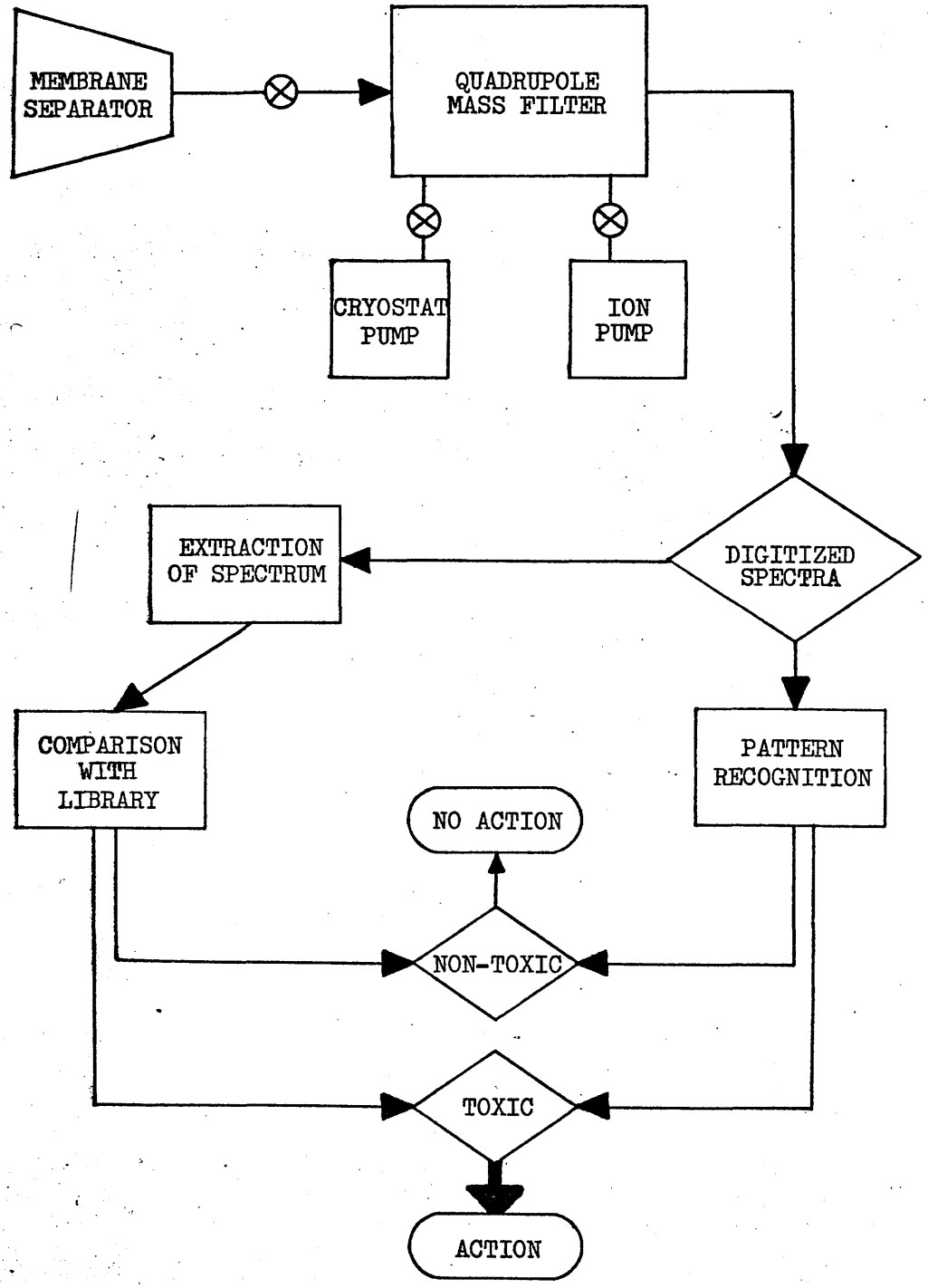
INTRODUCTION

The detection and identification of highly toxic atmospheric pollutants presents an interesting problem. Such pollutants, perhaps deliberately introduced into the atmosphere, would certainly be present in very low concentration, and would be required to be detected and identified, or at least classified, in the shortest time possible. The hardware for such a task should be small, rugged and portable.

A mass spectrometer would appear to be a suitable instrument because of its high detection sensitivity and rapid response. A quadrupole mass filter, in particular, is relatively small and can be made portable.

The problem of continuous introduction of atmospheric samples is solved by using a membrane separator[1] between the atmosphere and the quadrupole vacuum (10^{-3} Nm^{-2}, or 10^{-5} torr).[2] Such a separator allows concentration of minor components in the air because of the associated high enrichment factors of the inlet ($>10^{3}$). Also, since there is no dead space between the membrane and the atmosphere, the system responds rapidly to changes in composition of atmosphere. The low mass part of a spectrum is, however, obscured by the major components of the atmosphere.

FIG. 1


MEMBRANE SEPARATOR
QUADRUPOLE MASS FILTER
CRYOSTAT PUMP
ION PUMP
DIGITIZED SPECTRA
EXTRACTION OF SPECTRUM
COMPARISON WITH LIBRARY
PATTERN RECOGNITION
NO ACTION
NON-TOXIC
TOXIC
ACTION

This can be alleviated to some extent by using a two stage membrane separator.[1] However, apart from a reduction in sensitivity this would also increase the pumping problems.

Two problems are apparent with such a system:

a) The limited dynamic range of the mass spectrometer (10^6) would still make detection of very low concentration ($< 1 : 10^8$) material difficult.

b) Relevant mass spectral information about a pollutant must be abstracted from the large background, and the pollutant must be identified, or at least classified as toxic or non-toxic, as quickly as possible.

Two approaches to the second problem may be envisaged:

1. Successive spectra could be subtracted from each other or a ratio is taken. In this way, as in mixture analysis,[3] the mass spectrum of the compound which is being detected would appear, and this mass spectrum could then be compared in any of the usual ways[4] to a library of known toxics.

2. The spectra produced could be considered as an everchanging pattern and could be compared to known atmospheric situations to give an indication of the type of compound being detected.

Figure I describes the situation. The quadrupole is made portable by using a combination of a cryostat pump and an ion pump. The cryostat pump serves to create a rough vacuum. The pump can then be detached and the ion pump maintains the required quadrupole working vacuum.

The main work undertaken here is an investigation of the possibilities of using pattern recognition techniques for the purpose of abstraction of information from the spectra recorded in such a system. The pattern recognition technique investigated as a useful possibility is Cluster Analysis. The technique is investigated using mass spectral data and is shown to be useful. Other interesting uses are demonstrated.

PATTERN RECOGNITION[5-17]

Conceptualism is the philosophical theory that general ideas or patterns, separated from particular objects, exist in the mind. It is recognised that the decision making processes of a human being are related to the recognition of patterns. The purpose of 'pattern recognition' as used here is to elucidate the complex mechanisms of decision-making processes, to automate these operations using computers and hence to categorise or classify a given sample.

There are two stages in pattern recognition - deriving a decision rule and applying it. The pattern is defined by the labelled samples of that pattern. A decision rule is developed and is used to assign unique labels to new patterns. However, the concepts and algorithms of mathematical pattern recognition have other applications which evolve from this basic concept. Included are:-

a) The analysis of multivariate data to detect and extract complex qualitative and quantitative relationships among samples,

b) The determination, rather than the assignment, of 'natural' pattern classes,

c) The selection of measurements according to their usefulness in separating pattern classes,

d) The reduction of high-dimensional data to readily visualised two or three dimensions,

e) The extraction of a reduced data set of essential information from an unnecessarily large amount of data,

f) The labelling of a sample with a continuous number rather than a small finite set of class labels. The 'recognition' would then be the assignment of an appropriate continuous value to an unlabelled sample.

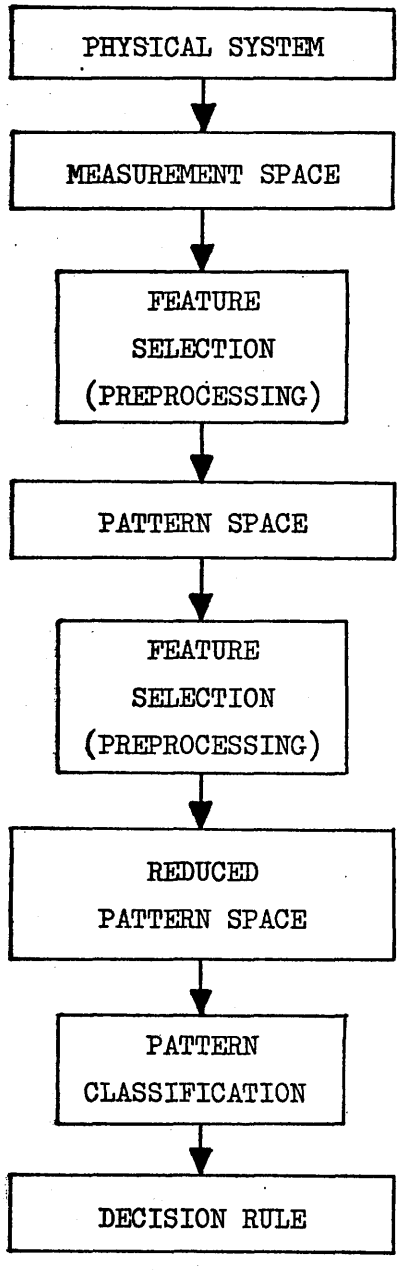
In fact, any problem for which exemplar data are available and a class label is required is a candidate for pattern recognition algorithms. A few examples of pattern recognition problems are:

visual,	character recognition, fingerprint classification, photograph interpretation,
auditory,	word or speaker identification,
engineering,	radar or sonar analysis, waveform classification,
biomedical,	electrocardiogram and electroencephalogram analysis, diagnosis based on tests and measurements,
other,	seismic data analysis, weather forecasting, prediction in economics and operations research, etc.

However, it should be remembered that an apparent application may be vetoed because there is not enough labelled samples, or there is no distinguishable pattern in the data, or there may simply be a quicker and better way to solve the problem.

FIG. 2

Stages in the derivation of the decision rule



In pattern recognition the stages involved in the derivation of a decision rule are shown in Figure 2.

The raw data which describes the physical system involved is called the measurement space. The pattern space must be finite-dimensional, of relatively low dimension, and must contain sufficient information to satisfactorily perform the classification. The pattern space and measurement space may be identical or processing of the measurement space may be necessary so that the pattern space fulfills the conditions given above.

In most physical systems the number of measurements, n, is high, often in the hundreds. This high dimensionality makes many pattern recognition problems difficult. A human being usually performs classification based on a small number of features. Obviously, as the number of inputs to a classifier becomes smaller, the design becomes simpler. Therefore, it is advantageous to select or extract important features from the observed samples. This problem of extraction is called feature selection or preprocessing. That is, it is the process by which a sample in measurement space is described by a finite and usually smaller set of numbers called features, say $x_1, x_2, \ldots\ldots, x_n$, which become components of the pattern space. Hence, a point in measurement space is transformed by the intermediate processing into a point $X = (x_1, x_2, \ldots\ldots\ldots, x_n)$ in pattern space.

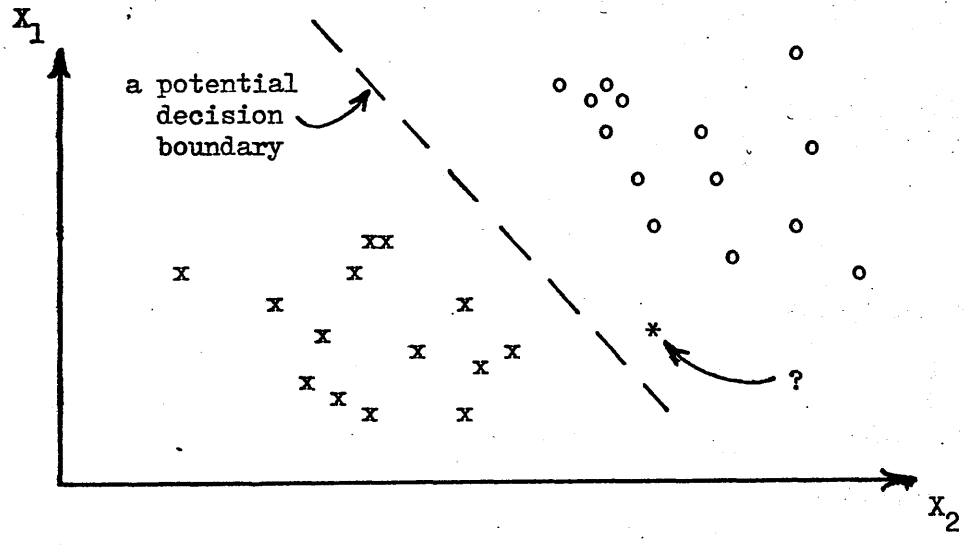
The process of deriving the decision rule from the finite set of labelled samples, with which a point in pattern space corresponding to an unlabelled sample might be classified, is called pattern classification.

Pattern recognition consists therefore of two main areas, which in practice may be difficult to separate. These are feature selection and classifier design.

BASIC CONCEPTS IN PATTERN CLASSIFICATION

A finite number of scalar variables or features, $x_1, x_2, \ldots\ldots, x_n$, describes a pattern. In the n - dimensional pattern space $\overline{X}$ which encloses the region in which patterns can occur, a particular pattern is a point $X = (x_1, x_2, \ldots\ldots, x_n)$. Points in the pattern space can be classified into a finite number of pattern classes $S_1, S_2, \ldots\ldots, S_N$ (S_i is "class i"). Except for a finite number of labelled samples of each class, the classes are unknown.

The geometrical concepts involved in pattern classification are deceptively simple. For example, consider the two-dimensional pattern space below where there are two pattern classes, S_1 and S_2, with labelled samples shown.



o samples of class S_1

x samples of class S_2

* unknown

The problem is to classify the unknown sample into S_1 or S_2. To do this a decision boundary, which separates the classes, is determined.

A decision rule or discriminant function will then assign any point on one side of the boundary to class S_1 and those on the other to S_2. Here, the problem of finding a good decision boundary is relatively easy. In higher dimensions, which occurs most often, theproblem becomes complex.

Discriminant functions are a convenient way to specify class boundaries. More meaningful, however, is the value of the function for a given sample. This characteristic discriminant function for a sample can indicate the degree of class membership - a more useful criterion since, in reality, class membership is not always obvious. A sample, therefore, will have a degree of membership in several classes and it can be called a member of the class to which it has the greatest degree of membership.

When applying pattern recognition algorithms it is generally assumed that most of the information about the nature of the system is to be found in the labelled samples. Therefore, a test of a decision rule is to apply it to a set of labelled samples and note the results to obtain an estimate of the error rate to be expected from that decision rule applied to unlabelled samples. For example, the set of labelled samples can be divided into two disjoint sets. The decision rule is derived from the design set and its accuracy is estimated using the test set. Alternatively, the 'leave one out' method determines a decision rule from the N total labelled samples minus one. The decision rule is tested on the isolated sample. The process is repeated leaving out each sample in turn (i.e. N times) and the error rate on the isolated samples gives an estimate of the performance of the decision rules. A final decision rule is chosen from the N found.

BASIC CONCEPTS IN FEATURE SELECTION

Feature selection is a difficult and only recently tackled problem in pattern recognition. The general requirements of a set of features is low dimensionality for a meaningful and computationally feasible classification algorithm, whilst retaining sufficient information to allow recognition of the patterns required. Each feature must also be determined consistently over all patterns, and they must maintain geometric consistency in the pattern space developed so that a small difference between two points corresponds to a small difference in the actual patterns in terms of the quality to be recognised.

The choice of features may be tackled in a number of ways.

If the system under examination is well understood features satisfying the above criteria may be defined intuitively. Alternatively, a large number of features may be defined which are supposed to be relevant to the quality to be recognised in the system. Using the samples, these features may then be quantitatively ranked to obtain a reduced set of the most relevant features. When a set of features is used to separate classes a linear transformation may be made of these features to optimise their classifying usefulness.

It should be noted that the division between pattern classification and feature selection is largely artificial and for convenience. Pattern classification is a form of feature selection and vice-versa.

SUPERVISED AND UNSUPERVISED LEARNING

In the above discussion it has been assumed that at least some of the points in the pattern space are labelled with a known classification which generally serve as a training set which is used to develop classification rules which are applied to unknown points. This type of method is called supervised learning.

However, if no labelled samples are known and the object is to determine useful and realistic densities or clusters of points in the pattern space $\underline{X}$ which will disclose meaningful relationships, no training set is used. This type of method is called unsupervised learning.

An unsupervised method of pattern recognition, Cluster Analysis, is employed in this work. However, labelled samples are also used to help to identify the natural pattern classes.

CLUSTER ANALYSIS

The classification of samples without the aid of a training set of labelled samples (as in supervised learning) is called cluster analysis or unsupervised learning. That is, the samples cannot be simply assigned class membership and cluster analysis is used to detect 'natural clusters' corresponding to 'natural classes' in the data. It is possible, therefore to break up the samples in a pattern classification problem into smaller groups so that a simpler pattern classification problem can be solved on each group. If a small set of labelled samples is also available, class membership may be extrapolated to unlabelled samples.

The number of clusters in a set of data and their location can give much information concerning the structure of the data. Hence, cluster analysis can be viewed as generalised correlation analysis (in certain applications). More, clustering also indicates anomalous points. That is, points which are widely separated from other sample points.

Clusters can be formed by grouping similar species and separating dissimilar ones. Similarity grouping is the most common form of clustering.

Clustering algorithms are generally applied to similarity matrices or correlation coefficient matrices. These matrices are computed using the recorded characteristics of the samples.

OUTLINE OF COMPUTATIONAL METHOD

Assuming reasonable feature selection, an original data matrix is set up of these characters of the samples, where the entries are of the form X_{ij} :-

	Samples							
Characters	1	2	3	...	...	t	Mean	S.D.
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	...	...	$X_{1,t}$	$\bar{X}_1$	s_1
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	...	...	$X_{2,t}$	$\bar{X}_2$	s_2
3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	...	...	$X_{3,t}$	$\bar{X}_3$	s_3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	$X_{n,1}$	$X_{n,2}$	$X_{n,3}$	...	...	$X_{n,t}$	$\bar{X}_n$	s_n

In order to compute the correlation or distance coefficients between each sample the characters must be standardised. To do this the mean and standard deviation of each character (i.e. of each row of the data matrix) is calculated.

The mean is calculated as:

$$\bar{X}_i = \frac{1}{t}\sum_{j=1}^{t} X_{ij}$$

The standard deviation is calculated as:

$$s_i = \left\{\frac{1}{t-1}\left[\sum_{j=1}^{t} X_{ij}^2 - \frac{\left(\sum_{j=1}^{t} X_{ij}\right)^2}{t}\right]\right\}^{\frac{1}{2}}$$

Standardised character states are then computed as:

$$X'_{ij} = \frac{X_{ij} - \bar{X}_i}{s_i}$$

Where X'_{ij} is the standardised character state code for character i and sample j, X_{ij} is the raw score for this character state, and $\bar{X}_i$ and s_i are the mean and standard deviation of character i, respectively. This produces a data matrix with characters standardised:

Characters	Samples 1	2	3	...	...	t
1	$X'_{1,1}$	$X'_{1,2}$	$X'_{1,3}$	...	...	$X'_{1,t}$
2	$X'_{2,1}$	$X'_{2,2}$	$X'_{2,3}$	...	...	$X'_{2,t}$
3	$X'_{3,1}$	$X'_{3,2}$	$X'_{3,3}$	...	...	$X'_{3,t}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	$X'_{n,1}$	$X'_{n,2}$	$X'_{n,3}$	...	...	$X'_{n,t}$

Correlation coefficients are now computed by:

$$r_{jk} = \frac{\sum_{i=1}^{n} X'_{ij}X'_{ik} - \frac{1}{n}\left(\sum_{i=1}^{n} X'_{ij}\right)\left(\sum_{i=1}^{n} X'_{ik}\right)}{\left\{\left[\sum_{i=1}^{n} X^2_{ij} - \frac{1}{n}\left(\sum_{i=1}^{n} X'_{ij}\right)^2\right]\left[\sum_{i=1}^{n} X'^2_{ik} - \frac{1}{n}\left(\sum_{i=1}^{n} X'_{ik}\right)^2\right]\right\}^{\frac{1}{2}}}$$

which is exactly equal to the formula:

$$r_{jk} = \frac{\sum_{i=1}^{n} (X'_{ij} - \bar{X}_j)(X'_{ik} - \bar{X}_k)}{\left[\sum_{i=1}^{n} (X'_{ij} - \bar{X}_j)^2 \sum_{i=1}^{n} (X'_{ik} - \bar{X}_k)^2\right]^{\frac{1}{2}}}$$

This produces a matrix of the product - moment correlation cofficients between all pairs of samples:

/samples:

Samples	Samples 1	2	3	4	5	...	t
1	*						
2	$r_{1,2}$	*					
3	$r_{1,3}$	$r_{2,3}$	*				
4	$r_{1,4}$	$r_{2,4}$	$r_{3,4}$	*			
5	$r_{1,5}$	$r_{2,5}$	$r_{3,5}$	$r_{4,5}$	*		
⋮	⋮	⋮	⋮	⋮	⋮	⋱	
t	$r_{1,t}$	$r_{2,t}$	$r_{3,t}$	$r_{4,t}$	$r_{5,t}$	...	*

Alternatively, a distance coefficient is calculated. That is, the Euclidean distance between two points in an n - dimensional space. The formula for such a distance, Δ_{jk}, between samples j and k is:

$$\Delta_{jk} = \left[\sum_{i=1}^{n} (X_{ij} - X_{ik})^2\right]^{\frac{1}{2}}$$

The square of this distance is often used i.e. Δ^2_{jk}, and since Δ^2_{jk} increases with the number of characters used in the comparison, an average distance is commonly computed. This is:

$$d_{jk} = \sqrt{\frac{\Delta^2_{jk}}{n}} \quad \text{or} \quad d^2_{jk} = \frac{\Delta^2_{jk}}{n}$$

This now produces a similarity matrix of average distance coefficients between all pairs of samples. (Other distance coefficients may be calculated which may suit a particular problem. For example, a weighted coefficient in which both the variance of separate characters and the correlations among characters are taken into account, may be computed).

That is,

That is,

Samples	Samples 1	2	3	4	...	t
1	*	$d^2_{1,2}$	$d^2_{1,3}$	$d^2_{1,4}$	...	$d^2_{1,t}$
2	$d_{1,2}$	*	$d^2_{2,3}$	$d^2_{2,4}$	...	$d^2_{2,t}$
3	$d_{1,3}$	$d_{2,3}$	*	$d^2_{3,4}$	...	$d^2_{3,t}$
4	$d_{1,4}$	$d_{2,4}$	$d_{3,4}$	*	...	$d^2_{4,t}$
⋮	⋮	⋮	⋮	⋮	⋱	⋮
t	$d_{1,t}$	$d_{2,t}$	$d_{3,t}$	$d_{4,t}$	...	*

Average distance coefficients between all pairs of samples.

CLUSTERING METHODS

There are a large class of numerical techniques for defining groups of related samples based on high similarity coefficients. Some methods may be more complex and more efficient than others depending upon the application. In general those samples most related are clustered and members are gradually admitted into the cluster by lowering the criteria of admission.

For example, in Single Linkage (Nearest Neighbour) clustering, the samples which are mutually related with the highest possible similarity coefficient are first clustered. The level of admission is then successively lowered by steps of equal magnitude. This method generally finds "straggling" clusters and usually fails with large populations because of chaining. With the Complete Linkage (Furthest Neighbour) method a given sample joining a cluster at a certain similarity coefficient must have relations at that level or above with every member of the cluster. This method generally finds spherical clusters but can produce irregular results since a comparison is effectively being made only between the individual seeking admission to the cluster and the individual in the cluster to which it is least similar. That is, no account is taken of group structure in the clusters.

The Average Linkage (Group Average; Unweighted Pair-group) method may be considered as intermediate between the two extremes of Single and Complete Linkage. Here the admission of any individual into a cluster is based on the average of similarities of that individual with the members of the cluster. Some account is therefore taken of group structure and reasonably spherical clusters tend to be found.

Other methods, such as Centroid Sorting (Weighted Group), Median (Gower's), Lance - Williams Flexible Beta, and McQuitty's Similarity Analysis try to take account of group structure to derive well behaved spherical clusters. However, currently recognised as perhaps the best method is Ward's Method.

Using distance coefficients, the method combines those two clusters whose fusion yields the least increase in the error sum of squares defined as,

$$\text{Ess} = \sum_{i=1}^{n} \phi_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} \phi_i\right)^2$$

where ϕ_i is the score of the i th individual. The method finds minimum variance spherical clusters.

It was recognised by Lance and Williams that, in fact, all of the conventional strategies noted above are simple variants of a single linear system defined by four parameters. That is, letting clusters p and q be fused, then the similarity, S(r,p+q), between cluster r and the new cluster (p+q) is obtained from the formula,

$$S(r,p+q) = \alpha_p S(r,p) + \alpha_q S(r,q) + \beta S(p,q) + \gamma \left| S(r,p) - S(r,q) \right|$$

where α_p, α_q, β and γ are assigned the values,

	α_p	α_q	β	γ
1 Single Linkage:	0.5	0.5	0	-0.5
2 Complete Linkage:	0.5	0.5	0	0.5
3 Centroid:	$k_p/(k_p+k_q)$	$k_q/(k_p+k_q)$	$-\alpha_p\alpha_q$	0
4 Median:	0.5	0.5	-0.25	0
5 Lance - Williams:	$(1-\beta)/2$	$(1-\beta)/2$	various	0
6 McQuitty's:	0.5	0.5	0	0
7 Average Linkage:	$k_p/(k_p+k_q)$	$k_q/(k_p+k_q)$	0	0
8 Ward's:	$(k_r+k_p)/(k_r+k_p+k_q)$	$(k_r+k_q)/(k_r+k_p+k_q)$	$-k_r/(k_r+k_p+k_q)$	0

where k_r, k_p and k_q are cluster sizes.

Hence all these methods are obtained from the same algorithm by varying the transformation parameters α_p, α_q, β and γ and can, therefore, be implemented in one computer programme.

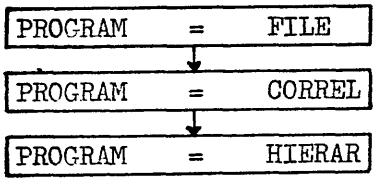
PROGRAMME[16,17]

The programme used for this study was CLUSTAN 1A, a package programme available from the Scientific and Social Sciences Programme Library of the University of Edinburgh (Cat. No. 19.200.200). The programme is written in Fortran IV and was run on the Edinburgh Regional Computing Centre (E.R.C.C.) I.B.M. 370/155 Computer through the satellite link in the Chemistry Department of the University of Glasgow. The author of the programme was D. Wishart and amendments for E.R.C.C. were by R.L. Middleton.

Broadly, CLUSTAN is a suite of programmes which are designed for the collective study of several methods of cluster analysis and other multivariate procedures.

The first program, FILE, creates a data file by reading observational data from cards or tape and storing it on a data file called CLUSDATA which is on magnetic disk. Thereafter, each CLUSTAN programme refers to the data file in order to evaluate an appropriate analysis.

For the purpose of this study, the programmes were called and used in the following order:



That is, the data file created by FILE is used by CORREL to compute the similarity matrix, distance coefficient matrices and other statistics. These results are required to execute further programmes. HIERAR uses these statistics and produces a fusion hierarchy by any of the eight clustering methods referred to above.

The programme prints the classification arrays so formed and dendograms can be drawn by hand or the array can be automatically punched on cards which can be submitted to PROGRAM = PLINK for automatic plotting.

PROGRAM = FILE, and hence all other programmes in the package, have a limitation on parameters. They are restricted to the following maxima:

Number of cases, individuals or samples	999,999
Number of continuous variables	120
Number of binary attributes	400

RESULTS

a) A study on the classification of ninety compounds of various types containing only carbon, hydrogen and sulphur was undertaken to:

i) Become familiar with the CLUSTAN package,

ii) Find the best form in which to use the mass spectral data,

iii) Test the eight hierarchical fusion options in the package to find the most useful for further application

The compounds, which are listed in Table 1, were carefully chosen from the classes,

i) Non-cyclic Thioethers (Sulphides)

ii) Cyclic Thioethers (Sulphides)

iii) Thiophenes

iv) Thiols

They were selected to give a broad range of structural types whilst incorporating some similar compounds, and to give a reasonable range of molecular weights. The mass spectra were taken from the A.P.I. collection where they are all normalised to 100% base peak.

The spectra were listed according to the intensity of their m/e values. For convenience the intensities were given an index according to the ranges

Range%	Index
0.00 - 9.99	0
10.00 - 19.99	1
20,00 - 29.99	2
⋮	⋮
80.00 - 89.99	8
90.00 - 100.00	9

TABLE I

Non-cyclic Thioethers (Sulphides)

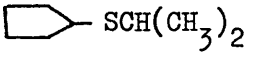
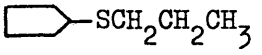
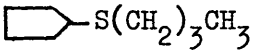
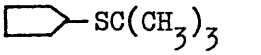
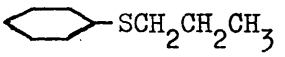
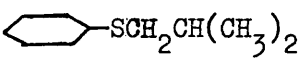
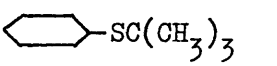
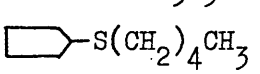
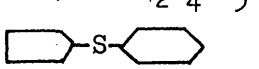
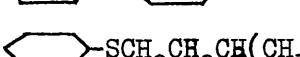
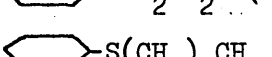
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
1	577	118	3-Thio heptane	$CH_3CH_2S(CH_2)_3CH_3$
2	922	118	2,2-Dimethyl-3-thio pentane	$CH_3C(CH_3)_2SCH_2CH_3$
3	727	146	2-Thio nonane	$CH_3S(CH_2)_6CH_3$
4	1025	118	5-Methyl-3-thio hexane	$CH_3CH_2SCH_2CH(CH_3)_2$
5	1230	146	2,2,4,4-Tetramethyl--3-thio pentane	$(CH_3)_3CSC(CH_3)_3$
6	1373	146	3,5-Dimethyl-4-thio heptane	$CH_3CH_2CH(CH_3)SCH(CH_3)CH_2CH_3$
7	1410	132	2,2,4-Trimethyl-3-thio pentane	$(CH_3)_3CSCH(CH_3)_2$
8	2147	144	(Isopropylthio)cyclopentane	
9	2148	144	(n-Propylthio)cyclopentane	
10	2149	158	(n-Butylthio)cyclopentane	
11	2158	158	(tert-Butylthio)cyclopentane	
12	2154	158	(n-Propylthio)cyclohexane	
13	2156	172	(sec-Butylthio)cyclohexane	
14	2157	172	(ter-Butylthio)cyclohexane	
15	2159	172	(n-Pentylthio)cyclopentane	
16	2160	184	(cyclopentylthio)cyclohexane	
17	2162	186	(Isopentylthio)cyclohexane	
18	2163	186	(n-Pentylthio)cyclohexane	

TABLE I (continued)

Non-cyclic Thioethers (Sulphides)

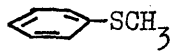
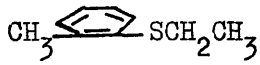
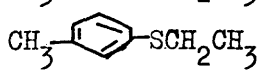
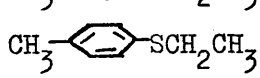
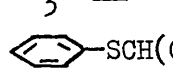
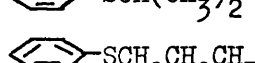
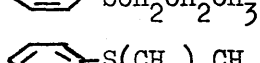
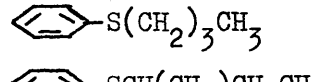
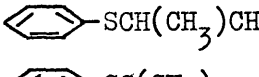
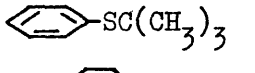
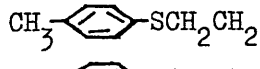
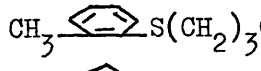
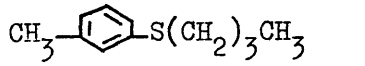
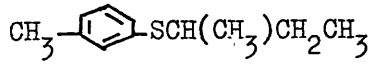
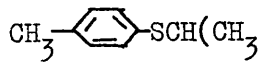
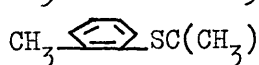
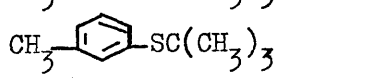
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
19	2195	124	(Methylthio)benzene	
20	2196	138	(Ethylthio)benzene	
21	2200	152	2-(Ethylthio)toluene	
22	2201	152	3-(Ethylthio)toluene	
23	2202	152	4-(Ethylthio)toluene	
24	2203	152	(Isopropylthio)benzene	
25	2204	152	(n-Propylthio)benzene	
26	2205	166	(n-Butylthio)benzene	
27	2206	166	(sec-Butylthio)benzene	
28	2207	166	(tert-Butylthio)benzene	
29	2214	166	4-(n-Propylthio)toluene	
30	2215	180	2-(n-Butylthio)toluene	
31	2216	180	3-(n-Butylthio)toluene	
32	2219	180	3-(sec-Butylthio)toluene	
33	2220	180	4-(sec-Butylthio)toluene	
34	2221	180	2-(tert-Butylthio)toluene	
35	2222	180	3-(tert-Butylthio)toluene	
36	2226	180	4-(Isobutylthio)toluene	

TABLE I (continued)

Cyclic Thioethers (Sulphides)

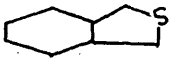
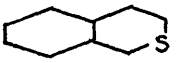
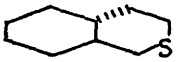
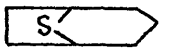
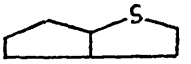
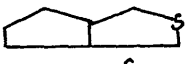
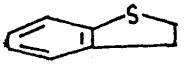
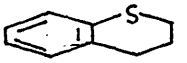
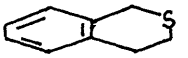
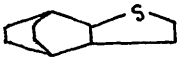
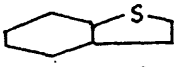
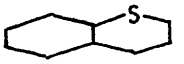
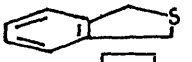
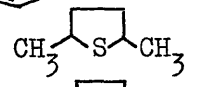
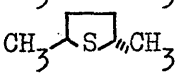
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
37	1048	142	Cis-2-thiohexahydroindan	
38	1049	142	Trans-2-thiohexahydroindan	
39	1050	156	2-Thio-(cis-decahydronaphtalene)	
40	1051	156	2-Thio-(trans-decahydronaphtalene)	
41	1792	128	8-Thiobicyclo(3.2.1) octane	
42	1879	114	7-Thiobicyclo(2.2.0) octane	
43	1881	128	Trans-2-thiobicyclo(3.3.0) octane	
44	1882	128	Cis-3-thiobicyclo(3.3.0) octane	
45	1883	136	1-Thio-(2,3-dihydroindene)	
46	1884	150	1-Thio-(1,2,3,4-tetrahydronaphtalene)	
47	1885	150	2-Thio-(1,2,3,4-tetrahydronaphtalene)	
48	1887	168	4,7-Ethano-cis-2-thiohexahydroindan	
49	1911	142	Cis-7- thiobicyclo(4.3.0) nonane	
50	1912	142	Trans-7-thiobicyclo(4.3.0) nonane	
51	1915	156	Trans-2-thiobicyclo(4,4.0) decane	
52	1918	136	2-Thio-(2,3-dihydroindene)	
53	572	116	2-Cis-5-dimethylthiocyclopentane	
54	573	116	2-Trans-5-dimethylthiocyclopentane	

TABLE I (continued)

Thiophenes

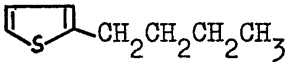
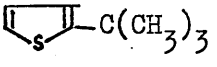
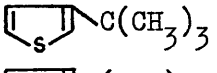
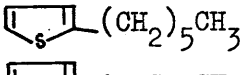
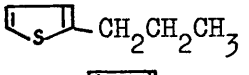
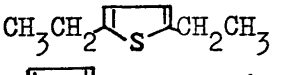
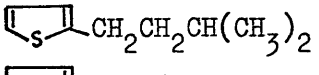
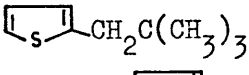
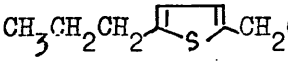
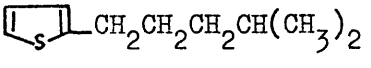
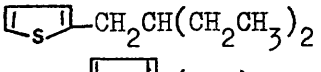
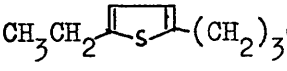
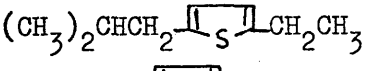
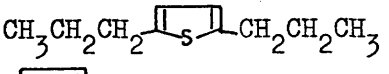
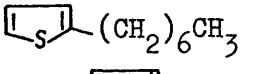
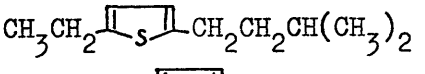
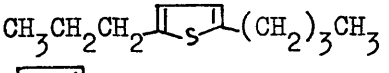
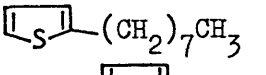
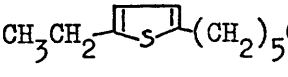
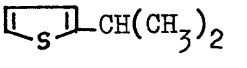
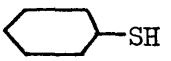
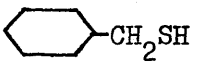
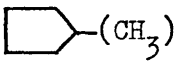
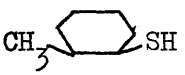
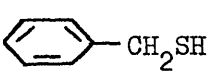
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
55	505	140	2-n-Butylthiophene	
56	506	140	2-tert-Butylthiophene	
57	507	140	3-tert-Butylthiophene	
58	1876	168	2-n-Hexylthiophene	
59	2042	126	2-n-Propylthiophene	
60	2044	140	2,5-Diethylthiophene	
61	2046	154	2-Isopentylthiophene	
62	2047	154	2-(2,2-Dimethylpropyl)thiophene	
63	2048	154	2-Ethyl-5-n-propylthiophene	
64	2049	168	2-(4-Methylpentyl)thiophene	
65	2050	168	2-(2-Ethylbutylthiophene)	
66	2051	168	2-Ethyl-5-n-butylthiophene	
67	2052	168	2-Ethyl-5-isobutylthiophene	
68	2053	168	2,5-Di-n-propylthiophene	
69	2054	182	2-n-Heptylthiophene	
70	2055	182	2-Ethyl-5-isopentylthiophene	
71	2056	182	2-n-Propyl-5n-butylthiophene	
72	2058	196	2-n-Octylthiophene	
73	2060	196	2-Ethyl-5-n-hexylthiophene	
74	2094	126	2-Isopropylthiophene	

TABLE I (continued)

Thiols

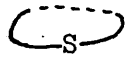
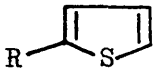
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
75	725	118	1-Hexanethiol	$CH_3(CH_2)_5SH$
76	726	132	1-Heptanethiol	$CH_3(CH_2)_6SH$
77	944	116	Cyclohexanethiol	
78	1229	130	Cyclohexylmethanethiol	
79	1235	118	4-Methyl-1-pentanethiol	$(CH_3)_2CH(CH_2)_3SH$
80	1236	116	1-Methylcyclopentanethiol	
81	1242	118	2-Methyl-1-pentanethiol	$CH_3CH_2CH_2CH(CH_3)CH_2SH$
82	1243	118	4-Methyl-2-pentanethiol	$(CH_3)_2CHCH_2CH(CH_3)SH$
83	1244	118	2-Methyl-3-pentanethiol	$CH_3CH_2CH(SH)CH(CH_3)_2$
84	1371	116	Cis-2-methylcyclopentanethiol	
85	1372	130	Cis-2-methylcylohexanethiol	
86	1414	116	Trans-2-methylcyclopentanethiol	
87	2084	118	2-Methyl-2-pentanethiol	$CH_3CH_2CH_2C(CH_3)_2SH$
88	2085	118	2,3-Dimethyl-2-butanethiol	$(CH_3)_2CHC(CH_3)_2SH$
89	2086	174	1-Decanethiol	$CH_3(CH_2)_9SH$
90	2087	124	Phenylmethanethiol	

In addition to these 'characteristics' of the spectra were listed:

i) The intensity of the molecular ion as a separate characteristic (index as above),

ii) The intensity of the fragment ions occurring by loss of 15, 29, 33, 34, 35, 43 and 46 mass units from the molecular ion (index as above),

iii) The degree of unsaturation of each compound,

		Index
e.g.	R—S—R'	1
		2
	 etc.	3

This process produced an unweildy number of characters since the highest molecular weight considered is 196 (2 - n - Octylthiophene, No. 72). Therefore, the data was extensively pruned of all characters which intuitively would contribute nothing to classification. The omitted characters were generally those mass numbers where no spectra contributed a substantial peak. This left a manageable data matrix with 42 characters.

	Characters	Spectra 1 2 3 90
1	Molecular Ion	
2	Degree of Unsaturation	
	m/e	
3	39	
4	41	
5	42	
6	43	
7	45	
8	47	
9	55	
10	56	
11	57	
12	60	
13	61	
14	67	
15	68	
16	69	
17	70	
18	75	
19	79	
20	81	
21	82	
22	83	
23	84	
24	85	
25	86	
26	91	
27	94	
28	95	
29	97	
30	99	
31	101	
32	110	
33	111	
34	124	
35	125	
36	139	
	Fragmentation Characters	
37	M - 15	
38	M - 29	

(over)

Characters	Spectra 1	2	3	...	...	90
Cont'd						
39 M - 33						
40 M - 34						
41 M - 43						
42 M - 46						

This data was punched on cards and the programme was activated. All eight hierarchical fusion options were called. The dendograms for the Group Average Method and Ward's Method are shown in Figures 3 and 4 respectively, where the compounds are listed in order.

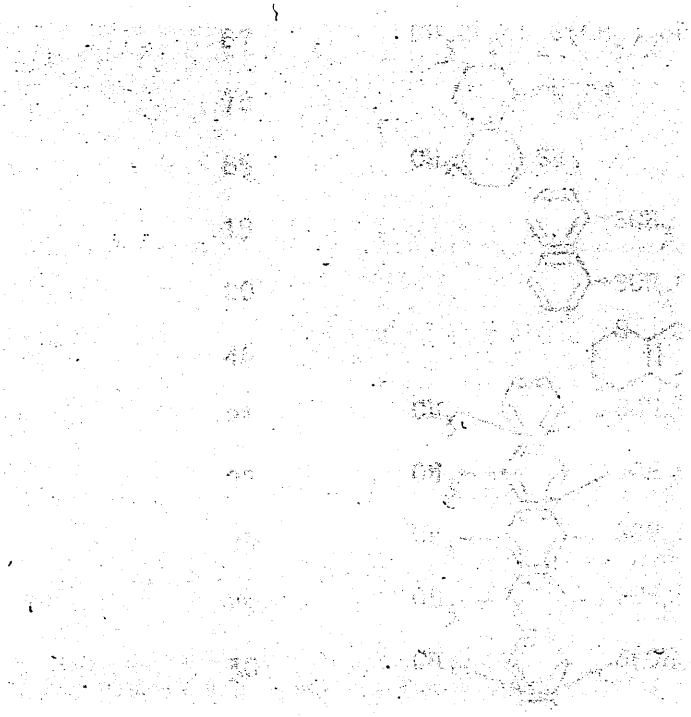
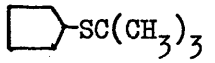
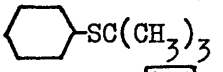
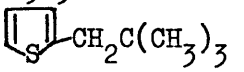
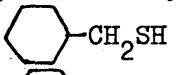
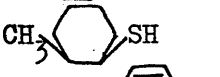
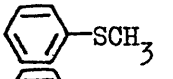
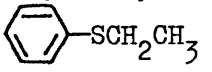
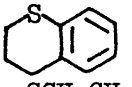
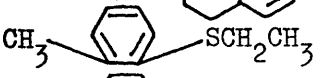
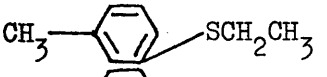
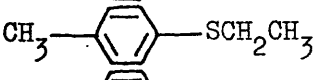
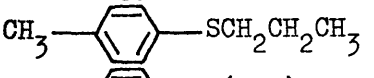
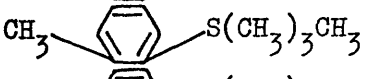
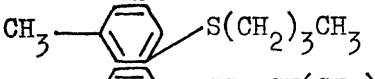
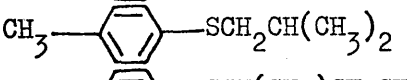
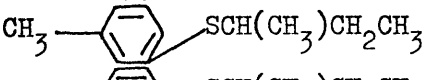
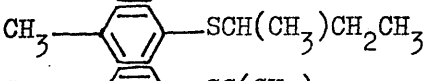
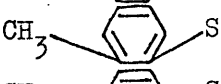
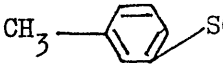
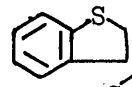
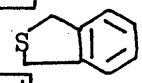


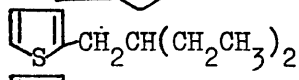
FIG. 3.

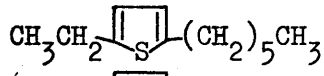
1	$CH_3CH_2SCH_2CH_2CH_2CH_3$
4	$CH_3CH_2SCH_2CH(CH_3)_2$
83	$CH_3CH_2CH(SH)CH(CH_3)_2$
88	$(CH_3)_2CHC(CH_3)_2SH$
2	$CH_3C(CH_3)_2SCH_2CH_3$
7	$(CH_3)_3CSCH(CH_3)_2$
5	$(CH_3)_3CSC(CH_3)_3$
11	
14	
62	
6	$CH_3CH_2CH(CH_3)SCH(CH_3)CH_2CH_3$
75	$CH_3(CH_2)_5SH$
81	$CH_3CH_2CH_2CH(CH_3)CH_2SH$
79	$(CH_3)_2CHCH_2CH_2CH_2SH$
87	$CH_3CH_2CH_2C(CH_3)_2SH$
78	
85	
19	
20	
46	
21	
22	
23	
29	
30	
31	
36	
32	
33	
34	
35	

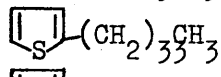
(over)

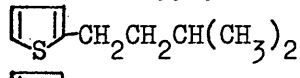
45 

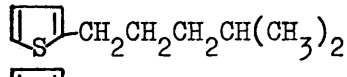
52 

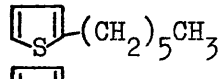
65 

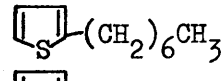
73 

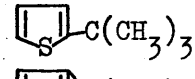
55 

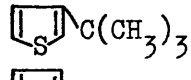
61 

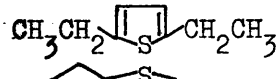
64 

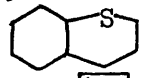
58 

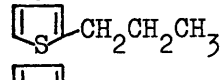
69 

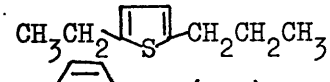
56 

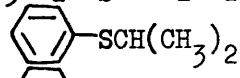
57 

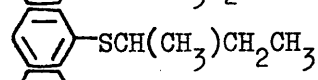
60 

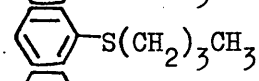
51 

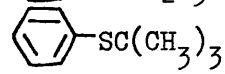
59 

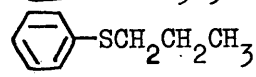
63 

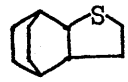
24 

27 

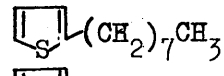
26 

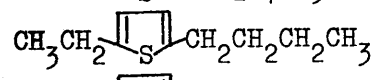
28 

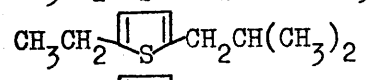
25 

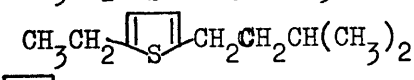
48 

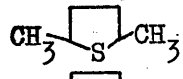
90

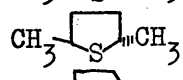
72 

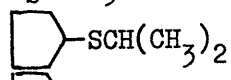
66 

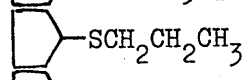
67 

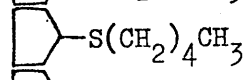
70 

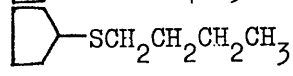
53 

54 

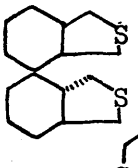
8 

9 

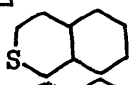
15 

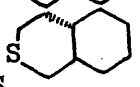
10 

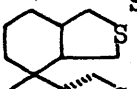
(over)

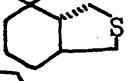
37 

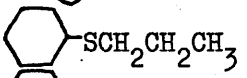
38

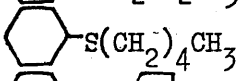
39 

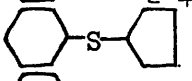
40 

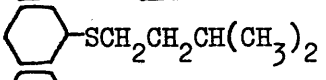
49 

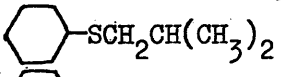
50 

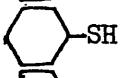
12 

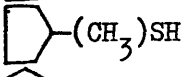
18 

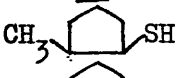
16 

17 

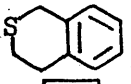
13 

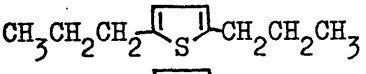
77 

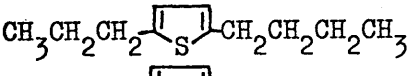
80 

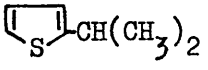
84 

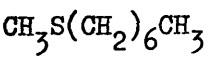
86 

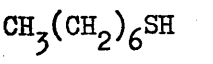
47 

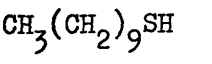
68 

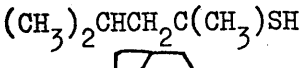
71 

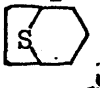
74 

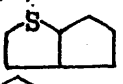
3 

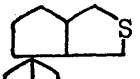
76 

89 

82 

41 

43 

44 

42 

(over)

FIG. 3

90 SULPHUR COMPOUNDS, 42 CHARACTERS, GROUP AVERAGE METHOD.

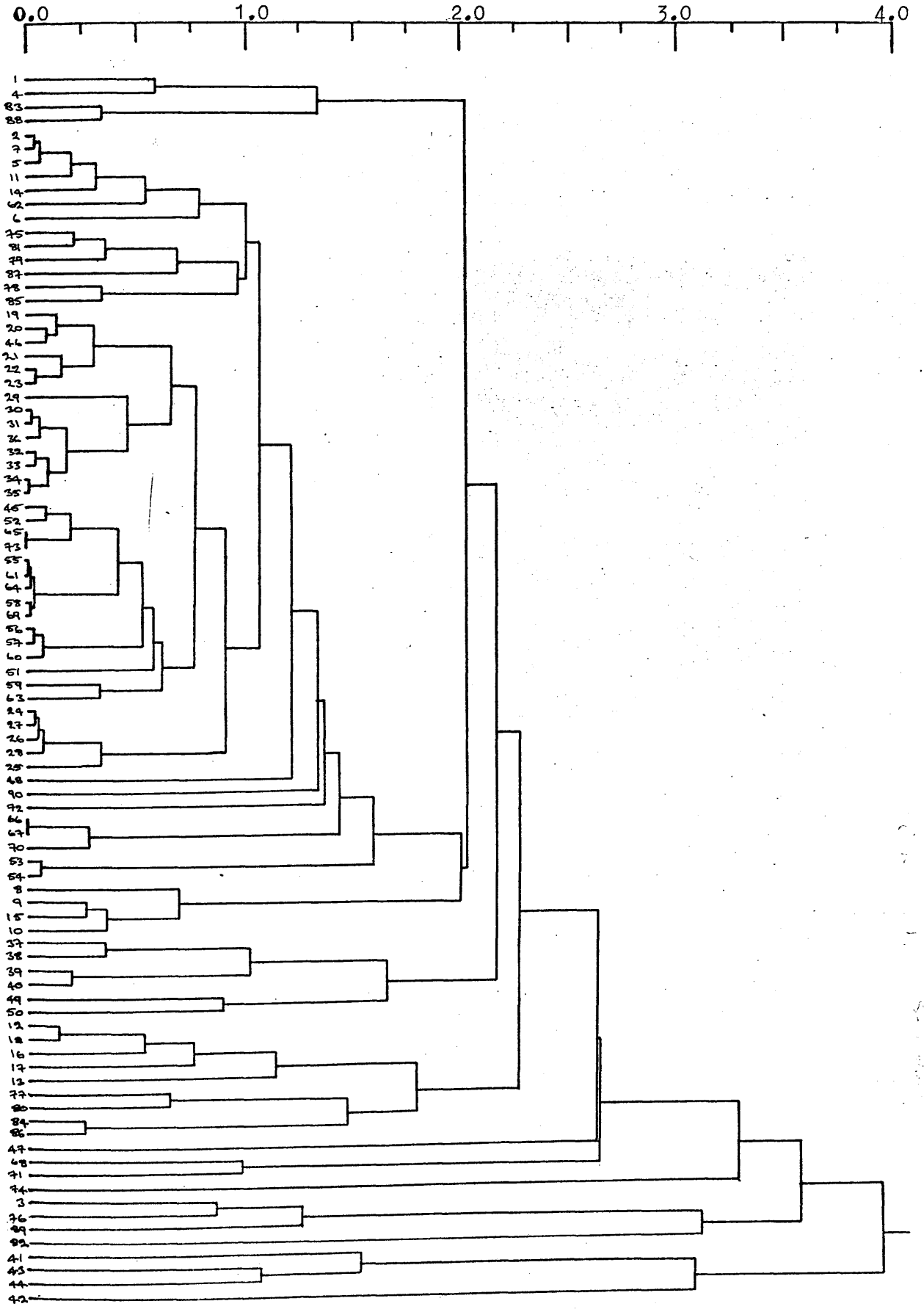
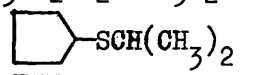
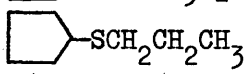
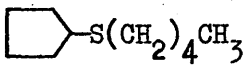
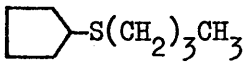
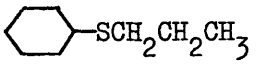
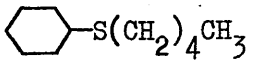
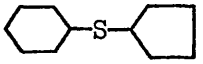
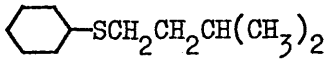
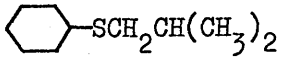
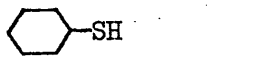
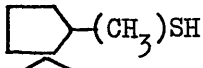
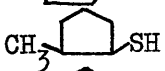
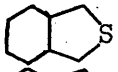
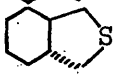
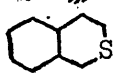
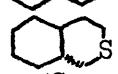
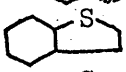
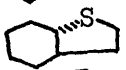
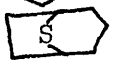
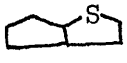
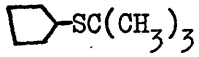
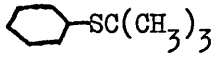
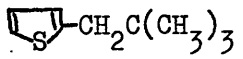
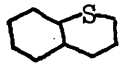
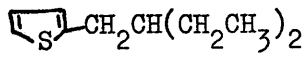
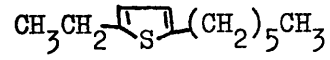
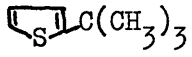
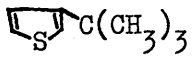
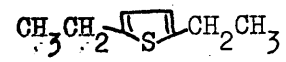
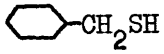
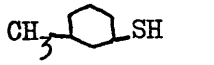
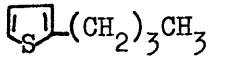
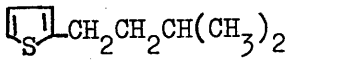
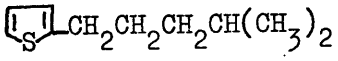
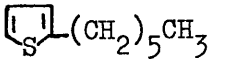
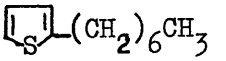
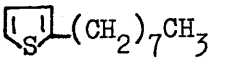
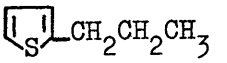
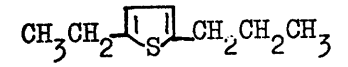
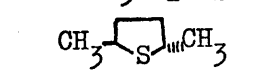
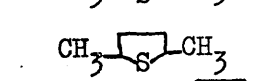
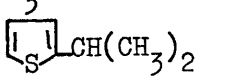


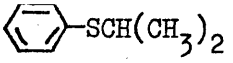
FIG. 4

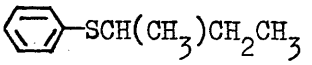
1	$CH_3CH_2S(CH_2)_3CH_3$
4	$CH_3CH_2SCH_2CH(CH_3)_2$
83	$CH_3CH_2CH(SH)CH(CH_3)_2$
88	$(CH_3)_2CHC(CH_3)_2SH$
87	$CH_3CH_2CH_2C(CH_3)_2SH$
8	
9	
15	
10	
3	$CH_3S(CH_2)_6CH_3$
76	$CH_3(CH_2)_6SH$
89	$CH_3(CH_2)_9SH$
82	$(CH_3)_2CHCH_2C(CH_3)SH$
12	
18	
16	
17	
13	
77	
80	
84	
86	
37	
38	
39	
40	
49	
50	
41	
43	
44	
42	

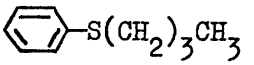
(over)

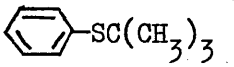
2	$CH_3C(CH_3)_2SCH_2CH_3$
7	$(CH_3)_3CSCH(CH_3)_2$
5	$(CH_3)_3CSC(CH_3)_3$
11	
14	
62	
6	$CH_3CH_2CH(CH_3)SCH(CH_3)CH_2CH_3$
51	
65	
73	
79	$(CH_3)_2CH(CH_2)_3SH$
56	
57	
60	
75	$CH_3(CH_2)_5SH$
81	$CH_3CH_2CH_2CH(CH_3)CH_2SH$
78	
85	
55	
61	
64	
58	
69	
72	
59	
63	
53	
54	
74	

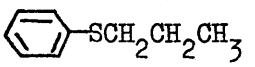
(over)

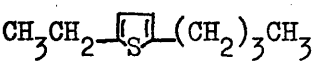
24 

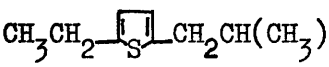
27 

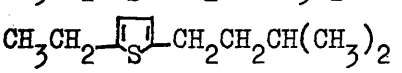
26 

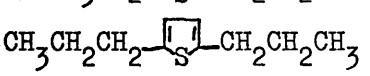
28 

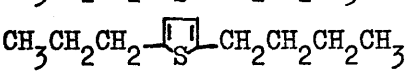
25 

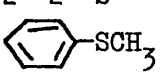
66 

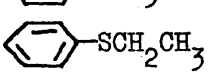
67 

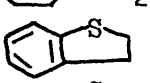
70 

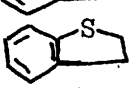
68 

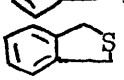
71 

19 

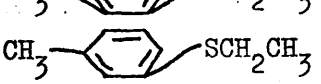
20 

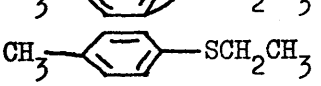
46 

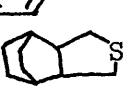
45 

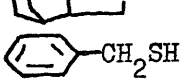
52 

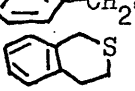
21 

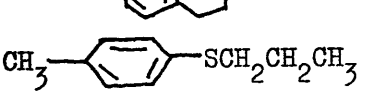
22 

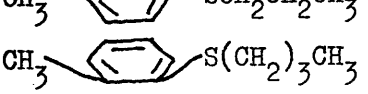
23 

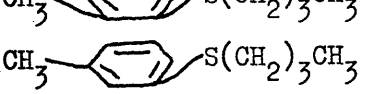
48 

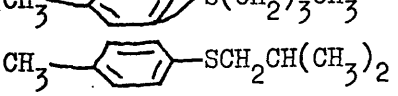
90 

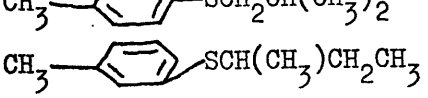
47 

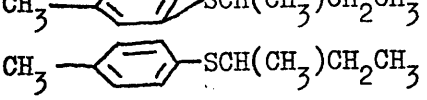
29 

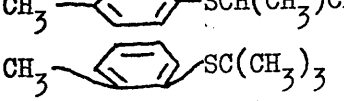
30 

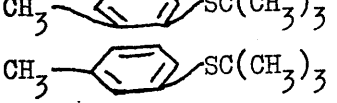
31 

36 

32 

33 

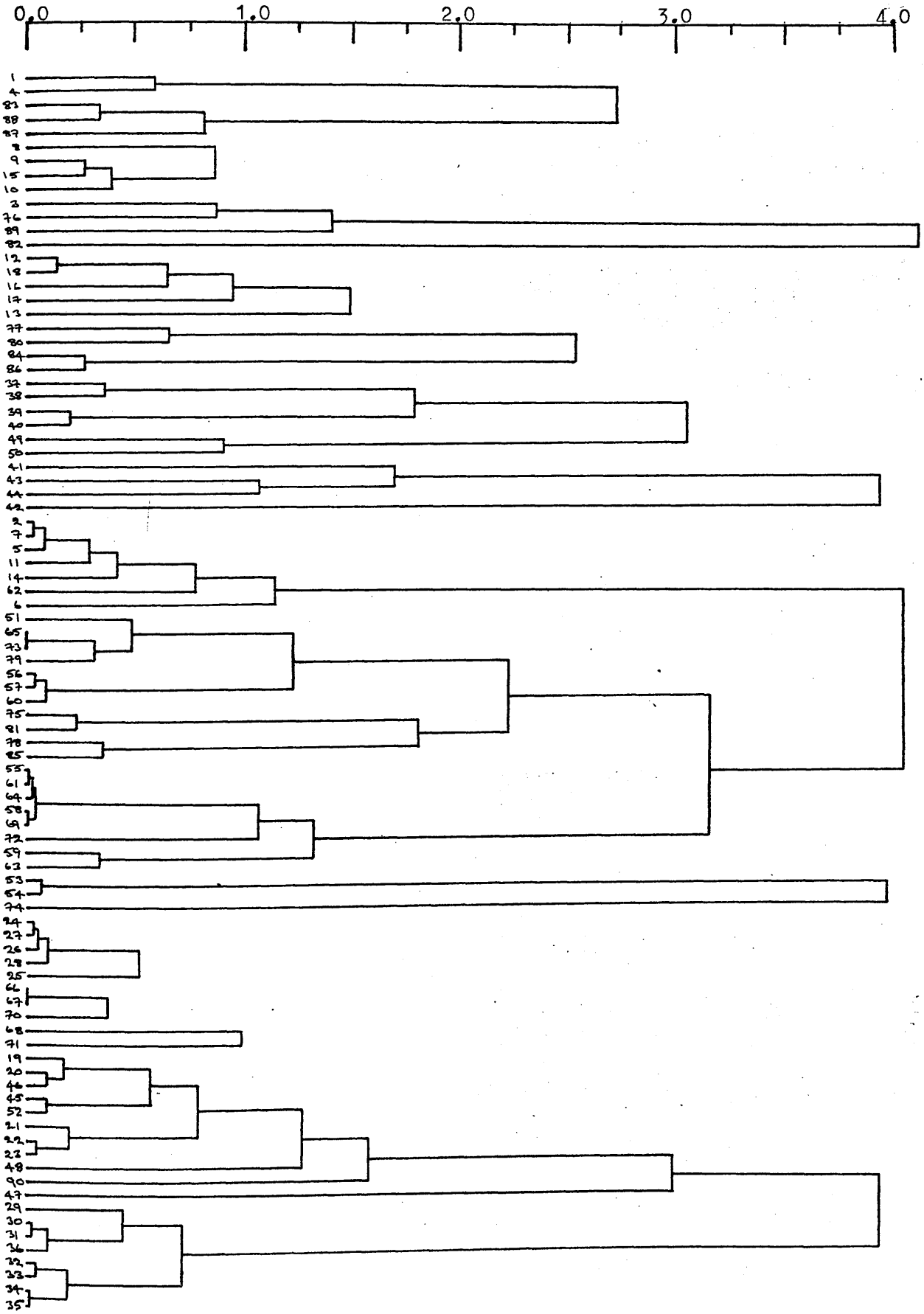
34 

35 

(over)

FIG. 4

90 SULPHUR COMPOUNDS, 42 CHARACTERS, WARD'S METHOD.



DISCUSSION

The two fusion options - Group Average Method and Ward's Method - give the most useful clustering results since,

i) Ward's Method, recognised as the best clustering method of the eight alternatives since it gives clear distinct clusters if clustering of the data is possible, appears to function well here - Figure 4.

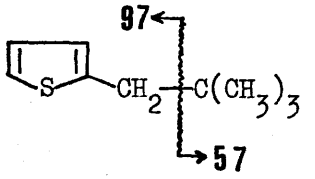
ii) The Group Average Method, Figure 3, gives good clustering but there is a slight tendency to form chains of clusters. However, for the possible application to mass spectrometry this feature may be desirable since those samples which are distinct from the others can be detected immediately from the dendogram. In Figure 3, this is the compounds appearing near the bottom of the dendogram.

In most of the results which follow both options are reproduced for comparison with these reasons in mind. None of the other clustering options were as effective as the two reproduced here and were not used again.

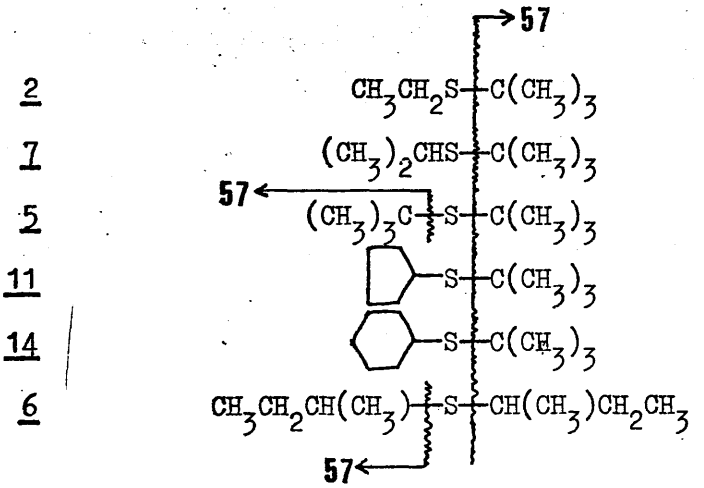
Considering the clustered compound lists of Figures 3 and 4 it is observed that tight clusters contain very similar compounds. That is, similar compounds are grouped closely together using the criteria chosen.

A striking anomaly is the inclusion of compound 62, 2 - (2, 2 - Dimethylpropyl) thiophene, with a group of compounds, Nos. 2, 7, 5, 11, 14 and 6, which are all non-cyclic thioethers.

However, the inclusion is based on the intense m/e 57 ion, $C_4H_9^+$, produced by cleavage of the bond β to the ring,



$C_4H_9^+$ is the parent ion in compounds 2, 7, 5, 11, 14 and 6 because of the cleavages shown,



The intense m/e 97 fragment of the thiophene (above) appears to be overshadowed by the m/e 57 ion plus the other features of the spectrum. Ideally, compound 62 should be grouped with the cluster of compounds 55, 61, 64, 58, 69, 72 which are all thiophenes with intense m/e 97 ions. This, of course, is not the fault of the clustering methods but of the information contained in the mass spectra and the criteria adopted to classify the spectra.

Table 2 lists the intensity indices for the features chosen for the 90 sulphur compounds, as described above. The compounds are arranged in the order of Figure 4 and the vertical lines in the table divide the tight clusters developed by the dendogram produced by Ward's Methods.

Considering the compounds discussed above, 2→6, in Table 2, the close association of compounds 2 and 7 is justified since their spectra are virtually identical.

TABLE 2

	1	4	8	3	8	8	7	8	9	1	5	1	0	3	7	6	8	9	8	2	1	2	1	8	1	6	1	7	1	3	7	7	8	0	8	4	8	6	3	7	3	8	3	9	4	0	4	9	5	0	4	1	4	3	4	4	2
M	6	3	3	2	1	6	5	4	5	3	3	2		5	4	4	5	5	4	2	5	5	5	3	9	8	8	9	6	6	8	5
Sat						1	1	1	1					1	1	2	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
39	1	2	3	2	1	4	4	3	2	3		3		3	2	3	3	4	4	3	4	6	5		7	6	2	2	5	6	7	4
41	3	4	9	8	5	9	9	9	9	7	8	9	8	7	7	9	9	9	7	5	8	8	4	3	9	8	2	3	5	5	6	2
42			1				1	1		4	6	3	1	1	1		1															
43		1	9	9	9	6	4	3		3	6	9	9	2	2		4															
45	1	1	1	1		2	2	1	2	1	1	1		2	1	2	1	2	2		3	4	3	2	4	3	2		4	5	6	4
47	4	4	4	1		1	2	1	2	1	3	3	1	1	1	1	1		1		1	1	1		2	1			1	1	1	
55	1	1	3					1	1	5	7	9	1	6	7	5	9	8	9	7	9	9	2	1	3	3	1	1	1	1	1	
56	6	4							2	6	9	7						1														
57	1	2	1	1	1				4	3	3	4	1	8	4			4														
60						1							1			1		1			8	7	1		1	1			1		4	3
61	3	1	1			5		1	1	9	2	1	6		1		2	8			1	1			2	2				1	
67						5	5	4	5						7	7	6	6	6	4	7	9	5	4	8	6	3	4	5	9	8	
68						7	9	7	9		1					2							1	1	1	1			2	1	1	
69			1	2	1	5	6	8	7	3	5	6	8		1	6	1															
70								2		5	9	7			1		7															
75	9	9	7	9	2					1			1					1														
79																							2	1	3	2	3	4	5	6	5	2
81														2	2	2	2	3			1	1	3	2	5	4	3	5		6	9	3
82												1		9	9	6	9	6	5		1	4								1	1	
83												5		4	5	5	6	6	6	9	9	8										
84												3	6																1			
85			3	4	4								1										1		1		2	1	9	2	4	9
86																											1	1	2	1	2	6
91									3									1														
94																							1		1	1			4	6	3	
95																							9	9	3	3	2	4	1	1	2	
97												2																				
99																									3	1	9	6	2	5	2	
101							5	5	6							3																
110																																
111												1																				
124																																
125																																
139																																
M-15						2				2															6	4						
M-29	2						2											9			1	1					1	1	2	5	2	9
M-33			3	4	4								1						6	9	9	8						1	1	1	2	3
M-34											1	1	6						5		1	4	3	4	2	1	4	5	4	6	3	5
M-43			7	9	2		5		1					2							1	1			2	2	9	6	9	2	4	
M-46	5																														1	1	

TABLE 2 (continued)

	2	7	5	1	1	1	4	6	2	6	5	1	6	5	7	3	7	9	5	6	5	7	6	0	7	5	8	1	7	8	8	5	5	5	6	1	6	4	5	8	6	9	7	2	5	9	6	3	5	3	5	4	7	4
M	2	1			2	2	3	1			3	2	2	2	2	2	3				2	1	2	2											1	1	2	1	2	2	2	3							4	3	3			
Sat					1	1	3				2	3	3			3	3	3					1	1											3	3	3	3	3	3	3	3							1	1	3			
39	1	2			2	2		1			1			1		2						2	2	2										1	1														2	2				
41	4	4	2		5	5		3			1			4	2	2					4	5	6	3																										3	2				
42														1							3	2																																
43		2												3							5	9					1																												
45					1	1	2								1	1	1					1	1	1		1	1							1	1					1	1							2						
47														1							2	2	1				1																										
55						2		1													3	2	9	9																										1	1				
56	2	2			1	1		2						9*							9	3																																
57	9	9	9		9	9	9	6																																													
60																								1																										2	1				
61								9													1																													2	2				
67					1	1				1													3	1																													
68					2																		1	1																														
69					4									1							2	1																																
70																						1																																
75	1		1	1			1																																														
79								1																																													
81					1			1															3	3																													
82					5																		1																														
83					3																		4																														
84																					1	2																														
85																1	1																																			
86																																																				
91																																																			
94																																																			
95																																																			
97							7				9*					1	1							1	5	9	9	9	9	9	9	9																				
99																																																			
101																																																	9	9			
110																																																			
111																		1																																		9	
124																																																			
125														9*																																						
139							1																																													
M-15							1										9	9	9																															1	9	9	9
M-29								4											1																														9	9			
M-33																								1	5																												
M-34																					1	2	3	1																												
M-43							2		9																	1																											
M-46																																																			

* omitted

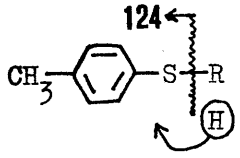
TABLE 2 (continued)

	24	27	26	28	25	66	67	70	68	71	19	20	46	45	52	21	22	23	48	90	47	29	30	31	36	32	33	34	35
M	3	3	4	1	7	2	2	2	2	2	9	9	9	9	6	9	9	9	9	3	9	9	6	6	7	4	4	1	1
Sat	4	4	4	4	4	3	3	3	3	3	4	4	5	5	5	4	4	4	3	4	5	4	4	4	4	4	4	4	4
39	1	1	1	1	2						1	2	3	1		1	2	1		1		2	1	2	2	1	1	1	1
41	1	1	1	2	1														5			1	1	1	2	1	1	2	2
42																													
43	1																					1							
45			1		3					1	2	3	2	1		5	3	3	3	1		4	3	2	3	1	1	2	1
47																													
55																			1										
56																													
57		1		3																					1	1	1	4	3
60																													
61																													
67														1					2										
68																													
69					1						1	1	1	1															
70																													
75																													
79																			5										
81																			2										
82																													
83																													
84																													
85																			2										
86																													
91											2		1	3	3	5	6	5	2	9	1	6	4	4	4	3	4	2	2
94																			1										
95																													
97									1	2																			
99																													
101																													
110	9	9	9	9	9				1	1																			
111																													
124																						7	9	9	9	9	9	9	9
125						9	9	9																					
139									9	9																			
M-15											4	6	5			3	5	5	2		2								
M-29		1			6				9	6		2	3			5	1	2			1	6				1	1		
M-33											2		1								4								
M-34													1								1								
M-43			2		1	9	9			9			1									1	2	3	6				
M-46	1										2		1								9								

Compound 5 associates closely with 2 and 7 and gives an indication that more subtle features in a spectrum are being overshadowed by very intense peaks. That is, virtually the only similarity 5 has with 2 and 7, apart from no intensity contribution at most of the peaks, is the m/e 57 peak. Similarly, compound 11, with perhaps significant contributions at many m/e values, associates closely with 2, 7 and 5 because m/e 57 is also its parent ion. Compound 62, as discussed above, correlates closely with this cluster because of the large contribution of the m/e 57 peak. Compound 6 has almost no justification for inclusion in the cluster <u>except</u> for its large (60 - 70%) m/e 57 peak.

It is apparent that more care must be taken in the selection of features ("preprocessing") to make less obvious contributions to the spectra influence the correlation result. An obvious first step is to use the square root of the intensity at each m/e value whereby very intense peaks will be reduced in importance and less intense peaks will become more significant.

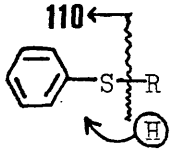
Compounds 29, 30, 31, 32, 33, 34, 35 and 36 form a tight cluster based mainly on the contribution of the large m/e 124 peak in all the spectra,



although the other spectral features <u>e.g.</u> m/e 91, certainly contribute.

The cluster 19, 20, 46, 45, 52, 21, 22 and 23 is tight. The spectra of the isomers 21, 22 and 23 are, not unexpectedly, almost identical. The high intensity of the molecular ion has a considerable influence in the association of the whole cluster but the saturation index, m/e 39, m/e 45, m/e 91, M - 15 and M - 29 must contribute.

The spectra of compounds 24, 25, 26, 27 and 28 are very similar especially the intense m/e 110 ion from,



The cluster 51, 65, 73, 79 and 56 contains some anomalies. Compounds 65 and 73 are considered identical. Compound 79 associated with 51, 65 an 73. Compound 65 does not associate with other mono-alkyl substitued thiophenes. Compound 73 does not associate with 66, 67 and 70. The anomalies are explained quite simply by human omission. That is, compound 65 has an intense m/e 97 ion. The index was omitted when the data was punched on cards. Similarly compound 73 has an intense m/e 125 ion and compound 79 has an intense m/e 56 ion. These two indices were also omitted. When these errors were corrected compound 65 associated closely with the other mono-alkyl substituted thiophenes, compound 73 associated closely with the other ethyl-alkyl substituted thiophenes, and compound 79 associated closely with compound 75, another thiol. Figures 3 and 4 are the dendograms with the errors included to demonstrate the usefulness of the clustering method for detecting errors in data for known or even unknown compounds. The anomalies here are very apparent.

The clusters formed in the top third of the Ward's Method dendogram are not as tight as in the rest of the dendogram. Considering Table 2, it is apparent that this is because there is a contribution to more characters by these compounds. In fact they are all aliphatic and show greater general fragmentation than the aromatics. Considering the whole dendogram it is apparent that the compounds have been broadly divided into aliphatic (top half) and aromatic (bottom half). There is some overlap at the centre, but remembering that compounds 65, 73 and 79 are displaced, the broad separation is encouraging.

In the aliphatic section the clustering on the criteria chosen is reasonable. For example, the close association of compounds 8, 9, 15 and 10, and also 12, 18, 16, 17 and 13 is noteable. All isomers, as expected, associate closely.

b) Encouraged by the potential utility of the cluster method for data reduction and compound identification and since the initial impetus for the study derived from air pollutant considerations, it was decided to use the cluster method on a real set of compounds to compare with the study on the artificial collection of sulphur compounds.

A GLC - MS investigation of C_6 - C_{20} organics in urban (Zurich) atmosphere was made by Grob and Grob.[18] They identified a number of components. The mass spectra of sixty compounds (Table 3) corresponding as closely as possible to those found by Grob and Grob, were obtained from catalogued data. In the same way as the sulphur compounds, these compounds were listed according to the intensity at each m/e value with the intensities of all molecular ions and fragment ions from M - 1 to M - 59 inclusive. It was thought that the degree of unsaturation would be a difficult quantity to extract automatically from a mass spectrum and this character was not included.

The characters were scanned and an abbreviated list of thirty-seven of the most important characters was retained,

	m/e	m/e	m/e
1. M	11. 39	21. 76	31. 118
2. -1	12. 41	22. 77	32. 119
3. -15	13. 43	23. 78	33. 120
4. -27	14. 51	24. 85	34. 134
5. -28	15. 55	25. 91	35. 142
6. -29	16. 56	26. 92	36. 154
7. -36	17. 57	27.105	37. 168
8. -43	18. 62	28.106	
9. -44	19. 69	29.115	
10. -57	20. 71	30.117	

TABLE 3

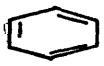
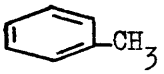
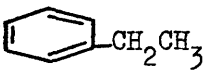
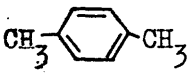
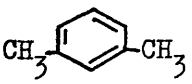
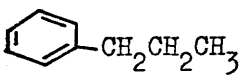
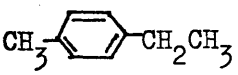
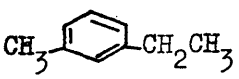
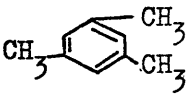
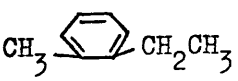
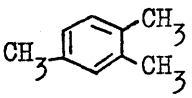
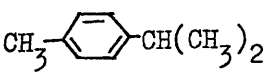
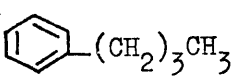
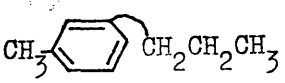
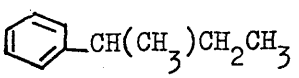
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
1	1612	78	Benzene	
2	3*	92	Toluene	
3	177	106	Ethylbenzene	
4	180	106	1,4-Dimethylbenzene	
5	179	106	1,3-Dimethylbenzene	
6	178	106	1,2-Dimethylbenzene	
7	256	120	n-Propylbenzene	
8	260	120	1-Ethyl-4-methylbenzene	
9	259	120	1-Ethyl-3-methylbenzene	
10	263	120	1,3,5-Trimethylbenzene	
11	258	120	1-Ethyl-2-methylbenzene	
12	262	120	1,2,4-Trimethylbenzene	
13	462	134	n-Isopropyl-4-methylbenzene	
14	494	134	n-Butylbenzene	
15	1431	134	Propylmethylbenzene	
16	460	134	sec-Butylbenzene	

TABLE 3 (continued)

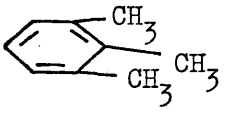
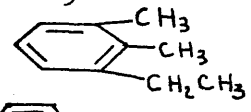
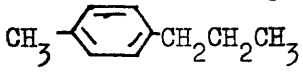
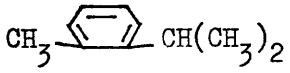
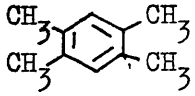
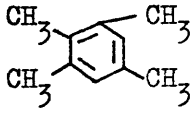
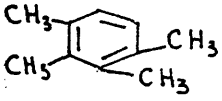
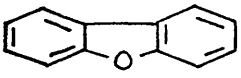
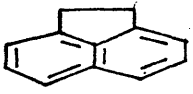
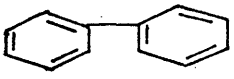
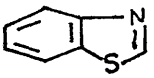
No	Cat. No	Mol. Wt.	Name	Semi-structural Formula
17	261	120	1,2,3-Trimethylbenzene	
18	264	134	Ethyldimethylbenzene	
19	1432	134	Propylmethylbenzene	
20	853	134	Isopropylmethylbenzene	
21	320	134	1,2,4,5-Tetramethylbenzene	
22	463	134	1,2,3,5-Tetramethylbenzene	
23	1957	134	1,2,3,4-Tetramethylbenzene	
24	705	282	n-Eicosane	$CH_3(CH_2)_{18}CH_3$
25	*61	166	Fluorene	
26	633	168	Dibenzofuran	
27	1436	154	Acenaphthene	
28	906	168	Methyldiphenyl	
29	1007	254	n-Octadecane	$CH_3(CH_2)_{16}CH_3$
30	613	154	Diphenyl	
31	1006	240	n-Heptadecane	$CH_3(CH_2)_{15}CH_3$
32	1759	135	Benzothiazole	

TABLE 3 (continued)

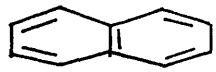
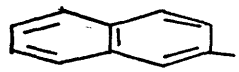
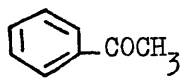
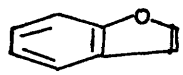
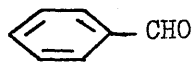
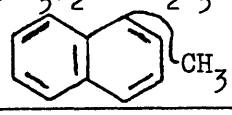
No	Cat. No	Mol. Wt.	Name	Semi-strucural Formula
33	1005	226	n-Hexadecane	$CH_3(CH_2)_{14}CH_3$
34	1004	212	n-Pentadecane	$CH_3(CH_2)_{13}CH_3$
35	410	128	Naphthalene	
36	894	142	2-Methylnaphthalene	
37	881	186	1-Dodecanol	$CH_3(CH_2)_{10}CH_2OH$
38	880	158	1-Decanol	$CH_3(CH_2)_8CH_2OH$
39	87*	120	Acetophenone	
40	1003	198	n-Tetradecane	$CH_3(CH_2)_{12}CH_3$
41	704	198	7-Methyltridecane	$CH_3(CH_2)_5CH(CH_3)(CH_2)_5CH_3$
42	1103	132	Methylindan	
43	86*	118	Benzofuran	
44	83	106	Benzaldehyde	
45	**	147	Dichlorobenzene	

TABLE 3 (continued)

No	Cat. No.	Mol. Wt.	Name	Semi-structural Formula
46	404	170	n-Dodecane	$CH_3(CH_2)_{10}CH_3$
47	523	184	n-Tridecane	$CH_3(CH_2)_{11}CH_3$
48	403	156	n-Undecane	$CH_3(CH_2)_9CH_3$
49	840	156	Isoundecane	$(CH_3)_2CH(CH_2)_7CH_3$
50	109	142	n-Decane	$CH_3(CH_2)_8CH_3$
51	484	140	Isodecene	$(CH_3)_2CHCH{=}CH(CH_2)_4CH_3$
52	89	99	Dichloroethane	$ClCH_2CH_2Cl$
53	479	142	Isodecane	$(CH_3)_2CH(CH_2)_6CH_3$
54	132	128	n-Nonane	$CH_3(CH_2)_7CH_3$
55	245	128	Isononane	$(CH_3)_2CH(CH_2)_5CH_3$
56	39	114	n-Octane	$CH_3(CH_2)_6CH_3$
57	40	114	2-Methylheptane	$(CH_3)_2CH(CH_2)_4CH_3$
58	14	100	n-Heptane	$CH_3(CH_2)_5CH_3$
59	15	100	2-Methylhexane	$(CH_3)_2CH(CH_2)_3CH_3$
60	990	142	1-Methylnaphthalene	

All Cat. Nos. (Catalogue Numbers) refer to the A.P.I. Collection except * which refers to the M.C.A.R.P. Collection and ** which refers to an Uncertified Spectrum.

The data was punched on cards and clustering was performed using the Group Average Method and Ward's Method. The appropriate dendograms are shown in Figures 5 and 6 respectively with the compounds listed in order.

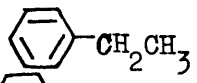
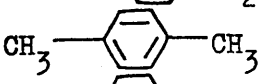
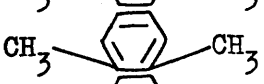
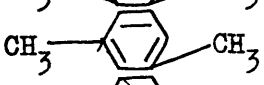
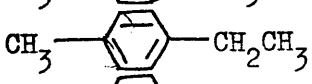
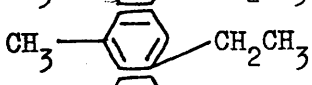
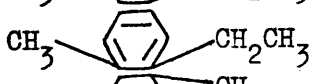
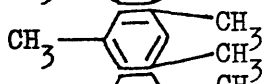
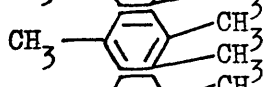
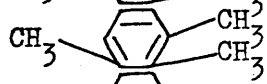
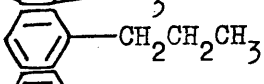
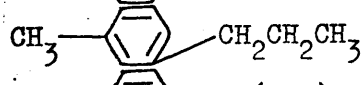
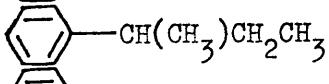
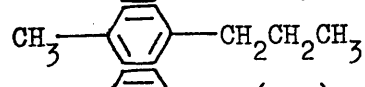
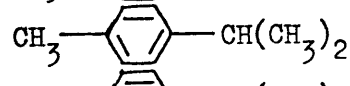
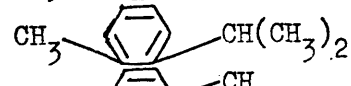
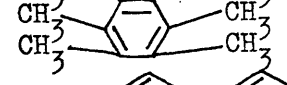
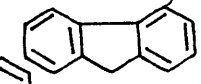
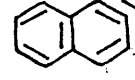
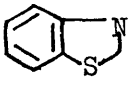
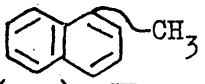
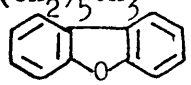
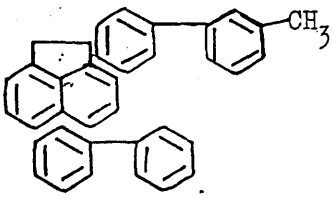
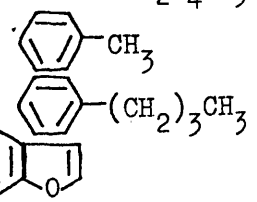
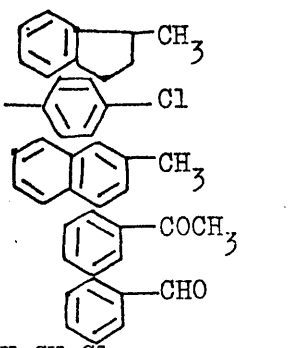
1	
3	
4	
6	
5	
8	
9	
11	
10	
12	
17	
7	
15	
16	
19	
13	
20	
18	
21	
22	
23	
25	
35	
32	
60	
24	$CH_3(CH_2)_{18}CH_3$
41	$CH_3(CH_2)_5CH(CH_3)(CH_2)_5CH_3$
46	$CH_3(CH_2)_{10}CH_3$
47	$CH_3(CH_2)_{11}CH_3$
48	$CH_3(CH_2)_9CH_3$
50	$CH_3(CH_2)_8CH_3$
49	$(CH_3)_2CH(CH_2)_7CH_3$
53	$(CH_3)_2CH(CH_2)_6CH_3$
55	$(CH_3)_2CH(CH_2)_5CH_3$
54	$CH_3(CH_2)_7CH_3$
56	$CH_3(CH_2)_6CH_3$
29	$CH_3(CH_2)_{16}CH_3$
31	$CH_3(CH_2)_{15}CH_3$
33	$CH_3(CH_2)_{14}CH_3$
34	$CH_3(CH_2)_{13}CH_3$
40	$CH_3(CH_2)_{12}CH_3$
57	$(CH_3)_2CH(CH_2)_4CH_3$
59	$(CH_3)_2CH(CH_2)_3CH_3$
58	$CH_3(CH_2)_5CH_3$
26	
28	
27	
30	
37	$CH_3(CH_2)_{10}CH_2OH$
38	$CH_3(CH_2)_8CH_2OH$
51	$(CH_3)_2CHCH{=}CH(CH_2)_4CH_3$
2	
14	
43	
42	
45	
36	
39	
44	
52	$ClCH_2CH_2Cl$

FIG. 5

60 'ZURICH' COMPOUNDS, 37 CHARACTERS, GROUP AVERAGE METHOD.

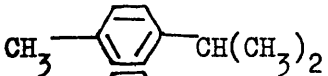
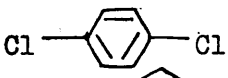
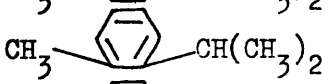
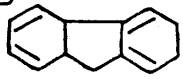
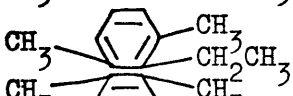
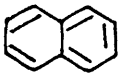
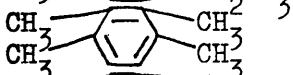
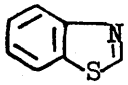
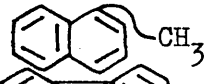
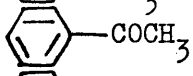
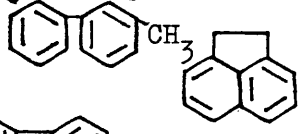
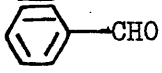
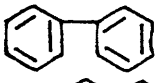
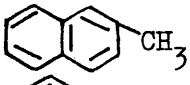
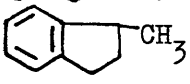
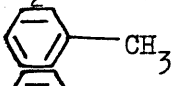
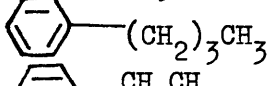
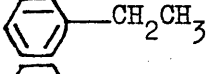
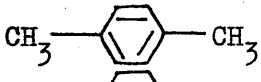
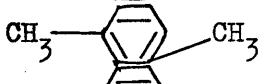
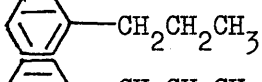
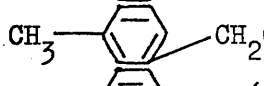
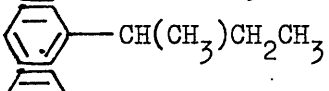
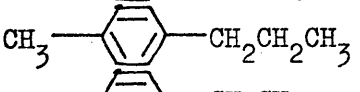
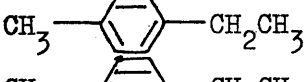
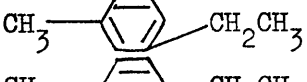
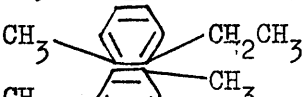
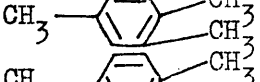
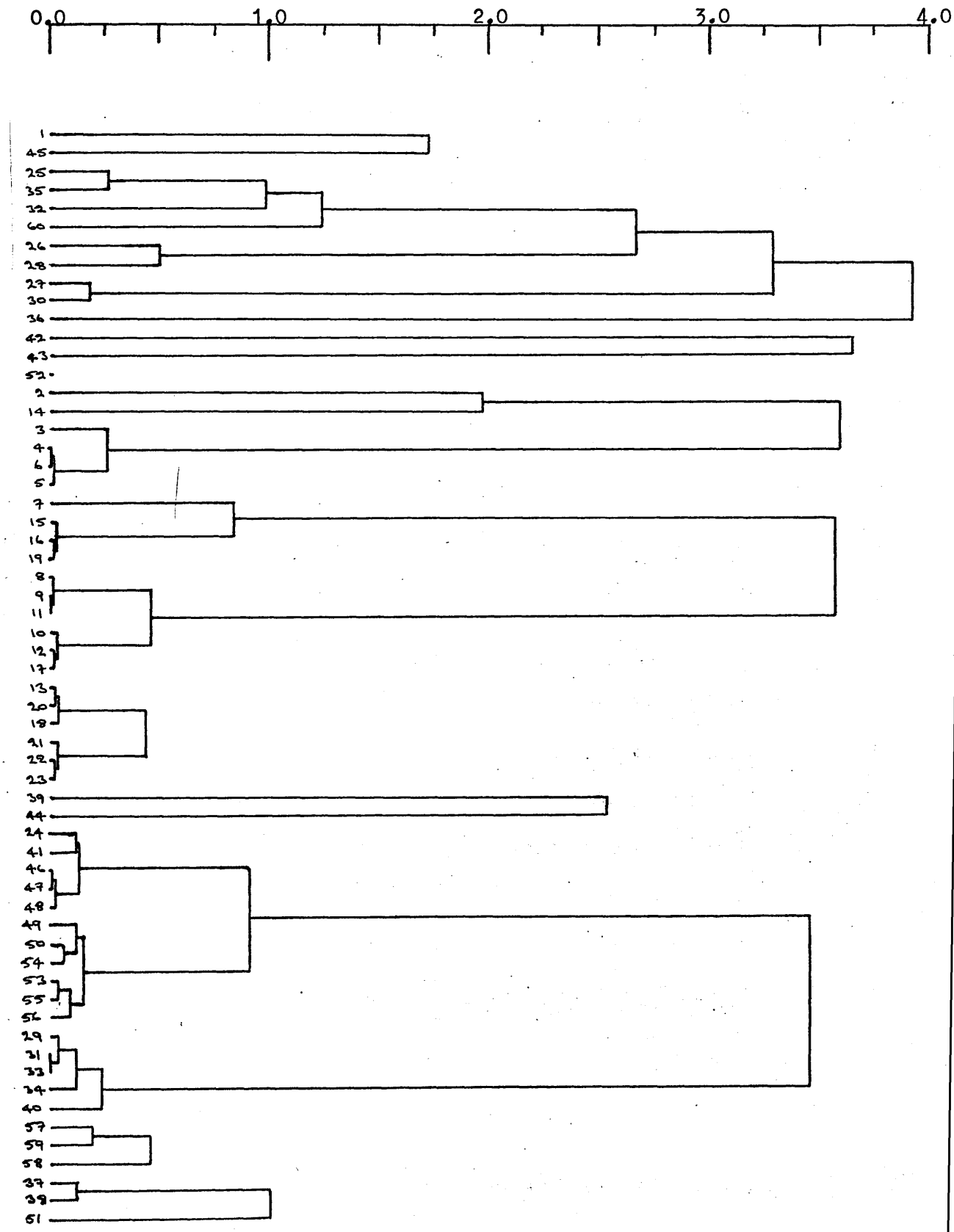
1		13	
45		20	
25		18	
35		21	
32		22	
60		23	
26		39	
28		44	
27		24	$CH_3(CH_2)_{18}CH_3$
30		41	$CH_3(CH_2)_5CH(CH_3)(CH_2)_5CH_3$
36		46	$CH_3(CH_2)_{10}CH_3$
42		47	$CH_3(CH_2)_{11}CH_3$
43		48	$CH_3(CH_2)_9CH_3$
52	$ClCH_2CH_2Cl$	49	$(CH_3)_2CH(CH_2)_7CH_3$
2		50	$CH_3(CH_2)_8CH_3$
14		54	$CH_3(CH_2)_7CH_3$
3		53	$(CH_3)_2CH(CH_2)_6CH_3$
4		55	$(CH_3)_2CH(CH_2)_5CH_3$
6		56	$CH_3(CH_2)_6CH_3$
5		29	$CH_3(CH_2)_{16}CH_3$
7		31	$CH_3(CH_2)_{15}CH_3$
15		33	$CH_3(CH_2)_{14}CH_3$
16		34	$CH_3(CH_2)_{13}CH_3$
19		40	$CH_3(CH_2)_{12}CH_3$
8		57	$(CH_3)_2CH(CH_2)_4CH_3$
9		59	$(CH_3)_2CH(CH_2)_3CH_3$
11		58	$CH_3(CH_2)_5CH_3$
10		37	$CH_3(CH_2)_{10}CH_2OH$
12		38	$CH_3(CH_2)_8CH_2OH$
17		51	$(CH_3)_2CHCH{=}CH(CH_2)_4CH_3$

FIG. 6

60 'ZURICH' COMPOUNDS, 37 CHARACTERS, WARD'S METHOD.



DISCUSSION

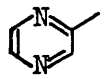
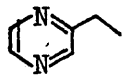
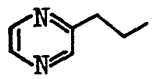
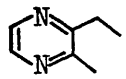
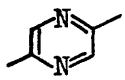
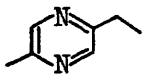
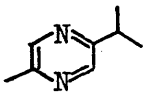
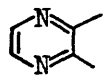
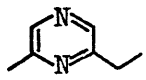
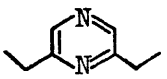
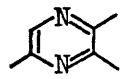
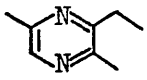
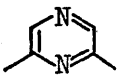
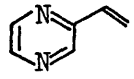
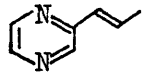
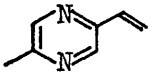
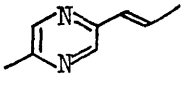
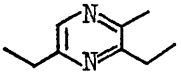
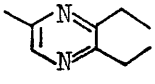
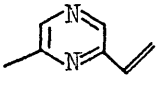
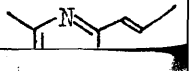
The clustering of the compounds and the general form of the profiles is again encouraging.

The separation of aro matics from saturated alkanes is particularly impressive as is the anticipated grouping of isomers.

The sifting of compounds 26, 28, 27, 30, 37, 38, 51, 2, 14, 43, 42, 45, 36, 39, 44 and 52 to the end of the chain in the Group Average Method is noteable. These compounds are fairly distinct from the aromatic and saturated hydrocarbons which make up the bulk of the compounds. However, the same effect is noted in the Ward's Method dendogram, where most of these different compounds are at the top of the figure. The reason for using the Group Average Method mentioned in a) may therefore be redundant and, since the overall structure of the Ward's Method dendogram is more balanced, the Group Average Method may be superfluous.

It is apparent that if the compounds were unknown then the identification of one compound in a cluster from its mass spectrum would be adequate to characterise the compounds in the cluster. Alternatively, the addition of a selection of known compounds to the clustering procedure would serve as labels to identify the class of compound in each cluster.

TABLE 4

No.	Mol. Wt.	Name	Formula
1	80	Pyrazine	
2	94	2-Methylpyrazine	
3	108	2-Ethylpyrazine	
4	122	2-Propylpyrazine	
5	122	2-Ethyl-3-methylpyrazine	
6	108	2,5-Dimethylpyrazine	
7	122	2-Ethyl-5-methylpyrazine	
8	136	2-Methyl-5-isopropylpyrazine	
9	108	2,3-Dimethylpyrazine	
10	122	2-Ethyl-6-methylpyrazine	
11	136	2,6-Diethylpyrazine	
12	122	Trimethylpyrazine	
13	136	2,5-Dimethyl-3-ethylpyrazine	
14	108	2,6-Dimethylpyrazine	
15	106	Vinylpyrazine	
16	120	2-(trans-1-Propenyl)pyrazine	
17	120	2-Methyl-5-vinylpyrazine	
18	134	2-Methyl-5-(trans-1-propenyl)pyrazine	
19	150	2,6-Diethyl-3-methylpyrazine	
20	150	2,3-Diethyl-5-methylpyrazine	
21	120	2-Methyl-6-vinylpyrazine	
22	134	2-Methyl-6-(trans-1-propenyl)pyrazine	

c) To test if the clustering programmes could effectively deal with more characters and to try and better understand what influence the characters had on clustering, a small study was carried out using the mass spectra of twenty-two various pyrazines.[19]

The pyrazines used are listed in Table 4. The characters used were the square root of the intensity at the ninety-six most significant m/e values.

m/e	38 - 123	inclusive	} 96
m/e	131 - 138	inclusive	
m/e	149 - 150	inclusive	

for one run using the Group Average and Ward's Method, and these ninety-six plus the square root of the intensity of twenty-four fragment ions,

M; M - 1, - 2, - 15, - 16, - 17, - 18, - 19, - 20, - 27, - 28, - 29, - 30, - 31, - 32, - 33, - 34, - 35, - 36, - 40, - 43, - 44, - 45 and 46, for another run using the same clustering methods. One hundred and twenty characters are used in this second run. This is the programme's present capacity.

Clustering was successfully carried out and the results are displayed in Figures 7 and 8 with the compounds listed in order.

DISCUSSION

It is noted that there is very little difference between the Group Average Method and Ward's Method in both the one hundred and twenty character dendograms and the ninety-six character dendograms. It appears that Ward's Method alone may be chosen for future studies and to develop consistent profiles.

Broadly, however, the dendograms give little information except to pick out the isomers; compounds 6, 12 and 9; 17 and 21; 11 and 13; 19 and 20; 18 and 22. Since the compounds used are very similar this is a reasonable result.

GROUP AVERAGE

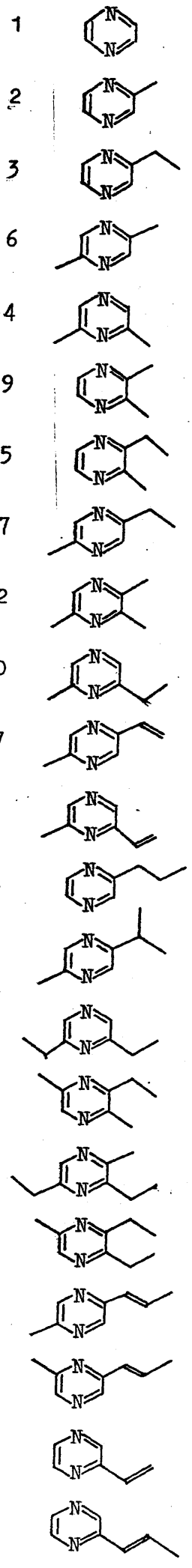


FIG. 7

PYRAZINE SPECTRA

22 - COMPOUNDS

120 - CHARACTERS

WARD'S

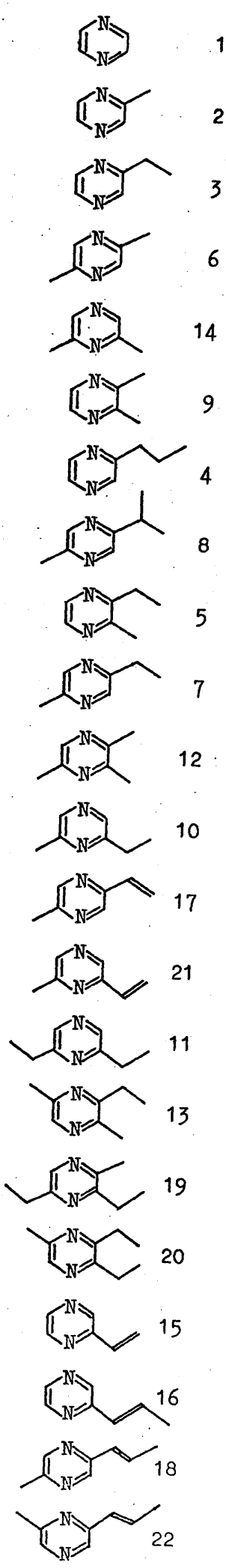
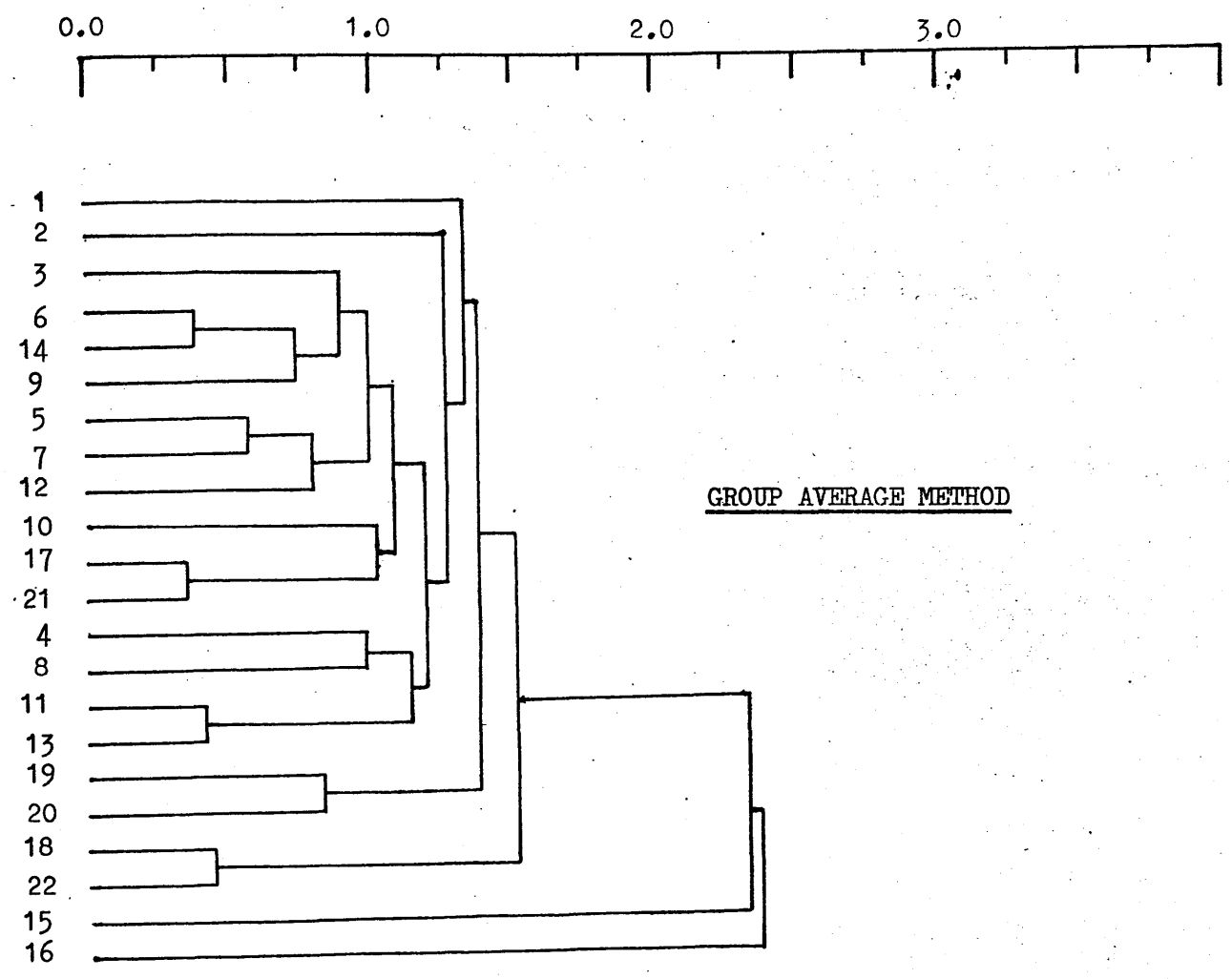
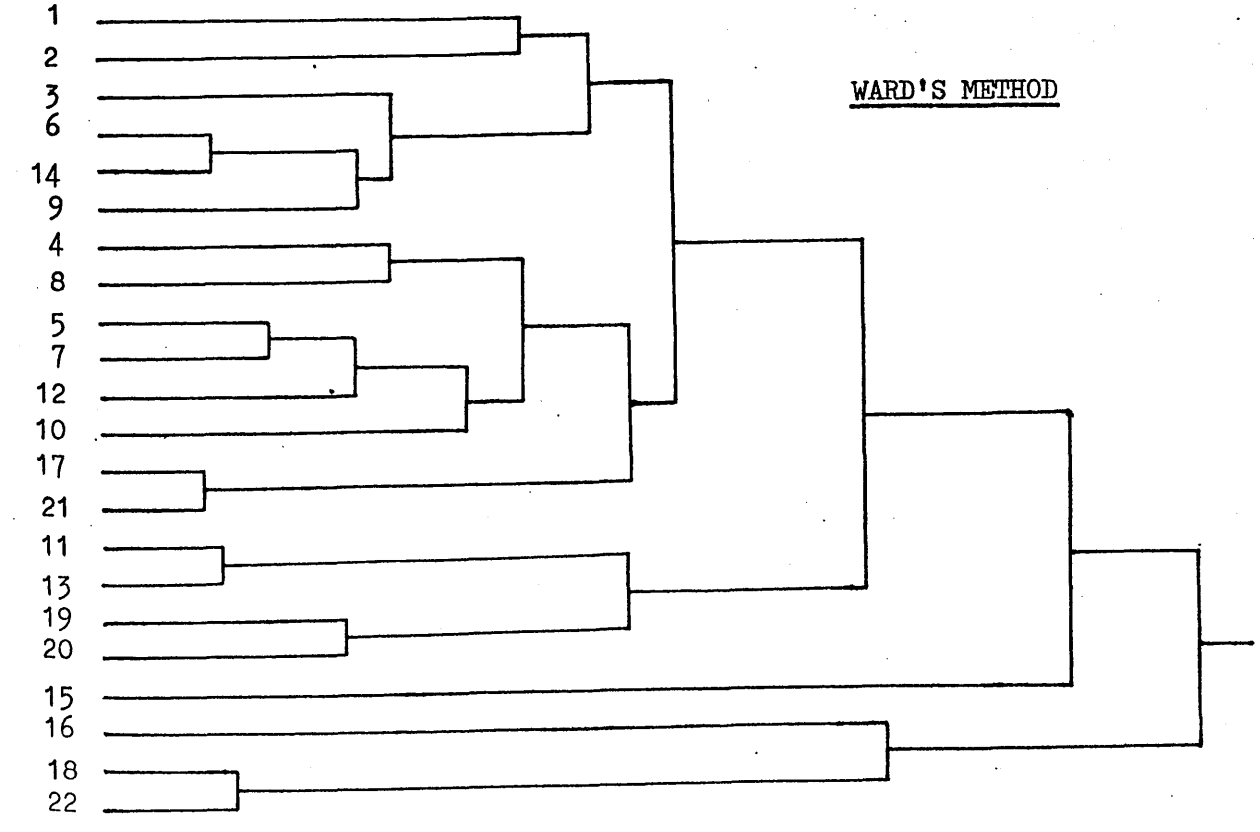


FIG. 7 PYRAZINE SPECTRA 22 - COMPOUNDS 120 - CHARACTERS





GROUP AVERAGE

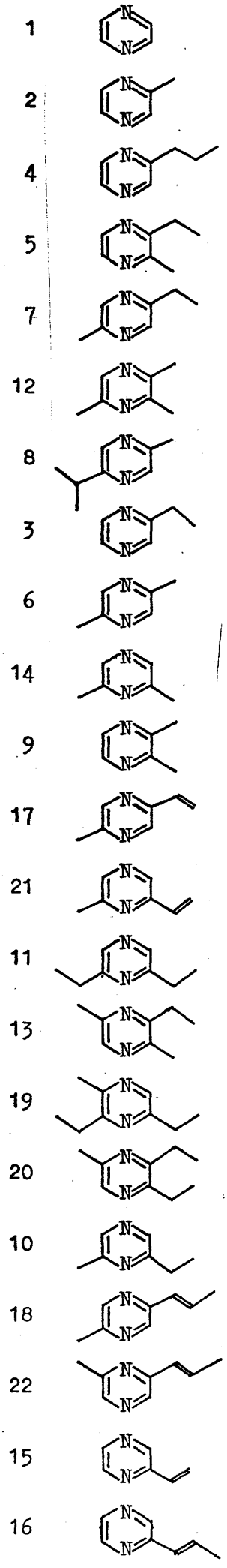


FIG. 8

PYRAZINE SPECTRA

22 - COMPOUNDS

96 - CHARACTERS

WARD'S

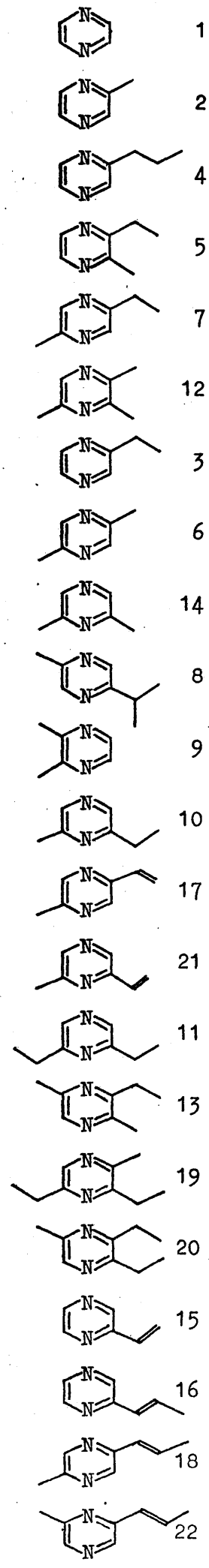


FIG. 8 PYRAZINE SPECTRA 22 - COMPOUNDS 96 - CHARACTERS

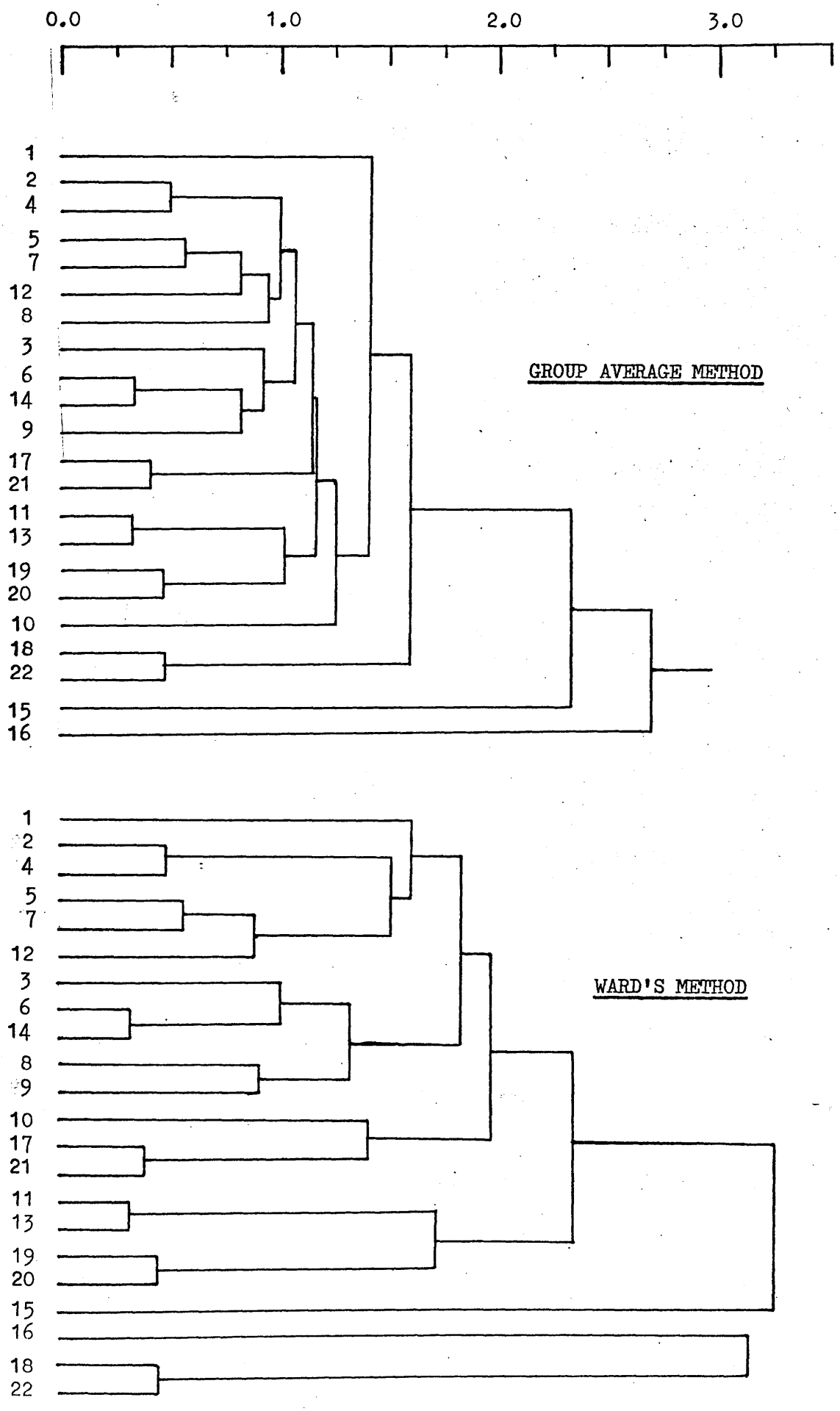
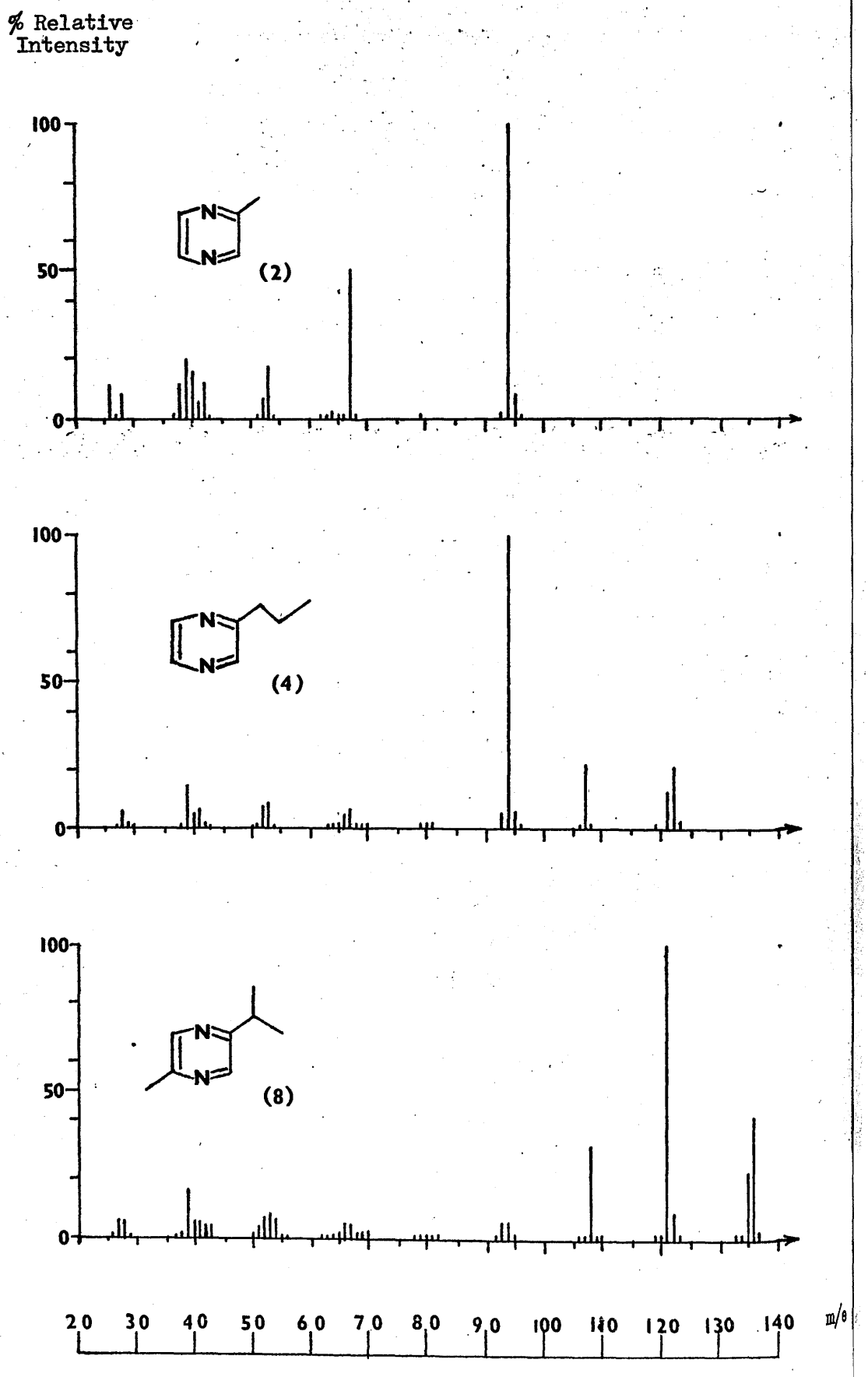


FIG. 9


% Relative Intensity
100
50
0
(2)
100
50
0
(4)
100
50
0
(8)
2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0
m/e

The changes noted between the ninety-six and the one hundred and twenty character dendograms is slight. Significant, however, is the association of compounds 2 and 4 using ninety-six characters, and 4 and 8 using one hundred and twenty characters. The three spectra are illustrated in Figure 9 where they are drawn with the peaks normalised to 100% of the base peak. The influence of the m/e 94 peak is apparently dominant when no fragmentation information is included and compounds 2 and 4 associate when only ninety-six characters are used. The influence of this peak is noteably overshadowed when fragmentation information is included. Compound 8 has a very small m/e 94 peak but associates with compound 4 because of contributions from M - 15 and M - 28 peaks. The 'pattern' of spectrum 4 certainly intuitively relates more closely to that of spectrum 8 than spectrum 2 and emphasises the need to include as much mass spectrometric information in the choice of characters as possible.

d) The potential of clustering methods for the reduction of large quantities of gas chromatography - mass spectrometry (GCMS) data was apparent from studies a), b) and c). A further study was undertaken, therefore, to test the applicability of the technique.

A paper[20] describing the volatile metabolites found by GCMS in the urine of normal subjects and subjects with diabetes mellitus presented a suitable test. The mass spectra of compounds, corresponding as closely as possible to those metabolites found, were collected from catalogued data. They are listed in Table 5. Where possible, alternative catalogued mass spectra of the same compound were used for normal (N) and abnormal (A) subjects. An X in the appropriate column (N or A) indicates that the compound was found in the normal or abnormal subject.

The characters chosen were the square root of the intensity at the ninety-six most significant m/e values,

m/e	38	-	123	inclusive	} 96
m/e	131	-	138	inclusive	
m/e	149	-	150	inclusive	

for some runs, and these ninety-six plus the square root of the intensity of twenty-four fragmentations, M; M - 1, - 2, - 15, - 16, - 17, - 18, - 19, - 20, - 27, - 28, - 29, - 30, - 31, - 32, - 33, - 34, - 35, - 36, - 40, - 43, - 44, - 45 and 46, making one hundred and twenty characters in total, for other runs.

TABLE 5

Cat. Nos.	Mol. Wt.	Name	Formula	N	A
293, 373	44	Acetaldehyde	CH_3CHO		X
294, 374	58	Propionaldehyde	CH_3CH_2CHO	X	X
DOW 3231	114	n-Heptanal	$CH_3(CH_2)_5CHO$	X	
85, 297	58	Acetone	CH_3COCH_3	X	X
298, 377	72	2-Butanone	$CH_3COCH_2CH_3$	X	X
299, 378	86	2-Pentanone	$CH_3COCH_2CH_2CH_3$	X	X
652	86	3-Methyl-2-butanone	$CH_3COCH(CH_3)_2$	X	X
380	100	4-Methyl-2-pentanone	$CH_3COCH_2CH(CH_3)_2$	X	X
663	100	3-Methyl-2-pentanone	$CH_3COCH(CH_3)CH_2CH_3$	X	X
662	100	3-Hexanone	$CH_3CH_2COCH_2CH_2CH_3$	X	X
826	114	2,4-Dimethyl-3-pentanone	$(CH_3)_2CHCOCH(CH_3)_2$	X	X
667	114	3-Heptanone	$CH_3CH_2CO(CH_2)_3CH_3$		X
666	114	2-Heptanone	$CH_3CO(CH_2)_4CH_3$	X	X
833	142	5-Nonanone	$CH_3(CH_2)_3CO(CH_2)_3CH_3$	X	X
75	46	Ethanol	CH_3CH_2OH	X	X
76	60	2-Propanol	$(CH_3)_2CHOH$		X
284	60	1-Propanol	$CH_3CH_2CH_2OH$		X
288	74	2-Methyl-1-propanol	$(CH_3)_2CHCH_2OH$		X
286	74	1-Butanol	$CH_3CH_2CH_2OH$	X	X
800	88	2-Methyl-1-butanol	$CH_3CH_2CH(CH_3)CH_2OH$		X
801	88	3-Methyl-1-butanol	$(CH_3)_2CHCH_2CH_2OH$		X
653	88	1-Pentanol	$CH_3(CH_2)_3CH_2OH$		X
1060	130	2-Octanol	$CH_3CHOH(CH_2)_5CH_3$		X
DOW 2963	130	2-Ethyl-1-hexanol	$CH_3(CH_2)_3CH(CH_2CH_3)CH_2OH$		X
42-m	130	4-Octanol	$CH_3CH_2CH_2CHOH(CH_2)_3CH_3$		X
DOW 820	130	1-Octanol	$CH_3(CH_2)_6CH_2OH$		X

TABLE 5 (continued)

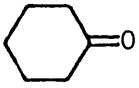
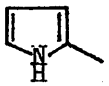
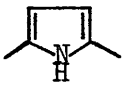
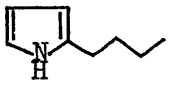
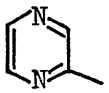
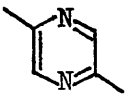
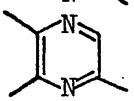
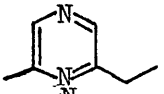
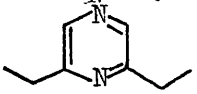
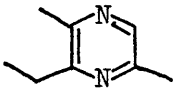
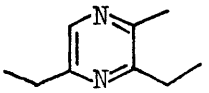
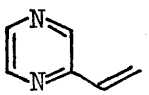
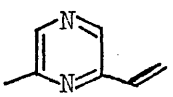
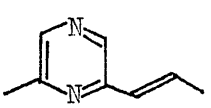
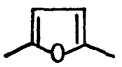
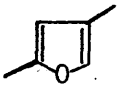
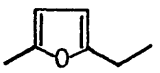
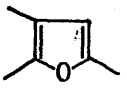
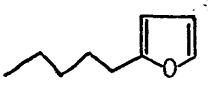
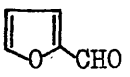
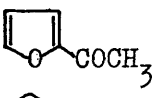
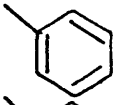
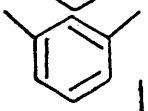
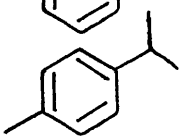
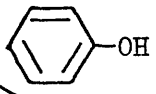
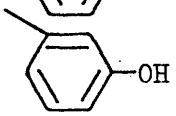
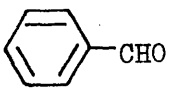
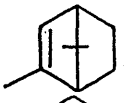
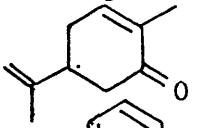
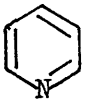
Cat. Nos.	Mol. Wt.	Name	Formula	N	A
660	98	3-Methylcyclopentanone		X	X
449	98	Cyclohexanone			X
381,DOW733	98	4-Methyl-3-penten-2-one	$CH_3COCH{=}C(CH_3)_2$	X	X
782	86	2,3-Butanedione	$CH_3COCOCH_3$	X	X
513, 616	67	Pyrrole		X	X
1719	81	1-Methylpyrrole		X	X
1536	81	2-Methylpyrrole		X	X
1418	95	2,5-Dimethylpyrrole		X	
624	123	1-Butylpyrrole		X	X
*	80	Pyrazine			X
1417, *	94	Methylpyrazine		X	X
250m, *	108	2,5-Dimethylpyrazine		X	X
*	122	2,3,5-Trimethylpyrazine		X	X
*	122	2-Methyl-6-ethylpyrazine		X	X
*	136	2,6-Diethylpyrazine			X
*	136	2,5-Dimethyl-3-ethylpyrazine			X
*	150	2,6-Dimethyl-3-methylpyrazine			X
*	106	Vinylpyrazine		X	X
*	120	2-Methyl-6-vinylpyrazine		X	X
*	134	2-Methyl-6-(trans-1Opropenyl)pyrazine			X

TABLE 5 (continued)

Cat. Nos.	Mol. Wt.	Name	Formula	N	A
815, 1829	96	2,5-Dimethylfuran		X	X
**	96	2,4-Dimethylfuran		X	X
**	110	2-Methyl-5-ethylfuran		X	X
**	110	2,3,5-Trimethylfuran		X	X
**	138	2-n-Pentylfuran		X	X
DOW 844	96	Furfural			X
DOW2932	110	Acetylfuran		X	
DOW 825	92	Toluene		X	X
178	106	Dimethylbenzene			X
462	134	1-Methyl-4-isopropylbenzene		X	
MCARP 66	94	Phenol			X
MCARP 69	108	β-Cresol		X	X
MCARP 83	106	Benzaldehyde		X	X
490	136	β-Pinene		X	X
MCARP 94	150	Carvone		X	X
619	79	Pyridine			X
604, 691	118	Chloroform	$CHCl_3$	X	X
476	94	Dimethyldisulphide	CH_3SSCH_3	X	X

All Cat. Nos. refer to A.P.I. collection except DOW, M.C.A.R.P.and

* H. A. Bondarovich et al, J. Agr. Food Chem., 15,1093(1967),

** K. Haynes et al Tet., 22,2223(1966).

Clustering was performed by the Group Average Method and Ward's Method on forty-two compounds corresponding to the metabolites found in normal urine, on sixty compounds corresponding to the metabolites found in abnormal urine, and on the one hundred and two compounds corresponding to the combination of normal and abnormal urine metabolites. Clustering was performed using both the ninety-six characters and the one hundred and twenty characters described above.

The dendograms produced for these combinations are shown in Figures 10, 11, 12, 13, 14 and 15 for Ward's Method only. The compounds are listed in order for the combined normal and abnormal dendogram i.e. Figure 12 and the numbers used to list the compounds in Figure 12 are used consistently in Figures 10, 11, 13, 14 and 15.

DISCUSSION

Figure 10 is a profile of the collection of the forty-two compounds identified in a normal subject. Some of the clusters are tight. Particularly noteable is the group of furans, compounds 9, 12, 13, 14 and 15; the group of ketones, compounds 16, 17, 18, 19, 21, 20, 22, 23 and 25; and the group of nitrogen aromatics, compounds 47, 48, 11, 58, 59, 51, 52, 53, 54, 56 and 57.

Figure 11 is a profile of the collection of sixty compounds identified in an abnormal subject. Not able is the clustering of the low molecular weight oxygen compounds, numbers 2, 1, 3, 4, 5 and 6; the group of ketones, compounds 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 and 26; the alcohols, compounds 29, 28, 30, 31 and 32 and also 37, 38 and 39; the unsaturated ketones, compounds 33, 34 and 35; the furans, compounds 8, 10, 12, 13 and 14; the group of nitrogen aromatics, compounds 49, 54, 47, 48, 58, 59, 56, 57, 50 and 51; and the phenyls, compounds 41, 43, 42 and 44.

FIG. 10, NORMAL METABOLITES, 42 COMPOUNDS, 120 CHARACTERS, WARD'S METHOD.

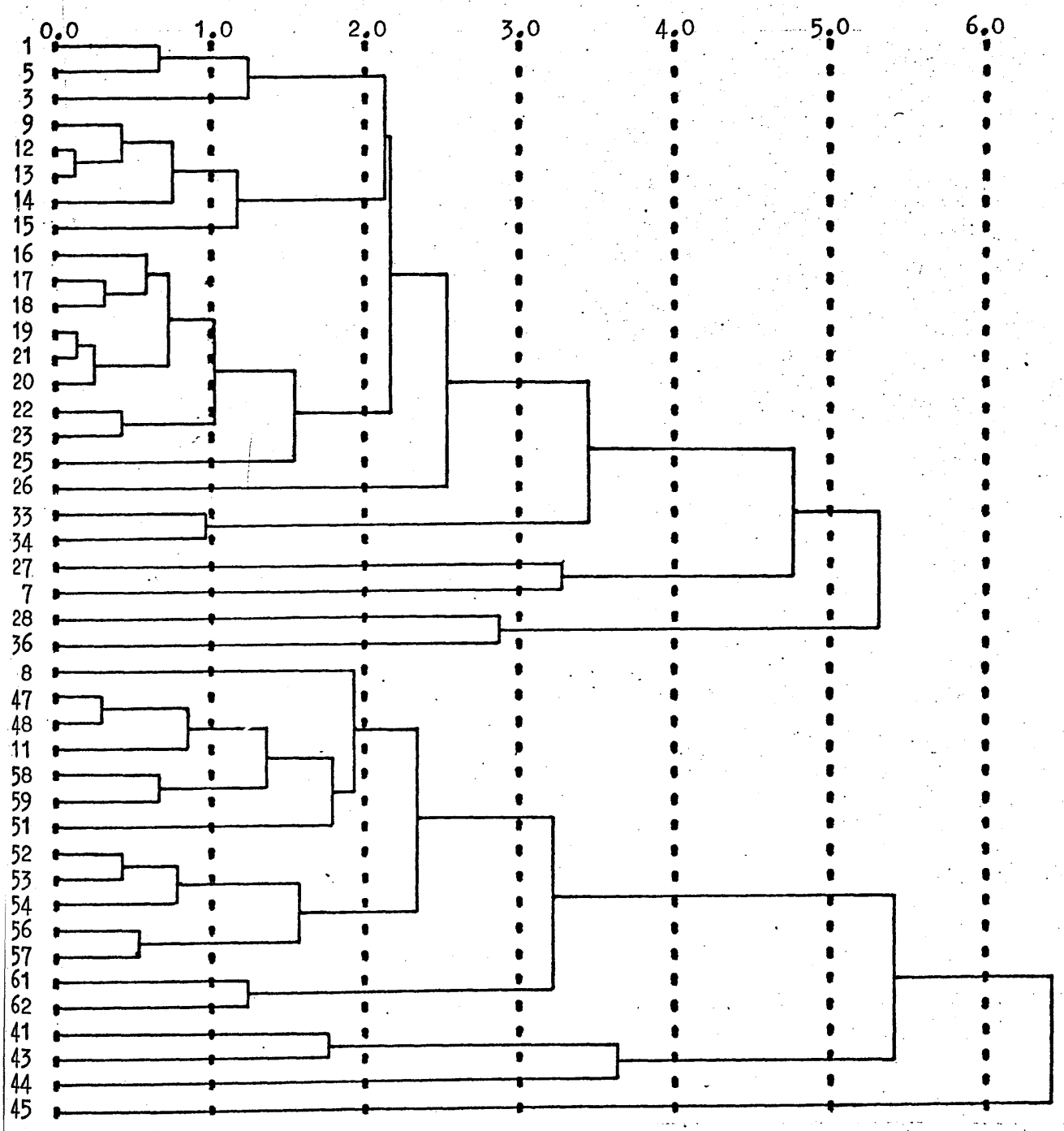


FIG. 11, ABNORMAL METABOLITES, 60 COMPOUNDS, 120 CHARACTERS, WARD'S METHO

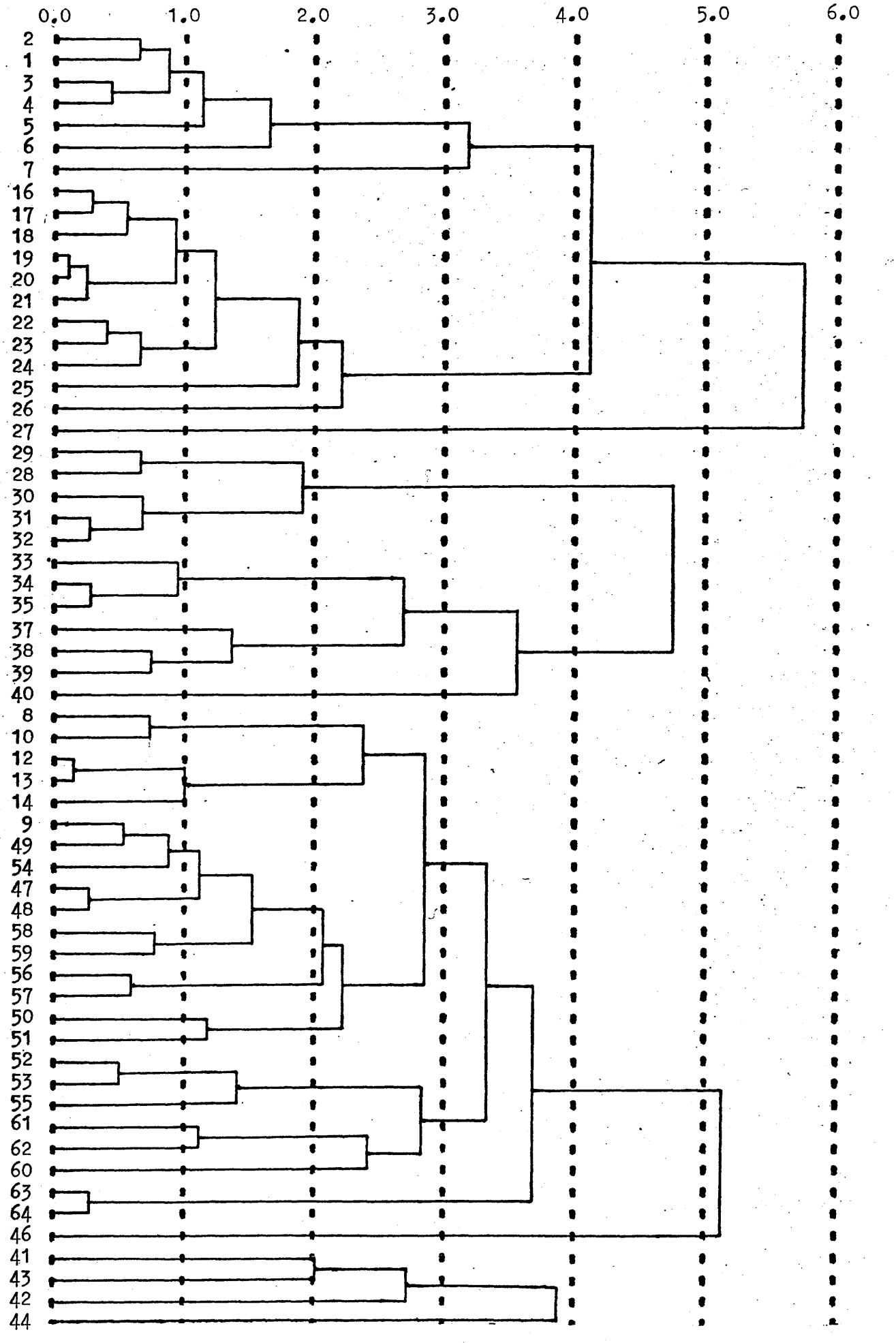
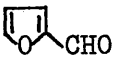
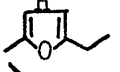
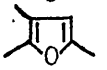
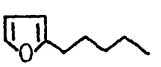
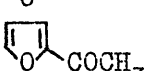


FIG. 12.

1	CH_3COCH_3
2	CH_3CHO
3	CH_3CH_2OH
4	$(CH_3)_2CHOH$
5	CH_3CH_2CHO
6	$CH_3CH_2CH_2OH$
7	CH_3SSCH_3
8	
9	
10	
11	
12	
13	
14	
15	

(over)

16 $CH_3COCH_2CH_3$

17 $CH_3COCH(CH_3)CH_2CH_3$

18 $CH_3COCH_2CH(CH_3)_2$

19 $CH_3COCH(CH_3)_2$

20 $CH_3COCH_2CH_2CH_3$

21 $CH_3COCOCH_3$

22 $CH_3CO(CH_2)_4CH_3$

23 $(CH_3)_2CHCOCH(CH_3)_2$

24 $CH_3CH_2COCH_2CH_2CH_2CH_3$

25 $CH_3(CH_2)_3CO(CH_2)_3CH_3$

26 $CH_3CH_2COCH_2CH_2CH_3$

27 $CHCl_3$

28 $CH_3CH_2CH_2CH_2OH$

29 $(CH_3)_2CHCH_2OH$

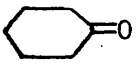
30 $CH_3CH_2CH(CH_3)CH_2OH$

31 $(CH_3)_2CHCH_2CH_2OH$

32 $CH_3(CH_2)_3CH_2OH$

33 $(CH_3)_2C{=}CHCOCH_3$

34 

35 

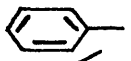
36 $CH_3(CH_2)_5CHO$

37 $CH_3(CH_2)_5CH(CH_3)OH$

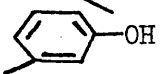
38 $CH_3(CH_2)_3CH(CH_2CH_3)CH_2OH$

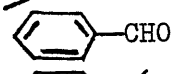
39 $CH_3(CH_2)_6CH_2OH$

40 $CH_3CH_2CH_2CHOH(CH_2)_3CH_3$

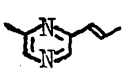
41 

42 

43 

44 

45 

46 

(over)

47 

48 

49 

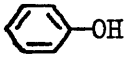
50 

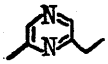
51 

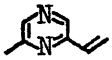
52 

53 

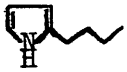
54 

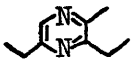
55 

56 

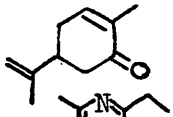
57 

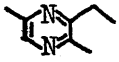
58 

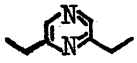
59 

60 

61 

62 

63 

64 

(over)

FIG. 12

COMBINED URINE METABOLITES, 102 COMPOUNDS, 120 CHARACTERS, WARD'S METHOD.

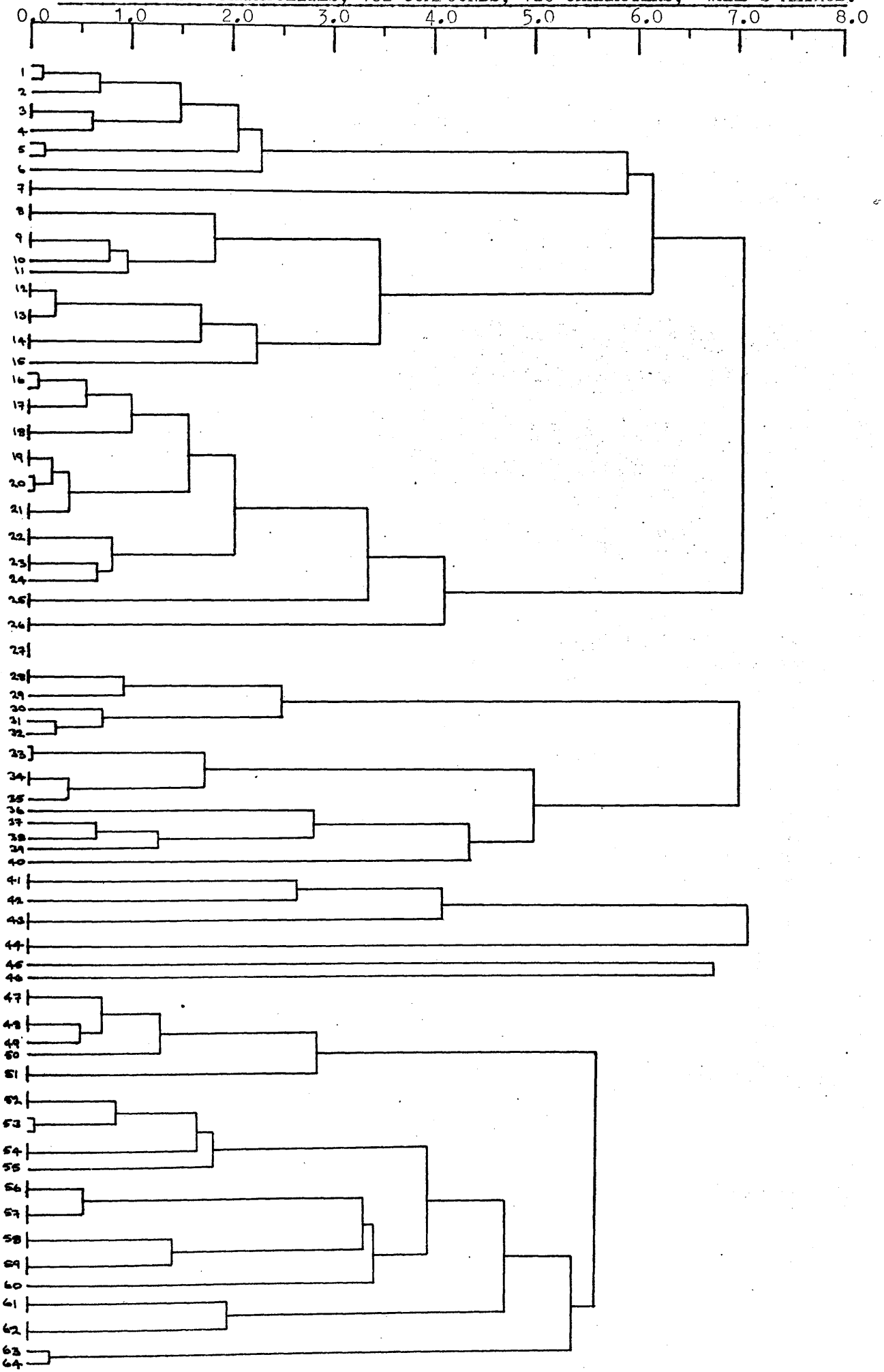


FIG. 13, NORMAL METABOLITES, 42 COMPOUNDS, 96 CHARACTERS, WARD'S METHOD.

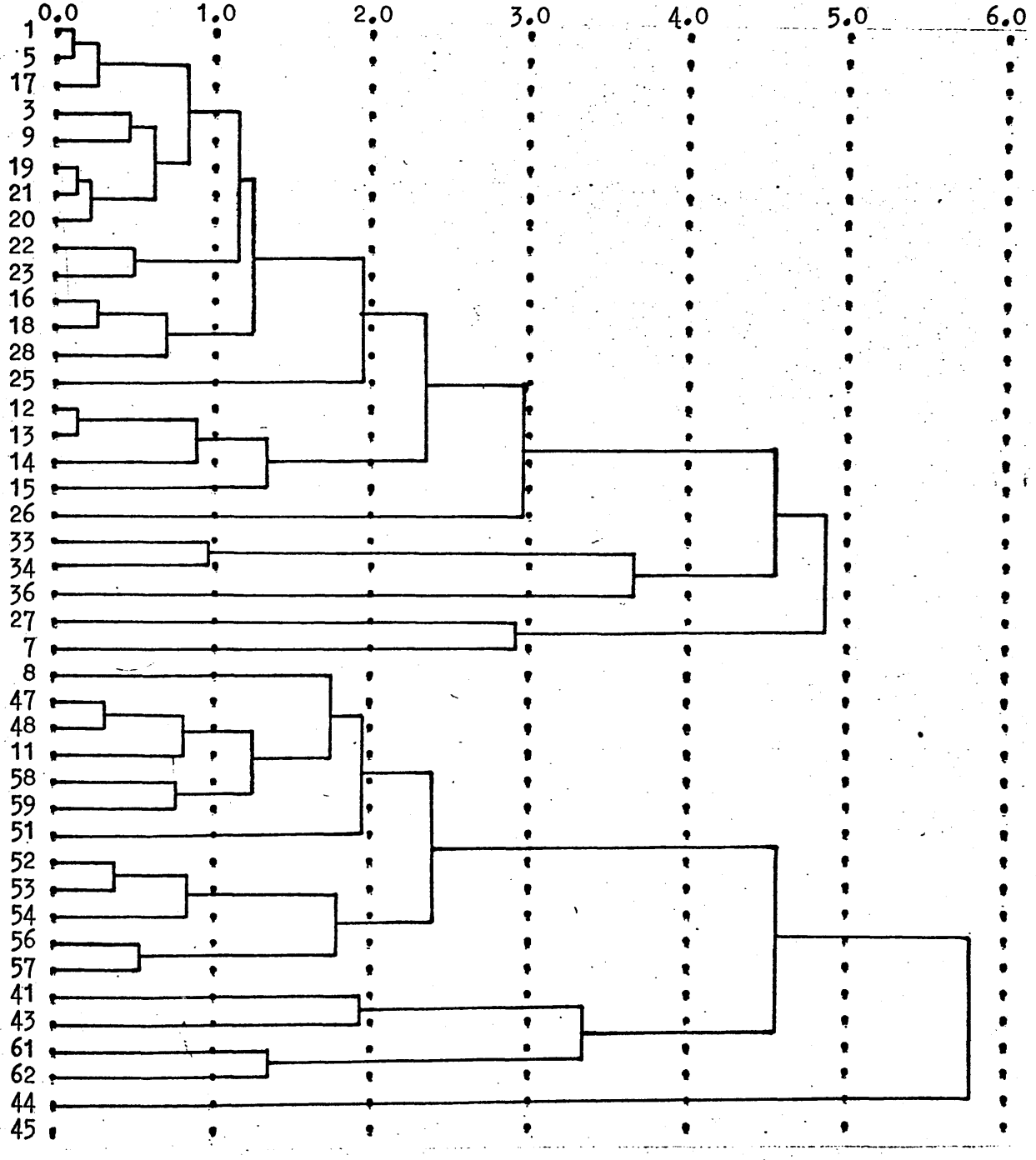


FIG. 14, ABNORMAL METABOLITES, 60 COMPOUNDS, 96 CHARACTERS, WARD'S METHOD.

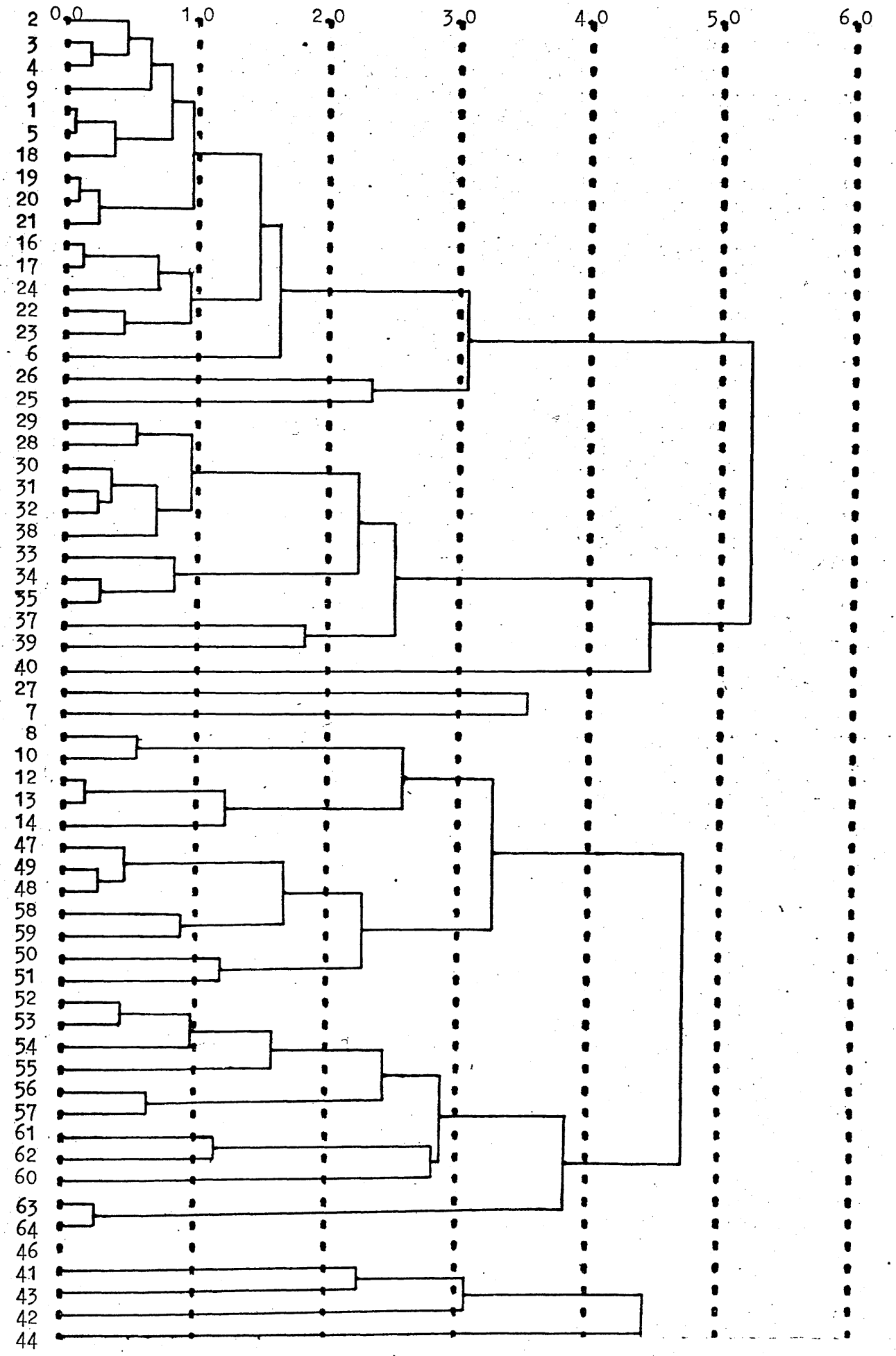
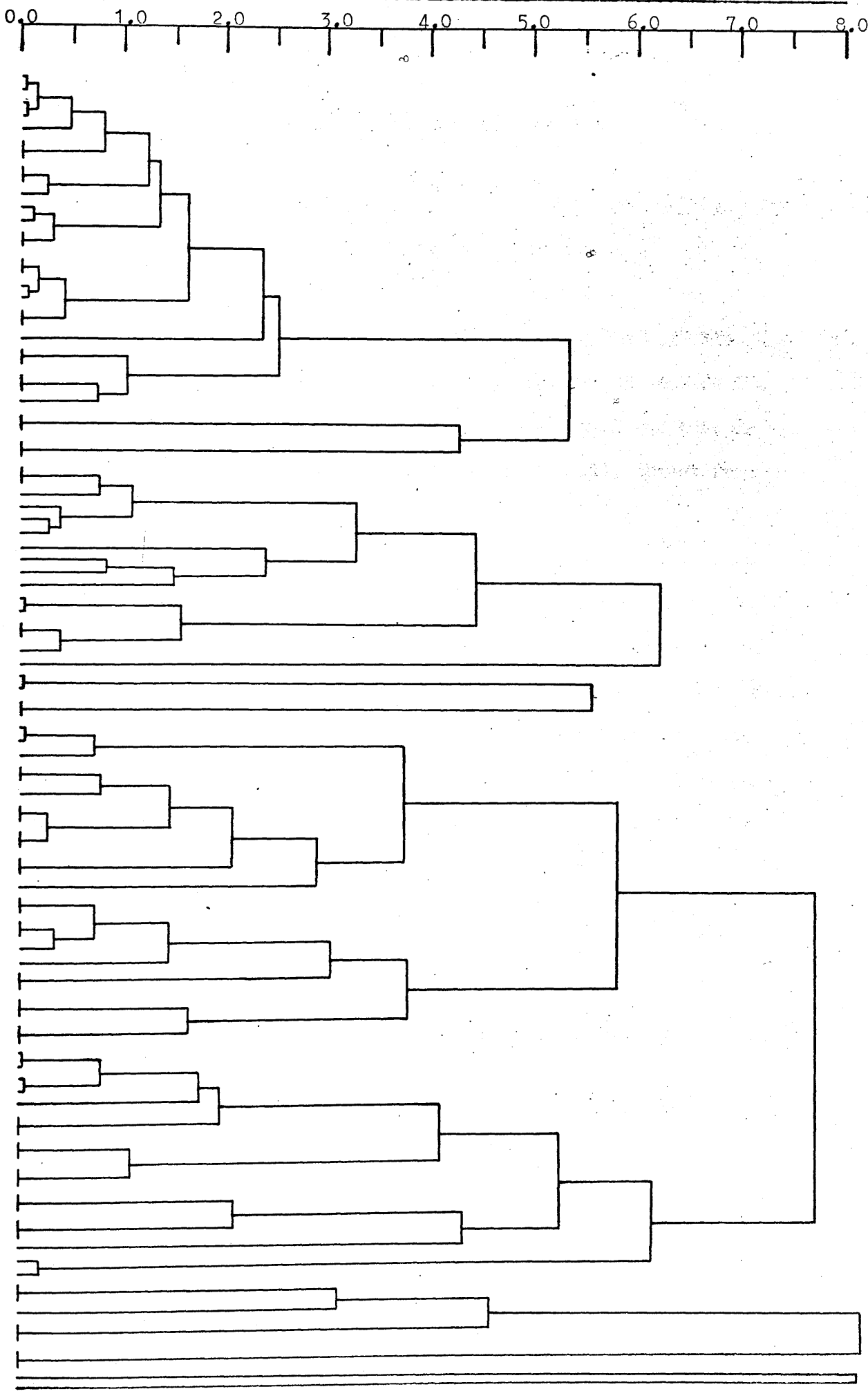


FIG. 15

COMBINED URINE METABOLITES, 102 COMPOUNDS, 96 CHARACTERS, WARD'S METHOD



The separation and grouping of these compounds is very encouraging from the point of view of data reduction. That is, with some more examples of profiles and a refinement of technique, described below, good predictive ability should be possible with unknown subjects.

Figure 12 presents a very useful direct comparison of a known normal subject with, in this case, a known abnormal subject. The clustering can be compared favourably with either the normal or the abnormal profiles, but, more important, is the fact that this direct comparison enables the compounds which are characteristically found in an abnormal subject, or alternatively, those compounds which are characteristically absent from an abnormal subject, to be pinpointed. That is, by simply running quickly through the dendogram from top to bottom, (this could be done automatically), the compounds 2, 4, 6, 10, 11, 15, 24, 29 - 32, 35 - 40, 42, 45, 46, 49, 50, 55, 60, 63 and 64 are immediately seen to be characteristic of normal or abnormal subjects. Since it is known which of these compounds came from which subject (N or A), the presence or absence of a metabolite in the abnormal subject is identified. Further, these compounds are already partially identified since they are clustered with (mass spectrometrically) similar compounds.

The dendograms produced by Ward's Method using the normal metabolites, the abnormal metabolites and the combination of both, without the inclusion of mass spectrometric fragmentation characters (i.e. only ninety-six characters) are shown in Figures 13, 14 and 15 for comparison. In this case clustering is very similar to the one hundred and twenty character examples.

CONCLUSIONS

The 'pattern recognition' approach to the analysis of mass spectral data adopted in this study is shown to be of considerable potential. The profiles of large collections of mass spectral data, obtained using what must be considered as a very crude attempt at selecting characters in this case, are interesting and useful applications can be suggested. The potential utility of the method for data reduction is demonstrated in study d). It is intended that the method be used to obtain profiles of the patterns of the mass spectra of the air samples obtained when the instrumentation mentioned in the introduction is operating.

Ward's Method appears to be the most suitable for further applications.

It is pertinent that the approaches made in the preprocessing stages of these studies, that is, intuitively selecting m/e values and fragmentation characteristics, is very elementary. Considerably better clustering is anticipated when more care is taken to select characters. For example, characteristic ions, series of ions, and the fragmentations associated with particular applications could be selected. There is also the further possibility of using high resolution mass spectrometric data and other parameters, such as gas-chromatographic retention times, to facilitate more specific clustering.

It is also noted that cluster analysis programmes are available, as in this case, in computer package form, therefore no programming is required by the mass spectroscopist.

REFERENCES.

1. P.M. Llewellyn, U.S. Patent 3,455,092(1969).

2. G.G. Collins and D. Utley, Chem. and Ind., 1972, 84.

3. L. Fraser Monteiro and R.I. Reed, Internat. J. Mass Spectrometry Ion Phys., 1969, 2, 265.

4. J. Hardy and I. Jardine, Proc. 6th International Mass Spectrometry Conference, Edinburgh, 1973.

5. K. Fukunaga, "Introduction to Statistical Pattern Recognition," Academic Press, London, 1972.

6. W.S. Meisel, "Computer - Oriented Approaches to Pattern Recognition," Academic Press, New York, 1972.

7. N.J. Nilsson, "Learning Machines," McGraw - Hill, New York, 1965.

8. R.R. Sokal and P.H.A. Sneath, "Principles of Numerical Taxonomy," W.H. Freeman and Company, San Francisco, 1963.

9. N. Jardine and R. Sibson, "Mathematical Taxonomy," John Wiley and Sons Ltd., London, 1971.

10. A.J. Cole (ed.), "Numerical Taxonomy," Academic Press, London, 1969.

11. R.C. Tryon and D.E. Bailey, "Cluster Analysis," McGraw - Hill, New York, 1970.

12. D.F. Morrison, "Multivariate Statistical Methods," McGraw - Hill, New York, 1967.

13. R.R. Sokal and C.D. Michener, Kans. Univ. Sci. Bull., 1958, 38, 1409.

14. J.H. Ward, J. Amer. Stat. Ass., 1963, 58, 236.

15. G.N. Lance and W.T. Williams, Comp. J., 1967, 9, 373.

16. D. Wishart, Biometrics, 1969, 25, 165; "Fortran programs for 8 methods of cluster analysis," Kans. Geol. Comp. Contr. No. 38, Kansas.

17. D. Wishart, "Clustan 1A," University of Edinburgh Scientific and Social Sciences Program Library, Cat. No. 19.200.200, 1969; Amendments by R.L. Middleton, Cat. No. 19.410.200, 1972.

18. K. Grob and G. Grob, J.Chromatogr., 1971, 62, 1.

19. H.A. Bondarovich, P. Friedel, V. Krampl, J.A. Renner, F.W. Shephard and M.A. Gianturco, J. Agr. Food Chem., 1967, 15, 1093.

20. A. Zlatkis, W. Bertsch, H.A. Lichtenstein, A. Tishbee, F. Shunbo, H.M. Liebich, A.M. Coscia and N. Fleischer, Analyt. Chem., 1973, 45, 764.

CHAPTER 3.

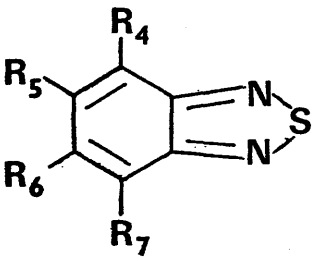
MASS SPECTRA OF BENZO - 2,1,3 - THIADIAZOLES

INTRODUCTION

Benzo - 2,1,3 - thiadiazoles show considerable herbicidal and fungicidal activities.[1] The mass spectra of a selection of benzo - 2,1,3 - thiadiazoles were obtained to determine and correlate the characteristics of their fragmentation under electron impact and especially to examine whether processes analogous to the benzyne formation postulated for benzo - 2,1.3 - selenodiazoles[2] were important. Although the mass spectra of the isomeric 1,2,3 - thiadiazoles[3,4] and the benzo - 1,2,3 - thiadiazole[5] have been reported and discussed, the mass spectrum of benzo - 2,1.3 - thiadiazole (1) has not been reported, except for the intensity of the M - 46 (NS) peak, which has been compared with the intensities of the analogous M - 30 (NO) and M - 28 (N_2) peaks from benzo - 2,1,3 - oxadiazole and benzo - 1,2.3 - thiadiazole respectively.[5] The mass spectra of the benzo - 2,1.3 - thiadiazoles listed in Table 1 are given in Table 2.

TABLE 1

Benzo - 2,1,3 - thiadiazoles (1 - 35)



No.	R_4	R_5	R_6	R_7
1	H	H	H	H
2	Cl	H	H	H
3	H	Cl	H	H
4	Cl	H	H	Cl
5	Cl	Cl	H	Cl
6	Cl	Cl	Cl	Cl
7	F	Cl	H	Cl
8	Br	H	H	Br
9	CH_2I	H	H	H
10	CH_2Br	Cl	H	Cl
11	CH_3	Cl	H	Cl
12	CH_3	H	Cl	Cl
13	CH_3	Cl	CH_3	Cl
14	OH	H	H	H
15	H	OH	H	H

TABLE 1 - continued

No.	R_4	R_5	R_6	R_7
16	Cl	OH	H	H
17	OH	Cl	H	Cl
18	OCH_3	H	H	H
19	H	OCH_3	H	H
20	OCH_3	H	H	Cl
21	Cl	OCH_3	Cl	Cl
22	OCH_2CH_3	Cl	H	Cl
23	OC_6H_5	NO_2	H	Cl
24	NO_2	H	H	H
25	H	NO_2	H	H
26	NO_2	Cl	H	H
27	NO_2	Cl	H	Cl
28	NO_2	Cl	Cl	NO_2
29	NO_2	Cl	NO_2	Cl
30	NH_2	NO_2	H	Cl
31	NH_2	NH_2	H	H
32	H	SO_2NH_2	H	H
33	CHO	Cl	H	Cl

TABLE 1 - continued

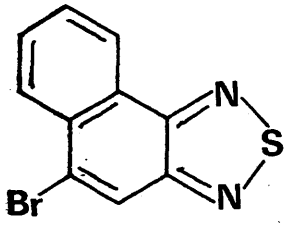
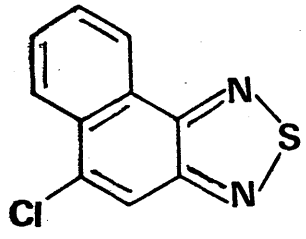
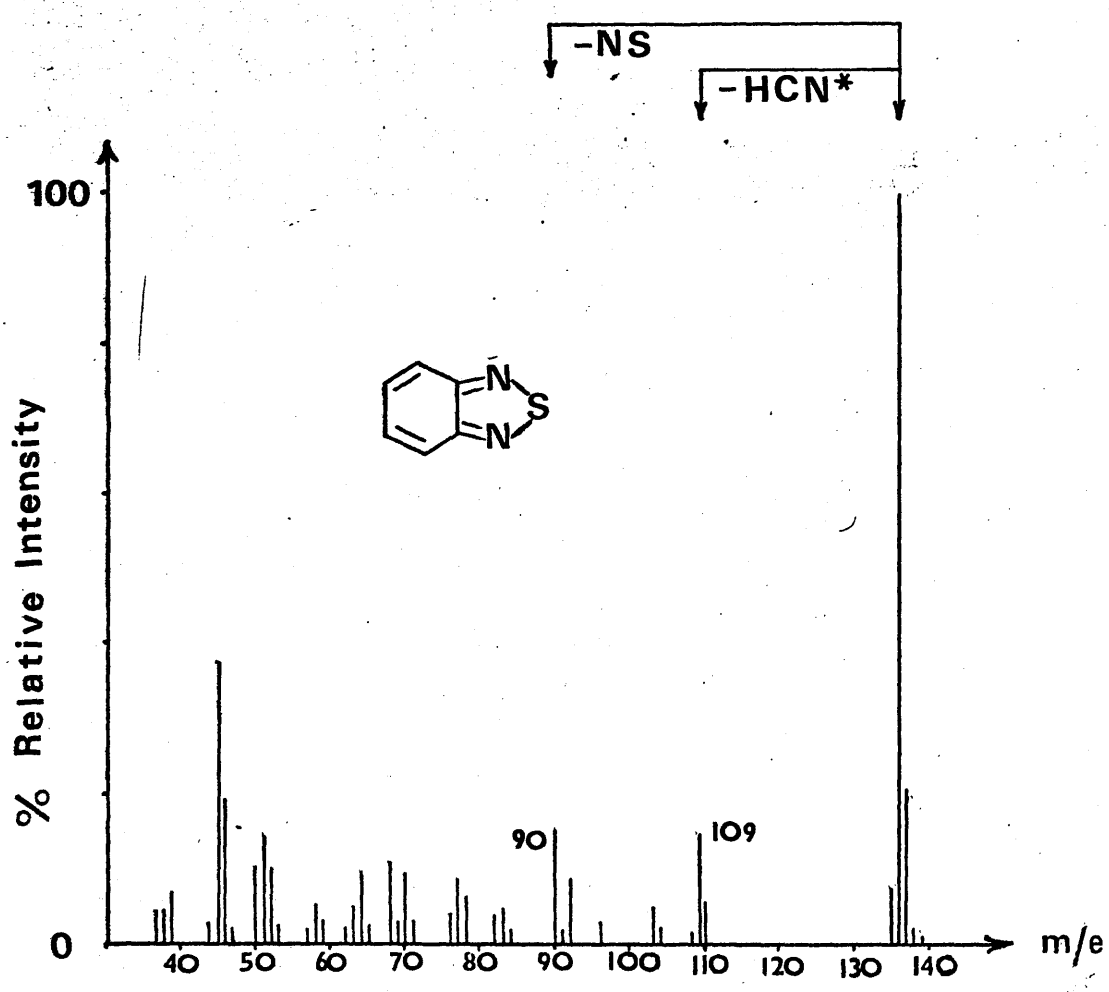
No.	R_4	R_5	R_6	R_7
34				
35				

FIG. 1

Mass Spectrum of Benzo-2,1,3-thiadiazole (1).



DISCUSSION

The mass spectrum (Fig. 1) of benzo - 2,1,3 - thiadiazole (1) displays the molecular ion as the base peak. The most important fragmentations are loss of H, HCN and NS.

As expected, the halogen - substituted benzo - 2,1,3 - thiadiazoles (2 - 8) underwent stepwise loss of halogen atoms rather than fragmentation of the ring system, although peaks from M - NS are present in all of the spectra, with M - HCN and M - halogen - HCN often occurring.

The halomethyl compounds (9,10) have low molecular ion abdundances with significant tropylium type ion peaks at M - halogen.[6] For the iodo compound the only other significant feature is the loss of HCN from this ion while for the bromo derivative Cl loss and Cl + NS loss from this tropylium type ion are significant.

The isomeric methyl benzo - 2,1,3 - thiadiazoles (11,12) yield virtually identical spectra which have high abundance molecular ions and M - Cl ions with small but significant M - 2Cl ions. The methyl benzo - 2,1,3 - thiadiazole (13) with one more chlorine has identical characteristics.

The hydroxy isomers (14,15) have similar spectra with the loss of HCN, CO and HCN + CO being significant.[7] The expulsion of carbon monoxide is more important when the hydroxyl is in the 5 - than in the 4 - position. Expulsion of carbon monoxide is important for the chlorinated hydroxy derivatives (16,17).

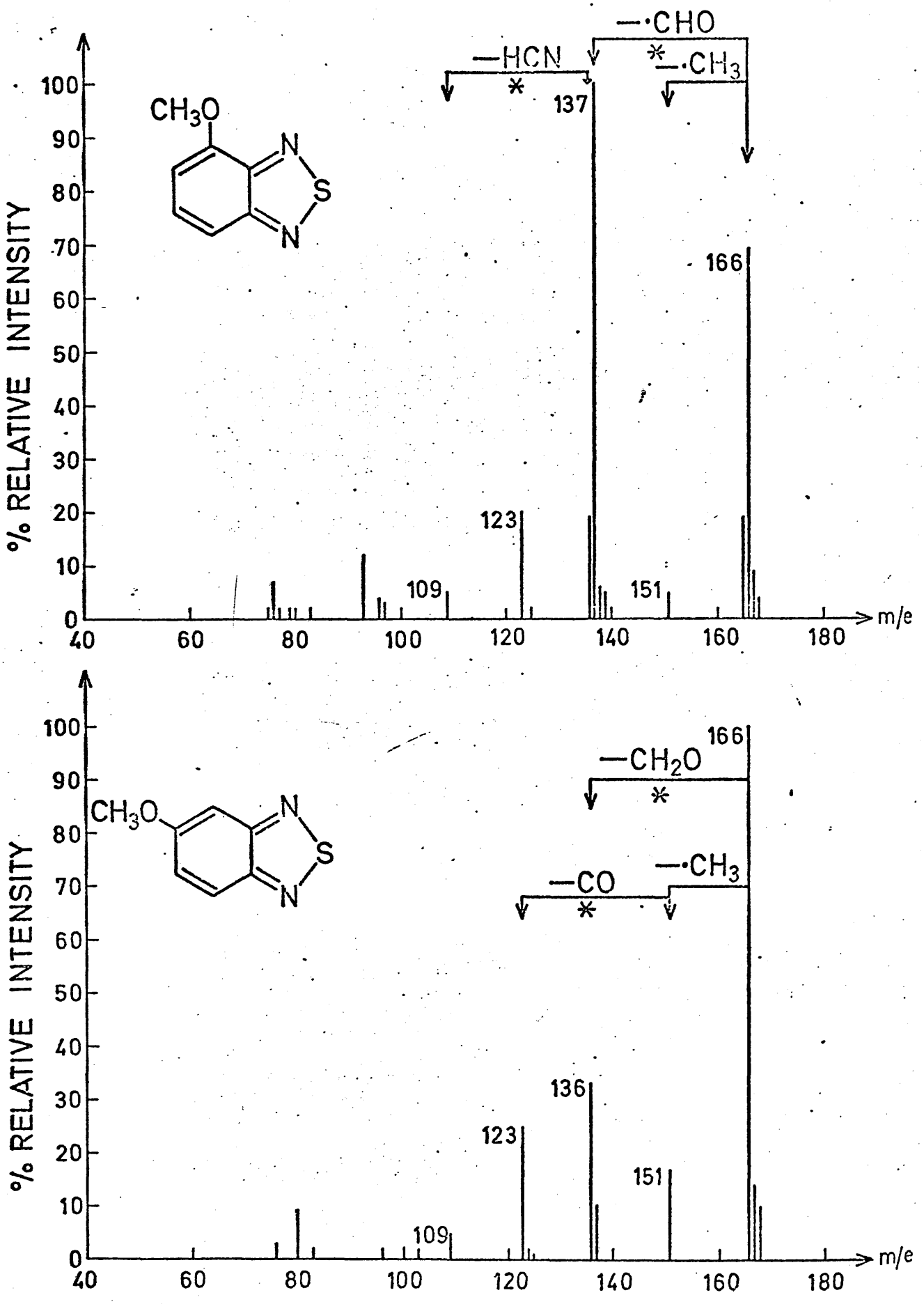


FIG. 2. MASS SPECTRA OF 4- AND 5-METHOXY BENZO-2,1,3-THIADIAZOLES (18,19)

The difference between the mass spectra of the isomeric methoxy compounds (18,19) is interesting (Fig. 2).

For both isomers loss of a methyl radical followed by expulsion of carbon monoxide is important as in the mass spectra of anisoles.[8] However, although the 5 - methoxy compound expels formaldehyde from the molecular ion to give an intense peak (33%) with a less significant loss of formyl radical (10%), the 4 - methoxy derivative displays the loss of formyl radical as the base peak with formaldyhyde expulsion being less important (19%). Intense metastables support these transitions in both spectra. Significantly the mass spectrum of the chloro substituted 4 - methoxy derivative (20) displays the loss of formyl radical as the base peak. Loss of methyl radical and then carbon monoxide is important for both chloro - substituted methoxy derivates (20,21).

In the spectrum of 4 - ethoxy - 5,7 - dichloro benzo - 2,1,3 - thiadiazole (22), the base peak is found to be formed by analogy to many phenetols.[9] The fragmentation of the phenoxy ion dominates the spectrum of the phenoxy substituted derivative (23).[10]

In the nitro substituted benzo - 2,1,3 - thiadiazoles (24 - 29), a change in the position of the nitro group does not change the fragmentation pattern significantly. The spectra of the 4 - and 5 - nitro benzo - 2,1,3 - thiadiazoles (24,25) are practically identical. The only important fragmentations are of the nitro group.[11] In the amino substituted compounds (30,31) the expected fragmentations of the amino group are found,[12] although where a nitro group is present (30) its fragmentation dominates.

The fragmentation of the sulphonamide (32) is completely analogous to that of benzene sulphonamide.[13] That is, expulsion of the NH_2 group gives a significant ion (12%) while further loss of SO affords the base peak. Expulsion of the SO_2NH_2 group from the molecular ion produces the second most abundant peak. Loss of carbon monoxide gives the base peak in the aldehyde (33) with further loss of chlorine radical producing a significant peak.[14]

The naphthalene derivatives (34,35) behave similarly. Loss of halogen and NS from the molecular ion base peaks produce the most significant ions in both spectra.

CONCLUSION

The dominating feature of the mass spectra of the benzo - 2,1,3 - thiadiazoles is that they almost all exhibit molecular ions of very great abundance (often the base peak as in the benzo - 2,1,3 - selenodiazoles[3]). The fragmentation of the ring system is far less important than fragmentation of the substituents. Thus it is often easy to deduce from the mass spectrum the nature, but not the positions of the substituents except for the hydroxy - and alkoxy - benzo - 2,1,3 - thiadiazoles. Fragmentations analogous to the benzyne formation postulated for the benzo - 2,1,3 - selenodiazoles[3] are not important.

EXPERIMENTAL

The benzo - 2,1,3 - thiadiazoles were synthesized in known ways.[1,15]

The mass spectra were determined on a GEC - AEI MS - 902 double focusing mass spectrometer with ionizing potential of 70 eV and an accelerating potential of 8 kV. The compounds were introduced via the direct insertion probe at an inlet temperature of 150°.

ACKNOWLEDGEMENTS

The samples for this work were provided by Drs. F.C.V. Larsson and S. -O. Lawesson with whom much discussion took place. The work has been submitted with Drs. F.C.V. Larsson, S. -O Lawesson and R.I. Reed for publication.

TABLE 2

Mass Spectra of Substitued Benzo - 2,1,3 - thiadiazoles (1 - 35)

m/e values are followed by intensity (%)

1. 138(2), 137(20), 136(100) $M^{+\cdot}$, 135(8), 110(6), 109(15), 108(2), 104(2), 103(5), 96(3), 92(7), 90(15), 84(2), 83(5), 82(4), 78(6), 77(9), 76(4), 71(3), 70(8), 69(3), 68(11), 64(10), 63(5), 58(5), 52(10), 51(15), 50(10), 46(19), 45(37).

2. 172(39), 170(100) $M^{+\cdot}$, 143(4), 135(22), 126(6), 124(17), 108(50), 103(39), 85(45), 76(6).

3. 172(37), 170(100) $M^{+\cdot}$, 143/5), 135(18), 126(6), 124(18), 108(5), 103(4), 85(5), 76(6).

4. 208(16), 207(5), 206(69), 205(9), 204(100) $M^{+\cdot}$; 179(2), 177(2), 171(14), 169(31), 160(7), 158(12), 142(3), 134(5), 110(9), 104(4), 103(5), 98(2), 83(2), 77(2), 76(3).

5. 244(5), 242(35), 241(9), 240(100), 239(8), 238(100) $M^{+\cdot}$, 207(5), 205(22), 203(35), 196(3), 194(6), 192(7), 170(5), 168(11), 144(4), 142(5), 120(5), 119(6), 100(3), 98(3), 75(3), 74(5), 59(6).

TABLE 2 - continued

6. 280(2), 278(14), 276(51), 274(100), 272(72) M $^{+\cdot}$,
241(11), 239(30), 237(31), 230(4), 228(8), 226(6),
207(9), 205(11), 204(18), 202(27), 195(9), 193(8),
181(4), 179(5), 178(5), 176(7), 172(6), 170(7),
167(5), 158(6), 155(10), 146(8), 144(12), 135(9),
134(8), 132(15), 123(8), 121(9), 109(20), 100(24),
86(12), 85(12), 79(9), 74(15), 70(11), 62(11),
47(11), 46(13).

7. 226(12), 225(6), 224(67), 223(8), 222(100) M $^{+\cdot}$,
189(10), 187(24), 178(3), 176(5), 166(3), 164(5),
155(4), 153(10), 143(6), 113(4), 111(4).

8. 297(5), 296(50), 295(8), 294(100), 292(50) M $^{+\cdot}$,
250(1), 248(2), 246(1), 220(5), 218(7), 215(28),
213(27), 184(7), 148(5), 134(22), 108(3), 84(5),
51(9).

9. 276(5) M $^{+\cdot}$, 151(6), 150(16), 149(100), 128(6),
127(12), 122(17), 105(7), 90(5), 64(8), 63(9),
55(6), 52(6), 51(5), 50(5), 46(7), 45(8).

10. 296(-) M $^{+\cdot}$, 221(20), 219(72), 218(14), 217(59), 216(13),
185(8), 184(45), 183(16), 182(100), 180(13), 166(15),
158(8), 156(8), 151(13), 149(13), 148(15), 147(8),
136(33), 135(12), 124(9), 123(12), 110(7), 109(10),
108(8), 88(7), 83(6), 78(7), 77(7), 76(12),
75(9), 70(7), 64(13), 51(12), 50(10), 46(11),
45(17), 44(10).

TABLE 2 - continued

11. 222(14), 221(8), 220(69), 219(14), 218(100) M $^{+\cdot}$, 217(17), 185(55), 184(14), 183(97), 182(7), 158(3), 156(6), 149(8), 148(8), 147(5), 124(5), 120(2), 109(5), 94(8).

12. 222(14), 221(8), 220(67), 219(17), 218(100) M $^{+\cdot}$, 217(11), 185(30), 184(10), 183(90), 182(5), 151(8), 148(10), 147(4), 124(5), 109(5).

13. 236(13), 235(8), 234(75), 233(15), 232(100) M $^{+\cdot}$, 231(8), 199(32), 197(100), 187(12), 183(8), 162(8), 161(12), 152(8).

14. 154(5), 153(9), 152(100) M $^{+\cdot}$, 125(13), 124(17), 97(10), 93(8), 91(3), 65(5), 64(5).

15. 154(6), 153(9), 152(100) M $^{+\cdot}$, 125(5), 124(37), 120(4), 106(4), 97(18), 91(6), 81(6), 80(25), 71(9), 70(11), 68(8), 64(10), 53(9), 52(27), 51(12).

16. 190(8), 189(30), 188(43), 187(76), 186(100) M $^{+\cdot}$, 185(18), 184(8), 183(40), 182(8), 181(22), 166(8), 164(14), 160(12), 158(26), 155(12), 152(38), 150(8), 149(18), 145(6), 143(16), 136(12), 134(12), 123(18).

TABLE 2 - continued

17. 224(12), 223(5), 222(57), 221(7), 220(83) M $^{+\cdot}$, 218(16), 208(6), 206(19), 204(26), 194(8), 192(9), 183(15), 181(7), 168(9), 156(8), 136(10), 135(6), 120(6), 119(6), 110(6), 98(11), 78(56), 77(11), 74(38), 59(67), 52(9), 51(9), 50(7), 45(38), 41(9).

18. 168(4), 167(9), 166(69) M $^{+\cdot}$, 165(19), 151(5), 139(5), 138(6), 137(100), 136(19), 135(21), 125(2), 123(20), 109(5), 97(3), 96(4), 93(12), 83(2), 79(2), 77(2), 76(7), 75(2).

19. 168(10), 167(14), 166(100) M $^{+\cdot}$, 151(17), 137(10), 136(33), 123(25), 109(4), 80(9).

20. 202(29), 201(11), 200(75) M $^{+\cdot}$, 199(11), 197(7), 187(32), 185(29), 183(16), 181(9), 173(34), 172(12), 171(100), 170(12), 159(12), 157(32), 152(11), 143(6), 137(8), 136(7), 135(10), 127(7), 111(6), 100(5), 98(4), 83(5), 76(8), 75(7), 73(6), 70(9), 64(16), 52(5), 51(13), 46(11), 45(10).

21. 274(9), 272(25), 271(6), 270(57), 269(13), 268(57) M $^{+\cdot}$, 257(11), 255(33), 253(33), 229(11), 227(30), 225(30), 222(20), 221(12), 220(71), 219(24), 218(71), 217(21), 185(43), 184(17), 183(100), 182(7), 181(26), 166(18), 158(6), 157(7), 151(12), 149(11), 148(13), 137(6), 136(30), 135(13), 132(8), 124(8), 123(11), 110(9), 109(11), 108(7), 79(6), 76(8), 74(71), 70(13), 64(8), 59(100), 52(7), 51(100), 50(7), 46(14), 45(71), 44(14), 43(13), 41(29).

TABLE 2 - continued

22. 252(3), 250(14), 248(22) M $^{+\cdot}$, 224(14), 223(6), 222(72), 221(9), 220(100), 206(7), 204(11), 169(5), 155(5), 98(5).

23. 309(8), 307(20) M $^{+\cdot}$, 292(8), 290(10), 262(10), 260(6), 216(7), 214(16), 206(5), 204(8), 186(9), 170(5), 119(9), 118(6), 94(6), 93(100), 86(8), 77(40), 70(5), 65(49), 51(34).

24. 183(7), 182(9), 181(100) M $^{+\cdot}$, 151(30), 136(5), 135(50), 123(24), 109(7), 108(23), 104(14), 91(7), 83(5), 82(5), 76(13), 75(5). 70(5), 64(20), 52(6), 51(6), 50(5), 46(5), 45(29).

25. 183(6), 182(6), 181(100) M $^{+\cdot}$, 151(11), 136(4), 135(42), 123(22), 109(5), 108(22), 91(5), 83(5), 77(5), 76(8), 64(16), 50(6).

26. 220(6), 218(8), 217(35), 216(9), 215(100) M $^{+\cdot}$, 204(5), 187(17), 185(50), 183(5), 181(7), 172(5), 171(7), 170(13), 169(20), 159(17), 142(10), 135(8), 134(25), 133(9), 125(5), 108(12), 107(6), 98(7), 89(5), 33(12), 76(10), 75(10), 70(17), 69(8), 64(9), 52(5), 51(7), 46(7), 45(7).

TABLE 2 - continued

27. 253(15), 252(11), 251(73), 250(10), 249(100) $M^{+\cdot}$,
233(5), 223(15), 222(13), 221(76), 220(20),
219(83), 206(16), 205(11), 204(22), 203(19),
195(8), 193(43), 192(5), 171(10), 170(24), 169(12),
168(43), 161(13), 159(17), 158(5), 142(18), 133(12),
117(9), 101(12), 100(12), 98(16), 83(10), 75(15),
74(14), 70(20), 59(19), 46(14), 45(15), 44(12).

28. 298(14), 297(7), 296(63), 295(9), 294(100) $M^{+\cdot}$,
248(5), 222(5), 220(8), 219(5), 206(9), 204(13),
202(9), 200(13), 186(5), 184(6), 172(8), 170(13),
169(6), 155(8), 136(7), 135(10), 134(8), 132(27),
123(9), 117(16), 109(10), 100(14), 99(16), 85(12),
83(11), 70(9), 46(11), 44(17).

29. 298(12), 297(5), 296(19), 295(7), 294(80) $M^{+\cdot}$,
265(12), 263(17), 250(9), 248(13), 240(7), 238(7),
222(41), 221(7), 220(69), 218(11), 208(12), 206(40),
205(7), 204(51), 202(11), 201(28), 200(9), 199(78),
171(25), 169(21), 167(16), 155(13), 132(31), 117(11),
109(13), 79(20), 52(13), 44(39), 30(100).

30. 232(34), 231(10), 230(100) $M^{+\cdot}$, 212(13), 202(14),
200(41), 186(27), 185(10), 184(70), 172(43), 171(9),
170(80), 152(14), 151(24), 150(40), 135(21), 124(16),
108(10), 103(10), 98(13), 94(11), 78(9), 77(19),
76(9), 70(11), 64(13), 63(11), 52(20), 51(14).

TABLE 2 - continued

31. 166(100) M $^{+\cdot}$, 165(16), 151(9), 149(9), 148(7), 139(6), 138(7), 136(27), 135(15), 134(9), 79(13), 64(9), 53(7), 52(11), 51(9), 50(6), 46(7), 45(9), 44(7).

32. 217(5), 216(5), 215(45) M $^{+\cdot}$, 199(12), 153(6), 152(20), 151(100), 137(5), 136(7), 135(72), 124(13), 123(6), 109(9), 108(27), 103(15), 91(10), 83(11), 80(7), 77(10), 76(12), 75(11), 70(10), 69(9), 64(35), 52(13), 51(12), 50(17), 46(11), 45(12).

33. 234(18), 233(11), 232(29) M $^{+\cdot}$, 231(12), 208(17), 207(8), 206(64), 205(16), 204(100), 203(10), 179(7), 177(11), 171(17), 170(11), 169(33), 168(20), 160(6), 158(7), 142(12), 110(11), 101(17), 100(12), 98(10), 75(21), 70(11).

34. 266(100), 264(100) M $^{+\cdot}$, 222(5), 220(16), 218(16), 186(10), 185(67), 183(10), 158(13), 141(10), 133(7), 132(7), 114(13), 77(5), 76(7), 69(7), 51(5), 50(5).

35. 224(5), 223(13), 222(93), 221(34), 220(100) M $^{+\cdot}$, 219(5), 193(4), 187(5), 186(9), 185(56), 184(7), 176(11), 174(20), 158(11), 153(11), 141(9), 128(5), 127(10), 114(10), 111(8), 110(20), 100(5), 75(5), 51(5), 50(5).

REFERENCES

1. J.J. van Daalen, J. Daams, H. Koopman and A. Tempel, Rec. Trav. Chim. Pays-Bas 36, 1159 (1967)

2. N.P. Buu-Hoi, P. Jacquignon and M. Mangane, Chem. Comm. 1965, 624

3. K.-P. Zeller, H. Meier and E. Muller, Tetrahedron 28, 1353 (1972)

4. idem, Org. Mass. Spectrom. 5, 373 (1971)

5. R. Lawrence and E.S. Waight, Ibid. 3, 367 (1970)

6. H. Budzikiewicz, C. Djerassi, and D.H. Williams, "Mass Spectrometry of Organic Compounds", Holden-Day, Inc., (San Francisco 1967) p. 443.

7. Ref. 6, p. 115

8. Ref. 6, p. 237

9. Ref. 6, p. 241

10. Ref. 6, p. 250

11. Ref. 6, p. 515

12. Ref. 6, p. 323

13. Ref. 6, p. 560

14. Ref. 6, p. 162

15. A. Michaelis, Ann. 274, 262 (1893).