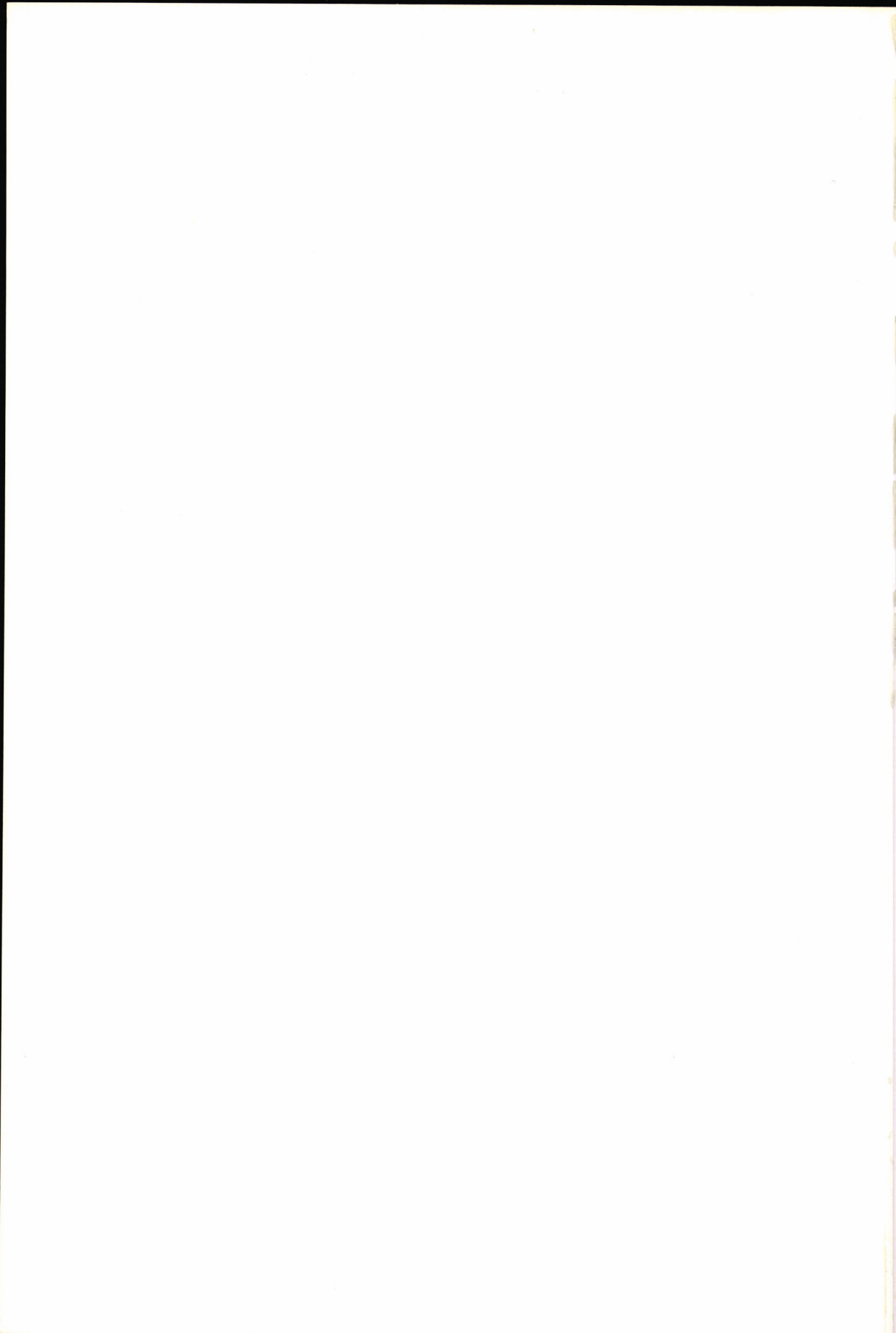# MASS SPECTROMETRIC FINGERPRINTING:
SOFT IONIZATION AND PATTERN RECOGNITION

Albert C. Tas

MASS SPECTROMETRIC FINGERPRINTING:
SOFT IONIZATION AND PATTERN RECOGNITION

Proefschrift

ter verkrijging van de graad van Doctor
aan de Rijksuniversiteit te Leiden,
op gezag van de Rector Magnificus
Dr. L. Leertouwer,
hoogleraar in de Faculteit der Godgeleerdheid,
volgens besluit van het college
van dekanen te verdedigen
op woensdag 6 maart 1991
te klokke 16.15 uur

door

Albert Cornelis Tas

geboren te Aalsmeer in 1941
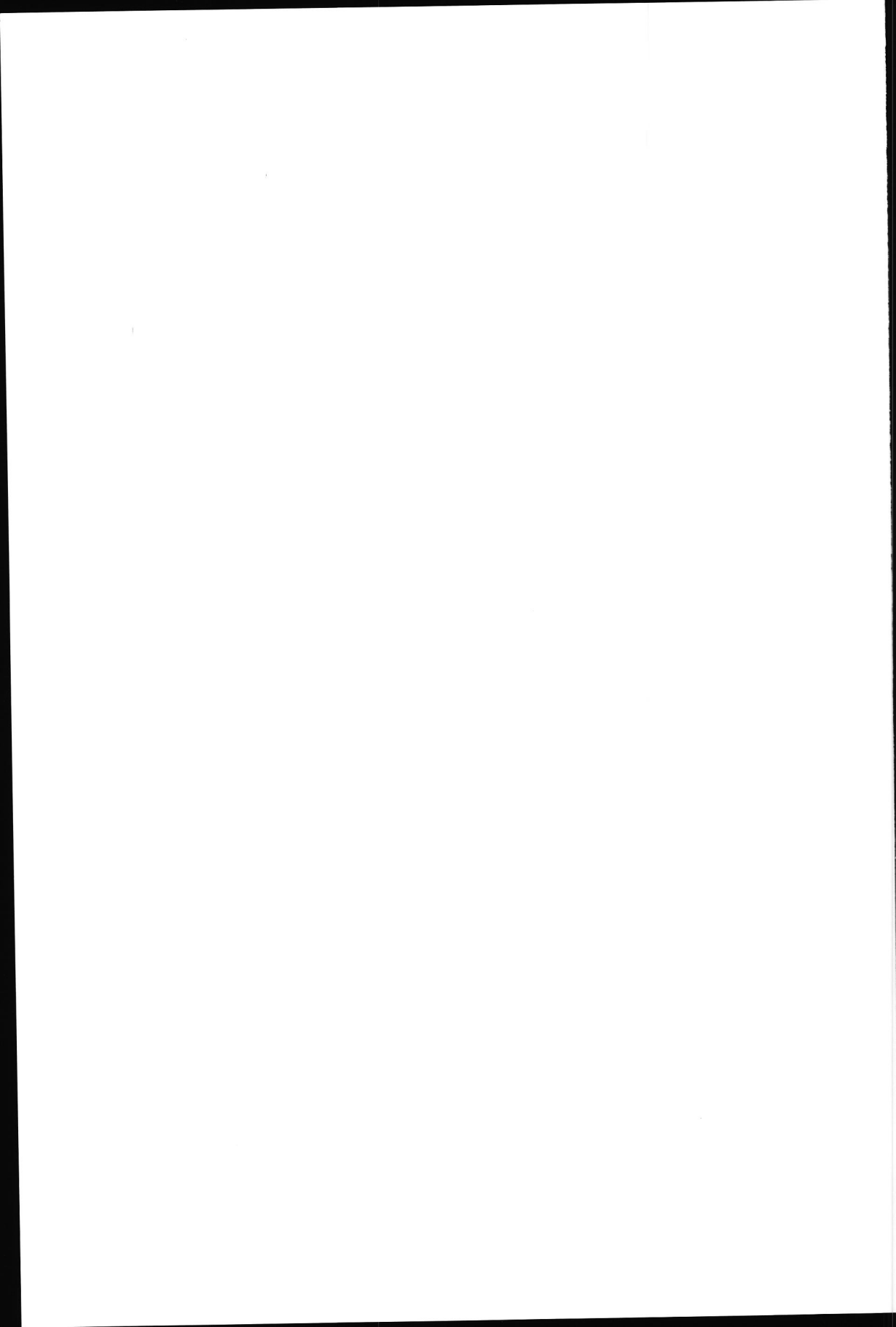
Promotiecommissie:

Promotor : Prof.dr. J. van der Greef

Referent : Prof.dr. N.M.M. Nibbering

Overige leden : Prof.dr. D.D. Breimer
Prof.dr. W. Soudijn
Prof.dr. G.J. Mulder

Aan Sytske
   Robert en Rebecca
   Niels
   Michiel

Contents

PART I
GENERAL INTRODUCTION

Chapter 1

INTRODUCTION TO PYROLYSIS AND BODY FLUID PROFILING

GENERAL INTRODUCTION

The challenges encountered by analytical chemists are increasingly focussed on the analysis of complex mixtures and trace analysis of compounds. In both cases sensitive and, even more important, selective analytical tools for the determination of individual components are mandatory. However, investigation of complex and often diffuse (bio)chemical systems requires in the first stage a broad analytical approach rather than a selective analysis of compounds, especially if no *a priori* knowledge exists of the kind of compounds involved in the properties of such systems. The analysis of mixtures of macromolecules or complex mixtures of chemical components with diverging physico-chemical properties can often hardly be performed by separating and determining individual components. Examples of such mixtures are micro-organisms, cells, body fluids, plant tissues, raw materials, food products, soil, humic substances and fossil deposits.

For analytical problems of this kind direct profiling methods, such as pyrolysis mass spectrometry (Py–MS), pyrolysis gas chromatography (Py–GC) and pyrolysis gas chromatography mass spectrometry (Py–GC/MS), offer a powerful alternative in that a great variety of compounds can be measured at the same time and usually limited sample preparation is needed. In Py–MS and Py–GC(MS) only small amounts of sample (typically 10-200 μg) are needed. Py–MS and Py–GC(MS) techniques and their applications have been reviewed extensively [1-5].

In general, mass spectrometry requires ions in vacuum. This involves formation of ions from volatile or volatilized compounds, or desorption of ions into the gas phase, for example by field desorption (FD) [6, 7] or charged aerosol droplets (thermospray, electrospray) [8]. For investigation of (bio)polymers or samples containing complex mixtures of macromolecules, such as micro-organisms, being involatile and beyond the mass range of the mass spectrometer, thermal degradation (pyrolysis) is often used for the generation of more volatile compounds of lower molecular weight which are amenable to this analytical technique. Also most chromatographic separation techniques are not capable of direct handling of macromolecular matrices and require compounds of lower molecular weight.

Py–MS [4, 9-12], Py–GC and Py–GC/(HR)MS [13, 14] are analytical techniques widely used for characterization of such matrices. Important applications are fingerprinting for characterization and differentiation of biomaterials and micro-organisms using pattern recognition methods [4, 5, 15], and identification

of individual compounds in pyrolysates for obtaining insight into the structure of the original (bio)macromolecules [3, 4, 16-18].

In the fingerprinting approach, differentiation is sometimes based on statistical evaluation rather than on chemical interpretation of differentiating components [15]. However, in most applications of this kind at least global indications are provided of biologically interesting chemical compounds on which differentiation is based. Comparison of the techniques most widely used in analyses of this type, Py–GC and Py–MS, shows that dedicated Py–MS equipment has a number of potential advantages over Py–GC [4, 9, 19], such as the high sample throughput, the handling of compounds differing to a large extent in polarity, the suitability of the data to computer handling for data analyses, an improved reproducibility and a better long-term stability. It is interesting to note that some recent reports show an apparent tendency for characterization of micro-organisms using well-defined chemical marker components only [20-22].

For identification of a large number of individual compounds in pyrolysates, for example extended series of isomeric or homologous compounds, Py–GC/(HR)MS [14, 17] and Py–GC/MS/MS [23] are frequently used. Serious limitations are the condensation of compounds of higher molecular weight and polar compounds on the inner wall of the glass liner, the elution problems for compounds of that type and the relatively long analysis time.

Another application of these techniques is the study of pyrolysis mechanisms via identification of pyrolysis products, especially in polymer analysis [3, 24-26]. For such investigations sample size dependence of pyrolysis product ratios has to be taken into account due to the involvement of secondary reactions [27].

Recent developments in pyrolysate analysis embrace the application of tandem mass spectrometry (Py–MS/MS) [18, 25, 28-31], Py–GC/MS/MS [23] and (off-line) Py–FAB/MS/MS [11]. Py–MS/MS is a powerful method for identifying molecular species directly in pyrolysates and has successfully been applied to the characterization of thermal decomposition products of polymers [25, 29], biopolymers [18, 30], bacteria [21, 28] and algae [31].

Some applications of fourier transform infrared spectroscopy (FTIR) in evolved product analysis during thermogravimetric analysis (TG–FTIR) and in Py–GC detection have been described [32, 33]. TG–FTIR has been applied to the determination of kinetic rates of compound evolution of coal and source rock samples under laboratory conditions. These rates were used to predict the conversion behaviour of fossil fuel in practical conversion processes or in geological processes, at higher and lower heating rates and temperatures, respectively [32]. Py–GC/FTIR was used in combination with Py–GC/MS in the investigation of chemical composition of particle size fractions of woody peat [33].

The application of HPLC and SFC to compounds of higher molecular weight and polar compounds from pyrolysates has been reported recently [34, 35]. In general,

the approach requires off-line pyrolysis and pyrolysate derivatization preceding HPLC [35] or SFC separation and detection, although some initial results of on-line pyrolysis under SFC conditions with subsequent SFC analysis (Py–SFC) have been reported [34]. Especially in SFC analysis of polar compounds, derivatization is required for reducing polarity, as was illustrated in the analysis of per-silylated oligomeric hexoses from maize syrup with SFC/MS [36].

PYROLYSIS–MASS SPECTROMETRY

In Py–MS four main pyrolysis approaches can be distinguished: direct probe Py–MS, laser Py–MS, inductively heated (Curie-point) filament Py–MS and resistively heated filament Py–MS.

Direct probe (DP) pyrolysis has found only limited applications, despite the fact that direct probes are standard equipment for most mass spectrometers [3, 10, 37-39]. Some disadvantages discouraged its widespread use as pyrolysis technique: (1) the occurrence of secondary reactions which are promoted during the relatively long residence time of the pyrolysate [27], which is more or less locked up in the direct probe cups; (2) a poorer reproducibility than direct chemical ionization (DCI) pyrolysis probably due to secondary reactions [39]; (3) for most of the devices the relatively low maximum pyrolysis temperature (400-500°C); and (4) the relatively low heating rates which often promote charring. Recently a DP device has been described which has been improved with respect to maximum attainable temperature and heating rate, resulting in better pyrolysis conditions [10].

For a number of techniques it is important to consider the balance existing between desorption and pyrolysis phenomena. One of these techniques, laser desorption mass spectrometry (LD–MS) [40-45] has a great potential for the analysis of macromolecules. In initial applications of LD–MS, cluster ions of polar compounds, such as sucrose, digoxin and digitonin, were observed, which were probably formed by ion-molecule reactions of desorbed compounds with metal cations [41]. For macromolecules pyrolytic fragmentation to compounds of higher molecular weight is often observed during laser desorption [40]. Recent investigations on KBr-doped polysaccharide samples, using laser desorption Fourier transform mass spectrometry (LD–FTMS), showed fragment series of oligomers (K^+ attachment ions) with ions up to m/z 1600 [40]. Fragmentation patterns originating from splitting of glycosidic linkages and of ring fragmentations were found to be distinctly different from the structurally related polysaccharides dextran, cellulose and starch, polyglucoses with different linkage types.

Three principal mechanisms of desorption can be distinguished: thermal [40-43], shock wave driven [44] and resonant [45]. Especially thermal desorption often

induces fragments from macromolecules, corresponding to those observed for other pyrolysis methods [2]. Thermal desorption is the main process occurring at lower power densities (10^6 W/cm^2) using CO_2 or Nd:YAG pulsed lasers. Doping of the samples with NaCl or KCl results in abundant $[M + Na]^+$ and $[M + K]^+$ ions, respectively. Desorption and cationization can be separate processes, the latter occurring in the gas phase [42].

Increasing the power density of laser pulses to 10^{11} W/cm^2 induces a typical 'energy sudden process' [46], a very short shock wave driven desorption process ejecting largely condensed microparticulate material, containing compounds of high molecular weight. As has been demonstrated by several flash heating methods, rapid heating often results in vaporization processes to be preferred to competitive decomposition processes [42, 44, 46, 47].

Interesting results have been obtained with resonant desorption [45]. Using a Nd:YAG laser at 266 nm at a power density of 10^8 W/cm^2 and applying an absorbing matrix, proteins with molecular weights up to 200000 Da were detected.

The short duration of ion currents necessitates trapping of ions with FTMS [40, 48, 49] or an ion trap detector (ITD), or their detection with time-of-flight mass spectrometry (TOFMS)[43-45]. Detection of (parts of) LD–MS spectra is also possible using a sector-field mass spectrometer with electro-optical array or photoplate detection [41].

In Py–MS, Curie-point pyrolyzers are frequently applied. Curie-point wires [4, 16, 19] as well as thin V-shaped foils [9] are inductively heated to specific temperatures, depending on the alloy composition (Curie-points). Short temperature rise times of typically 0.1 s are mostly attained. Pyrolyzers of this type are often connected to the ion source of low-resolution fast-scanning mass spectrometers via a heated, gold-coated, expansion chamber. This chamber acts as a buffer volume for the pyrolysate, allowing sufficient scans to be taken with a scanning mass spectrometer and diminishing contamination of the ion source. Low-voltage electron impact ionization of 14 to 20 eV are mostly used [4, 9, 16] to minimize ionization-induced fragmentation of pyrolysis products. It is a dedicated, highly automated equipment with respect to sample introduction and data handling. These aspects make this equipment very suitable for fingerprinting of large numbers of samples. Major drawbacks are the condensation of compounds of higher molecular weight and polar compounds on the walls of the expansion chamber and the relatively high degree of ionization-induced fragmentation. This results in profiles extending over relatively small mass ranges of typically m/z 20-250. An improved experimental set-up for Curie-point Py–MS has been described with respect to pyrolysate transfer and limitation of ionization-induced fragmentation by Schulten and Görtz [50], who combined near-source pyrolysis with field ionization (FI/(HR)MS). Pyrograms were obtained from polysaccharides, containing more structural information.

Pyrolysis–direct chemical ionization–mass spectrometry (Py–DCI/MS) was recently introduced as a pyrolysis technique for the characterization of complex (bio)macromolecular samples [12, 39] and for analysis of biopolymers [18, 30, 51-53]. A major characteristic of this technique is pyrolysis in the ion plasma, using a filament which is resistively heated by current programming. DCI is widely applied as a desorption technique for non-volatile and polar compounds [54]. However, due to the heating applied, desorption and pyrolysis are often competing processes. In a review on structural elucidation of complex carbohydrates, the aspects of desorption and pyrolysis in DCI are considered [55]. Spectra taken during early desorption show enhanced high-mass fragments, while in subsequent spectra acquired at higher wire temperatures an increase in lower-mass fragments is observed due to thermal degradation. Pyrolysis becomes a major process in the analysis of more polar compounds and compounds of higher molecular weight. If DCI is used as a desorption technique, pyrolysis processes are omitted as much as possible. Derivatization of compounds to lower the polarity and coating of the DCI wire [55] with a protective surface to reduce the energy needed to achieve desorption are approaches to diminish pyrolysis [56]. A third approach, rapid heating, may also be effective in changing the rate of desorption relative to the rate of decomposition, resulting in detection of compounds of higher molecular weight [47]. Workers using pyrolysis in the characterization of macromolecular systems usually try to preserve larger fragments from the original macromolecules as characteristic building blocks. Simply stated, they have to apply thermal degradation, but try to limit it as much as possible. This is clearly reflected in the emphasis on pyrolysate detection of compounds of higher molecular weight and polar compounds in pyrolysates, which is an apparent tendency in many recent investigations [10, 18, 30, 51-53].

Desorption in DCI is thought to proceed along two mechanistic routes, superheating of a condensed phase and particle bombardment of the surface [55]. Superheating on the metal surface induces the formation of bubbles of gaseous pyrolysis products, disrupting the sample surface. In addition, particle bombardment may assist in disrupting this surface. Possibly a spray of droplets or clusters will be ejected into the ion plasma of the CI source. Raising the heating rate of the wire may result in desorption of compounds of higher molecular weight [47]. This is illustrated during the pyrolysis of cellulose, where a sudden short-lasting pyrolysate evolution is observed with formation of oligomers up to anhydrocelluhexaose [52]. These phenomena show analogy to 'energy sudden processes' [46]. It can be inferred from the processes described that, if DCI is applied as a pyrolysis technique, conditions are favourable for detection of pyrolysis products of higher molecular weight.

Numerous articles have been published on the ion chemistry of chemical ionization [57-59]. Of special importance in the context of this thesis are the reviews by

Westmore and Alauddin on ammonia chemical ionization mass spectrometry [59] and by H. Budzikiewicz on mass spectrometry of negative ions [60].

OUTLINE AND SCOPE OF THE THESIS

The aim of this thesis is to describe the potential of DCI/MS as a profiling/pyrolysis technique for the characterization of complex bio(macro)molecular mixtures and for structural investigations of biopolymers. Because this potential is strongly dependent on the ability to detect characteristic components which reflect the structure of the original macromolecules or the composition of the mixture, much emphasis has been given to the interpretation of the chemical information provided by the pyrograms. In initial investigations high-resolution mass spectrometry (HR/MS) was used; in later studies both HR/MS and tandem mass spectrometry (MS/MS) were applied to the direct structural determination in the pyrolysate of a variety of components. Both techniques were found to be essential in obtaining a better understanding of pyrolysate composition and of the processes determining this composition. Markers representing a great variety of compounds were identified in the pyrograms.

Multivariate analysis (MVA) techniques are of great importance in the evaluation of pyrolysis data. These techniques were used to display profile vectors, to detect trends in data, to trace relevant compounds, to compare pyrolysis techniques and ionization conditions, and to assess longer-term reproducibility.

Aspects of potential importance in Py–DCI/MS were considered, such as: (1) pyrolysate transfer of compounds of higher molecular weight and polar compounds; (2) limitation of ionization-induced fragmentation processes; (3) application of positive and negative ion detection, enabling a more specific detection of (groups of) compounds; (4) the large storage accommodation of DCI pyrograms, which is necessary to reflect in more detail complex macromolecular systems; (5) the necessity to apply accurate mass measurements (HR/MS) and/or tandem mass spectrometry (MS/MS), because of the absence of specific fragmentation patterns, which may be indicative for the type of components involved.

## REFERENCES

1. W.J. Irwin, J. Anal. Appl. Pyrolysis, 1 (1979) 3-25 and 89-122.

2. W.J. Irwin, The Analyst, 103 (1978) 673-704.

3. H.-R. Schulten and R.P. Lattimer, Mass Spectrom. Rev., 3 (1984) 231-315.

4. H.L.C. Meuzelaar, J. Haverkamp and F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials: Compendium and Atlas, Elsevier, New York, 1982.

5. G. Wieten, H.L.C. Meuzelaar and J. Haverkamp in: Gas Chromatography Mass Spectrometry, Applications in Microbiology, G. Odham, L. Larsson and P.-A. Mardh (Eds), Plenum Press, New York, 1984.

6. N.M.M. Nibbering, J. Chromatogr., 251 (1982) 93-104.

7. J. van der Greef, Thesis, University of Amsterdam, 1980.

8. C.R. Blakley, J.J. Carmody and M.L. Vestal, Anal. Chem., 52 (1980) 1636-1641.

9. R.E. Aries and C.S. Gutteridge, Intern. Analyst, 1 (1987) 7-12.

10. H.-R. Schulten, N. Simmleit and R. Müller, Anal. Chem., 59 (1987) 2903-2908.

11. R.E. Winans, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

12. A.C. Tas, J. de Waart, J. Bouwman, M.C. ten Noever de Brauw and J. van der Greef, J. Anal. Appl. Pyrolysis, 11 (1987) 329-340.

13. A.D. Pouwels, G.B Eijkel and J.J. Boon, J. Anal. Appl. Pyrolysis, 14 (1989) 237-280.

14. M.E.L. Kohnen, B. Horsfield, J.S. Sinninghe Damsté and J.W. de Leeuw, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

15. L.A. Shute, C.S. Gutteridge, J.R. Norris and R.C.W. Berkeley, J. Gen. Microbiology, 130 (1984) 343-355.

16. J.J. Boon, W.R. de Boer, F.J. Kruyssen and J.T.M. Wouters, J. Gen. Microbiology, 122 (1981) 119-127.

17. E.W. Tegelaar, J.W. de Leeuw, C. Largeau, S. Derenne, H.-R. Schulten, R. Müller, J.J. Boon, M. Nip and J.C.M. Sprenkels, J. Anal. Appl. Pyrolysis, 15 (1989) 29-54.

18. A.C. Tas, A. Kerkenaar, G.F. La Vos and J. van der Greef, J. Anal. Appl. Pyrolysis, 15 (1989) 55-70.

19. G.S. Gutteridge and J.R. Norris, J. Appl. Bacteriology, 47 (1979) 5-43.

20. S.L. Morgan, B. Watt, K. Ueda and A. Fox, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

21. S. DeLuca, E.W. Sarver, P. de B. Harrington and K.J. Voorhees, Anal. Chem., 62 (1990) 1465-1472.

22. J.P. Dworzanski and H.L.C. Meuzelaar, Proceedings of the 38th ASMS Conf. on Mass Spectrom. and Allied Topics, Tucson, 1990.

23. T.O. Munson and D.D. Fetterolf, J. Anal. Appl. Pyrolysis, 11 (1987) 15-24.

24. A. Venema, J. Th. Jelink and J. Kamphuis, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

25. R.P. Lattimer, H. Muenster and H. Budzikiewicz, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

16

26. S. Tsuge, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

27. S.A. Groves and R.S. Lehrle, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

28. K.J. Voorhees, S.L. Durfee, J.R. Holtzclaw, C.G. Enke and M.R. Bauer, J. Anal. Appl. Pyrolysis, 14 (1988) 7-15.

29. A.P. Snyder, J. Anal. Appl. Pyrolysis, 17 (1990) 127-141.

30. R.J. Helleur and R. Guevremont, J. Anal. Appl. Pyrolysis, 15 (1989) 85-95.

31. J. de Waart, A.C. Tas, G.F. La Vos and J. van der Greef, J. Anal. Appl. Pyrolysis, in press.

32. P.R. Solomon, R.M. Carangelo and R. Bassilakis, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

33. G.D. Calvert, J.S. Esterle and J.R. Durig, J. Anal. Appl. Pyrolysis, 16 (1989) 5-25.

34. K.A. Roberts, R.R. Ristenbatt, P.A. Pizzola, D.C. Locke and P.R. De Forest, Proceedings of the 9th Intern. Conf. on Fundam. Aspects, Anal. Techn., Process. and Appl. of Pyrolysis, J. Anal. Appl. Pyrolysis, in press.

35. P.W. Arisz, J.A. Lomax and J.J. Boon, Anal. Chem., 62 (1990) 1519-1522.

36. V.N. Reinhold, D.M. Sheely, J. Kuei and G.-R. Her, Anal. Chem., 60 (1988) 2719-2722.

37. T.H. Risby and A.L. Yergey, J. Phys. Chem., 80 (1976), 2839-2845.

38. J.P. Anhalt and C. Fenselau, Anal. Chem., 47 (1975) 219-225.

39. A.C. Tas, J. van der Greef, J. de Waart, J. Bouwman and M.C. ten Noever de Brauw, J. Anal. Appl. Pyrolysis, 7 (1985) 249-255.

40. M.L. Coates and C.L. Wilkins, Anal. Chem., 59 (1987) 197-200.

41. M.A. Posthumus, P.G. Kistemaker, H.L.C. Meuzelaar and M.C. ten Noever de Brauw, Anal. Chem., 50 (1978), 985-991.

42. M.P. Chiarelli and M.L. Gross, Int. J. Mass Spectrom. Ion Process., 78 (1987) 37-41.

43. J.-C. Tabet and R.J. Cotter, Anal. Chem., 56 (1984), 1662-1667.

44. B. Lindner and U. Seydel, Anal. Chem., 57 (1985) 895-899.

45. M. Karas and F. Hillenkamp, Anal. Chem., 60 (1988) 2299-2301.

46. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong and C.M. Whitehouse, Mass Spectrom. Rev., 9 (1990) 37-70.

47. R.J. Beuhler, E. Flanigan, L.J. Greene and L. Friedman, J. Am. Chem. Soc., 96 (1974) 3990-3999.

48. D.A. Laude, C.L. Johlman, R.S. Brown, D.A. Weil and C.L. Wilkins, Mass Spectrom. Rev., 5 (1986) 107-166.

49. N.M.M. Nibbering, Nach. Chem. Tech. Lab., 32 (1984), 1044-1050.

50. H.-R. Schulten and W. Görtz, Anal. Chem., 50 (1978) 428-433.

51. J. Metzger, Fresenius Z. Anal. Chem., 308 (1981), 29-30.

52. A.D. Pouwels, G.B. Eijkel, P.W. Arisz and J.J. Boon, J. Anal. Appl. Pyrolysis, 15 (1989) 71-84.

53. M.A. Scheijen and J.J. Boon, Rapid Commun. Mass Spectrometry, 3 (1989) 238-240.

54. M.A. Baldwin and F.W. McLafferty, Org. Mass Spectrom., 7 (1973) 1353-1356.

55. V.N. Reinhold, Structural Elucidation of Complex Carbohydrates, in Mass Spectrometry in Biomedical Research, S.J. Gaskell, Ed., John Wiley & Sons, New York, 1986.

56. V.N. Reinhold and S.A. Carr, Anal. Chem., 54 (1982) 499-503.

57. B. Munson, Anal. Chem., 49 (1977) 772A-776A.

58. A.G. Harrison, Chemical Ionization Mass Spectrometry, CRC Press, Boca Raton, FL., 1983.

59. J.B. Westmore and M.M. Alauddin, Mass Spectrom. Rev., 5 (1986) 381-465.

60. H. Budzikiewicz, Angew. Chem. Int. Ed. Engl., 20 (1981) 624-637.
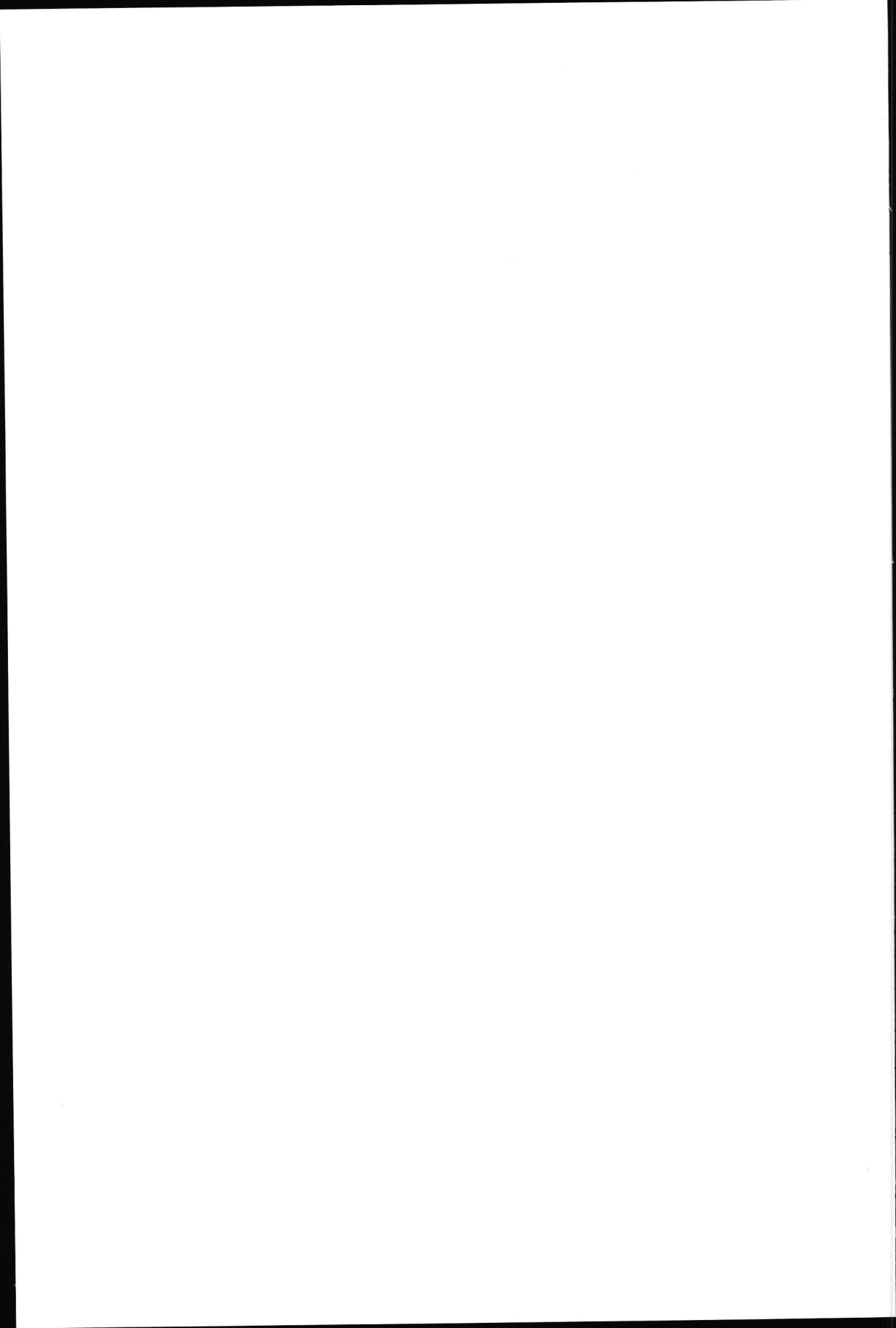
Chapter 2

INTRODUCTION TO PRINCIPAL COMPONENTS AND DISCRIMINANT ANALYSIS

1. PRINCIPAL COMPONENTS ANALYSIS

1.1 General description

In mixture spectra, constituents are often represented by multiple intensities with high correlations, especially if these signals are unique for a particular component. Furthermore, screening of complex (bio)chemical samples containing a wide range of components may also reflect the underlying correlations between such components, for instance based on biological interrelations. Still another source of interrelation exists, the so-called pseudo-correlation among variables, which may occur after spectrum normalization. Typically, this phenomenon is caused by intensities dominantly present in mixture profiles in fluctuating amounts. As a consequence, data sets may contain different types of correlation and it is very important to differentiate between them for a poper interpretation.

In general, collinearities in the data enable the representation of profiles by a limited number of composite variables, which can be linear combinations of the original variables. Principal components analysis (PCA) is a method extensively used for this purpose [1-3]. PCA often provides information on the number of factors (dimensions) influencing the profiles. In the case of mixture profiles they correspond to compounds that vary in the investigated samples, if no other sources of systematic variation are present [3]. By selecting those principal components (PCs) which represent the systematic information of the data set, a considerable data reduction can be attained, which means a compression of relevant information into a limited number of composite variables. PCs which mainly represent noise in the data can be skipped, which results in noise reduction.

Geometrically, PCs form the basis for the display of profiles as vectors in a reduced space instead of in the original q-dimensional space (q correspond to the number of measured variables). Representation of the data by PCs often provides an improved survey compared to display of the original variables. Due to their favourable property of representing the axes of maximum spread (variance) of the object vectors in the q-dimensional space, only a limited number of these variables are needed for display.

Display methods are often the starting point for exploratory data analysis: by plotting the factor scores—the coordinates of the profiles on the factors—clusters of similar profiles, systematic trends in the data and single outliers are observed; by plotting the factor loadings [4] or applying the variance maximization technique [5]

the directions of constituent axes are determined by detecting clusters of correlated variables in PC space.

PCs, also called principal factors [3], obtained by PCA can be rotated into the direction of the constituent vectors, defining the constituent space in which profiles are located according to their mixture ratio [6]. Target factor analysis is a comparable approach to determine whether a certain compound is a basic factor of the constituent space, taking into account experimental error and possible non-linearities in the data [7, 8]. In general, application of these linear models is based on the assumption that mixture profiles are linear combinations of the pure constituent spectra [7], a condition which is often not fulfilled. Reactions between chemical components, for example occurring during pyrolysis of chemical mixtures, may be the cause of non-linearities in the data resulting in an increase of the number of PCs [8].

1.2 Geometrical representation

Objects (samples, profiles, spectra, etc.) measured on q variables can be represented as n points or vectors in a q-dimensional space or in a lower-dimensional sub-space embedded in this total space. The dimensionality of the space in which these vectors are located is dependent on a number of conditions. Suppose objects consist of k components and are normalized, then the dimensionality of the mixture space equals $q - 1$, $k - 1$ or $n - 1$, whichever is the smallest [3, 7]. If the number of objects and the number of variables exceeds the number of components k, the data vectors are located in a $k - 1$ dimensional sub-space which is determined (spanned) by the k pure components. Objects are fitted in this sub-space according to their mixture ratio [6]. For example, if profiles consist of various linear combinations of 3 pure profiles, normalization of the intensities to a constant sum has the geometrical effect of fitting the data vectors into a 2-dimensional plane, irrespective of the number of variables (intensities) measured (Fig. 1).

These geometrical considerations are based on the assumption of a linear model in which profile intensities of the pure components add linearly according to Beer's law:

$$\mathbf{y} = \mathbf{S}^T\mathbf{c} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is a $q \times 1$ column vector of the mixture's profile intensities, $\mathbf{c}$ is a $k \times 1$ column vector of known concentrations of the constituents, $\mathbf{S}^T$ is a $q \times k$ matrix of k pure spectra (columns) measured on q intensities and $\mathbf{e}$ is a $q \times 1$ column vector of errors. As already mentioned, secondary reactions often occur during pyrolysis and
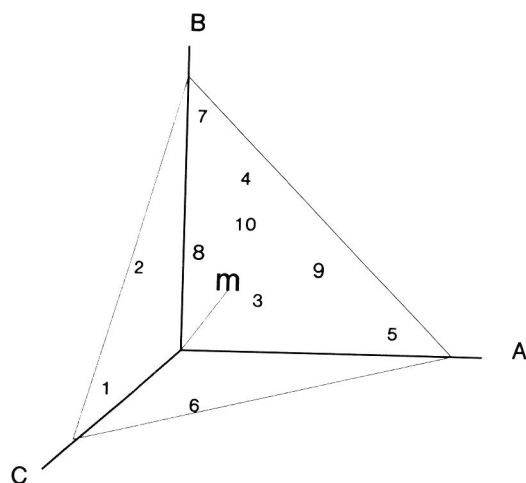
Fig. 1   Normalized object vectors of the pure components (A, B and C) and of 10 mixtures thereof. The mixtures are fitted in the plane which is spanned by the pure components
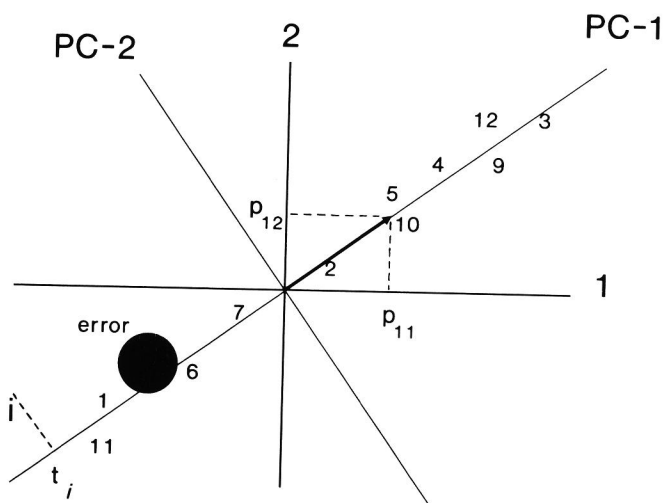

Fig. 2   Eigenvector of length 1 (unit vector) in the direction of maximum variance (spread) of n profile vectors. The direction of the vector in the original space is defined by $\cos a_1 = p_{11}$ and $\cos a_2 = p_{12}$. The projection of spectrum vector i is t_i, the ith element of scores vector $\mathbf{t}$, which is expressed in eigenvector units

cause non-linearities in the data, resulting in an increased number of chemical components and an increased dimensionality of the factor space [7, 8].

In addition, error, for example measurement error, causes the sub-space to be less well defined and may complicate the determination of the number of components present in mixtures [3, 9].

The geometrical explanation of PCA can be based on the representation of n profile vectors in an orthogonal q-dimensional space (exemplified for 2 dimensions in Fig. 2). In this q-dimensional space new orthogonal variables are defined, the so-called PCs t_h, which represent the variance or spread of the n vectors more efficiently than the original variables. They consist of a linear combination of the original q variables: PC-1 axis in the direction of the highest variance; PC-2 in the next highest variance direction, orthogonal to the first PC, etc. For the h-th PC t_h the linear combination of the original $\mathbf{x}$ vectors is:

$$PC_h = \mathbf{t}_h = p_{h1}\mathbf{x}_1 + p_{h2}\mathbf{x}_2 + \dots + p_{hq}\mathbf{x}_q \tag{2}$$

The vector of the loadings $p_{h1}, p_{h2} \dots p_{hq-1}, p_{hq}$ of PC_h on q variables is the eigenvector ($\mathbf{p}_h$). The eigenvectors define a new orthogonal set of basis axes of unit length, the so-called orthonormal basis. The loadings, the angle cosines between the PCs and the original q variables, are equal to the projection of the unit vector on the original axis (Fig. 2, p_{11} and p_{12}). They represent the directions of the PCs in the original space. Projection of the n profile vectors on the h-th eigenvector provides the n elements scores vector $\mathbf{t}_h$, their coordinates being expressed in eigenvector units (unit length). The variance of the subsequent PCs is equal to the eigenvalues. It has to be noted that the PCs $\mathbf{t}$ are latent variables which are not directly observed by measurements. It is important to realize that the Euclidean distances between object vectors are invariant upon orthogonal rotation of the axes. Therefore, if all eigenvectors with eigenvalues larger than zero are retained, distances between object vectors remain unchanged.

If substantial collinearity exists in a data matrix, the object vectors are located near or in a subspace of k ($k < q$) dimensions, which is spanned by the first k eigenvectors. In that case only a relatively small proportion of the variance, often mainly consisting of error, is represented by the last $q - k$ eigenvectors, which can be deleted accordingly. Therefore PCA may be useful in determining the number of intrinsic factors which influence the profiles, resulting in data compression and noise reduction [3]. In the example represented in Fig. 2 the variables x_1 and x_2 are correlated. The object vectors 1-12 are found near a one-dimensional subspace (a line) closely to PC-1 which clearly represents the systematic information in the data. If it is assumed that these data contain random error, PC-2 may well represent mainly error. However, error will be only partially removed by skipping PC-2 and part of it will be included in PC-1. As can be inferred from Fig. 2, the variance

component parallel to PC-1 will be embedded (embedded error); the variance component orthogonal to PC-1 will be removed (extracted error). The selection of the number of significant eigenvectors (real eigenvectors) is often not straightforward (cf. Chapter 11, section 1.2).

If PCA is applied to normalized objects before mean substraction (Fig. 1) the first PC is mainly determined by the total intensity of the objects, while subsequent PCs provide relevant information on the mixture plane [5]. As can be judged from this example, after mean substraction only two PCs are needed to describe the plane, irrespective of the number of variables (Fig. 3). However, the PCs are not the only possible factors to span the mixture plane. It can also be spanned by an infinitive number of orthogonal factors and, as stated before, by the pure component factors, the so-called basic factors. These are found in directions in the factor space where the pure component vectors are located, the directions $M - A$, $M - B$ and $M - C$ (Fig. 3). In general, such component vectors are not orthogonal [6]. Detection of component vectors may be one of the aims of factor analysis [6, 10, 11]. It is important to note that typical vectors which span the space do not necessarily represent pure components. Samples of complex composition, such as biological and geological samples, may be linear combinations of a number of samples of a specific composition [12]. For example in Fig. 1, samples can be considered as mixtures of the samples 1, 5 and 7. In Q-mode analysis these typical samples are indicated as end-members [12].

2. DISCRIMINANT ANALYSIS

2.1 General description

For discrimination between profiles of samples (objects) arranged into *a priori* defined groups, which often differ only slightly in mixture composition, the application of sophisticated multivariate analytical techniques is mandatory. Especially in the presence of several sources of variation, such as measurement error, instrumental drift, natural within-group and between-group variation, techniques have to be used that maximize the between-group variance relative to the within-group variance ($W^{-1} B$). Important objectives of discriminant analysis are: (1) differentiation of groups of samples, (2) classification of unknown samples, (3) determination of (groups of) variables that are important in differentating groups, and (4) display of objects and variables.

The discussion in this chapter will be focused on linear discriminant analysis preceded by PCA, a technique that seeks linear combinations of the factor score vectors instead of the original features for optimization of $W^{-1} B$. Application of usual discriminant analysis techniques to mixture profiles may fail because the
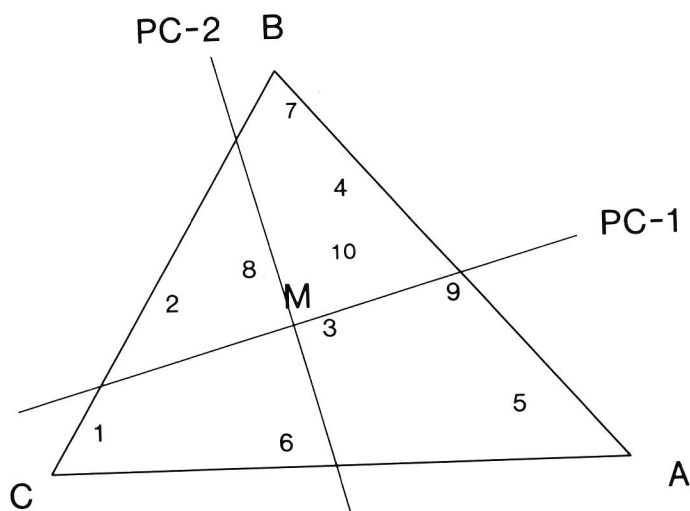
Fig. 3   PCA of pure profile vectors (A, B, C) and profile vectors of mixtures, after normalization and mean profile substraction (two PCs)
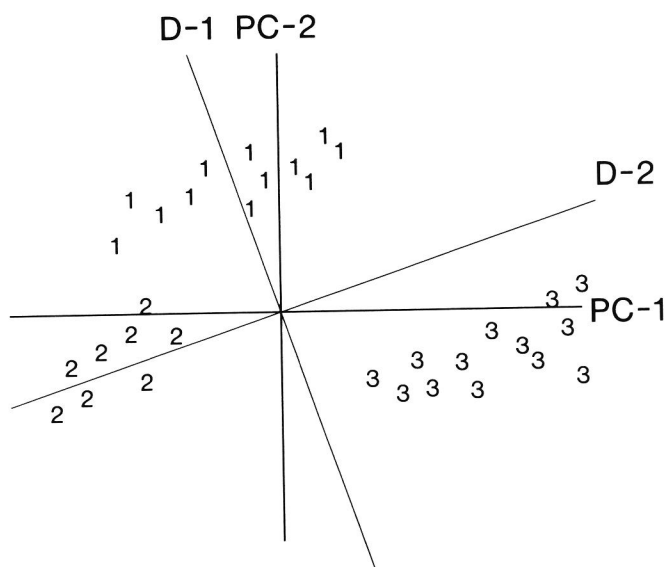

Fig. 4   Rotation of PCs to a direction of maximum ratio of between-group to within-group variance (D functions) for display of the optimum separation between the groups (three categories)

number of variables often approaches or exceeds the number of samples, thus causing unstable discriminant functions and/or matrix inversion problems [7]. Therefore, in Py–MS applications discriminant analysis is often preceded by PCA to condense the relevant information into a limited number of real PCs. This also means that part of the error, the so-called extracted error (XE), is removed from the data, which adds to the stability of the obtained discriminant functions. The latter approach is known as principal components-discriminant analysis (PC-DA) [13, 14] or principal components-canonical variate analysis (PC-CVA) [7, 15]].

Profiling by chromatography or spectrometry, for example pyrolysis-mass spectrometry (Py–MS), often requires replicate measurements per sample to keep control of the different sources of variation, such as instrumental error and drift or sample inhomogeneities. Discriminant analysis (DA), applied to data sets after defining these replicates as groups, can be considered as an unsupervised approach. This must be distinguished from repeated sampling to monitor, for example, biological variation or process variation which is required in a proper experimental set-up.

Replicates of similar samples tend to cluster in discriminant space, provided that also dissimilar samples are part of the data set. In general, it can be noted that for dissimilar samples with large between-group to within-group variance ratios display of the object scores on the main D functions is comparable to factor score plots of replicate means. Comparison of target factor analysis and discriminant analysis of replicates revealed an improvement of results of estimation of mixture composition using the latter approach, as was shown in a Py–MS study of artificial mixtures of the polyhexoses glycogen and dextran and the protein bovine serum albumin [7, 8]. PC-DA was found to be far superior to PCA, because of the optimization of the between-group to within-group variance ratio. Consequently, also the reliability of correlations between the D functions and the original variables will be improved by applying replicate measurements. This is important for the interpretation of variables with major contributions to relevant D functions.

In a recent paper a comparable method has been introduced, called discriminant principal components analysis (DPCA) [16]. The approach consists of a rescaling of all variables in the raw data set with the inverse of its pooled within-group standard deviations, followed by performing PCA on the between-group sum-of-squares-and-cross-product groups (SSCP) matrix.

As already stated for PCA, application of PC-DA is based on the assumption of linear response models.

2.2 *Geometrical representation*

Suppose n normalized objects, measured on q variables, are known to belong to k independent groups ($k < q$). If the variation within these groups is relatively small the profile vectors are located approximately in a $k - 1$ dimensional subspace, as will be indicated by PCA. With comparatively large within-group variance the subspace is of a higher dimensionality, and consequently more eigenvectors are needed to span this subspace. Part of the error, XE, present in the data set can be removed by PCA. Subsequent DA, using the factor scores matrix $\mathbf{T}$, will maximize the ratio of the variances $\mathbf{W}^{-1} \mathbf{B}$. This optimization is carried out by rotation of the factor scores vectors to new directions, the so-called discriminant scores vectors $\mathbf{d}$, contained in matrix $\mathbf{D}$ (Fig. 4). It is important to note that the Euclidian distances between objects in factor space do not change upon rotation of factor scores vectors to discriminant scores vectors, provided the number of $\mathbf{T}$ and $\mathbf{D}$ vectors is equal.

REFERENCES

1. R.J. Rummel, Applied Factor Analysis, Northwestern University Press, Evanston, Ill., 1970.

2. W.R. Dillon and M. Goldstein, Multivariate Analysis, Methods and Applications, John Wiley & Sons, New York, 1984.

3. E.R. Malinowski and D.G. Howery, Factor Analysis in Chemistry, John Wiley & Sons, New York, 1980.

4. K.R. Gabriel, Biometrika, 58 (1971) 453-467.

5. W. Windig, Chemomet. and Intell. Lab. Systems, 4 (1988) 201-213.

6. W. Windig, W.H. McCLennen and H.L.C. Meuzelaar, Chemomet. and Intell. Lab. Systems, 1 (1987) 151-165.

7. L.V. Vallis, H.J.H. MacFie and C.S. Gutteridge, Anal. Chem., 57 (1985) 704-709.

8. D. van de Meent, J.W. de Leeuw, P.A. Schenk, W. Windig and J. Haverkamp, J. Anal. Appl. Pyrolysis, 4 (1982) 133-142.

9. E.R. Malinowski, J. of Chemometrics, 1 (1987) 33-40.

10. B.M.G. Vandeginste, W. Derks and G. Kateman, Anal. Chim. Acta, 173 (1985) 253-264.

11. P.J. Gemperline, J. Chem. Inf. and Comp. Science, 24 (1984) 206-212.

12. K.G. Jöreskog, J.E. Klovan and R.A. Reyment, Geological Factor Analysis, Elsevier Scientific Publishing Company, Amsterdam, 1976.

13. R. Hoogerbrugge, S.J. Willig and P.G. Kistemaker, Anal. Chem., 55 (1983) 1711-1712.

14. W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81-88.

15. P.A. Lachenbruch, Discriminant Analysis, Hafner Press, New York, 1975.

16. P.W. Yendle and H.J.H. MacFie, J. of Chemometrics, 3 (1989) 589-600.
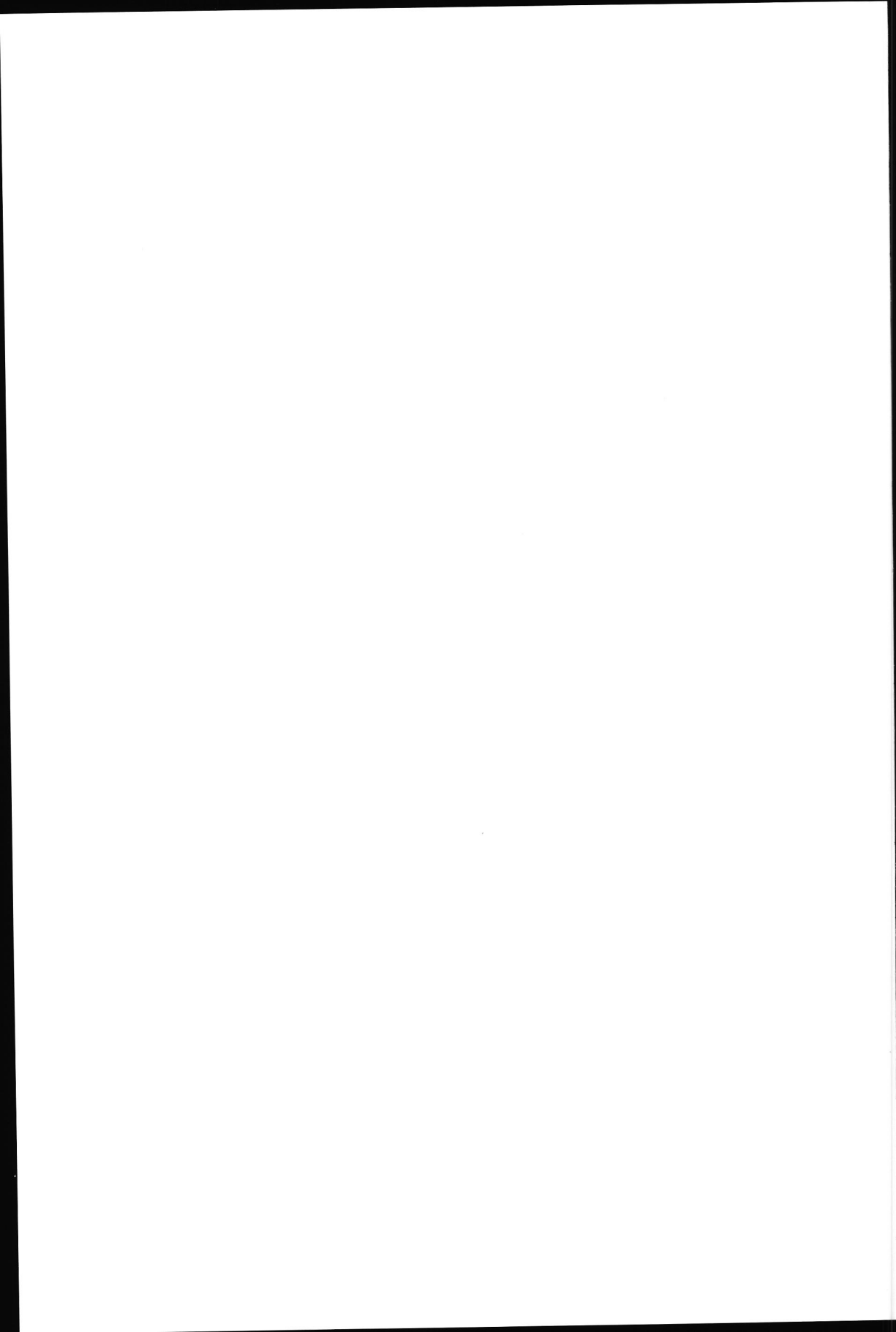
PART II
EXPERIMENTS
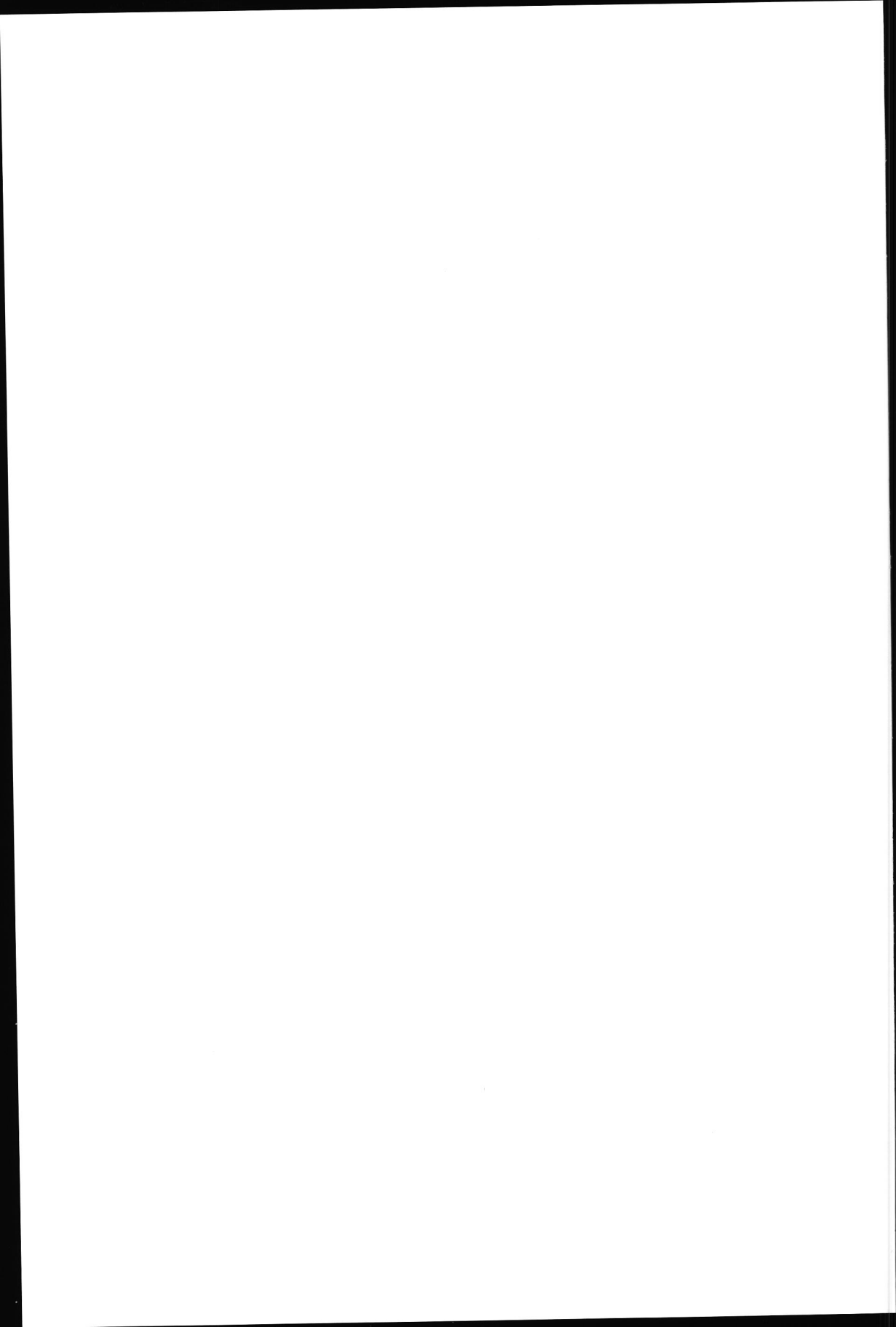
Chapter 3

COMPARISON OF DIRECT CHEMICAL IONIZATION AND DIRECT-PROBE ELECTRON IMPACT / CHEMICAL IONIZATION PYROLYSIS FOR CHARACTERIZATION OF *PSEUDOMONAS* AND *SERRATIA* BACTERIA

A.C. TAS, J. VAN DER GREEF *, J. DE WAART, J. BOUWMAN and M.C. TEN NOEVER DE BRAUW

CIVO-TNO Institutes, P.O. Box 360, 3700 AJ Zeist (The Netherlands)

SUMMARY

Pyrograms of *Serratia* and *Pseudomonas* have been measured with direct chemical ionization (DCI) and direct-probe electron impact and chemical ionization. Using pattern recognition, it was found that DCI yields pyrograms with the highest specificity. Furthermore, DCI pyrolysis was found to be more sensitive than the other two methods and also allows faster analysis.

INTRODUCTION

In recent years, interest in the analysis of non-volatile compounds has greatly increased. The development of "soft" ionization methods has broadened the molecular weight range of compounds that can be studied to approximately 15,000. Macromolecules of biological materials, however, are still not amenable to mass spectrometric analysis without pre-treatment. Chemical transformation to smaller and more volatile products by pyrolysis is a universal approach in this respect.

Various methods have been developed for the pyrolysis of microorganisms in combination with mass spectrometry, gas chromatography or a combination of both techniques [1,2]. Pyrolysis by linear programming of a direct insertion probe is an attractive method [3–5] because it is available on almost every commercial instrument and in addition allows the recording of time profiles of the pyrolysate. Direct pyrolysis in the ion source has the advantage that compounds with low volatility can also be recorded, but a disadvantage is the stronger source contamination. Various parameters have to be optimized in order to obtain a high specificity, including culturing of

the microorganisms, sampling methods from culture plates, temperature profile during pyrolysis, pyrolysate transfer to the ionization region and sample size [2,6].

The specificity of the pyrogram can be increased by reducing the fragmentation of the ions generated from the pyrolysate. In this respect, low-energy (12 eV) electron impact has turned out to be very useful [2]. Other methods that have been applied are chemical ionization [7–9], field ionization–field desorption [10–12] and photoionization [13]. Both chemical ionization and photoionization enable a high selectivity to be achieved by using different reagent gases and photon energies, respectively.

Data processing by multivariate analysis is essential for the optimal extraction of the important features from the pyrogram [2,14–16]. Multivariate analysis has a clear potential not only for tracing the dissimilarities but also as a means of chemical interpretation of the observed differences [17].

The aim of this study was to evaluate the suitability of direct chemical ionization (DCI) [18] as a rapid method for pyrolysing bacteria and to make a comparison with direct-probe electron impact/chemical ionization (EI/CI) pyrolysis. For this purpose two different bacterial genera were selected: *Pseudomonas* and *Serratia*. The genus *Pseudomonas* has been subjected to extensive studies [3,11,19–23]. The origin of many specific features of different strains has been determined, using electron-impact, field ionization in the high-resolution mode and gas chromatography–mass spectrometry (GC–MS). *Serratia* has been studied only occasionally.

EXPERIMENTAL

A Finnigan-MAT 8230 mass spectrometer equipped with a combined EI/CI source was used. EI (70 eV was preferred instead of 20 eV because of the higher sensitivity and the better long-term reproducibility expected), ammonia chemical ionization and DCI were performed with a source temperature of 240°C. Pyrolysis was carried out by direct-probe analysis in combination with EI and CI by increasing the crucible temperature linearly at 60°C/min from 20 to 400°C and by direct chemical ionization, using a linear current programme (8 mA/s). The scan speeds for the direct probe and the DCI method were 2.0 and 0.7 s/decade, respectively. The spectra obtained in the temperature range of pyrolysate formation were averaged.

After normalization with respect to the total ion current, data processing was performed with the ARTHUR software package (University of Seattle, Laboratory for Chemometrics) after selecting the 70 most relevant masses by calculating Fisher weights between the entire groups and not between sample replicates. Principal component analysis after autoscaling followed by discriminant analysis was performed with a program developed by the FOM-Institute, Amsterdam [15].

Pseudomonas aeruginosa and *Serratia marcescens* strains, originating from the CIVO-TNO bacteria collection, were cultured on a plate count agar (PCA) growth medium. Living bacteria were removed from the culture medium by careful scraping with the DCI wire or a small loop and used for pyrolysis–MS analysis without any pre-treatment.
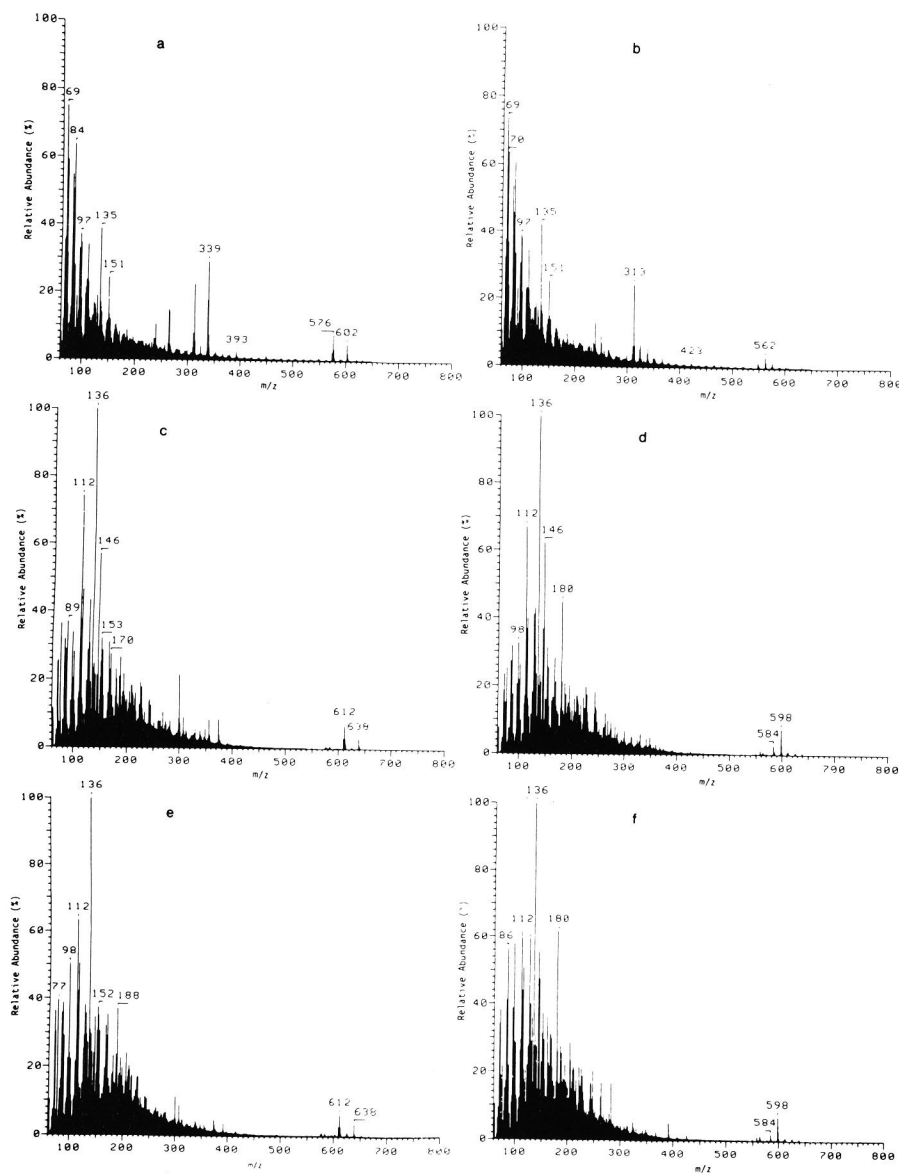

Fig. 1. Direct-probe electron impact/chemical ionization and direct chemical ionization pyrograms of *Pseudomonas* (a, c and e) and *Serratia* (b, d and f).
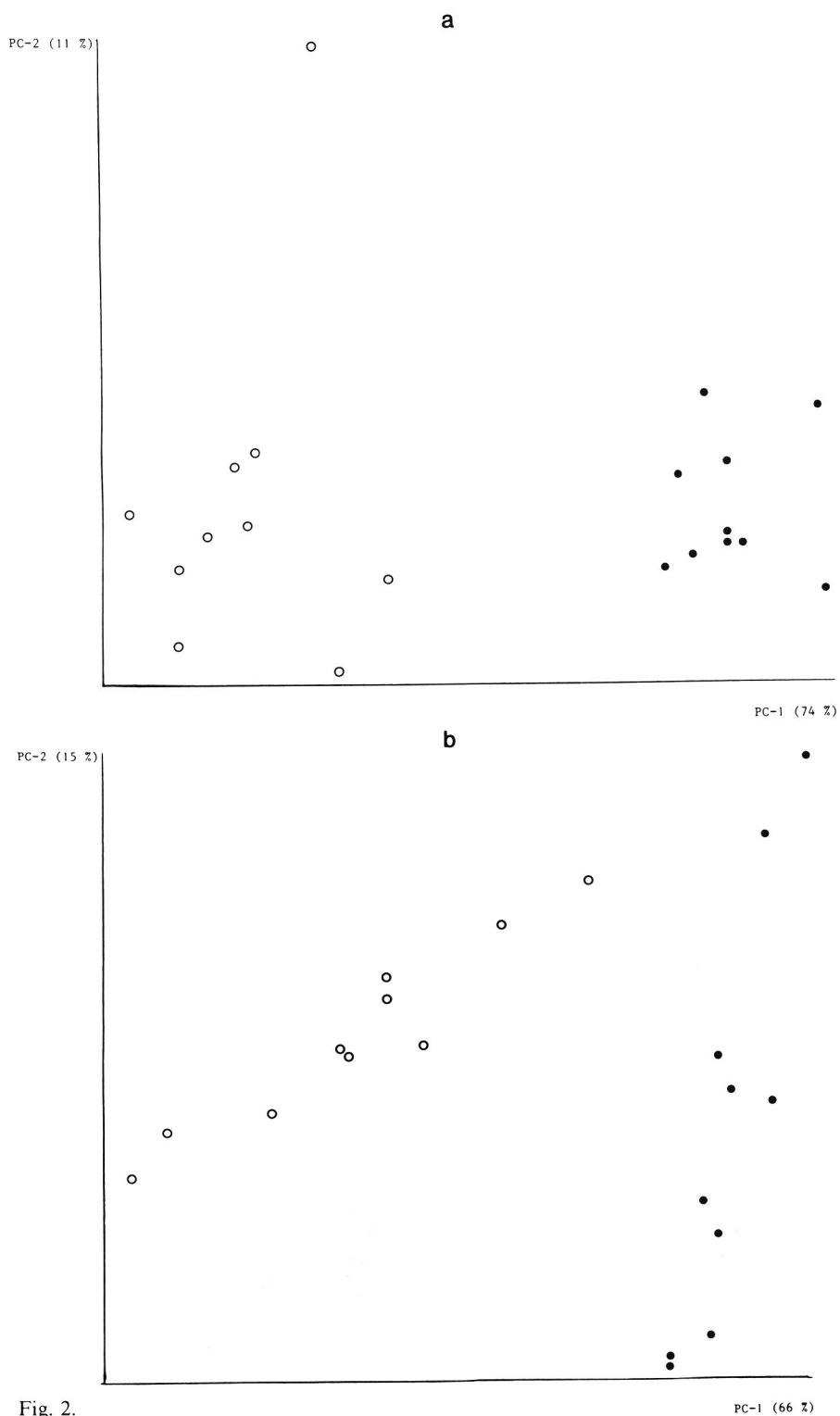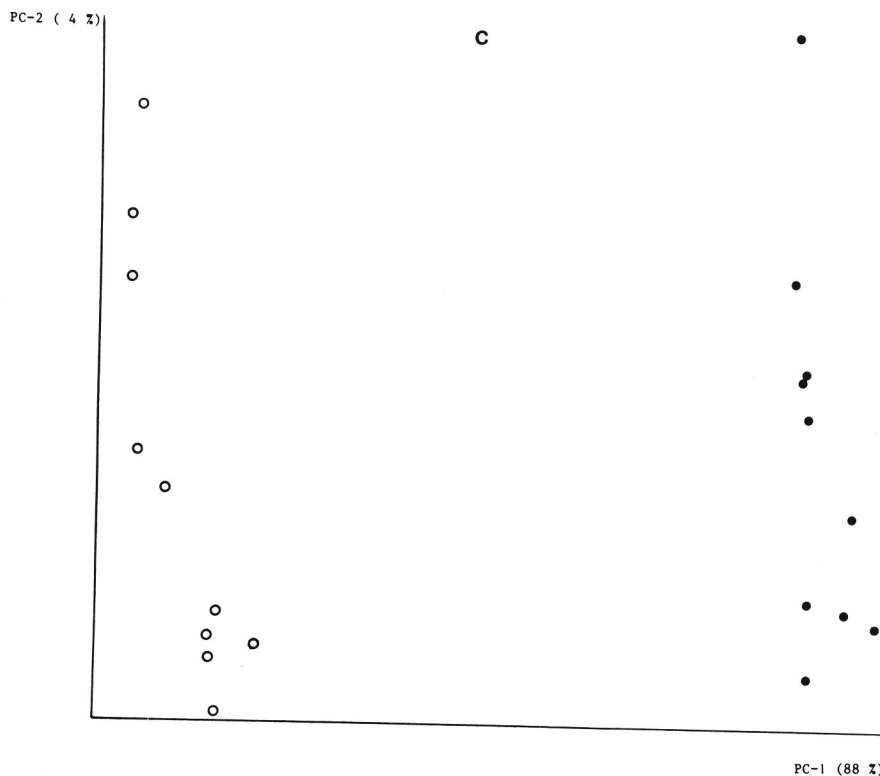
Fig. 2.

Fig. 2. Principal component analysis of pyrograms of *Pseudomonas* (open dots) and *Serratia* (black dots) obtained by direct probe electron impact (a), chemical ionization (b) and direct chemical ionization (c).

RESULTS

Averaged pyrograms obtained by direct-probe analysis in combination with EI and CI and also DCI pyrograms are given in Fig. 1. Visual inspection of these spectra reveals that with CI and DCI the amount of fragmentation is reduced in comparison with EI (70 eV). Further, clear differences are observed between *Pseudomonas* and *Serratia* pyrograms with all ionization methods, especially in the higher mass region. In the DCI mode, using medium resolution ($R = 7000$) and high-voltage scanning, it was found that the masses at m/z 598 and 612 were $[M + NH_4]^+$ ions of compounds with the elemental composition $C_{36}H_{68}O_5$ and $C_{37}H_{70}O_5$. These compounds are probably pyrolysis products of lipids [3].

In order to gain an insight into the specificity of the different methods, principal component analysis was carried out on a data set which was reduced to 70 masses, selected on basis of their Fisher weights. The results of the principal component analysis are shown in Fig. 2, where the objects are plotted in the plane formed by the first and second principal components. A

measure of the inner and outer group variances in each instance was obtained by discriminant analysis preceded by principal component analysis [15]. In all instances the first discriminant function showed by far the highest discriminating power. For the individual methods the following Fisher weights for the first discriminant function were calculated: EI 26, CI 13 and DCI 310. This is in line with the observation that the variances of the data set explained by the first principal component (see Fig. 2) were EI 74%, CI 66% and DCI 88%. The high discrimination power of DCI pyrolysis originates mainly from a higher overall reproducibility compared with CI. Direct-probe EI (70 eV) has a higher discriminating power than CI, despite the greater extent of fragmentation as a result of a higher reproducibility.

It is known that fast heating rates and rapid pyrolysate transfer from the pyrolysis regime to the ionization region enhance the reproducibility. Both effects might explain the higher reproducibility observed for DCI pyrolysis. Other advantages of DCI over the other two methods are the higher sensitivity and the higher speed of analysis.

CONCLUSIONS

DCI pyrolysis is more sensitive and faster than analysis with direct-probe EI and CI. Moreover, DCI pyrograms exhibit a higher power to discriminate between the genera *Pseudomonas* and *Serratia*. Possible reasons are a steeper temperature/time specificity and a better pyrolysate transfer for DCI. Long-term reproducibility has still to be evaluated for DCI pyrolysis.

REFERENCES

1 C.S. Gutteridge and J.R. Norris, J. Appl. Bacteriol., 47 (1979) 5.
2 G. Wieten, H.L.C. Meuzelaar and J. Haverkamp, in G. Odham, L. Larsson and P.A. Mårdh (Editors), Gas Chromatography–Mass Spectrometry. Applications in Microbiology, Plenum Press, New York, 1984, Ch. 10.
3 J.P. Anhalt and C. Fenselau, Anal. Chem., 47 (1975) 219.
4 T.H. Risby and A.L. Yergey, J. Phys. Chem., 80 (1976) 2839.
5 C.S. Gutteridge and D.J. Puckey, J. Gen. Microbiol., 128 (1982) 721.
6 W. Windig, P.G. Kistemaker, J. Haverkamp and H.L.C. Meuzelaar, J. Anal. Appl. Pyrol., 1 (1979) 39.
7 K. Jankowski, J.R.J. Pare, D. Goudin and H. Virelizier, Anal. Chem. Symp. Ser., 14 (1983) 443.

8 A. Zsolnay and B. Little, J. Anal. Appl. Pyrol., 4 (1983) 335.
9 A. Zsolnay, J. Anal. Appl. Pyrol., 4 (1982) 47.
10 H.-R. Schulten, in C.E.R. Jones and C.A. Cramers (Editors), Analytical Pyrolysis, Elsevier, Amsterdam, 1977, pp. 17–28.
11 H.-R. Schulten, H.D. Beckey, H.L.C. Meuzelaar and A.J.H. Boerboom, Anal. Chem., 45 (1973) 191.
12 H.-R. Schulten, J. Anal. Appl. Pyrol., 6 (1984) 251.
13 W. Genuit and J.J. Boon, J. Anal. Appl. Pyrol., 8 (1985) 25.
14 W. Eshuis, P.G. Kistemaker and H.L.C. Meuzelaar, in C.E.R. Jones and C.A. Cramers (Editors), Analytical Pyrolysis, Elsevier, Amsterdam, 1977, pp. 151–166.
15 R. Hoogerbrugge, S.J. Willig and P.G. Kistemaker, Anal. Chem., 55 (1983) 1710.
16 W. Windig and P.G. Kistemaker, J. Anal. Appl. Pyrol., 3 (1981) 213.
17 W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 88.
18 R.J. Cotter, Anal. Chem., 52 (1980) 1589A.
19 H.J.H. MacFie, C.S. Gutteridge and J.R. Norris, J. Gen. Microbiol., 104 (1978) 67.
20 C.S. Gutteridge and J.R. Norris, Appl. Environ. Microbiol., 40 (1980) 462.
21 E. Reiner, in C.E.R. Jones and C.A. Cramers (Editors), Analytical Pyrolysis, Elsevier, Amsterdam, 1977, pp. 49–56.
22 G.L. French, C.S. Gutteridge and I. Phillips, J. Appl. Bacteriol., 49 (1980) 505.
23 P.G. Simmonds, Appl. Microbiol., 20 (1970) 567.
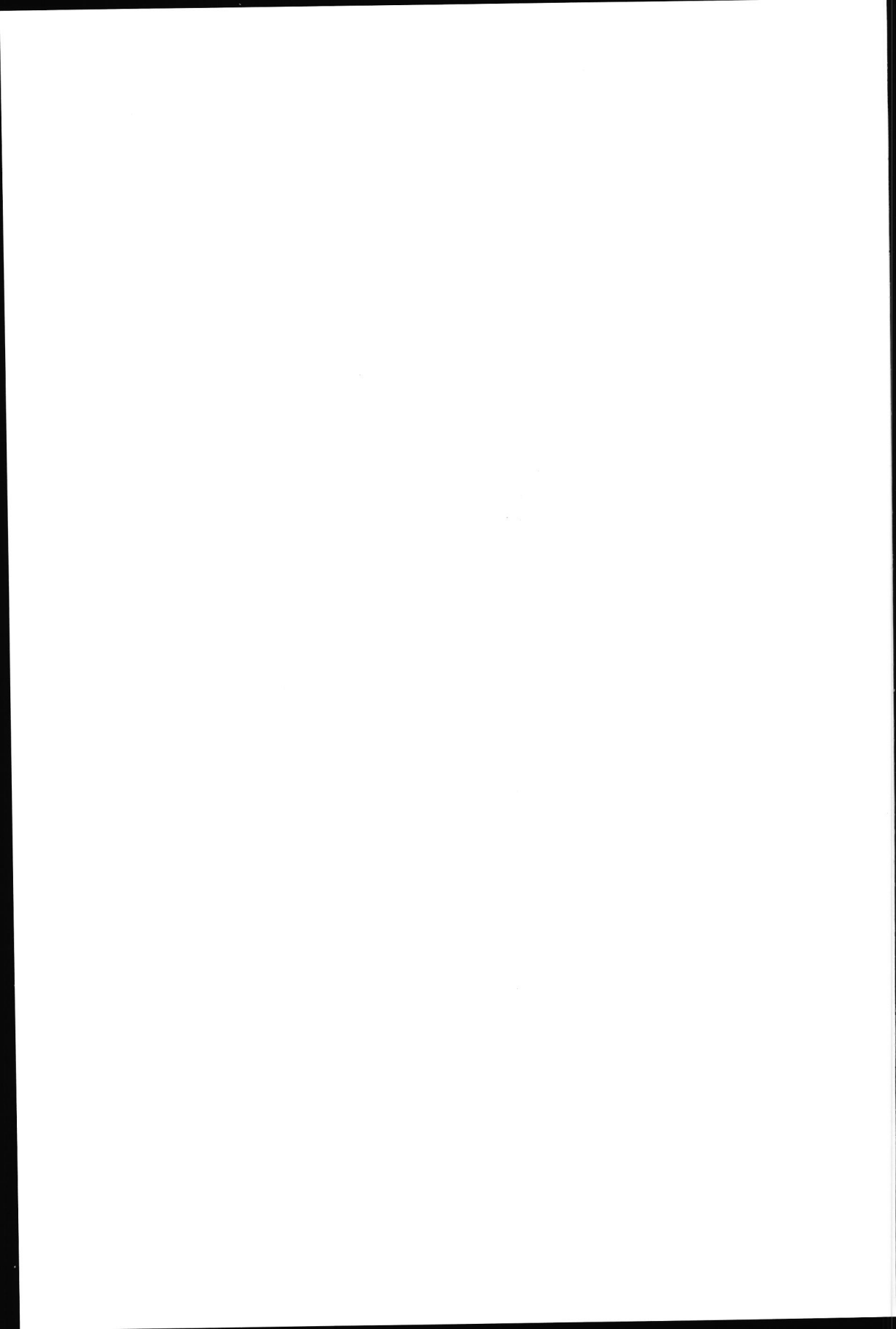
Chapter 4

RAPID CHARACTERIZATION OF *SALMONELLA* STRAINS WITH DIRECT CHEMICAL IONIZATION PYROLYSIS

A.C. TAS *, J. DE WAART, J. BOUWMAN, M.C. TEN NOEVER DE BRAUW and J. VAN DER GREEF

TNO-CIVO Food Analysis Institute, P.O. Box 360, 3700 AJ Zeist (The Netherlands)

ABSTRACT

Direct chemical ionization pyrolysis–mass spectrometry was applied in differentiating a number of epidemiologically important *Salmonella* strains from related strains which often interfere during microbiological identification procedures.

Diglycerides (degradation products of phospholipids) were found to be especially important in the characterization of the strains. Longer term reproducibility was hampered by systematic variations in intensity of the higher mass range part of the pyrograms.

Bacteria; direct chemical ionization; discriminant analysis; mass spectrometry; principal components analysis; pyrolysis; *Salmonella*.

INTRODUCTION

In conjunction with mass spectrometry (MS) various pyrolysis (Py) methods have been developed for the analysis of highly complex biopolymeric materials. These Py–MS methods include direct probe pyrolysis [1,2], laser pyrolysis [3] and Curie-point pyrolysis [4], which are the most widely used. Much attention has been paid to aspects like pyrolysate transfer and to ionization methods in order to maintain the characteristicity of the pyrolysate.

Pyrolysate transfer to the ionization region faces some problems when pyrolysis is carried out outside the ion source, especially for the high molecular weight compounds (> 500) and the more polar fractions of the pyrolysate. This transfer could be improved by heating the expansion chamber or by shortening the transfer line by omitting such a chamber [4].

In general, ionization methods which induce a reduced amount of fragmentation are preferred. This experience has led to the use of methods such as low energy (12 eV) electron impact ionization [4], photo ionization [5], field ionization [6] and chemical ionization [7].

In a recent paper we reported on the application of direct chemical ionization pyrolysis–mass spectrometry to the microorganisms *Pseudomonas aeruginosa* and *Serratia marcescens* [8]. The technique combines two important features: the sample is directly introduced into the ion source, avoiding pyrolysate transfer problems, and the relatively soft ionization process enables the detection of larger molecules. Moreover time-dependent pyrolysate formation [4] can also be studied with this technique. Differentiation of microorganisms has been carried out by various pyrolysis techniques in combination with gas chromatography (GC) [9], MS [10,11] and GC–MS [11]. One of the applications of Py–GC in microbiology was the characterization of *Salmonella* organisms on serotype level by pyrolysis of complete cells [12] and cell fragments [13]. Direct probe Py–MS applied, among others, to a *Salmonella* strain and some other bacteria of the Enterobacteriaceae family, revealed the importance of the detection of larger molecules in differentiating [10].

In this paper the potential of direct chemical ionization (DCI) MS pyrolysis for rapid characterization of a number of epidemiologically important *Salmonella* strains and some related strains is reported. The significance of masses in the higher mass range (m/z 550–650) is described and suggestions for their structure are given. Principal component–discriminant analysis has been used for extracting relevant features from the pyrograms [14].

EXPERIMENTAL

Pyrolysis–mass spectrometry

DCI pyrolysis was performed on a Finnigan-MAT 8230 mass spectrometer coupled to a Finnigan-MAT SS300 data system. Ammonia was used as a reactant gas at an indicated source pressure of 0.4 Pa; the source temperature was maintained at 240°C. During linear current programming (8 mA/s) spectra were acquired over the mass range m/z 60–1000, at a scan speed of 0.7 s/decade. Spectra obtained during evaporation and pyrolysate formation were averaged. All bacterial strains were analysed in triplicate. DCI Py–high-resolution MS for exact mass measurement was conducted at a resolution of 10,000 (10% valley definition). Unless otherwise stated, identifications and structure proposals are supported by exact mass measurements.

In order to study the longer term reproducibility, three identical experiments were carried out under these conditions with intervals of 2 and 4 months respectively (longer term reproducibility experiments).

Data analysis

In view of the available computer capacity, each spectrum was reduced to a subset of 70 mass intensities. This selection was based on Fisher weights. It was applied to full-scale spectra and to the m/z 550–650 mass range. Spectra were normalized to 100% total ion current in order to correct for differences in sample size. With full-scale spectra ten ions with the highest intensity were excluded from normalization. The data set was transformed by autoscaling in order to perform principal component–discriminant analysis on a "correlation about the mean" matrix.

Data processing was performed using the pattern recognition package ARTHUR [15], extended with the principal component–discriminant analysis routines developed by the FOM-Institute (Amsterdam, The Netherlands) as well as self-written routines for data extraction from spectra, input in ARTHUR and conversion of factor spectra into SS300 spectrum format. Longer term classification was evaluated by projecting the group mean data vectors of the first and third experiment on the discriminant axis of the second experiment, which represents the difference between *Salmonellas* and the remaining strains.

Bacteria and growth conditions

Fifteen strains, belonging to the family of Enterobacteriaceae, all being fresh field isolates, were investigated. The following *Salmonella* strains, being epidemiologically important to food chemistry, were included: *senftenberg, infantis, typhimurium* 220, *typhimurium* 510, *panama* A and G, *eimsbuettel, agona, bandaka, montevideo.* The remaining strains were *Escherichia coli, Proteus mirabilis, Citrobacter freundii, Arizona* sp., *Serratia marcescens*, which were selected because of their interfering behaviour during microbiological differentiation. All strains were cultivated under standardized conditions on Brilliant Green Agar solid medium for 20 h at a temperature of 37°C.

RESULTS AND DISCUSSION

Analysis of full scale spectra; all groups based on triplicate measurements

Discriminant analysis applied to groups consisting of triplicate measurements has a strongly unsupervised character. When using this type of category arrangement Fisher weights are comparable to characteristicity weights [16]. This approach enables the handling of data bases consisting of a relatively small number of objects and a large number of variables.
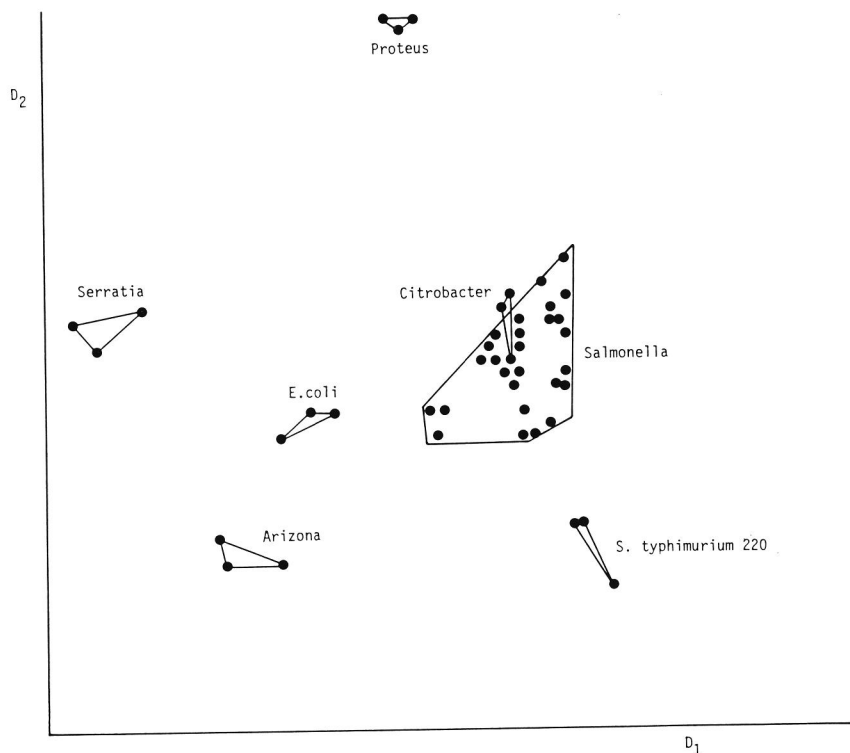
Fig. 1. Discriminant plot of *Salmonella* strains and remaining strains often interfering during microbiological differentiation; groups based on triplicate measurements.

Clusters of groups detected in the discriminant plots under these circumstances are in general meaningful.

In the discriminant plot (Fig. 1) most of the *Salmonella* strains cluster; only *S. typhimurium* 220 occupies a separate position. *Citrobacter*, an organism belonging to the same tribe and being serologically difficult to distinguish, is part of the *Salmonella* cluster. However this strain can be distinguished from *Salmonella* by the third discriminant function. The four remaining species are found in separate positions in the plot, but shifted to a different position compared with *S. typhimurium*.

It can be concluded from this analysis that *Salmonella* strains possess characteristic properties by which they can be distinguished from the interfering strains.

Full scale spectra; Salmonella strains in one group; remaining strains: groups based on triplicate measurements

With this approach characteristic differences between the group of *Salmonella* strains and the remaining strains can be investigated. Except for
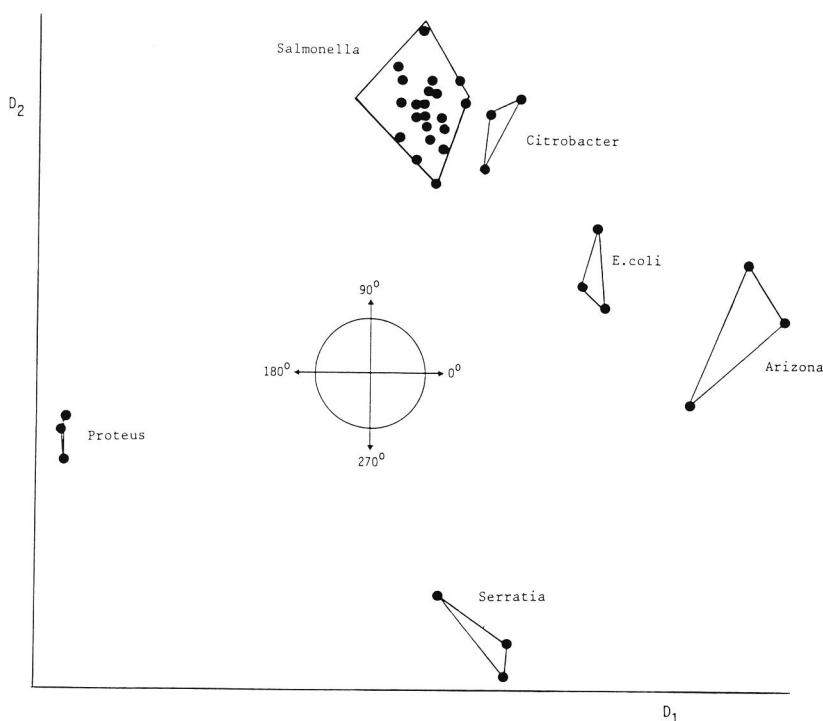
Fig. 2. Discriminant plot of *Salmonella* strains and remaining strains; *Salmonella* strains in one group; remaining strains: groups based on triplicate measurements; analysis of full-scale spectra.

Citrobacter, all strains can be clearly differentiated from the *Salmonella* group, the $150°/330°$ and $40°/220°$ being the main axis of discrimination (Fig. 2). Again *Citrobacter* can be distinguished more clearly from the *Salmonella* group on the third discriminant function.

Discriminant spectra are represented here in the form of correlation patterns. In doing so masses are displayed in accordance with their importance to characterization rather than with their intensity in the pyrograms. Also minor components, like diglycerides, are strongly emphasized in this form of representation.

The discriminant spectra obtained by counterclockwise rotation of the first discriminant function in the $150°/330°$ directions show a strong influence of mass intensities in the range m/z 550–650 (Fig. 3). Most of these masses can be ascribed to diglycerides ($[M + HN_4]^+$ ions), the degradation products of phospholipids. Correlation coefficients of these ions with the rotated functions are mainly in the range of 0.6–0.9. In the mass range m/z 300–400 monoglycerides ($[M + NH_4 - H_2O]^+$ ions) can be expected. Intensities at m/z 328, 330, 342, 356 and 370 emanate from these
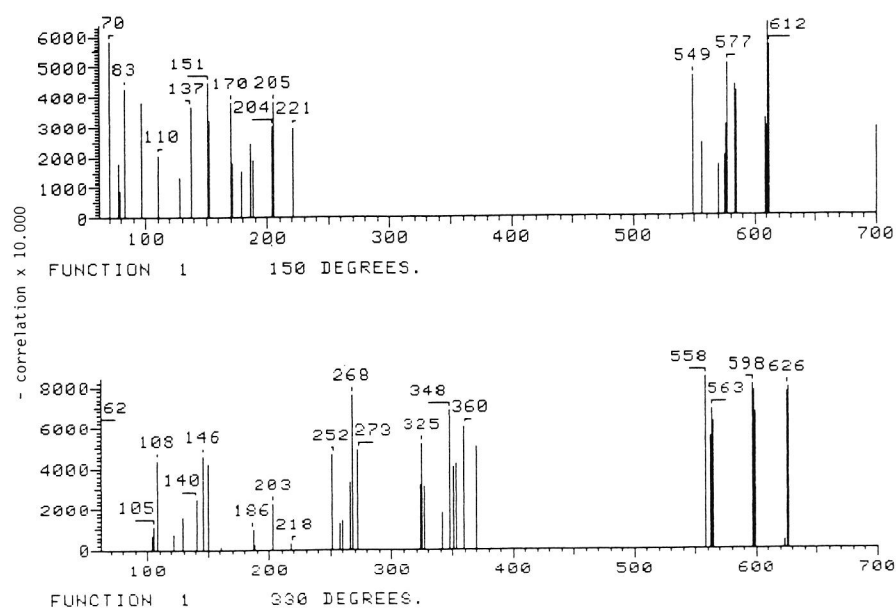
Fig. 3. Factor spectra (mass–discriminant function correlations) after counterclockwise rotation of the first discriminant function over 150° and 330° (see Fig. 2).
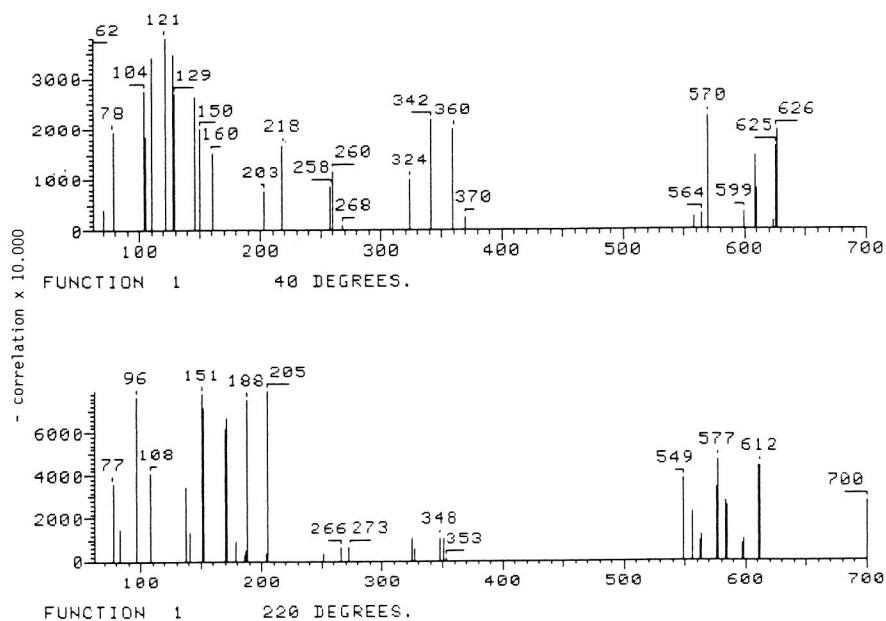

Fig. 4. Factor spectra (mass–discriminant function correlations) after counterclockwise rotation of the first discriminant function over 40° and 220° (see Fig. 2).

products. They originate from 16:1, 16:0, 17:cyclo, 18:1 and 19:cyclo fatty acid esters of glycerol respectively. Only a few of these masses are found in the discriminant spectra (Fig. 3). In general lower mass–discriminant function correlations occur in the lower mass range, which is an indication of the less important contribution of this range to discrimination along this axis. Elemental compositions determined for some of the compounds in this range give an indication of the structure and origin: m/z 146 a deoxyhexose fragment $[M + NH_4]^+$ ion; m/z 150 a pentose fragment $[M + NH_4]^+$ ion; m/z 180 a hexose fragment $[M + NH_4]^+$ ion.

In the 40°/220° directions in which the *Salmonella* group and *Proteus mirabilis* can be differentiated, a group of masses is found with a relatively high abundance in *Proteus*: m/z 151, 152, 170, 171, 188 and 205 (Fig. 4). The elemental composition of two of these masses, m/z 170 and 152, determined by high resolution measurements appears to be $C_8H_{12}NO_3$ and $C_8H_{10}NO_2$, after correction for NH_4, $C_8H_8O_3$ and $C_8H_6O_2$. This points to a highly unsaturated compound which is subject to loss of water.
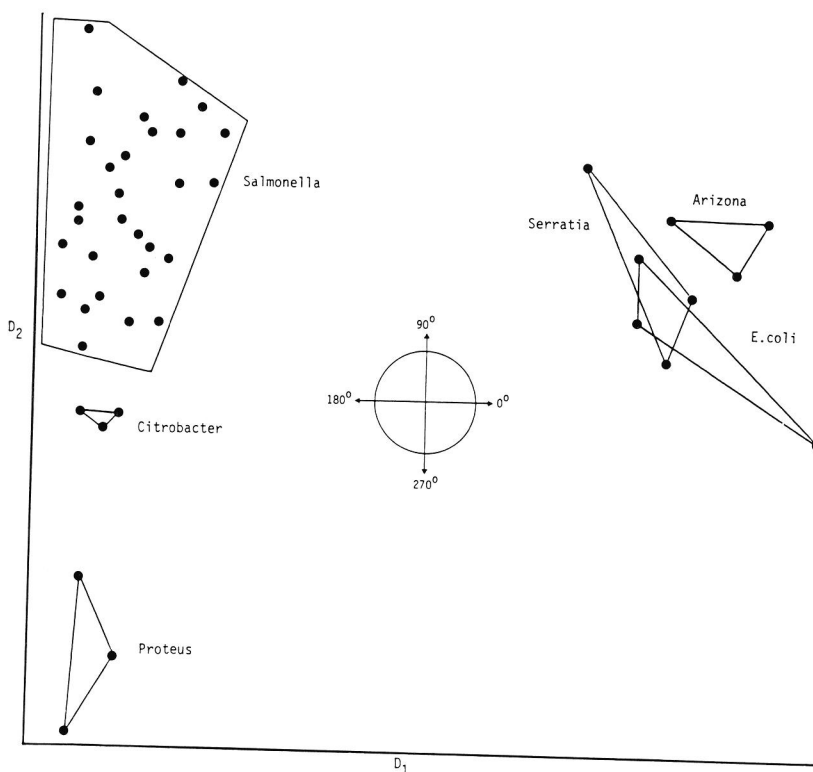

Fig. 5. Discriminant plot of *Salmonella* strains and remaining strains; *Salmonella* strains in one group; remaining strains: groups based on triplicate measurements; analysis of mass range 550–650.

Analysis of mass range 550–650; Salmonella strains in one group; remaining strains: groups based on triplicate measurements

From the discriminant plot (Fig. 5) the significance of this mass range in differentiating is clearly visible. Again *Arizona, Serratia, E. coli* and *Proteus* can be distinguished from the *Salmonella* strains, while *Citrobacter* is found near the *Salmonella* cluster. The discriminant spectrum in the 180° direction indicates the presence of relatively high levels of diglycerides with the m/z values 582, 584, 610, 612 and 638 in the *Salmonella, Citrobacter* and *Proteus* bacteria (Fig. 6). This series of masses can be ascribed to combinations of 16:0, 16:1 and 18:1 fatty acid esters of glycerol in the form of $[M + NH_4]^+$ ions as indicated in Fig. 7; combinations of saturated–unsaturated and unsaturated–unsaturated fatty acids especially occur. The spectrum taken in the direction (0°) of *E. coli, Serratia* and *Arizona* shows the m/z values 558, 598 and 626 indicating relatively high contents of combinations of 14:0, 16:0, 17:cyclo and 19:cyclo fatty acids (Figs. 6 and 7). Apart from the $[M + NH_4]^+$ ions, other degradation products of phospholipids, present in very low intensities in the pyrograms, are also visible; two examples are given in Fig. 7. The intensities m/z 563, 575, 577, 591 and 603, which can
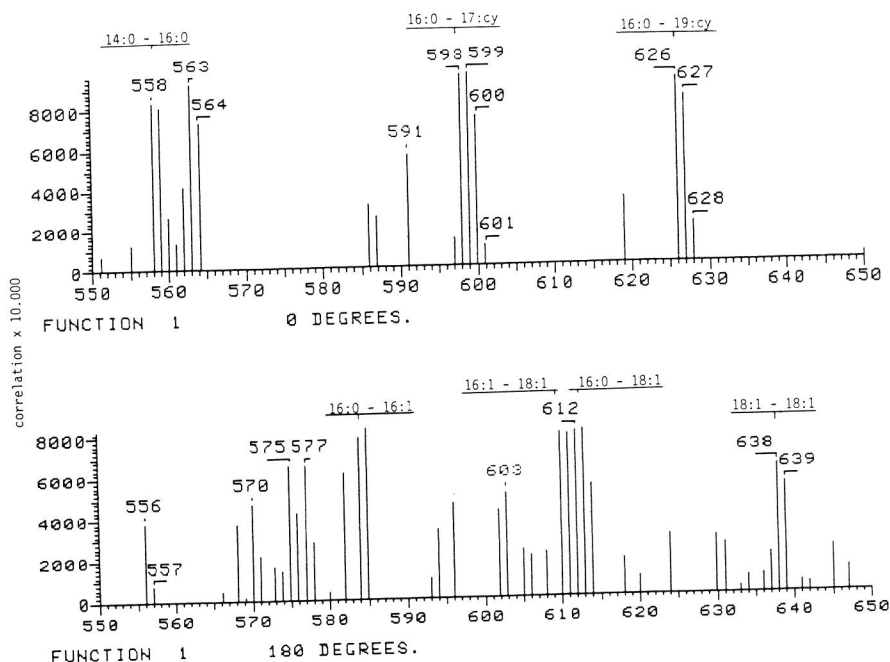

Fig. 6. Factor spectra (mass–discriminant function correlations) in the 0° and 180° directions (see Fig. 5). Fatty acid compositions of diglycerides (ammoniated molecules) are indicated. cy = cyclopropyl unit.
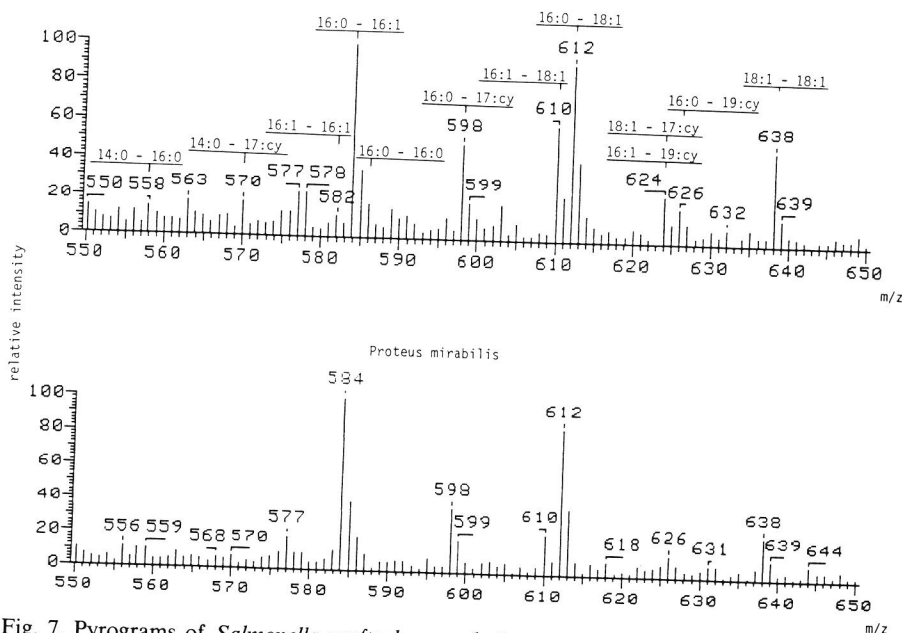
Fig. 7. Pyrograms of *Salmonella senftenberg* and *Proteus mirabilis*, presented over the mass range 550–650. Probable fatty acid compositions of diglycerides (ammoniated masses) are indicated.
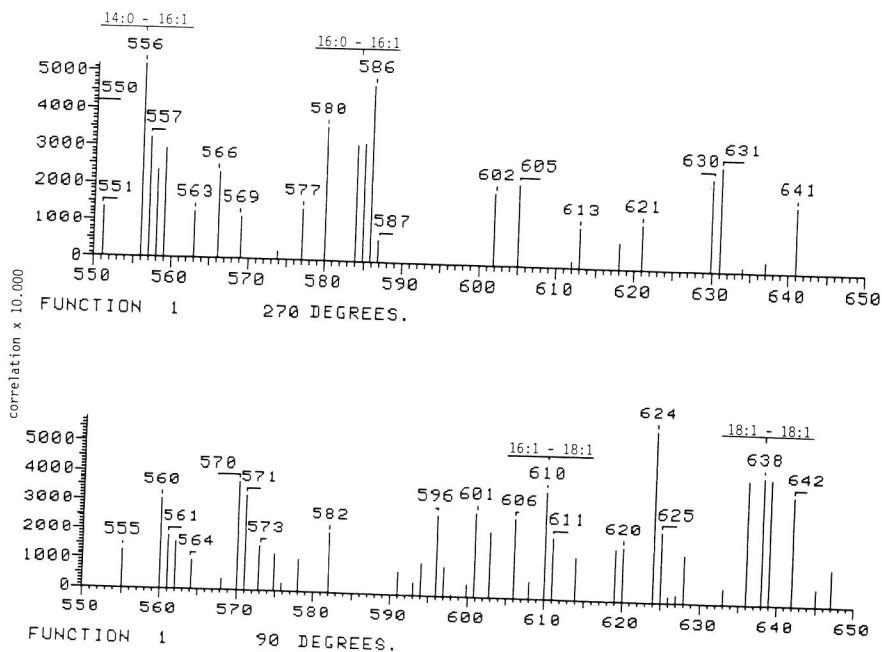

Fig. 8. Factor spectra (mass–discriminant function correlations) after counterclockwise rotation of function D_1 over 270° and 90° (see Fig. 5). Fatty acid compositions of diglycerides (ammoniated masses) are indicated.

48

also be observecd in the factor spectra in the $0°$ and $180°$ directions, are possibly $[M - H_2O + H]^+$ ions in the same series of diglycerides, found 35 mass units lower in the pyrograms than their corresponding $[M + NH_4]^+$ ions. Because of the low intensities of these compounds no high-resolution measurements have been carried out for confirmation.

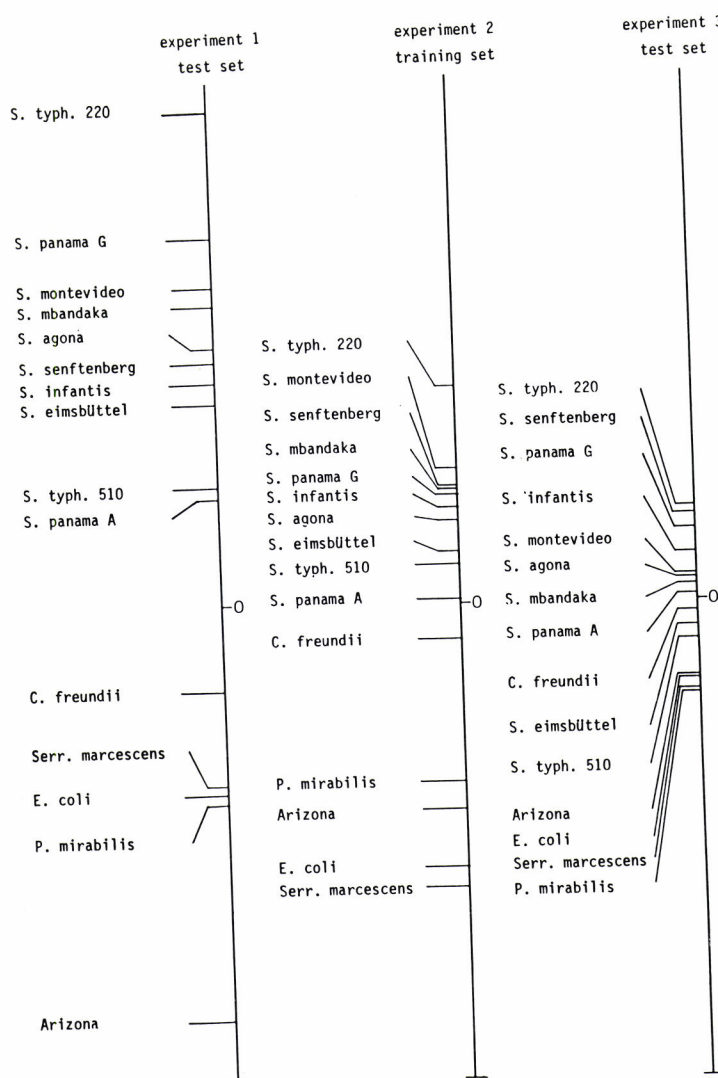The position of *Proteus* is mainly determined by compounds with the

Fig. 9. Longer term prediction ability. Projections of the group mean data vectors of the first and third experiment (test sets) on the discriminant axis of the second experiment (training set), developed for discrimination of two groups: *Salmonella* strains and remaining strains.

m/z values 556, 584 and 612, which result from the combinations of 14:0, 16:0 and 16:1 fatty acids present at a relatively high level compared with the *Salmonella* group (Fig. 8). A survey of combinations of fatty acid esters of glycerol is presented in Fig. 7. The combinations indicated are the most probable ones, as can be concluded from the occurrence of monoglycerides in the pyrograms (m/z 300–400 range) and gas chromatographic data of fatty acid methyl esters obtained after alkaline hydrolysis/extraction/methylation of comparable Gram negative bacteria [17]. A number of these combinations have also been detected by direct probe electron impact Py–MS [10].

Longer term reproducibility

Determination of the prediction ability provides a good impression of the reproducibility. The three identical experiments [longer term reproducibility (LTR) experiments] conducted at two- and four-month intervals were used for this investigation. The prediction ability was calculated by projecting the group mean data vectors of the first and the third experiment on the discriminant function of the second experiment, developed for discrimination of the *Salmonella* group and the group of interfering strains (Fig. 9).

In the first experiment the prediction ability appeared to be excellent. Even *Citrobacter* could be distinguished correctly. In the third experiment
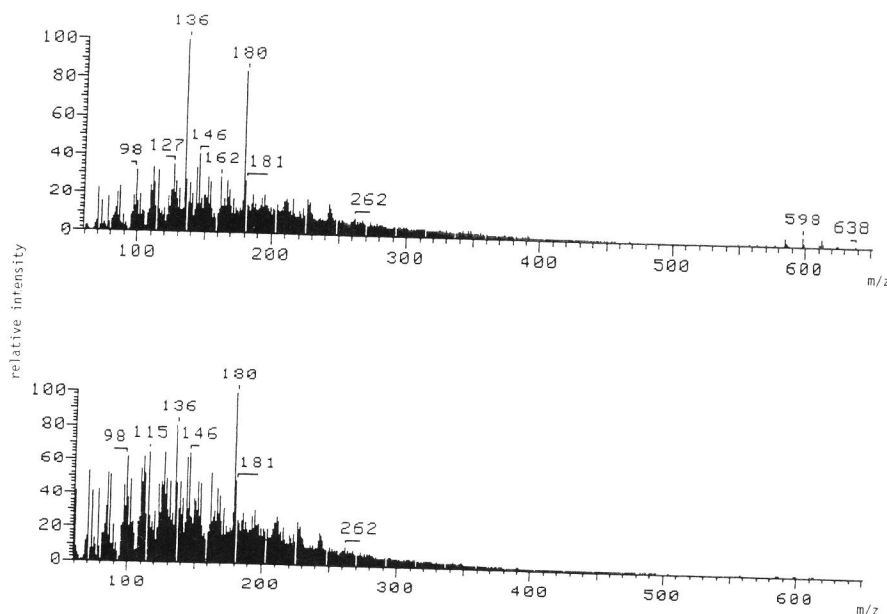

Fig. 10. Longer term reproducibility. Comparison of pyrograms of *Salmonella agona* of the first and third experiment.

however *Citrobacter* was misclassified, whereas the remaining strains could still be distinguished. As can be concluded from these LTR experiments, the discriminating information shows a downward tendency from the first to the third experiment.

A possible explanation for this is provided by the mass intensities in the range m/z 550–650. These intensities of the same strains in the LTR experiments show a decrease, absolute and relative to the lower mass range, for the pyrograms of the third experiment. A typical example for the *Salmonella agona* strain is given in Fig. 10. A decrease of peak intensities in a mass range containing vital characteristic information leads to a loss of discriminating power. This decrease in sensitivity can be attributed to instrumental parameters such as ion source contamination rather than to biological variations.

REFERENCES

1 H.G. Boettger and A.M. Kelly, Proceedings of the 17th Annual A.S.M.S. Conference on Mass Spectrometry and Allied Topics, Dallas, U.S.A., 1969, p. 333.
2 J.L. Wiebers, Anal. Biochem., 51 (1973) 542.
3 F.J. Vastola, A.J. Pirone and B.E. Knox, Proceedings of the 14th Annual Conference on Mass Spectrometry and Allied Topics, ASTM Committee E-14, Philadelphia, 1966, p. 78.
4 S.M. Huff, H.L.C. Meuzelaar, D.L. Pope and C.R. Kjeldsberg, J. Anal. Appl. Pyrolysis, 3 (1981) 95.
5 W. Genuit and J.J. Boon, J. Anal. Appl. Pyrolysis, 8 (1985) 25.
6 U. Bahr and H.-R. Schulten, J. Anal. Appl. Pyrolysis, 5 (1983) 27.
7 A. Zsolnay, J. Anal. Appl. Pyrolysis, 4 (1982) 47.
8 A.C. Tas, J. van der Greef, J. de Waart, J. Bouwman and M.C. ten Noever de Brauw, J. Anal. Appl. Pyrolysis, 7 (1985) 249.
9 E. Reiner and W.H. Ewing, Nature (London), 217 (1968) 191.
10 J.P. Anhalt and C. Fenselau, Anal. Chem., 47 (1975) 219.
11 G. Wieten, H.L.C. Meuzelaar and J. Haverkamp, in G. Odham, L. Larsson and P.A. Mårdh (Editors), Gas Chromatography–Mass Spectrometry, Applications in Microbiology, Plenum Press, New York, 1984, Ch. 10.
12 E. Reiner, J.J. Hicks, M.M. Ball and W.J. Martin, Anal. Chem., 44 (1972) 1058.
13 B.S. Emswiler and A.W. Kotula, Appl. Environm. Microbiol., 35 (1978) 97.
14 W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81.
15 Pattern recognition package ARTHUR 81, Infometrix, Inc., Seattle, WA, U.S.A.
16 W. Eshuis, P.G. Kistemaker and H.L.C. Meuzelaar, in C.E.R. Jones and C.A. Cramers (Editors), Analytical Pyrolysis, Elsevier, Amsterdam, 1977, pp. 151–166.
17 N.A. Machtiger and W.M. O'Leary, J. Bacteriol., 114 (1973) 80.

Chapter 5

Characterization of *Salmonella* and possible interfering strains using GC profiling and factor analysis

A. C. Tas[1], G. Wieten[2], J. de Waart[1], L. Berwald[2] and J. van der Greef[1]

[1]*TNO-CIVO Institutes, Toxicological Analysis Department, Zeist, The Netherlands and* [2]*National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands*

Summary

Fatty acid methyl ester data of *Salmonella* and potentially interfering strains obtained by derivatization gas chromatography were investigated with principal components and factor-discriminant analysis. *Salmonella* strains could be distinguished using both data-handling techniques. The biochemical basis for this characterization was found to be in strong agreement with the higher mass range part of pyrolysis – direct chemical ionization mass spectrometric data representing diglyceride information. The quality of the analytical procedure as judged by longer term reproducibility and component assignment was found to be good. In combination with sophisticated data-handling methods this derivatization gas chromatographic approach can be a useful diagnostic tool for *Salmonella* characterization.

Introduction

For a considerable time gas chromatography (GC) has proven to be a useful analytical method for rapid and objective characterization of bacteria. Two main approaches can be identified, pyrolysis – gas chromatography (Py-GC) [1, 2] and methods based on extraction/derivatization [3 – 5]. Py-GC is suitable for the detection of a wide range of volatile compounds, providing a more general overview; while derivatization-GC is directed toward the detection of specific types of compounds.

Reproducibility and resolution, aspects which are essential for profiling purposes, have been considerably improved with the introduction of modern equipment. Py-GC is highly demanding with respect to chromatographic conditions due to the wide range

Correspondence to: A. C. Tas, TNO-CIVO Institute, Toxicological Analysis Department, P.O. Box 360, 3700 AJ Zeist, The Netherlands

of polarities and differences in volatility of pyrolysis products. The development of new low dead volume Curie point pyrolysers [6] and the introduction of thermostable polar stationary phases has led to the extension of the range of applications [7, 8]. In situations where sensitive detection of a specific group of compounds is required, the use of derivatization-GC is recommendable. The potential of this approach is illustrated by recent work on the specific detection of rhamnose and muramic acid as chemical markers of bacterial cell walls as well as bacterial characterization by fatty acid methyl ester profiling [3, 4, 9, 10]. Despite the relatively long analysis time required for GC analysis, the automation of sample injection, system calibration and data handling allows a high sample throughput.

Recent experiments carried out with pyrolysis-direct chemical ionization mass spectrometry (PY-DCI/MS) on a number of *Salmonella* and potentially interfering strains indicate that long-chain fatty acid diesters of glycerol (diglycerides), the pyrolysis products of phospholipids, are important in the characterization of the strains [11]. These results prompted us to investigate a number of these strains with an automatic GC system developed specifically for the routine determination of fatty acid methyl ester profiles [12]. The GC data obtained were subjected to explorative data analysis with factor-discriminant routines in order to trace the important biochemical variables for strain differentiation [13]. The results of this analysis were compared with diglyceride data obtained by Py-DCI/MS [11]. In a separate paper aspects related to the use of automated procedures for bacterial identification will be discussed.

Materials and Methods

Gas chromatography

Gas chromatography was performed using the Hewlett Packard 5898 Microbial Identification System, consisting of an HP 5790A gas chromatograph coupled to an HP 3390A integrator and a series 9000 workstation. The analysis was conducted with a 25 m × 0.2 mm methyl phenyl silicone-fused silica capillary column (HP Ultra), film thickness 0.33 μm using hydrogen as carrier gas, under the conditions indicated by Hewlett Packard [12]. These conditions were: injector temperature 250 °C; detector temperature 300 °C; initial oven temperature 170 °C; temperature ramp 5 °C/min; and final oven temperature 270 °C. All bacterial extracts except one (*Salmonella mbandaka* standard strain, which was lost after the first analysis) were analyzed in duplicate with an interval of 14 d.

Test samples of bacterial fatty acid methyl esters supplied by the manufacturer for system calibration purposes were analyzed by gas chromatography-mass spectrometry (GC-MS). This analysis was used to confirm the fatty acid identification by the HP system based on retention time indices. One of the test samples was particularly appropriate for comparison, as it contained a mixture of methyl esters of long-chain saturated, mono-unsaturated, cyclopropyl and hydroxy fatty acids which normally occur in Gram-negative bacteria. GC-MS analysis was performed on a Varian 3700 gas chromatograph, equipped with a 25 m × 0.32 mm DB-1 (film thickness, 0.25 μm) fused silica column, coupled to a Finnigan MAT 112S mass spectrometer.

Extraction/derivatization

The extraction/derivatization of the fatty acids from bacteria was carried out according to the standardized procedure indicated by Hewlett Packard [12]. This procedure consisted of the following steps:

1. saponification with sodium hydroxide (approx. 4 M) in methanol water (1:1) at 100 °C for 30 min;
2. methylation with methanol-HCl (approx. 3 M) at 80 °C for 10 min, followed by rapid cooling to room temperature;
3. extraction with hexane-methyl-t-butylether for 10 min;
4. after removal of the water layer, the organic layer was washed once with dilute sodium hydroxide (approx. 0.3 M) for 5 min;
5. the organic phase was used for analysis.

For 2 samples, *Salmonella agona* and *Serratia marcescens*, this procedure was carried out in duplicate.

Bacteria and culturing conditions

Twenty-nine strains consisting of both fresh-field isolates (FF) and standard strains (SS) were investigated. In view of their epidemiological importance with respect to food microbiology the following *Salmonella* strains were selected: *S. eimsbuettel* (FF and SS), *S. typhimurium 220* and *510* (FF and SS), *S. panama A* and *G* (FF and SS), *S. infantis* (FF and SS), *S. mbandaka* (FF and SS), *S. agona* (FF and SS), *S. senftenberg 87* (FF and SS), *S. senftenberg* isolated from cacao powder and *S. montevideo* (FF and SS). Additional strains selected because of their interference in microbiological isolation and identification were: *Citrobacter freundii* (FF and SS), *E. coli* (FF and SS), *Serratia marcescens* (FF and SS), *Arizona* (FF) and *Proteus mirabilis* (FF). All strains were cultivated under standard conditions on Brilliant Green Agar solid medium for 18 h at a temperature of 37 °C.

Data analysis

Twelve peaks were selected from the chromatograms to be included in the data analysis. Peaks with more than 3 missing values in the data set were rejected. Chromatograms were normalized to 100% total area in order to correct for differences in sample concentration. The data set was transformed by autoscaling. Principal component and discriminant analyses were therefore performed on a 'correlation about the mean' matrix.

Principal component analysis was performed on a training set of 31 patterns which were generated first, and a test set consisting of 30 duplicate patterns (1 pattern, *S. mbandaka*, was missing) generated 14 d later. Principal component-discriminant analysis was carried out on this data set. The following categories were defined: all *Salmonella* strains were combined in 1 group; the remaining groups were based on duplicate measurements of the strains. The pattern recognition program ARTHUR was used [13], extended with self-written routines for data input and BIPLOT display. Discriminant analysis and graphical rotation routines were developed by the FOM-Institute, Amsterdam [14].

Results and Discussion

Principal component analysis

Principal component (PC) plots, which display a major part of the variance contained in the original variables, provide important information about the position of objects in the multidimensional space. The plots provide an unsupervised result, because no category information has been used. Objects which are found to form clusters in such plots are likely to be closely related. The reliability of such information increases if a relatively larger portion of the variance is displayed in the plot and if the same cluster information is found in several orthogonal projections. In the plot of object scores on the first 2 principal components representing 65% of the variance contained in the original variables, the group of *Salmonella* strains shows a wide cluster (Fig. 1a, Table 1). The *S. senftenberg* standard strain is the only member of the *Salmonella* group which occupies an isolated position in this plot.

According to DNA hybridization studies, the investigated non-*Salmonella* strains are all closely related to *Salmonella* [15]. The position of these non-*Salmonella* strains indicates that only *Citrobacter* strains cannot be differentiated with certainty from the *Salmonella* cluster in view of their position just outside this cluster. The remaining strains, however, are found at sufficiently remote positions in the plot to be distinguished.

From the plot of the same principal components, now displaying the variables, it can be inferred which variables are important in distinguishing the strains in the plot (Fig. 2). In these plots variables are represented using their loadings on the principal components as their coordinates (BIPLOT technique). Variables, displayed in the same direction as a cluster of objects, are present in larger amounts in those objects. A survey of semiquantitative assignments based on these directions in the plots (Figs 1a and 2) is presented in Table 2.

The *Salmonella* strains in general contain larger amounts of 12:0, 15:0, 16:1, 17:0 and 18:1 fatty acids and as a minor component an unknown (UNKN) fatty acid. The same is true for the *Citrobacter* strains. However, compared to the *Salmonella* strains,

TABLE 1

ABBREVIATED *SALMONELLA* NAMES USED IN FIGURES 1 AND 3

| Name	Abbreviation	Name	Abbreviation
S. eimsbuettel-FF	SEIF	*S. eimsbuettel*-SS	SEIS
S. typhimurium 220-FF	TY2F	*S. typhimurium 220*-SS	TY2S
S. typhimurium 510-FF	TY5F	*S. typhimurium 510*-SS	TY5S
S. panama A-FF	PAAF	*S. panama A*-SS	PAAS
S. panama G-FF	PAGF	*S. panama G*-SS	PAGS
S. infantis-FF	INFF	*S. infantis*-SS	INFS
S. mbandaka-FF	MBAF	*S. mbandaka*-SS	MBAS
S. agona-FF	AGOF	*S. agona*-SS	AGOS
S. senftenberg 87-FF	SE8F	*S. senftenberg 87*-SS	SE8S
S. montevideo-FF	MONF	*S. montevideo*-SS	MONS
S. senftenberg cacao	SECA		
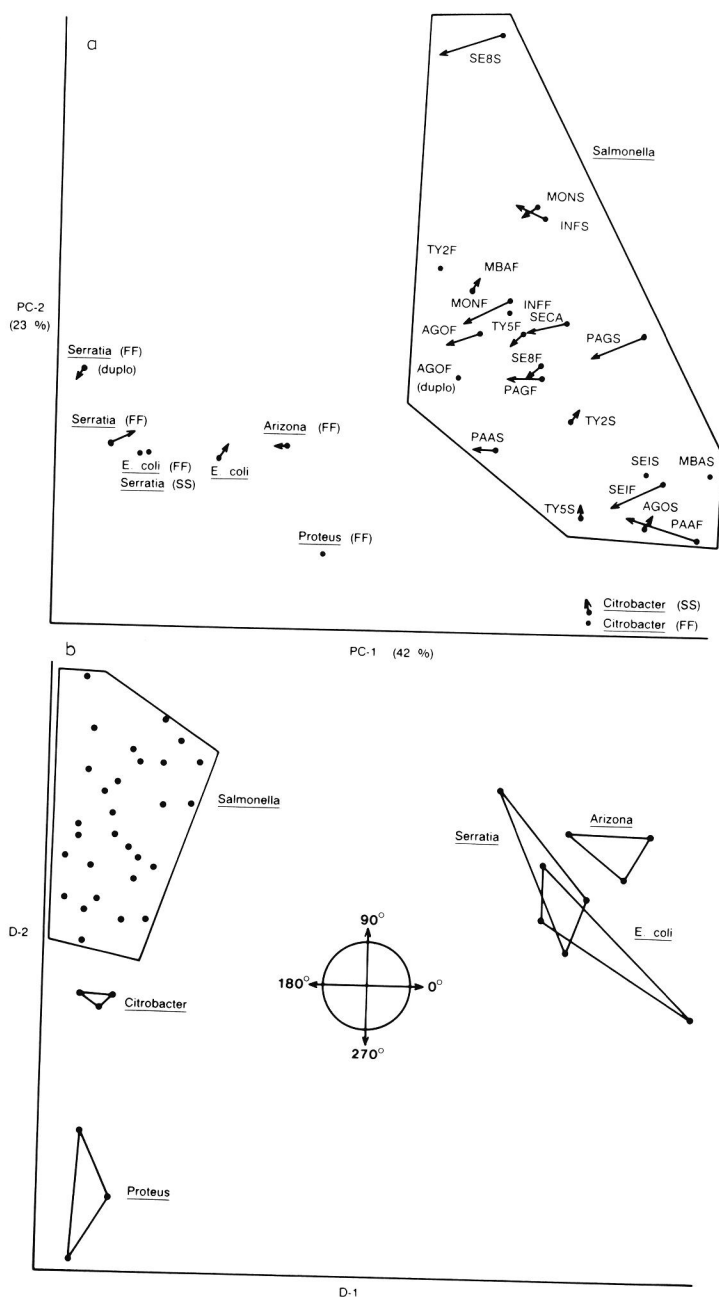
Fig. 1.   (a) Principal component object plot PC 1 vs 2 of *Salmonella* and potentially interfering strains (FF and SS). Duplicate measurements are connected by arrows unless both points coincide. (b) Discriminant object plot DA 1 vs 2 of *Salmonella* and potentially interfering strains (FF) based on diglyceride data acquired with pyrolysis-direct chemical ionization mass spectrometry (from [11]).
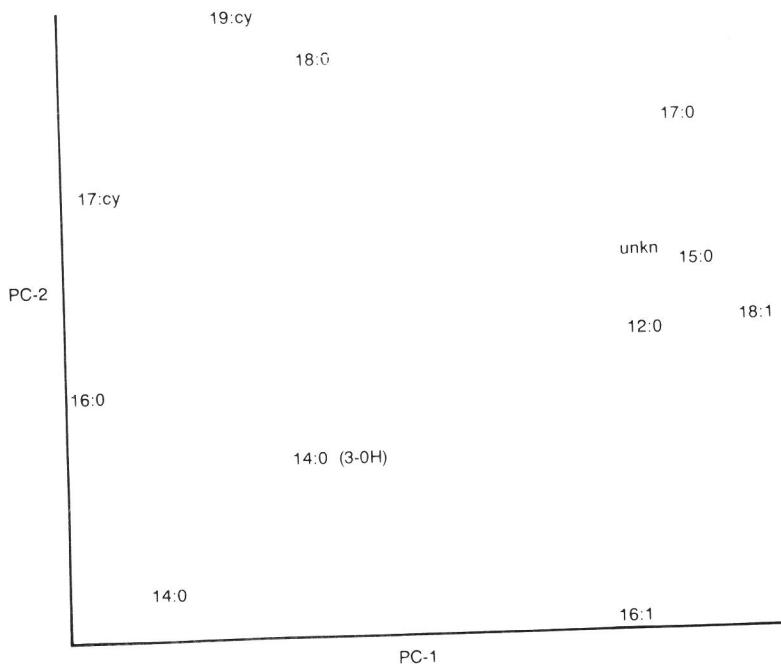
Fig. 2. Principal component variable plot PC 1 vs 2 corresponding to object plot Fig. 1a.

Citrobacter strains contain larger amounts of 16:1 fatty acid, which determines their position just below the *Salmonella* cluster. At the lower right-hand side of the plot in Fig. 1a, a sub-cluster of *Salmonella* strains, consisting of *S. mbandaka* SS, *S. eims-buettel* FF and SS, *S. typhimurium 510* SS, *S. agona* SS and *S. panama A* FF, can be observed which also contain larger amounts of 16:1 fatty acid. The remaining strains located at the left-hand side of the object plot (Fig. 1a) contain larger amounts of 14:0, 17:cycl, 16:0 and to a minor extent 19:cycl fatty acids.

The discriminant (DA) plot in Fig. 1b is reproduced from the previously reported article on DCI-pyrolysis [11]. This plot was based on mass intensities in the higher mass region (m/z 550–700) of pyrograms obtained from a similar set of strains as the one presently analyzed. A remarkable similarity can be observed between the PC and DA plots (Figs 1a and 1b), being almost each other's mirror image. This indicates that the 2 sets of mass spectral and gas chromatographic information were essentially derived from the same set of chemical components. In fact, the presently observed biochemical basis for the object plot in Fig. 1a corresponds to a large extent to the DCI data which showed larger amounts of the combinations of 16:0–16:1, 16:0–18:1, 16:1–18:1 and 18:1–18:1 fatty acid esters of glycerol in *Salmonella* and *Citrobacter* strains and, on the other hand, larger amounts of the combinations of 14:0–16:0, 16:0–17:cycl and 16:0–19:cycl fatty acids in *Serratia*, *Arizona* and *E. coli* strains [11]. In the DCI diglyceride data, mainly originating from the pyrolysis of cellular phospholipids, the

TABLE 2

RELATIVE FATTY ACID CONTENTS OF *SALMONELLA* AND POTENTIALLY INTERFERING STRAINS

	12:0	14:0	unkn	15:0	14:0 3-OH	16:1	16:0	17:cy	17:0	18:1	18:0	19:cy
Salmonella strains												
S. eimsbuettel-FF	+	–	–	–	–	++		–	–	++	–	–
S. eimsbuettel-SS	+	–	+	+	–	++		–	–	++	–	–
S. typhimurium 220-FF	+	–	+	+	–	–		–	–	++	–	–
S. typhimurium 220-SS	++	–	+	+	–	+		–	+	+	+	–
S. typhimurium 510-FF	+	–	+	+	–	+		–	–	++	–	–
S. typhimurium 510-SS	+	–	–	–	–	++		–	–	++	–	–
S. panama A-FF	+	–	–	–	–	++		–	–	++	–	–
S. panama A-SS	+	–	+	+	–	++		–	–	++	–	–
S. panama G-FF	+	–	+	+	–	+		–	–	+	–	–
S. panama G-SS	++	–	+	+	–	+		–	–	+	–	–
S. infantis-FF	+	–	+	+	–	+		–	–	++	–	–
S. infantis-SS	+	–	+	+	–	–		–	+	+	+	–
S. mbandaka-FF	+	–	+	+	–	–		–	+	+	+	–
S. mbandaka-SS	+	–	+	+	–	–		–	+	+	+	–
S. agona-FF	+	–	+	+	–	++		–	–	++	–	–
S. agona-FF (duplo)	+	–	+	+	–	+		–	–	+	–	–
S. agona-SS	+	–	–	–	–	++		–	–	++	–	–
S. senftenberg 87-FF	+	–	+	+	–	+		–	–	+	–	–
S. senftenberg 87-SS	–	–	+	+	–	–		–	–	+	–	–
S. senftenberg cacao	+	–	+	+	–	+		–	–	++	+	+
S. montevideo-FF	+	–	+	+	–	+		–	–	+	–	–
S. montevideo-SS	+	–	+	+	–	–		–	+	+	+	–
Non-*Salmonella* strains												
Serratia-FF	–	+	–	–	+	–	++	+	–	–	–	–
idem (duplo)	–	+	–	–	+	–	++	+	–	–	–	–
Serratia-SS	–	+	–	–	+	–	++	+	–	–	–	–
E. coli-FF	–	+	–	–	+	–	++	+	–	–	–	–
E. coli-SS	–	+	–	–	+	–	++	+	–	–	–	–
Arizona-FF	–	+	–	–	++	–	+	+	–	–	–	–
Proteus-FF	–	+	–	–	++	–	+	–	–	–	–	–
Citrobacter-FF	+	–	–	+	–	++	–	–	+	+	–	–
Citrobacter-SS	+	–	–	+	–	++	–	–	+	+	–	–

–: relatively low contents; +: relatively medium contents; ++: relatively high contents.

combination with 15:0 and 17:0 fatty acids was not observed. This can be due to the low abundances of these compounds or to the possibility that such combinations do not occur in the phospholipids of these bacteria and have a different cellular origin.

In the plot of PC 1 vs 3, which accounts for 52% of the variance, a *Salmonella* subcluster is observed at the right-hand side of the plot consisting of those strains containing extra high levels of 18:1 fatty acid often accompanied by a relatively low 17:cycl content (Figs 3 and 4). In this plot the *S. mbandaka* SS is found in an outlier position. By the same analytical approach the relationship between fresh-field isolates and standard strains can also be investigated. According to the distances in the plot (Fig. 1a)
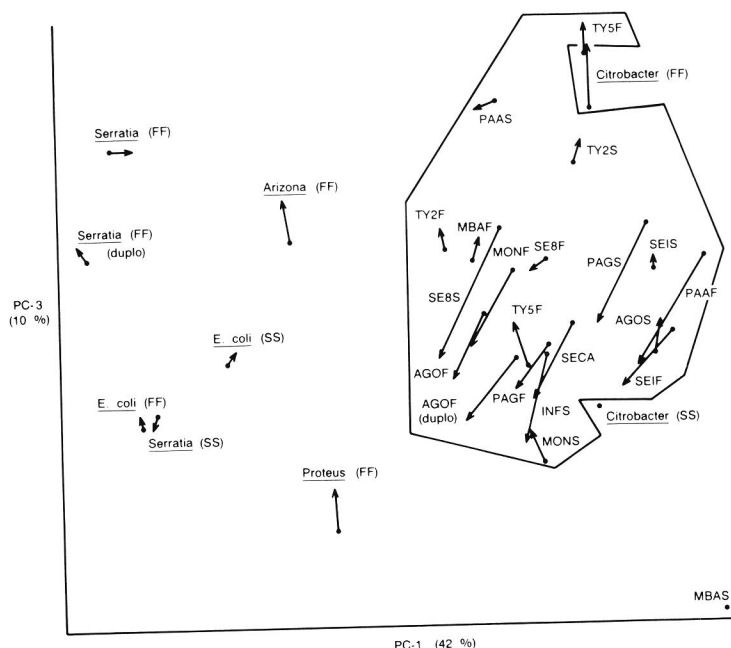
Fig. 3. Principal component object plot PC 1 vs 3 of *Salmonella* and potentially interfering strains. Duplicate measurements are connected by arrows unless both points coincide.

for many of the strains, these effects play only a minor role. However, for some strains considerable differences exist. Especially between the *Salmonella* strains *S. senftenberg*, *S. mbandaka*, *S. panama A*, *S. typhimurium 510* and *220* are these differences significant and cause an increase of the cluster size. From the variable plot (Fig. 2), the origin of these shifts can be inferred. For example, the *S. senftenberg* standard strain shows higher contents of 18:0 and 19:cycl fatty acids; in the *S. panama A* fresh-field isolate the higher contents of 16:1, 18:1 and the lower contents of 17:cycl and 19:cycl fatty acids compared to the corresponding standard strain cause the shift in position. GC-MS measurements on the test mixture revealed that the automated structural assignments based on retention indices were correct for the compounds used in this investigation.

The reproducibility of the GC analyses as judged from the measurements conducted on the same set of samples with an intermediate period of 14 d was found to be good. Projection of the test set data, consisting of duplicate measurements, on the space of the training set data (Figs 1a and 3) shows that most of the test and training set objects are similar. However, a number of *Salmonella* species show a larger systematic shift mainly along the first and third principal component axis, the test set objects being displayed in approximately the 240 °C direction (Fig. 3).

The influence on pattern reproducibility as introduced by the extraction/derivatiza-
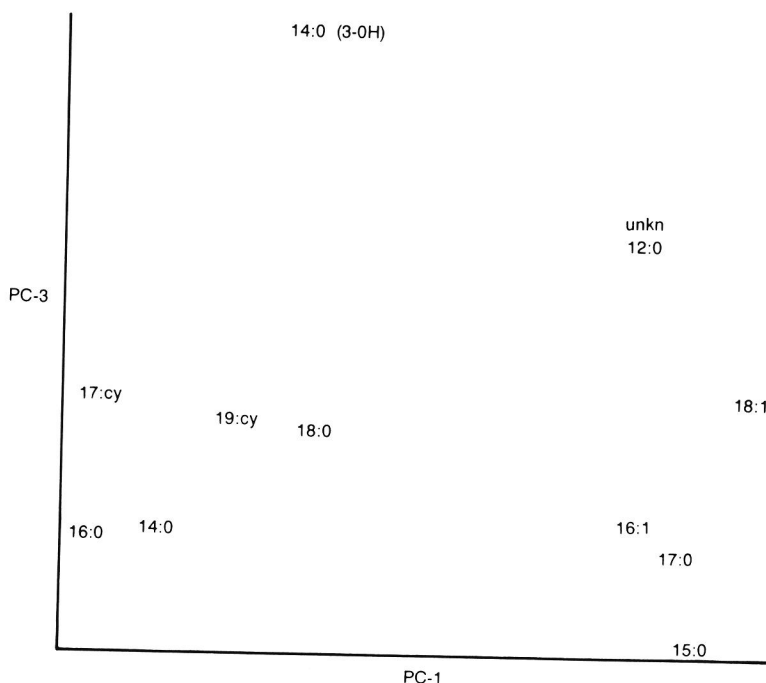
Fig. 4.   Principal component variable plot PC 1 vs 3 corresponding to object plot Fig. 3.

tion procedure is difficult to evaluate. According to the PC plots a slightly larger dupli-
cate variation of the extraction/derivatization procedure compared to the experimen-
tal GC variation can be observed in projections of *Serratia marcescens* on the first
2 principal components, while only a minor variation is observed with *S. agona*.

Principal component – discriminant analysis

In order to investigate in more detail the characteristic differences in fatty acid pat-
terns between *Salmonella* strains and potentially interfering strains, principal
component – discriminant analysis was applied. All *Salmonella* strains were defined as
a group, while for the remaining strains the same was done for each duplicate measure-
ment. As can be expected the *Salmonella* cluster is now more compact and the dis-
tances between this cluster and the remaining strains have increased (Fig. 5). However,
the *Citrobacter* standard strain is still located in the vicinity of this cluster, while the
fresh-field isolate is better distinguished in the latter plot. All *Serratia* strains, standard
strains, fresh-field isolates as well as the duplicate measurements from the derivatiza-
tion/extraction procedure almost coincide. However, standard strains and fresh-field
isolates of *Citrobacter* and *E. coli* are found at different positions in the plot.

In analogy to the principal component analysis, the variable contributions to the
discriminant functions were investigated by the BIPLOT technique. Investigation of
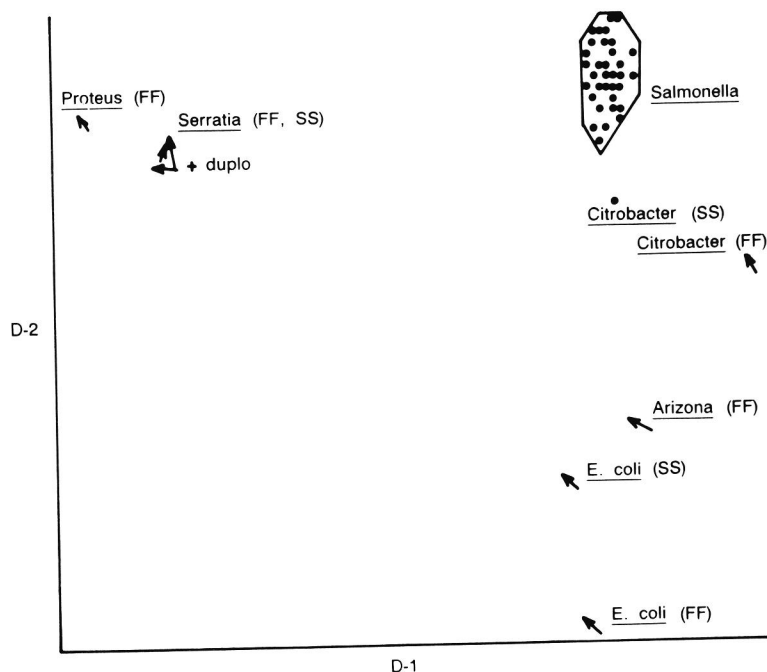
Fig. 5. Discriminant object plot DA 1 vs 2 of *Salmonella* and potentially interfering strains. Duplicate measurements are connected by arrows unless both points coincide.

the variable plot of the first 2 discriminant functions indicates mainly the same trends as already observed for the principal component analysis: higher contents of 18:1, 16:1, 17:0, 15:0 and 12:0 fatty acids are characteristic for the *Salmonella* cluster and the *Citrobacter* strains (Fig. 6).

Despite the fact that the non-*Salmonella* strains appear in 2 separate areas in the plot, common chemical factors can be indicated. The fatty acids 16:0, 14:0 and 17:cycl occur in larger amounts in these strains as is shown by their intermediate position in the lower left-hand side of the plot (compare Figs 5 and 6) between the 2 subclusters. The subcluster at the left-hand side of the plot, consisting of *Serratia* and *Proteus* strains, is characterized by low contents of 12:0 fatty acid as compared to the groups of *E. coli, Arizona* and *Citrobacter* found at the right-hand side. The object plot of the first and third discriminant function is also suitable in distinguishing *Citrobacter* from *Salmonella* strains (Fig. 7). Moreover, the third discriminant function displays the scatter between *Proteus* and the *Serratia* strains. The *Serratia* strains contain higher levels of the 14:0 3-OH and 17:cycl fatty acids (Fig. 8). The observations with respect to 12:0, 14:0 3-OH and 17:cycl fatty acids are additional aspects emphasized by this discriminant analysis approach.
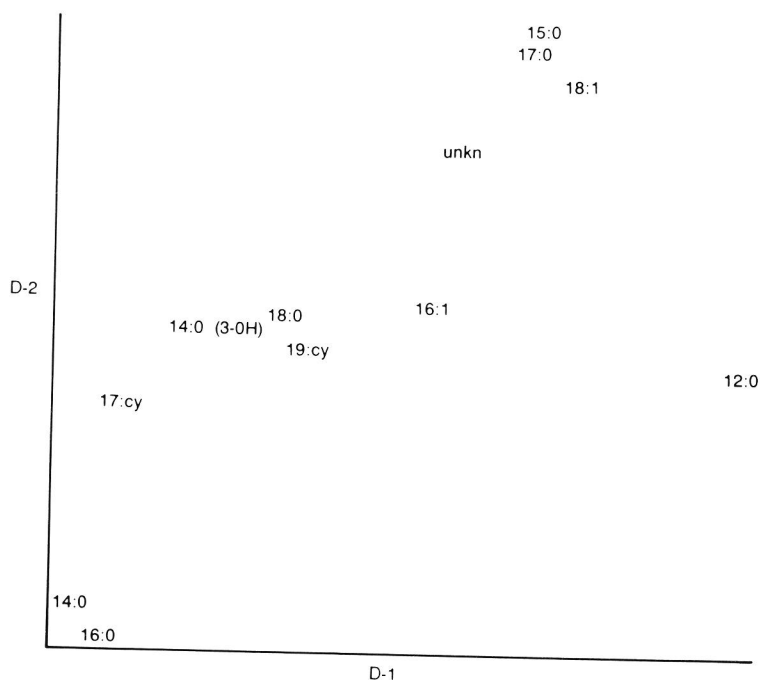
Fig. 6. Discriminant variable plot DA 1 vs 2 corresponding to object plot Fig. 5.
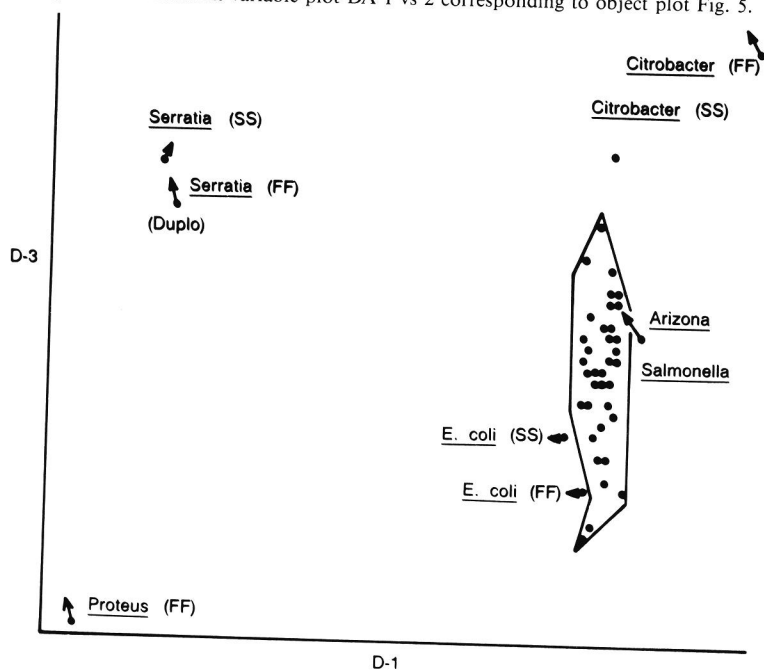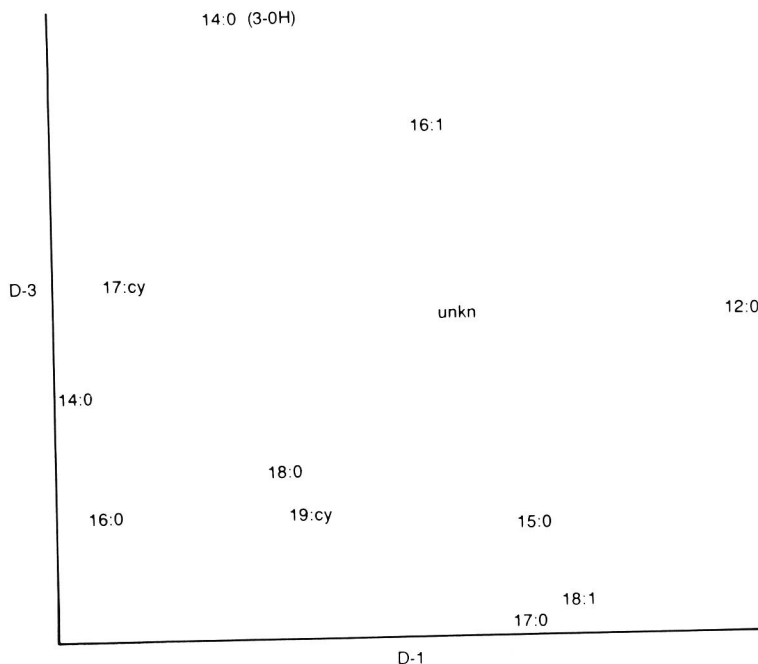

Fig. 7. Discriminant object plot DA 1 vs 3 of *Salmonella* and potentially interfering strains. Duplicate measurements are connected by arrows unless both points coincide.

Fig. 8. Discriminant variable plot DA 1 vs 3 corresponding to object plot Fig. 7.

Acknowledgements

References

1 Gutteridge, C. S. and Norris, J. R. (1979) The application of pyrolysis techniques to the identification of micro-organisms. J. Appl. Bacteriol. 47, 5 – 43.
2 Wieten, G., Meuzelaar, H. L. C. and Haverkamp, J. (1984) Analytical Pyrolysis in Clinical and Pharmaceutical Microbiology. In: Gas Chromatography-Mass Spectrometry, Applications in Microbiology, (Odham G., Larsson, L. and Mardh P. -A, Eds.), pp. 335 – 380, Plenum Press, New York.
3 Goodfellow, M. and Minnikin, D. E. (1985) Chemical Methods in Bacterial Systematics, Academic Press, London.
4 Moss, C. W. (1981) Gas-liquid chromatography as an analytical tool in microbiology. J. Chromatogr. 203, 337 – 347.
5 Fox, A. and Morgan, S. L. (1985) The Chemotaxonomic Characterization of Microorganisms by Capillary Gas Chromatography and Gas Chromatography-Mass Spectrometry. In: Instrumental Methods for Rapid Microbiological Analysis, (Nelson, W. H., Ed.), pp. 135 – 164, VC Publishers, Deerfield Beach, U. S. A.

6   Boon, J. J., Pouwels, A. D. and Eykel, G. B. (1987) Pyrolysis high resolution gas chromatography-mass spectrometry studies on beech wood: capillary high resolution mass spectrometry of a beech lignin fraction. Biochem. Soc. Trans. 15, 170–174.

7   Wieten, G., Boon, J. J., Groothuis, D. G., Portaels, F. and Minnikin, D. E. (1984) Rapid detection of mycobacterial contamination in batches of whole cells of purified *M. leprae* by pyrolysis gas chromatography. FEMS Microbiol. Lett. 25, 289–293.

8   Fox, A. and Morgan, S. M. (1986) Identification of microorganisms by GC/MS. Anal. Chem. 58, 1310A–1316A.

9   Gilbart, J., Fox, A., Whiton, R. S. and Morgan, S. L. (1986) Rhamnose and muramic acid: chemical markers for bacterial cell walls in mammalian tissues. J. Microbiol. Methods 5, 271–282.

10 Brondz, I. and Olsen, I. (1986) Microbial Chemotaxonomy. Chromatography, Electrophoresis and Relevant Profiling Techniques. J. Chromatogr. Biomed. Appl. 379, 367–411.

11 Tas, A. C., de Waart, J., Bouwman, J., ten Noever-de Brauw, M. C. and van der Greef, J. (1987) Rapid characterization of *Salmonella* strains with direct chemical ionization pyrolysis. J. Anal. Appl. Pyrol. 11, 329–340.

12 Miller, L. (1984) Gas-liquid chromatography of cellular fatty acids as a bacterial identification aid. Gas Chromatography, Hewlett Packard Application Note 228-37, 1–8.

13 Pattern recognition package ARTHUR 81, Infometrix, Inc., Seattle, WA, U. S. A.

14 Windig, W., Haverkamp, J. and Kistemaker, P. G. (1983) Chemical interpretation of sets of mass spectra by discriminant analysis and graphical rotation. Anal. Chem. 55, 81–88.

15 Fitts, R., Diamond, M., Hamilton, C. and Neri, M. (1983) DNA Hybridization assay for Detection of *Salmonella* spp. in Foods. Appl. Environ. Microbiol. 46, 1146–1151.
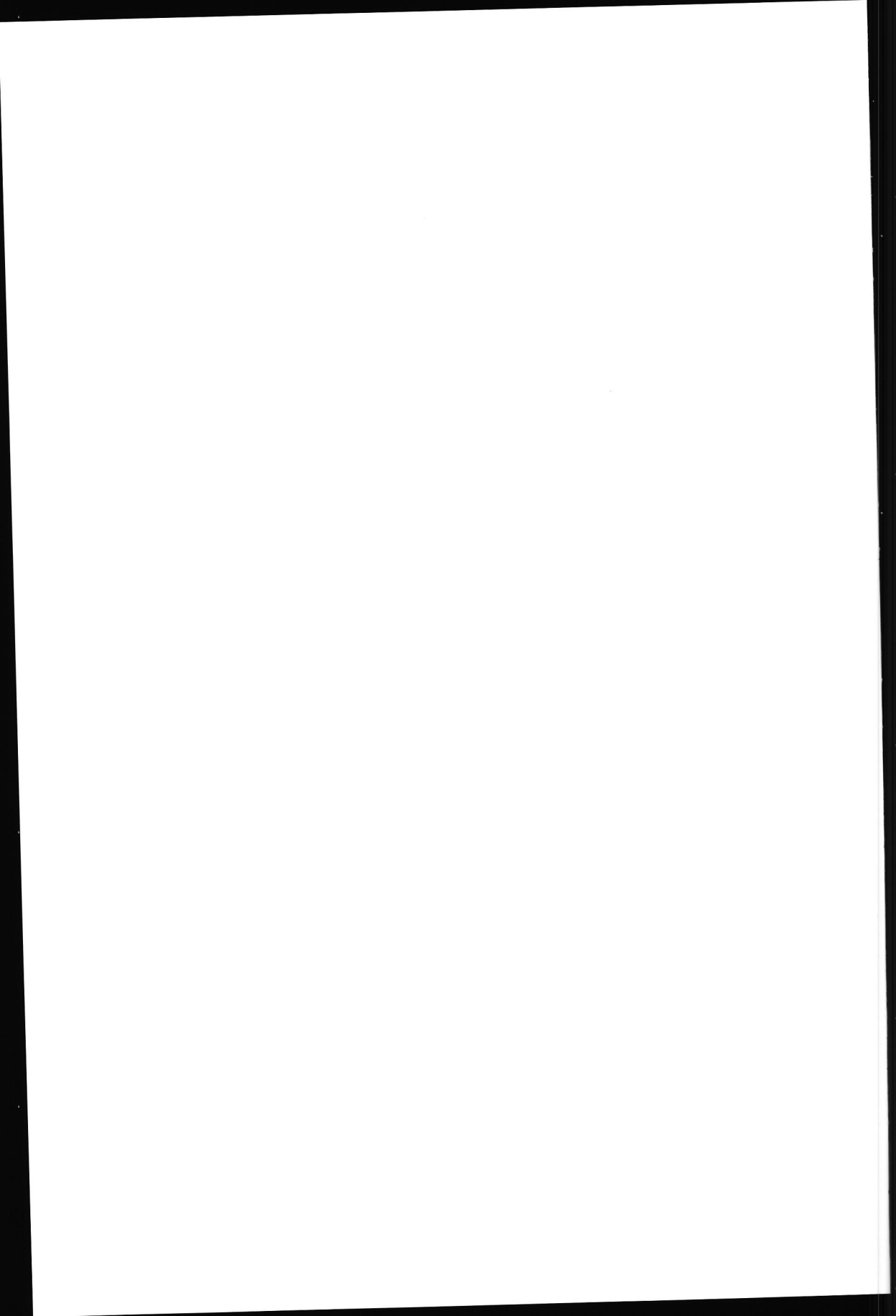
Chapter 6

Direct chemical ionization–mass spectrometric profiling of urine in premenstrual syndrome

A. C. TAS,*§ H. VAN DEN BERG,† J. ODINK,† H. KORTHALS,† J. T. N. M. THISSEN‡ and J. VAN DER GREEF*

Departments of * Toxicological Analysis and † Clinical Biochemistry, TNO–CIVO Institutes, Postbus 360, 3700 AJ Zeist, The Netherlands
‡ TNO Institute of Applied Computer Science, Postbus 15, 6700 AA Wageningen, The Netherlands

Abstract: In search of the pathophysiological background of premenstrual syndrome, direct chemical ionization–mass spectrometric profiling was applied to urine samples from 17 patients and 18 control subjects, collected on days 11 and 25 of the menstrual cycle. Oestrogenic compounds and hippuric acid were found to be involved in differences between these groups. Quotient profiles calculated for each subject from their profiles of days 11 and 25 showed cycle day-dependent group differences. Masses accounting for these differences between patients and control subjects indicate the involvement of unsaturated heterocyclic compounds in premenstrual syndrome.

Introduction

Premenstrual syndrome (PMS) is characterized by physical, psychological and be-havioural symptoms, occurring during the late luteal phase of the menstrual cycle and disappearing at the onset of, or during the first days of menstruation. After its first description [1], considerable effort has been directed towards the clinical characteriz-ation and pathophysiological delineation of this syndrome. The available information is mainly of a descriptive nature. Although there is strong evidence that it is related to ovarian function [2, 3], the pathophysiological background of PMS still remains unclear.

Direct chemical ionization–mass spectrometric (DCI–MS) profiling of biological fluids promises to be a powerful tool in the classification of diseases and in the investigation of their pathogenesis. It combines soft ionization with desorption and pyrolysis, and thus enables profiling of matrices containing compounds of different polarities and volatilities over a wide molecular weight range. So far, DCI–MS has successfully been used in the characterization of bacteria [4, 5], algae [6], biopolymers [7], wines and sherries [8], and viral proteins [9]. For evaluation of the complex DCI–MS profiles explorative multivariate data analysis techniques such as factor-discriminant analysis [5, 10] and quotient profile analysis [11] are needed.

§ To whom correspondence should be addressed.

In search of the (bio)chemical basis of PMS, we decided to apply DCI–MS profiling to urine samples of women suffering from PMS.

Experimental

Subjects

Patient selection and the study protocol have been described elsewhere [12–14]. In short, women volunteering in the study were recruited by advertisements and selected on the basis of strictly defined criteria. Classification was based on the increase of complaints in the luteal phase versus those in the follicular phase. Women were considered as being PMS patients if there was a significant difference in symptom scores, using a PMS questionnaire, for at least five out of 10 symptom clusters. The following symptom clusters were regarded as being important for the diagnosis of PMS: crying; depression; headache; irritability; lowered school/work performance; mood swings; painful breasts; feeling swollen or bloated; tension; weight gain. Diagnosis was confirmed by a gynaecologist, who excluded the occurrence of somatic pathology. Controls were matched for age, education or profession, and civil state. In the control group score differences were observed for not more than three of the 10 PMS symptom clusters. Seventeen PMS patients and 18 controls were investigated.

Sample collection

During days 11 and 25 of the menstrual cycle 24-h urine was collected into polyethylene bottles. Throughout the collection period the bottles were kept in a cool and dark place. No preservatives were used. Urine samples were stored at $-20°C$ until analysis.

Sample pretreatment

Urine samples were adjusted to pH 2 with 1 M HCl after centrifugation at 2000g for 5 min. A portion of 4–8 ml, depending on the concentration, was applied to C-18 columns (Betron Scientific, No. 607303) which were successively preconditioned twice with 2 ml methanol and twice with 2 ml 10 mM HCl. After absorption of the urine components the column was rinsed twice with 1 ml 10 mM HCl. The components were subsequently eluted with five portions of 0.5 ml methanol. Samples were lyophilized and then dissolved in 200 µl methanol.

Mass spectrometry

Direct chemical ionization–mass spectrometry was performed with a Finnigan MAT 8230 double-focusing mass spectrometer coupled to a Finnigan MAT SS300 data system. Ammonia was used as a reactant gas at an indicated source pressure of 53 Pa; the source temperature was set at 240°C. About 1 µl of the sample solution was applied to the DCI wire. Heating of the DCI wire was performed by linear current programming at a rate of 8 mA s^{-1}. The mass spectrometer operated at a resolution of 1000 and a scan rate of 0.7 s/decade. All samples were measured in duplicate. The spectra obtained in the range of desorption and pyrolysate formation were averaged to one profile; duplicate profiles were subsequently averaged to one individual profile. High-resolution measurements ($R = 10\ 000$) were carried out by applying the acceleration voltage scan technique, with tris-(pentafluoroethyl)-s-triazine as a reference compound.

Data analysis

The pattern recognition program ARTHUR [15] extended with routines for principal component-discriminant (PC-D) analysis and graphical rotation (FOM Institute, Amsterdam) [10], was used for data analysis, as well as home-written routines for BIPLOT facilities, for conversion of factor spectra to SS300 mass spectrum formate and for the calculation of quotient spectra [11].

Between-group differences were analysed by averaging individual profiles of days 11 and 25 to one individual profile. Interactions between group and cycle day (cycle day-dependent group differences) were analysed with quotient profiles [11], calculated by dividing the individual profiles of day 11 by those of day 25.

Averaged and quotient profiles were reduced to subsets of 70 variables, the highest Fisher weights being the selection criterion. The resulting patterns were normalized to total ion current for adjustment of differences in sample size, and then transformed by autoscaling (zero mean and unit standard deviation).

Subsequently, PC-D analyses were performed on the transformed data sets. For both data sets the first nine principal components with the highest variance contributions were used as input variables for the discriminant analysis. For display purposes two-dimensional plots were constructed by plotting the discriminant function against an arbitrarily chosen original variable.

Factor spectra obtained after graphical rotation [10], displaying correlations between masses and the discriminant function, were used to investigate the differences in patterns between PMS patients and controls.

The classification performances of the discriminant analyses were evaluated as follows: five subjects (two patients and two controls) were randomly selected from the training set and placed into a test set. Subsequently, feature selection and PC-D analysis were carried out as described above. The classification of the test set subjects was evaluated with a 1-Nearest Neighbour (1NN) routine [16]. These cycles were repeated 30 times, and the mean classification performance was determined.

Results

Group differences

Typical DCI–MS profiles for both categories were obtained by averaging all PMS (Fig. 1A) and control profiles (Fig. 1B). PC-D analysis, performed on the basis of the individual PMS and control patterns, yielded a discriminant function on which both groups showed only minor overlap (Fig. 2). 1NN classification of the test set subjects resulted in a mean correct classification of approximately 63%.

Factor spectra in the $0°$ (PMS) and $180°$ (controls) directions revealed some remarkable mass series (Fig. 3). In the $180°$ spectrum the series m/z 288 and 306 may correspond to oestrone ($[M+NH_4]^+$ ion) and estriol ($[M+NH_4]^+$ ion), respectively. The ions at m/z 290 may originate from oestradiol ($[M+NH_4]^+$ ion). In addition, the intensity found at m/z 304 may point to the corresponding ion of hydroxylated oestrone. These compounds are mainly present in urine as conjugates which decompose during heating of the DCI wire. The masses observed are indicative of relatively high concentrations of oestrogens in urines of the control group.

The intensity at m/z 194 in the factor spectrum in the $0°$ direction can be ascribed to the $[M+H]^+$ of hippuric acid methyl ester, which probably has been formed from the corresponding acid during extraction with and storage in methanol. The intensities at m/z
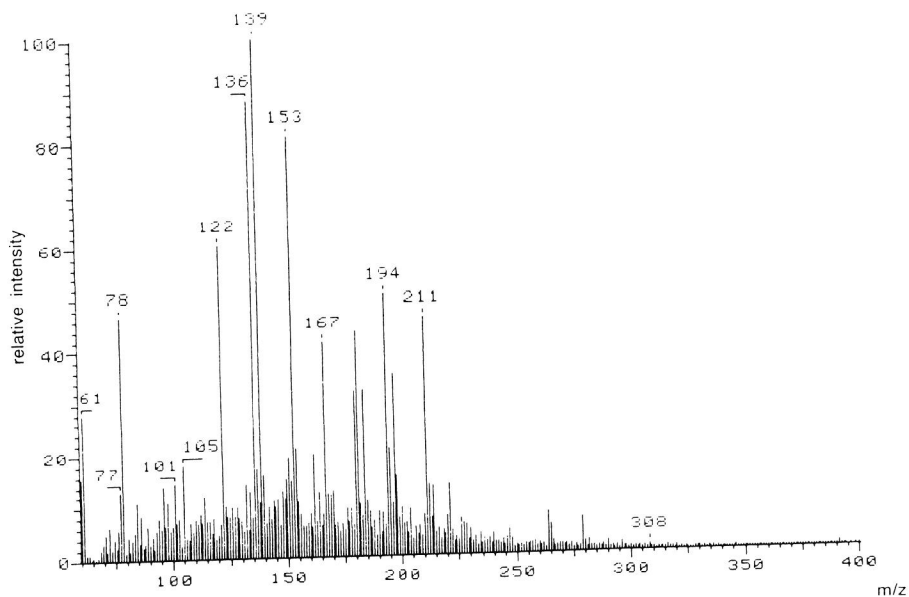
68

Figure 1A
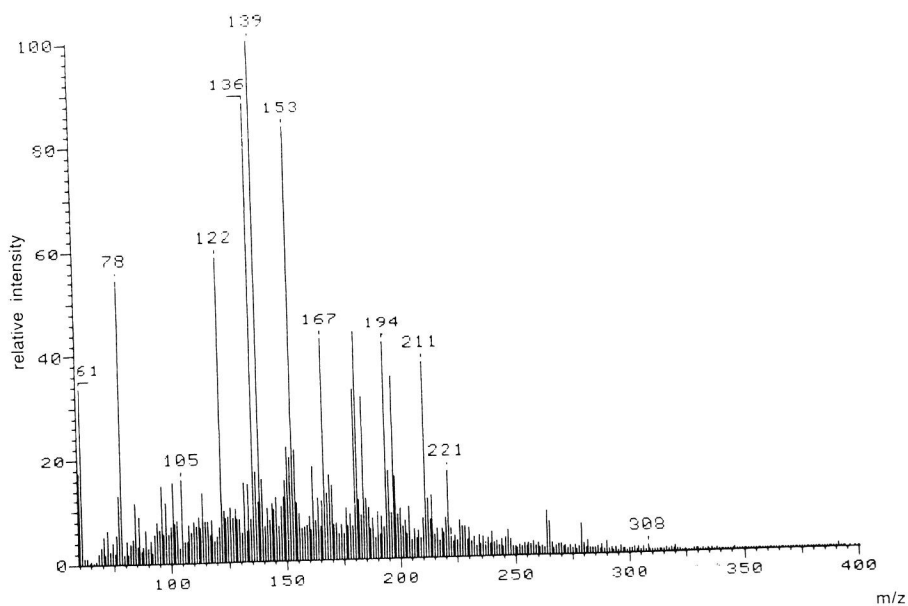Averaged DCI–MS urine profile of PMS patients.

Figure 1B
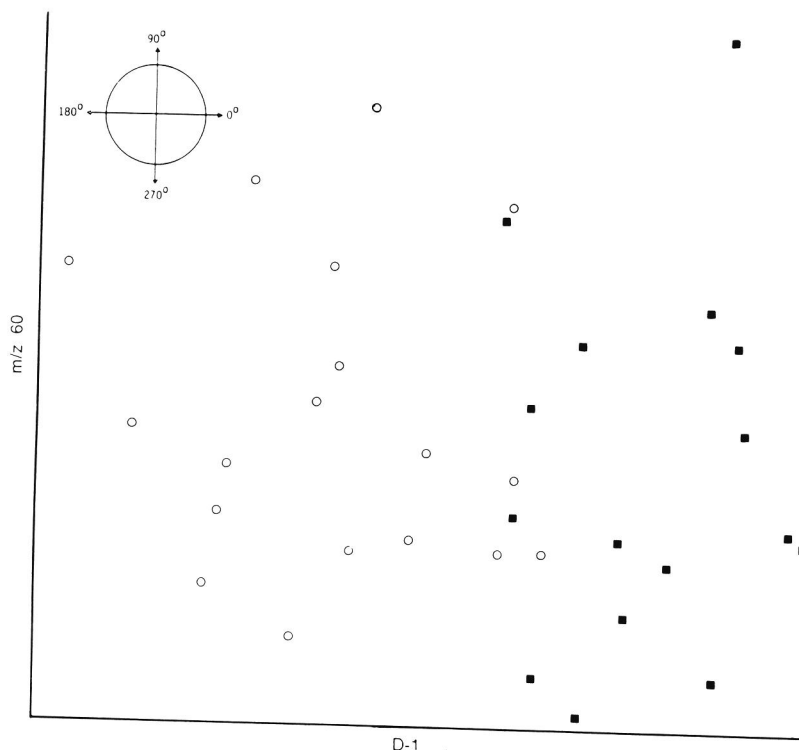Averaged DCI–MS urine profile of controls.

Figure 2
Object plot of a discriminant function versus a random variable introduced for creating a plane. ■, PMS patients; ○, controls.

195 and 196 may be caused by the ^{13}C and ^{18}O isotopic contributions of this compound, represented in the spectrum according to their correlations with the discriminant function. The mass at m/z 105 can also be ascribed to the same compound being a benzoyl fragment. This result indicates that this compound was present in relatively high amounts in urine of PMS patients.

A number of masses with comparable correlations with the D-axis in the 180° direction have not been identified.

Interaction between group and cycle day (quotient profiles)

To investigate cycle day-dependent differences between PMS patients and controls, the quotient spectrum approach was used [11]. PC-D analysis, applied to the quotient patterns of the control and PMS groups, yielded a discriminant function (Fig. 4), which separated both groups. The 1NN classification performance of the discriminant analysis was approximately 67%.

The factor spectra in the 0° and 180° directions (Fig. 5) revealed a series of masses partly differing from the analysis of the averaged profiles (Fig. 3). A substantial contribution to the spectrum was observed for masses in the higher mass range.

Some of these masses were investigated with high-resolution mass spectrometry providing the following probable elemental compositions: m/z 358.140, $C_{18}H_{20}N_3O_5$;
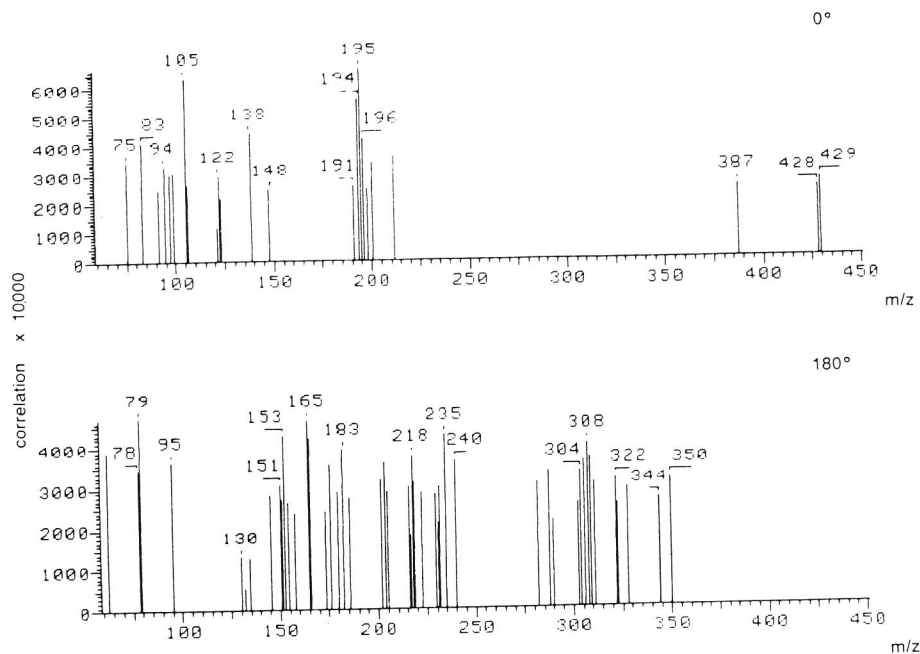
Figure 3
Factor spectra (mass-discriminant function correlations) in the 0° and 180° direction, going together with Fig. 2.

m/z 390.166, $C_{19}H_{24}N_3O_6$; m/z 401.173, $C_{18}H_{29}N_2O_6S$; m/z 402.167, $C_{20}H_{24}N_3O_6$; m/z 415.193, $C_{19}H_{31}N_2O_6S$; m/z 420.182, $C_{20}H_{26}N_3O_7$; m/z 436.191, $C_{20}H_{28}N_4O_7$.

Because all these compounds contain nitrogen and therefore probably have a higher proton affinity than NH_3, it is likely that preferentially the $[M+H]^+$ ions have been formed. Therefore, the actual elemental compositions contain one hydrogen less. Some chemically relevant differences appeared in the series of elemental compositions: m/z 436–420, NH_2; m/z 420–402, H_2O; m/z 390–358, CH_4O; m/z 415–401, CH_2. The elemental compositions point to highly unsaturated, probably aromatic, heterocyclic compounds.

Discussion

In general, the investigation and monitoring of biochemical processes requires a broad analytical approach, because of the complexity and coherence of the phenomena observed. Often a variety of chemical compounds strongly differing in chemical and physical properties is involved in these processes. Most analytical methods are designed for sensitive and specific detection of classes of compounds rather than for providing a general overview of complex processes. For the latter approach the application of profiling techniques can provide new leads to the understanding of pathogenesis. DCI–MS is a profiling technique, that is suitable for the detection of compounds over an extended molecular weight range and with different polarities.

In connection with a non-specific profiling technique, body fluid profiling for the investigation of unknown complex phenomena requires that a sample pretreatment
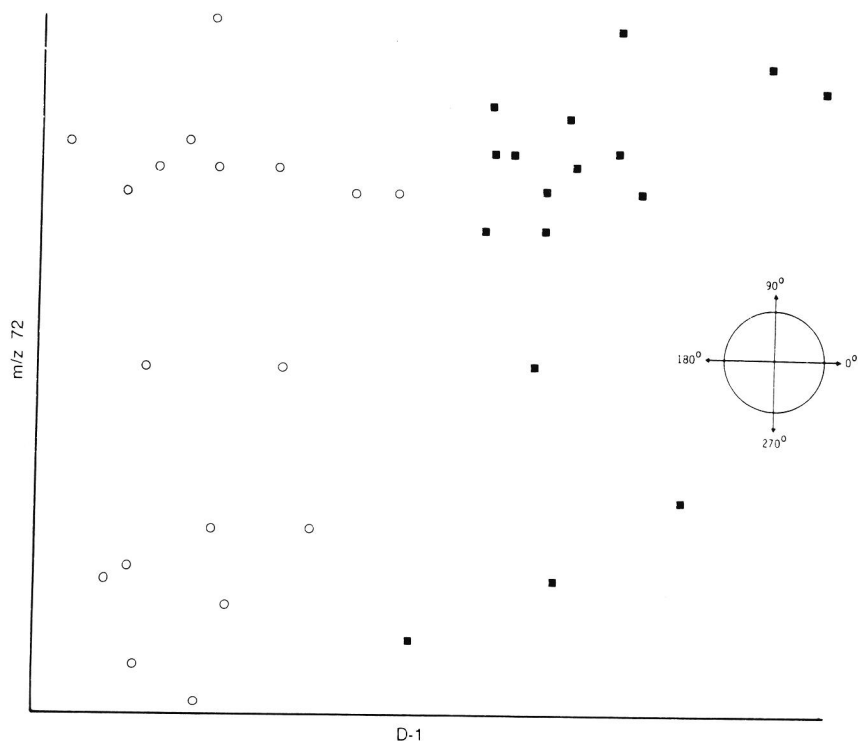
Figure 4
Object plot of a discriminant function, based on quotient profiles, versus a random variable introduced for creating a plane. ■, PMS patients; ○, controls.

method is not very compound-selective. The pretreatment method used was aimed at removing the bulk of salts and urea, while retaining most of the less polar compounds. Therefore it is not unlikely that losses of amphoteric compounds have occurred.

The DCI–MS pattern recognition approach was found to be successful in differentiating between the PMS and control groups, in spite of some overlap due to within-group differences in symptom intensity, symptom heterogeneity and other sources of variation, such as variations in cycle stage and dietary factors.

The main differences observed between PMS and control groups, independent of the cycle day, suggest the involvement of oestrogens and hippuric acid; levels of oestrogens were lower and levels of hippuric acid were higher for the PMS group. The first observation is in accordance with the subnormal peak oestrogen levels that have been reported for PMS patients in the follicular phase [17]. On the other hand, no consistent differences in levels for these compounds were found in the luteal phase [17]. In serum, no significant differences in oestradiol levels were observed between PMS patients and controls [13].

Higher concentrations of hippuric acid in the urine of PMS patients cannot be explained by known metabolic deviations and justify further investigation. The same holds for the coherent series of unsaturated, probably heterocyclic compounds, which were found to be important for differentiation of PMS and control quotient profiles.
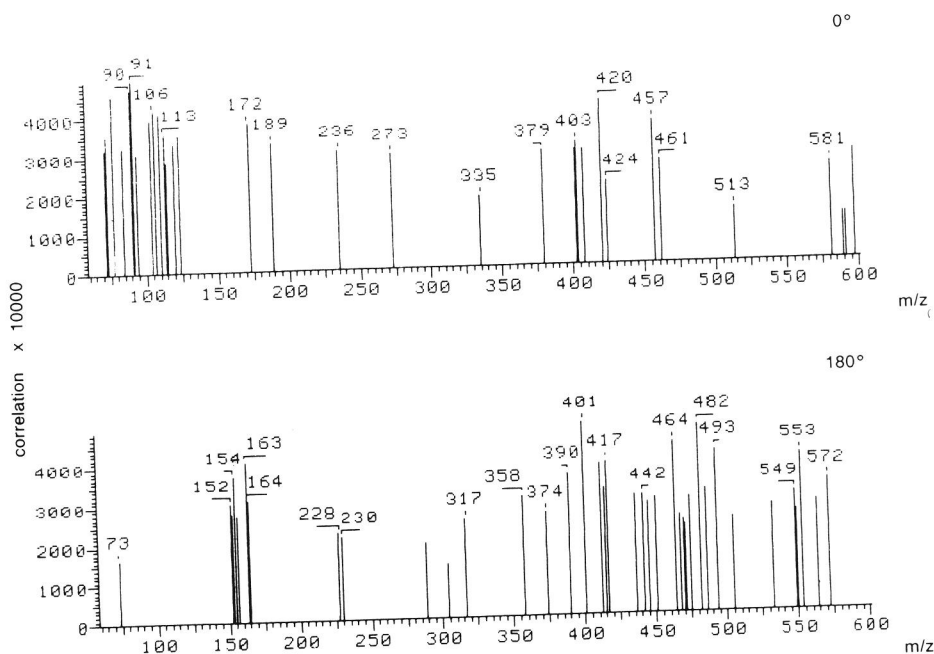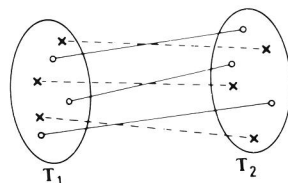
72

Figure 5
Factor spectra (mass-discriminant function correlations) in the 0° and 180° directions, going together with Fig. 4.

Figure 6
Cluster of objects in a two-dimensional space at two points of time. Although at both events categories are overlapping a clear separation based on trends is possible. The same objects are connected.

Quotient profiles, which can reflect changes in individual profiles with time, are suitable to investigate differences between PMS and controls due to menstrual cycle effects. The principle of this approach is explained in Fig. 6 on the basis of a hypothetical case in which two variables are observed as a function of time for two categories. Apparently, both at time T_1 and time T_2 there is an overlap of both categories, resulting in an incomplete differentiation. However, by focusing on "trend" differences, one can clearly see that all objects of class "X" have decreasing values with time, whereas the class "O" objects have increasing values. In multidimensional data comparable situations exist which can be investigated by this approach.

We conclude that DCI–MS profiling indeed enabled differentiation between women with and without PMS complaints. The profiling technique offers a broad analytic approach which might be useful in the classification of diseases and the investigation of their complex pathogenesis. Additional studies using specific methods focused on the compounds of interest, are required for interpretation of these analytical results.

References

[1] R. T. Frank, *Archs Neurol. Psychiat.* **26**, 1053–1057 (1931).
[2] R. L. Reid and S. S. C. Yen, *Am. J. Obstet. Gynecol.* **85**, 139 (1981).
[3] J. Bancroft and T. Backstrom, *Clin. Endocr.* **22**, 313–336 (1985).
[4] A. C. Tas, J. van der Greef, J. de Waart, J. Bouwman and M. C. ten Noever de Brauw, *J. Anal. Appl. Pyrol.* **7**, 249–255 (1985).
[5] A. C. Tas, J. de Waart, J. Bouwman, M. C. ten Noever de Brauw and J. van der Greef, *J. Anal. Appl. Pyrol.* **11**, 329–339 (1987).
[6] J. van der Greef, J. de Waart and A. C. Tas, *Proceedings of the Second Workshop of the COST 48 Subgroup 3: Biomass Conversion, Removal and Use of Nutrients* (J. de Waart and P. H. Nienhuis, Eds), pp. 34–36. TNO–CIVO, Zeist, The Netherlands, October (1988).
[7] A. C. Tas, H. B. Bastiaanse, J. van der Greef and A. Kerkenaar, *J. Anal. Appl. Pyrol.* **15**, 55–70 (1989).
[8] J. van der Greef, A. C. Tas, J. Bouwman and M. C. ten Noever de Brauw, *Adv. Mass Spectrom.* **10**, 1227–1228 (1986).
[9] A. C. Tas, J. Odink, M. D. Ferrari, W. Boogerd, L. Van Ekdom, A. C. B. Peters and J. van der Greef, in press.
[10] W. Windig, J. Haverkamp and P. G. Kistemaker, *Analyt. Chem.* **55**, 81–88 (1983).
[11] J. van der Greef and D. C. Leegwater, *Biomed. Mass Spectrom.* **10**, 1–4 (1983).
[12] H. M. van der Ploeg, *Person. Individ. Diff.* **8**, 95–100 (1987).
[13] H. van den Berg and J. Schrijver, CIVO-Report V. 85.174/140373 (1985).
[14] J. Schrijver, A. J. Speek and W. H. P. Schreurs, *Int. J. Vit. Nutr. Res.* **51**, 216–222 (1981).
[15] Pattern recognition package ARTHUR 81, Infometrix, Inc., Seattle, WA, USA.
[16] D. Coomans and D. L. Massart, *Analytica Chim. Acta* **138**, 153–165 (1982).
[17] L. Dennerstein, C. Spencer-Gardner, J. B. Brown, M. A. Smith and G. D. Burrows, *J. Psychosom. Obstet. Gynaecol.* **3**, 37–51 (1984).
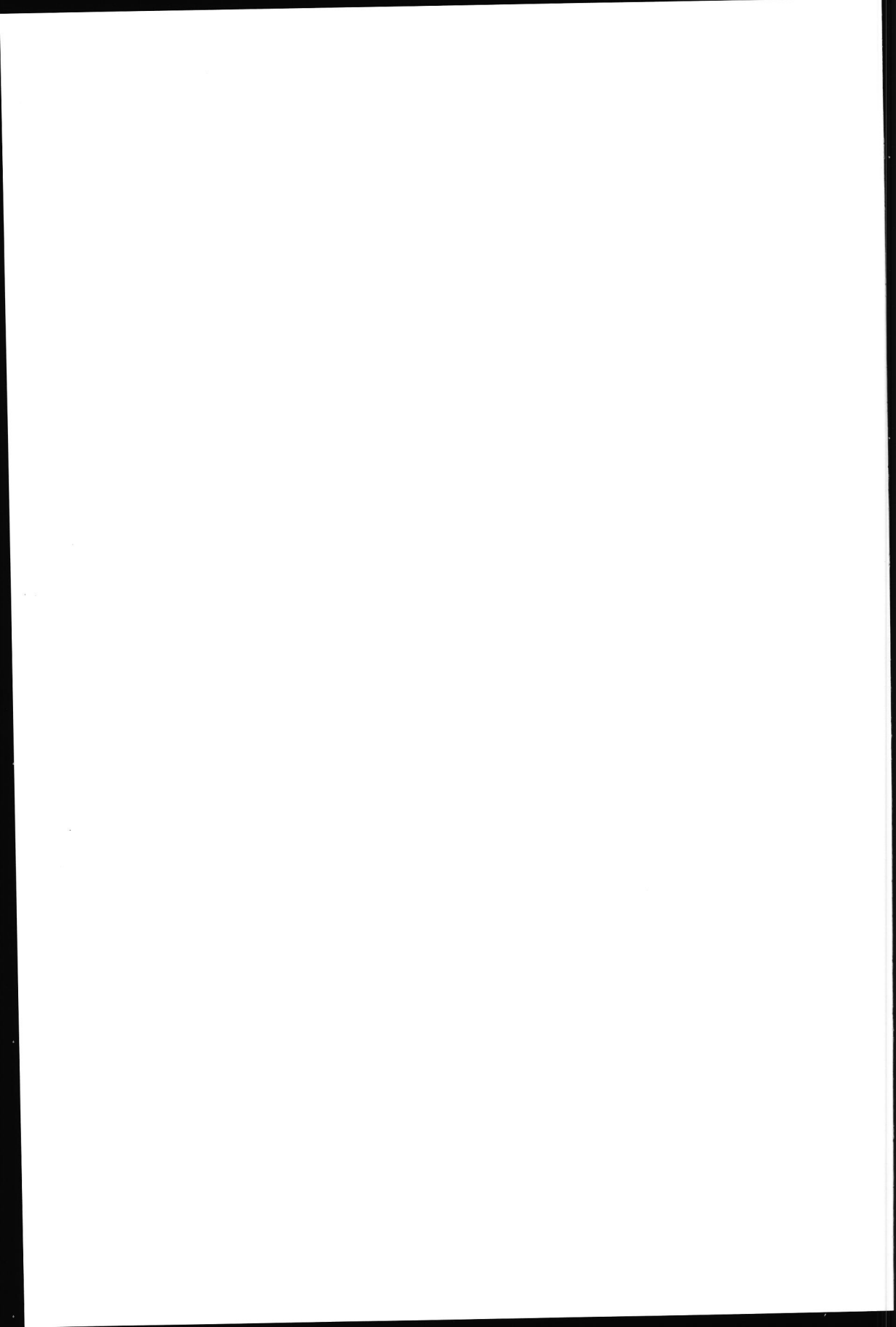
Chapter 7

CHARACTERIZATION OF VIRUS INFECTED CELL CULTURES BY PYROLYSIS—DIRECT CHEMICAL IONIZATION MASS SPECTROMETRY

A.C. Tas, J. Odink and J. van der Greef
TNO-CIVO Institutes, P.O. Box 360, 3700 AJ Zeist, The Netherlands

M.D. Ferrari, L. van Ekdom, A.C.B. Peters
State University Leiden, P.O. Box 9600, 2300 RC Leiden, The Netherlands

W. Boogerd
Slotervaart Hospital, Louwesweg 6, 1066 EC Amsterdam, The Netherlands

ABSTRACT

The supernatants of Vero cell cultures after infection with a herpes simplex virus or a poliomyelitis virus as well as a blank were analyzed by pyrolysis—direct chemical ionization mass spectrometry (Py—DCI/MS). Informative pyrograms were obtained and used for characterization of viral proteins by applying pattern recognition methods. Differentiation of viral proteins was evaluated by analyzing "blind" samples. Herpes viruses could be classified correctly but the observed differences between the blank and the polio virus supernatants were too small for reliable classification of the polio viruses. Purification of the samples seems to be a prerequisite for further studies. The potential value of Py—DCI/MS as a rapid noninvasive diagnostic method for viral meningoencephalitis is stressed.

INTRODUCTION

Early diagnosis of viral meningoencephalitis is essential for various reasons. First, differentiation from other conditions with a similar clinical picture, which requires a rapid and radically different therapy, is of direct importance. Second, an early identification of the pathogen will spare the patient extensive diagnostic procedures while, on the other hand, it may necessitate a rapid instillation of a specific antiviral treatment. This is especially true for herpes simplex encephalitis (HSE), the most frequent cause of (sporadically occurring) fatal encephalitis [1]. Particularly in the early phase of HSE clinical signs as well as laboratory tests are unspecific [2], whereas morbidity and mortality are significantly lowered when a specific antiviral treatment is started early in the course of infection [3]. At present, however, an early diagnosis of HSE still requires a brain biopsy [2, 4]. As an alternative, cerebrospinal fluid (CSF) has been used in attempts to develop non-invasive methods of HSE diagnosis. However, viral cultures from CSF usually are negative and a significant antibody response generally develops too late for diagnosis in the early phase of HSE [1, 2]. Assays for detection of either viral antigens by immunofluorescens [5, 6], radioimmunoassay (RIA) [7], enzyme-linked immunosorbent assay (ELISA) [8] and immunoblotting [9], or nucleic acids by dot-blot hybridization [10] have been developed, but their value has not been established in the clinical situation. Obviously, there is still a need for a rapid and accurate non-invasive diagnostic test for viral meningoencephalitis. A new approach for the detection of early excreted viral proteins in CSF could be pyrolysis–direct chemical ionization mass spectrometry (Py–DCI/MS) [11]. Recently, this technique was used for the characterization of bacteria [11-13] and fungi [14].

Py–DCI/MS is based on rapid heating of a sample inside the chemical ionization source of a mass spectrometer, which results in an optimum transfer of products of higher molecular weight and polar pyrolysis products to the ionization region. Moreover, application of chemical ionization conditions reduces the amount of mass spectrometric fragmentation. The rapid screening capabilities, about 10 minutes per sample, and the good reproducibility [11, 12] makes Py–DCI/MS attractive as a bioanalytical method.

Because the pyrograms contain a wealth of information, pattern recognition is an indispensable tool for optimum information extraction. The combination of PCs and discriminant analysis (PC-D) has been shown to be very effective in mass spectral data handling [12, 15, 16]. The large parameter/object ratio requires PCA as a data compression step prior to discriminant analysis. Strong correlation structures which are present in the pyrograms are effectively handled by PCA in the first stage and are in fact the basis for data compression. After the discrimination-oriented approach interpretation of the data is effectively realized

by factor rotation [17] in combination with high-resolution mass spectral measurement of important components.

In this study supernatants of infected cell cultures have been investigated by Py−DCI/MS. We have chosen herpes simplex virus type-I (HSV-I), the pathogen of HSE, for reasons outlined above, and poliovirus type I as an example of the enterovirus group, the most frequent pathogens of human viral meningitis.

EXPERIMENTAL

Cell cultures

Vero cells infected with either HSV-1 (McKrae; Herpes group), poliovirus type 1 (Sabin I; Polio group) and Vero cells without infection (Blank group) were cultured for 24 hours at 37°C, in EagleMEM containing (in mg/l): 264.9 $CaCl_2.2H_2O$, 400 KCl, 200 $MgSO_4.7H_2O$, 6800 NaCl, 2000 $NaHCO_3$, 158.3 $NaH_2PO_4.2H_2O$, 1000 glucose and 17.0 sodium phenol red. Ten Herpes, 9 Polio and 7 Blanc samples were cultured under these conditions. In addition, 6 coded samples belonging to one of these groups were cultured under the same conditions and analyzed blindly. After being centrifuged for 10 min at 1000 g the supernatants containing the released viral proteins and the major part of the substances mentioned were isolated and stored at −70°C. After lyophilization samples were suspended in 100 µl of methanol.

Sample preparation

One millilitre of supernatants were lyophilized overnight. Just before measurement, the remaining solids were suspended in 0.2 ml methanol.

Mass spectrometry

DCI pyrolysis was performed on a Finnigan-MAT 8230 mass spectrometer. Data were accumulated, stored and processed on a Finnigan-MAT SS300 data system. Ammonia was used as a reactant gas at an indicated source pressure of 53 Pa; the source temperature was set at 240°C. During heating of the DCI wire by linear current programming (8 mA/s) spectra were acquired over a mass range of m/z 60–1000, at a scan speed of 0.7 s/decade. Spectra acquired during pyrolysate formation were averaged to one profile. All samples were analyzed in duplicate over a period of two days; both series of analyses were performed in the same random order.

Pattern recognition

Data analysis was performed with the pattern recognition package ARTHUR [18], extended with discriminant analysis and graphic rotation routines developed by the FOM institute, Amsterdam, as well as home-written routines for data extraction from spectra, input in ARTHUR, and display of objects and variables in comparable plots (BIPLOT) [19]. Prior to preprocessing and data analysis duplicate measurements were averaged to one profile.

In view of the available computer capacity, each spectrum was reduced to a pattern of 70 mass intensities, using Fisher weights [20] as a selection criterion. Differences in sample size were adjusted for by normalization to total ion current. Prior to PC-D analysis the data set was transformed by autoscaling (features with zero mean and unit standard deviation).

The classification stability was evaluated by six blind samples, classified according to a KNN procedure based on the discriminant scores.

RESULTS AND DISCUSSION

In the total ion current profile of the Py–DCI/MS experiments with the viral samples a second strong peak was observed. Investigation of this phenomenon revealed a series of masses at m/z 81/83, 98/100, 139/141/143, 156/158/160, 197/199/201 and 255/257/259. From the observed isotopic patterns it could be concluded that the observed series is in accordance with the general composition Na_nCl_{n-1} with n = 2, 3, 4 and 5. For n = 2 and 3 NH_3 adducts were observed as well. The salt impurities dominate an averaged profile based on all spectra. Sodium chloride could not be removed by dialysis in a sufficiently selective way, compounds of lower molecular weight being removed as well. The latter was found to have a detrimental effect on differentiation.

The DCI pyrograms for the Herpes (A), Polio (B) and Blank (C) viral samples are given in Fig. 1. The profiles contain information in the mass range m/z 60 - 300. At masses above m/z 200 the information content is low. Besides the salt cluster peaks mentioned a peak at m/z 180 is observed, probably originating from glucose ($[M + NH_4-H_2O]^+$ ion), which was a constituent of the culture medium. Obviously, the averaged profiles are very similar, and pattern recognition analysis is mandatory to extract relevant information.

Three categories were defined containing 10 (Herpes), 9 (Polio) and 7 (Blank) profiles. PC-D analysis applied to the 26 patterns, obtained from the pyrograms by selecting 70 masses with the highest Fisher weight, resulted in two discriminant functions (Fig. 2), which show the position of the patterns in this two-dimensional space. Examination of this plot reveals that the Herpes group is quite well separated from both other groups and that differentiation between the Polio group
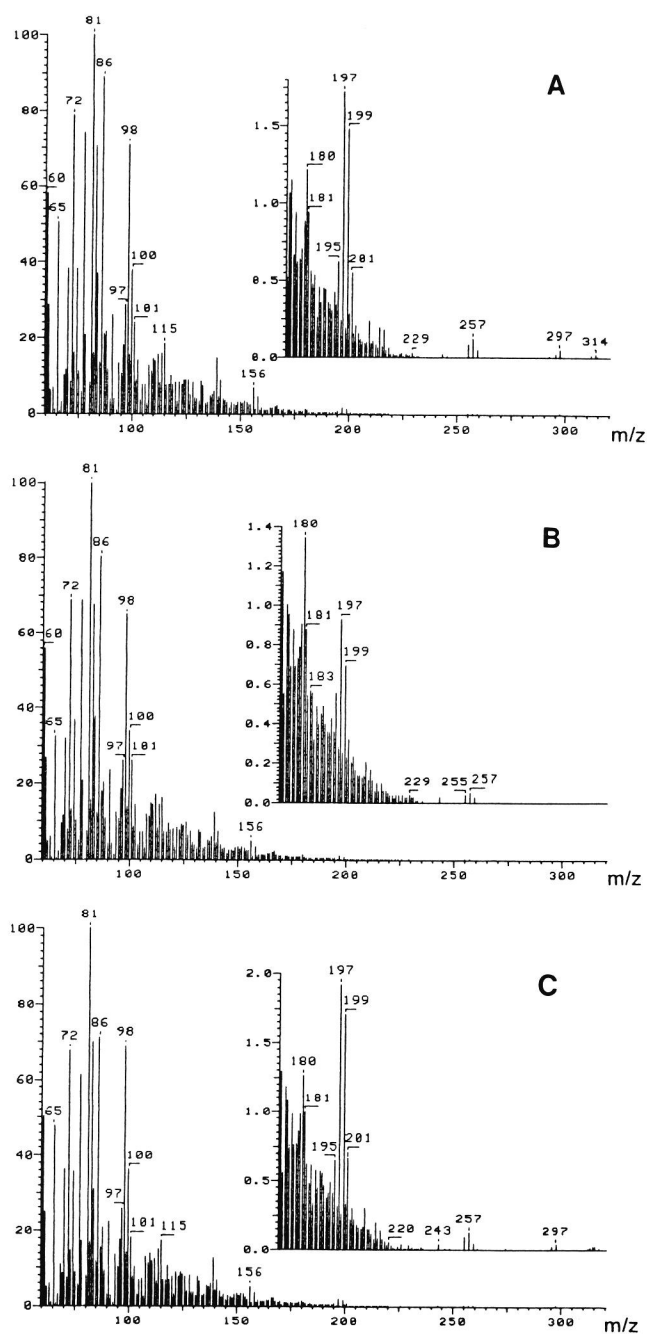
Fig. 1   Pyrograms of lyophilized supernatants of
          A: Herpes simplex virus type 1
          B: Polio virus type 1
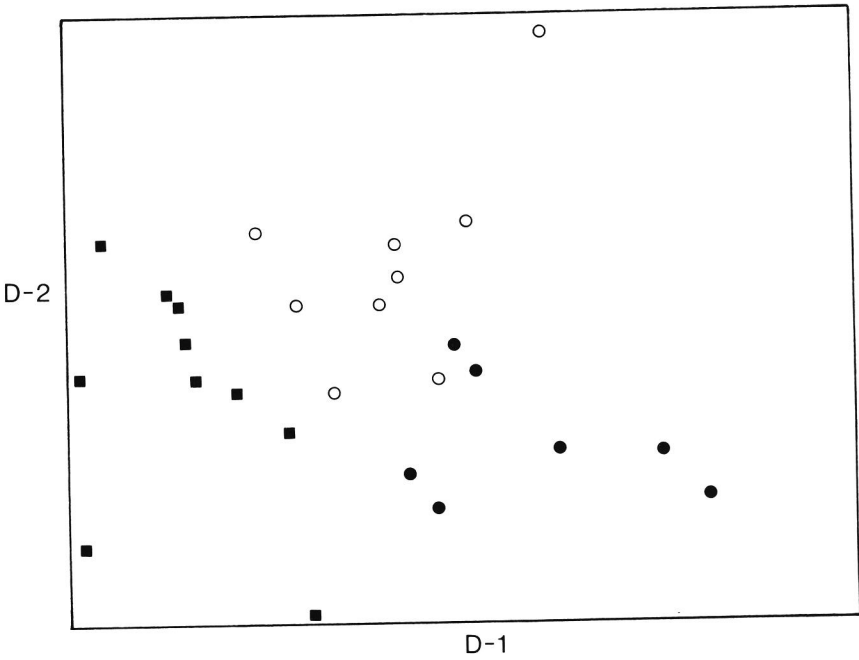   C: Vero cells without infection

Fig. 2   Discriminant scores plot of Herpes (■), Polio (o) and Blank cells (●) without infection
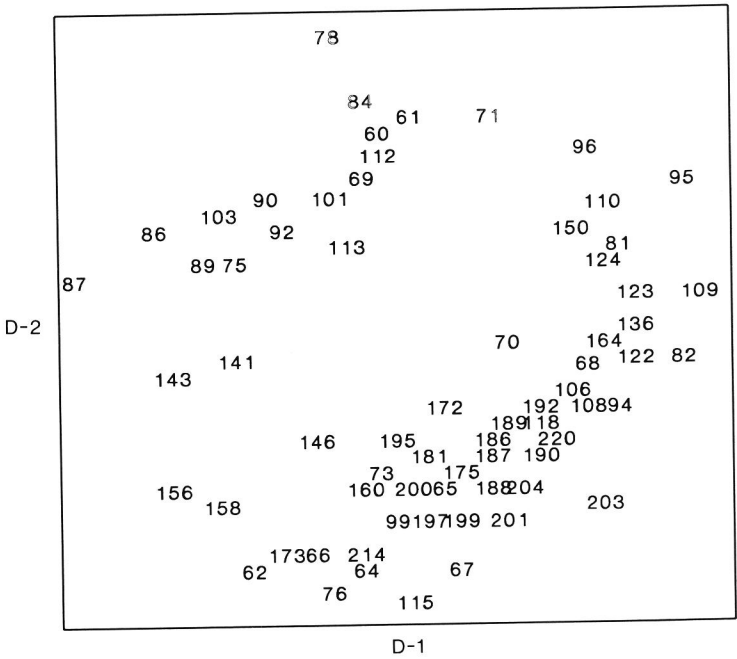

Fig. 3   Discriminant variable plot, displaying the masses according to their correlation coeffi-
cients with the discriminant axes (cf. Figure 2)

and the Blank group is much less pronounced. Most of the discriminative power is found in the D-1 direction and only a limited discrimination, mainly between the Polio and Blank groups, is observed in the D-2 direction.

In Fig. 3 a variable plot (BIPLOT) of the PC-D analysis shows the positioning of the variables (masses) in the discriminant space, according to their correlation coefficient with the D-axes. Different clusters are observed among which a salt cluster containing peaks at m/z 141 and 156/158. Apparently the Herpes samples contain more salt impurities. In the broad band located in the Blank object region interesting correlations are found. Besides the strong isotopic correlations as observed throughout the whole plot, series of peaks differing by 14 mass units, possibly CH_2 groups, are observed: m/z 62/76, 68/82, 94/108/122/136, 95/109/123, 96/110, 146/160, 175/189, 181/195, and 172/186/200/214. Discrimination between Herpes samples and the other two categories is to a large extent caused by the mass intensities at m/z 86/87 and 95/109, the former two mass intensities being more abundant in the Herpes profiles and the latter ones in the Blank profiles.

All blind samples were uniformly predicted by 2NN to 6NN and were classified as Herpes (2), Polio (1) and Blank (3). The two Herpes samples and the Polio sample were located correctly within the training set clusters in the discriminant plot whereas the three Blank samples were positioned on the border of the blank cluster just below the Polio training set. Two of the three Blank samples were mistaken for Polio samples. Discrimination between the Herpes group and both other groups seems to be stable, as illustrated by the correct classification of the Herpes samples and the fact that none of the other ones were misclassified as a Herpes sample.

DCI/MS profiling in combination with pattern recognition methods is capable of producing and extracting relevant information from such complex matrices as supernatants of cell cultures. Although simple extrapolation to human disease is not possible, these results suggest the potential applicability of Py–DCI/MS for non-invasive diagnosis of human meningo-encephalitis.

REFERENCES

1. J. Boos and M.M. Esiri, Sporadic Encephalitis I. in Viral Encephalitis, Pathology, Diagnosis and Management, p. 55. Blackwell Scientific Publications, London (1986).

2. A.J. Nahmias, R.J. Witley, A.N. Visintini, et al., J. Infect. Dis. 145 (1982) 829.

3. B.K. Skoldeberg, L. Alestig, A. Burman, et al., Lancet II (1984) 707.

4. D.F. Hanley, R.T. Johnson and R.J. Whitley, Arch. Neurol. 44 (1987) 1289.

5. A.D. Dayan, M.I. Stokes, Lancet I (1973) 177.

6.   R.H. Boerman, A.C.B. Peters, M. Verhey et al., J. Neurol. Sci., in press.

7.   A.B. Chen, T. Ben-Porat, R.J. Whitley and A.S. Kaplan, Virology, 91 (1978) 234.

8.   R.M. Coleman, P.D. Bailey, R.J. Whitley, et al., J. Virol. Methods, 7 (1983) 117.

9.   F.D. Lakeman, J. Koga and R.J. Whitley, J. Infect. Dis. 155 (1987) 1172.

10. R.H. Boerman, A.K. Raap, M. van der Ploeg, et al., Lancet II (1987) 94.

11. A.C. Tas, J. van der Greef, J. de Waart, J. Bouwman and M.C. ten Noever de Brauw, J. Anal. Appl. Pyrolysis, 7 (1985) 249.

12. A.C. Tas, J. de Waart, J. Bouwman, M.C. ten Noever de Brauw and J. van der Greef. J. Anal. Appl. Pyrolysis, 11 (1987) 329.

13. J. van der Greef, A.C. Tas and M.C. ten Noever de Brauw, Biomed. Environ. Mass Spectrom. 16 (1988) 45.

14. A.C. Tas, H.B. Bastiaanse, J. van der Greef and A. Kerkenaar, J. Anal. Appl. Pyrolysis, 14 (1989) 309.

15. W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem. 55 (1983) 81.

16. L.V. Vallis, H.J. McFie and G.S. Gutteridge, Anal. Chem. 57 (1985) 704.

17. H.L.C. Meuzelaar, W. Windig, A.M. Harper, S.M. Huff, W.H. McClennen and J.M. Richards, Science, 226 (1984) 268.

18. Pattern recognition package ARTHUR 81, Infometrix, Inc. Seattle.

19. K.R. Gabriel, Biometrika, 58 (1971) 453.

20. W. Eshuis, P.G. Kistemaker, H.L.C. Meuzelaar, J. Anal. Appl. Pyrolysis, 1 (1977) 199.

Chapter 8

PYROLYSIS–DIRECT CHEMICAL IONIZATION MASS SPECTROMETRY OF THE DIMORPHIC FUNGUS *CANDIDA ALBICANS* AND THE PLEOMORPHIC FUNGUS *OPHIOSTOMA ULMI*

A.C. TAS [1]*, H.B. BASTIAANSE [2], J. VAN DER GREEF [1] and A. KERKENAAR [2]

[1] *TNO-CIVO Food Analysis Institute, P.O. Box 360, 3700 AJ Zeist (The Netherlands)*
[2] *TNO Institute of Applied Chemistry, P.O. Box 108, 3700 AC Zeist (The Netherlands)*

ABSTRACT

Pyrolysis–direct chemical ionization mass spectrometry was applied to investigate dimorphism-related differences in the fungi *Candida albicans* and *Ophiostoma ulmi*. In *O. ulmi* cell walls the yeast form contained more deoxyhexose (probably rhamnose) residues. This is in contrast to the results described for the related fungus *Sporothrix schenckii*. In *C. albicans* differences observed in the mass pyrograms of the yeast and hyphal forms decrease after application of the sterol biosynthesis inhibitor lombazole. This is in accordance with the observation that both forms tend to converge to a more common morphology. Besides the effect of the inhibitor on the cell biopolymer composition the primary effect on sterol biosynthesis could also be detected from the mass pyrograms. The presence of higher methylated sterols together with lower levels of ergosterol in spectra of the treated forms reflects the inhibition of the C-14 demethylation process by lombazole.

Candida albicans; direct chemical ionization; discriminant analysis; fungi; mass spectrometry; *Ophiostoma ulmi*; pyrolysis.

INTRODUCTION

Pyrolysis–direct chemical ionization mass spectrometry is an important tool for the investigation of biopolymers. The method allows direct measurements on these polymers without extended sample pre-treatment. A number of pyrolysis methods has been used in this area the most important ones being Curie-point pyrolysis with low-voltage electron impact ionization [1,2], laser pyrolysis [3], oven pyrolysis with electron impact ionization [4], temperature-programmed oven pyrolysis and Curie-point pyrolysis with field ionization (FI) detection [5]. The latter approach, using a soft ionization method, has the advantage of inducing only limited mass spectrometric

fragmentation of ionized compounds [6]. Investigations carried out with this method showed the advantages of detecting larger molecules in pyrolysates, affording more information about the original composition of biopolymers [7].

The recently introduced pyrolysis–direct chemical ionization mass spectrometry (Py–DCIMS) also is a potentially powerful method in biopolymer research [8]. Pyrolysis in the source guarantees an optimal pyrolysate transfer to the ionization region where relatively soft chemical ionization with ammonia results in detection of larger molecules, mainly in the form of their $[M + H]^+$ and $[M + NH_4]^+$ ion.

Application of this method to bacteria [9,10], wines and sherries [11] showed that polar and relatively non-volatile compounds such as diglycerides, monoglycerides and monosaccharides can be detected. Information on the disaccharide level could also be obtained. Important additional advantages of the method are the ease of operation, the high sample throughput of about 6 min/sample and the high reproducibility [8].

This method was used for the investigation of the dimorphic fungus *Candida albicans* and the pleomorphic fungus *Ophiostoma ulmi*. *C. albicans* is a dimorphic fungus which can grow in either a budding yeast or a hyphal form, the latter being the most active form in infections [12]. A number of chemically related imidazole derivatives, sterol biosynthesis inhibitors (SBIs), are known to affect cell membrane functions resulting in inhibition of the yeast–hyphal mutual conversion. Fluorescence microscopy of yeast cells stained with diethanol shows an irregular deposition of chitin and frequent absence of septa in the cell walls of *C. albicans* cultured in the presence of an SBI [12]. An important aspect of the SBI-induced inhibiton [13] of hyphal cell growth is the formation of clustered yeast-like cells, a less pathogenic form of this fungus.

The effect of lombazole, an SBI, was illustrated by gas chromatographic comparison of sterol profiles in untreated and treated *C. albicans* showing the inhibition of sterol C-14 demethylation, which resulted in diminished ergosterol contents with a concomitant accumulation of sterols retaining the C-14 methyl group. Both aspects of the effects of lombazole (Fig. 1), the structure of the cell walls and the sterol profile, were investigated by Py–DCIMS in combination with factor-discriminant analysis [9,14]. In addition, the influence of some cultivating media on cell composition was examined.
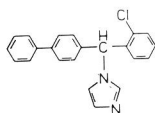

Fig. 1. Structure of lombazole.

The second part of this pyrolysis study concerned the detection of differences between the cell walls of the yeast form of *O. ulmi* and those of the hyphal form. This fungus, the cause of Dutch elm disease, can appear under certain growth conditions either as budding cells or as hyphae [15,16]. Both forms are important in pathogenesis: the budding form is passively transported through individual xylem vessels, while the pathogenic hyphal form propagates from one vessel to another by disrupting vessel walls. In *S. schenckii*, a fungus with similar dimorphism-related differences, a higher level of rhamnose residues was observed in the hyphal form [17,18]. Py–DCIMS can be a suitable technique for determining such differences in *O. ulmi* cell walls.

Because of the similarities in chemical structure between *O. ulmi* cell walls and rhamnomannan, the carbohydrate part of a toxic glycopeptide, the latter was also subjected to Py–DCIMS. The glycopeptide is another factor in the elm disease pathogenesis [19].

EXPERIMENTAL

Pyrolysis–mass spectrometry

Py–DCIMS was performed on a Finnigan-MAT 8230 mass spectrometer coupled to a Finnigan-MAT SS300 data system. Ammonia was used as reactant gas at an indicated source pressure of 53 Pa; the temperature of the ion source was held at 240 °C. During heating of the DCI wire by linear current programming (8 mA/s) spectra were acquired over the mass range m/z 60–1000, at a scan speed of 0.7 s/decade. Spectra obtained during pyrolysate formation were averaged to one profile. All isolates from whole cells (*C. albicans*) and cell walls (*O. ulmi*) were analysed in triplicate and duplicate, respectively.

Data analysis

In view of the available computer capacity, each spectrum was reduced to a subset of 77 mass intensities. Fisher weights were used as selection criterion; triplicate (*C. albicans*) and duplicate (*O. ulmi*) measurements were defined as categories. Differences in sample size were corrected by normalization to 100% total ion current. The data sets were transformed by autoscaling to perform principal component-discriminant analysis on a 'correlation about the mean' matrix [14].

Data analysis was performed with the pattern recognition package ARTHUR [20]. This package was extended with discriminant analysis and graphical rotation routines developed by the FOM Institute, Amsterdam as well as home-written routines for conversion of spectral data into ARTHUR

format, for conversion of factor spectra into SS300 spectrum format and for construction of comparable plots of objects and variables (BIPLOT routines) [10,21]. Factor spectra, calculated after graphical rotation [14], represent the correlations between the selected masses and the rotated discriminant functions.

Culturing conditions

The identity and growth conditions of *C. albicans* as well as the source of lombazole and its applications have been described previously [22]. The organism has been cultured in its yeast form in Kimmig's broth [22], mainly in its hyphal form in Eagle's medium [12], and in both morphologies in SSV medium [23]. Separation of the morphologic forms was carried out by microfiltration as described by Ghannoum et al. [23]. The fungus was cultured in these media in the absence or presence of 5 mg/l lombazole (Fig. 1) with the exception of the hyphal form cultured in SSV, which was not subjected to treatment with this SBI.

O. ulmi, isolate H6, a gift from D.M. Elgersma, was maintained on malt agar slants [24]. Induction of the mycelial state and the yeast morphology was performed according to the nutritional parameters defined by Kulkarni and Nickerson [16]. Cells were harvested after 18 hours, washed, and cell walls isolated according to the procedure of Gorin and Spencer [17] after separation of the yeast cells and hyphae as described by Dewerchin and Van Laere [25]. The phytotoxic glycopeptide of *O.ulmi* [19] was a gift from R.J. Scheffer [26].

Chemical standards

$(1 \rightarrow 4)$-α-Glucan (amylose) was obtained from Avebe; $(1 \rightarrow 6)$-α-glucan (dextran), $(1 \rightarrow 4)$-β-glucan (cellulose) and $(1 \rightarrow 3)$-β-$(1 \rightarrow 4)$-β-glucan (β-glucan) were purchased from Pharmacia Fine Chemicals, Fluka and Biocon, respectively.

RESULTS

Candida albicans

Under the ammonia chemical ionization conditions applied a great variety of pyrolysis products are ionized. The more polar compounds are reactive, but also relatively apolar compounds such as olefins can be detected. Protonated and ammoniated ions are frequently formed while, in the case of aromatic compounds, charge exchange reactions may take place.

Fragmentation, which is reduced under these conditions, often occurs in the form of losses of neutral molecules. For example, in the case of
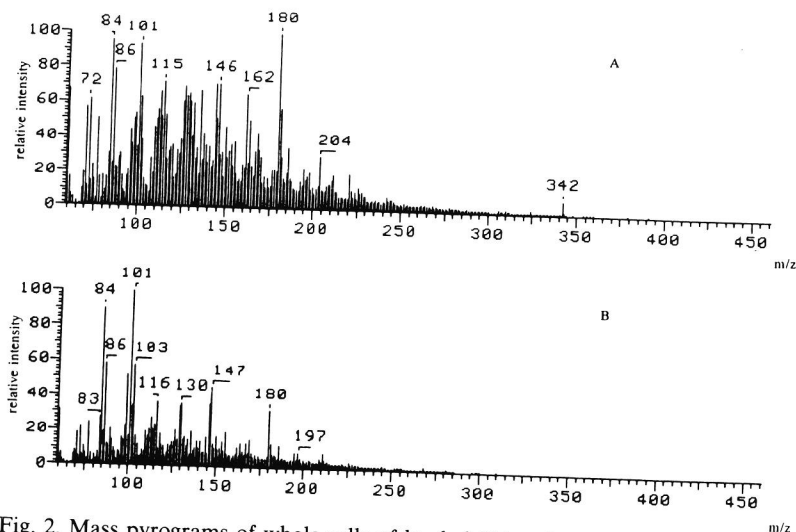
Fig. 2. Mass pyrograms of whole cells of hyphal (A) and yeast (B) form of *C. albicans*.

polyhydroxy compounds such as saccharides loss of water molecules occurs frequently.

In the mass pyrograms of whole cells of the yeast and hyphal forms of *C. albicans* cultured in Eagle's medium a number of components can be recognized (Fig. 2A and B). Hexose residues are indicated in the mass pyrograms by the mass series m/z 180, 162 and 144, being the $[M + NH_4 - H_2O]^+$, $[M + NH_4 - 2 H_2O]^+$ and $[M + NH_4 - 3 H_2O]^+$ ions, respectively; deoxyhexose residues by m/z 164 and 146, the $[M + NH_4 - H_2O]^+$ and $[M + NH_4 - 2 H_2O]^+$ ions respectively. These compounds are more abundant in the hyphal form (Fig. 2A) than in the yeast form. An additional mass series can be observed in higher intensities in the pyrogram of this form: m/z 221, 204, 186 and 168. The series points to N-acetylglucosamine residues originating from chitin in the form of the $[M + NH_4 - H_2O]^+$, $[M + H - H_2O]^+$, $[M + H - 2 H_2O]^+$ and $[M + H - 3 H_2O]^+$ ions respectively. This observation was substantiated by the pyrolysis of shrimp chitin.

The ion in the higher mass range at m/z 342 is caused by dihexose units ($[M + NH_4 - H_2O]^+$ ion) and indicates that the mass pyrograms provide information on disaccharides.

Principal components-discriminant analysis applied to the mass pyrograms of the various forms of *C. albicans* was used to investigate effects of morphology, medium conditions and SBI. The first four discriminant functions obtained, showed $\lambda/(1 - \lambda)$ values of 44, 16, 7 and 4, representing 21%, 10%, 11% and 11%, respectively, of the variance contained in the selected variables. This multivariate analysis of the isolates of *C. albicans* cultured in three media, with and without SBI, revealed indeed three main

88
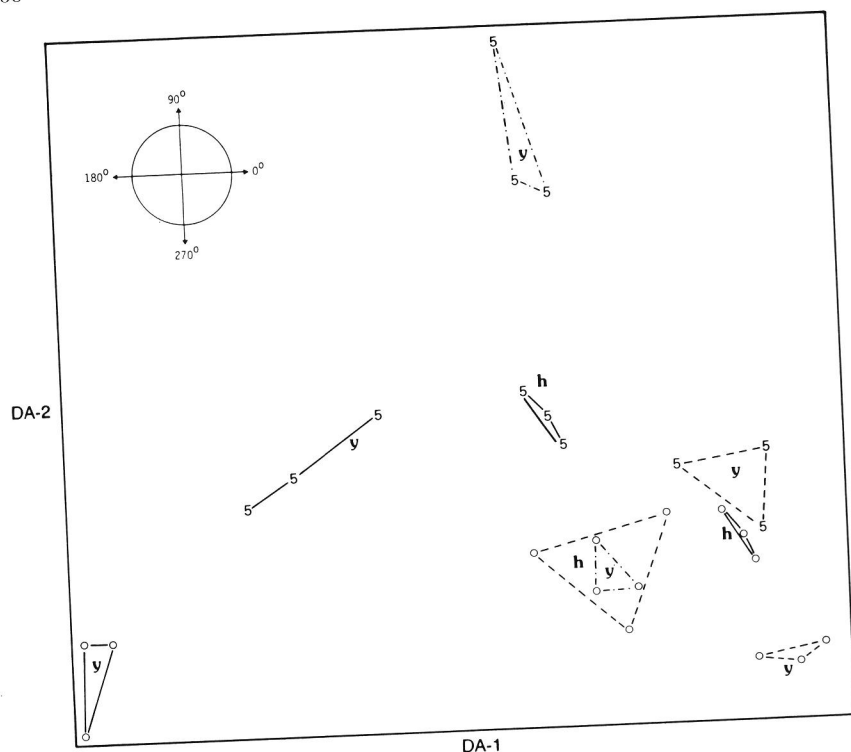
Fig. 3. Discriminant plot of yeast (y) and hyphal (h) forms of *C. albicans*, cultured in Eagle's (———), SSV (— — —) and Kimmig's (·–·–·) medium, without (○) and with (5) 5 μg/ml lombazole.

tendencies in the data, which especially could be observed in the plot of the first two discriminant functions (Fig. 3): the effects of the media on either form, the differences between yeast and hyphal forms within the same media, and the effect of the SBI on the position of the triplicates in the plot. When attention is focused on forms cultured in Eagle's medium the main difference between the yeast and hyphal forms cultured without SBI is found along the first discriminant axis (0° and 180°). The factor spectrum of this axis in the 0° direction is clearly dominated by hexose and deoxy-hexose masses, in accordance with the direct observations made on the mass pyrograms of both forms cultured in Eagle's medium. Differences observed for these forms clearly decrease for the SBI-treated ones. Apparently both forms tend to converge to a similar morphology (cf. Eagle's 0 and 5). This observation is in agreement with the Nomarski interference, scannning electron microscopy, and phase contrast microscopy of the yeast and hyphal form, respectively, in the absence or presence of an SBI. Both forms yield clustered yeast-like cells after treatment with an SBI [12]. The effect of SBI treatment is also reflected along the second discriminant axis (90° and
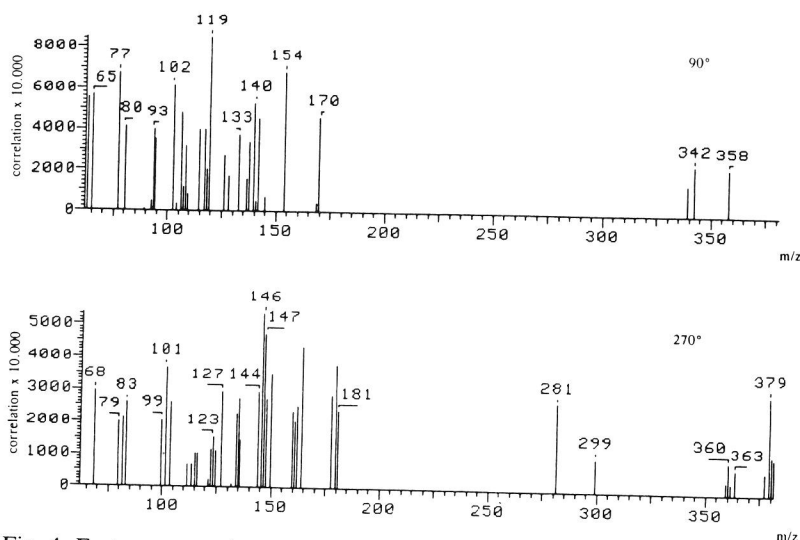
Fig. 4. Factor spectra (mass-discriminant function correlation) after counterclockwise rotation of the first discriminant function over 90° and 270° (corresponding to Fig. 3).

270°). Apparently the size of the effect is dependent on the medium, being most pronounced for the yeast form cultured in Kimmig's medium.

Analysis of the factor spectrum in the 270° direction reveals, among others, intensities at m/z 164 and 146 which point to the $[M + NH_4 - H_2O]^+$ and $[M + NH_4 - 2 H_2O]^+$ ions of deoxyhexose (probably rhamnose) residues, present in higher contents in strains cultured without the SBI (Fig. 4). The same holds for hexose residues indicated by the $[M + NH_4 - H_2O]^+$, $[M + NH_4 - 2 H_2O]^+$ and $[M + NH_4 - 3 H_2O]^+$ ions at m/z 180, 162 and 144. In the higher mass range the m/z value 379 is a characteristic indicator of the $[M + H - H_2O]^+$ ion of ergosterol, present in larger amounts in the uninhibited strains. This observation was confirmed by measurement of the standard ergosterol. The mass pyrograms apparently contain information on cell biopolymers as well as sterols.

The factor spectrum in the 90° direction contains signals at m/z 154 and 170 and in the higher mass range at m/z 342 and 358 (Fig. 4). These signals can be ascribed to lombazole and its metabolites: m/z 154 and 170 to a biphenyl and hydroxylated biphenyl fragment; m/z 342 and 358 to dehydrolombazole and its hydroxylated analogue. Especially the assignment of the mass signal at m/z 342 is tentative, due to interference with the nominal mass of dihexose units.

A close examination of the mass pyrograms of the untreated and SBI-treated yeast cells from Kimmig's medium shows the influence of the SBI on the sterol pattern (Fig. 5). In the treated yeast cells the m/z value 379 has strongly decreased, while a number of higher methylated sterols appear at higher intensities. The m/z values 407, 409 and 423 can be ascribed to the
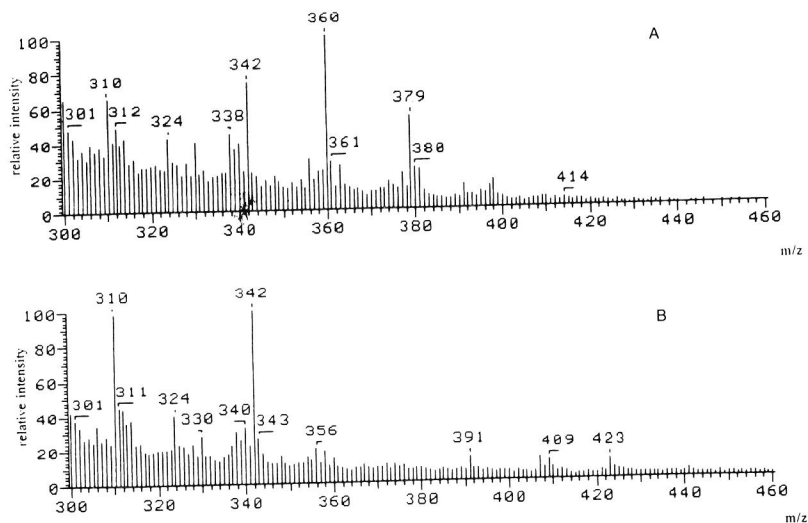
Fig. 5. Mass pyrograms of control (A) and treated yeast (B) cells cultured in Kimmig's medium.

$[M + H - H_2O]^+$ ions of dehydrolanosterol, lanosterol and 24-methylene-24,25-dihydrolanosterol, respectively. The inhibition of the ergosterol biosynthesis through blocking of a demethylation process is clearly demonstrated by these mass pyrograms.

Ophiostoma ulmi

DCI–pyrolysis of *O. ulmi* hyphal and yeast cell wall fractions yields spectra in which hexose and deoxyhexose residues are dominant, according to the mass series at m/z 180, 162, 144 and 164, 146 (Fig. 6A and B). This is in agreement with the results of recent investigations in which several polysaccharides such as rhamnomannans, galactomannans and glucans have been characterized [15,17]. In the mass series belonging to the hexose residues the low intensity of the m/z 162 $[M + NH_4 - 2 H_2O]^+$ ion as compared to the m/z 180 ion is striking. The relative abundance of this ion is higher in the mass pyrogram of the glycopeptide rhamnomannan. The mass pyrogram of this biopolymer shows a high degree of similarity with those of the *O. ulmi* cell wall fractions (Fig. 7).

In the mass range m/z 200–250 the intensities at m/z 240, 222 and 204 may represent structural features related to the linkage type between monosaccharide units. This series, being the $[M + NH_4]^+$, $[M + NH_4 - H_2O]^+$ and $[M + NH_4 - 2 H_2O]^+$ ions respectively, can be ascribed to a hexose monosaccharide coupled to an acetaldehyde unit formed by a retro-aldolization fragmentation of the second ring. This type of compounds may be
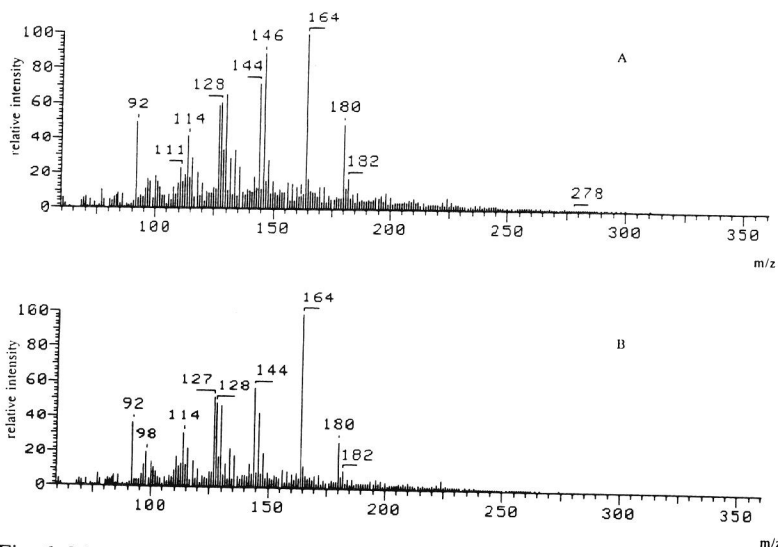
Fig. 6. Mass pyrograms of cell wall fractions of hyphal (A) and yeast (B) forms of *O. ulmi*.

indicative of the presence of 1,6-linkages between these monosaccharide units as has been found by Curie-point Py–FIMS [5]. It was observed by Py–DCI of the reference compounds dextran and amylose, predominantly $(1 \rightarrow 6)$-α- and $(1 \rightarrow 4)$-α-linked glucans, that the intensities at m/z 240 and 222 were indeed relatively abundant in the $(1 \rightarrow 6)$-α-type and almost absent in the $(1 \rightarrow 4)$-α-polysaccharide. However, Py–DCI of cellulose $(1 \rightarrow 4$-β-$)$ and β-glucan $(1 \rightarrow 4$-β, $1 \rightarrow 3$-β-$)$ revealed the presence of the same m/z values in smaller abundances, probably originating from the same type of compounds. Apparently different linkage types are reflected in different intensities of these ions.

Py–DCIMS/MS measurements of m/z 240 and 222 formed during pyrolysis of dextran provided additional evidence for the suggested structures. These ions exhibited loss of ammonia followed by repeated water losses, and additional fragmentation revealed specific fragments from the hexose ring. A more detailed report on these Py–DCIMS/MS experiments
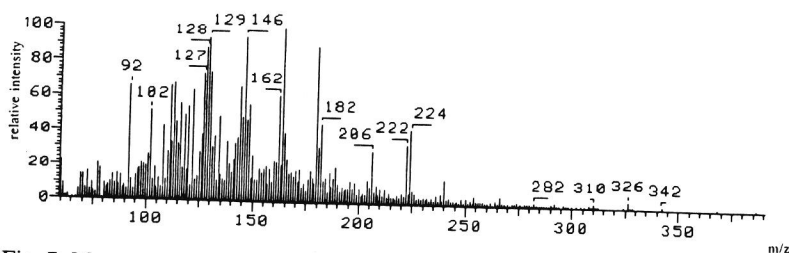
Fig. 7. Mass pyrogram of rhamnomannan.

92
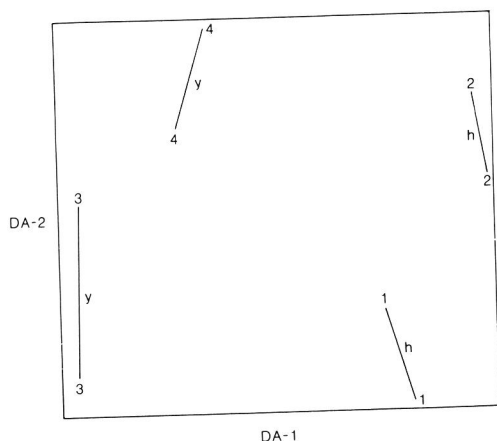
Fig. 8. Discriminant plot of hyphal (categories 1 and 2) and yeast (categories 3 and 4) forms. Categories 2 and 4 represent forms cultured 6 months after those of categories 1 and 3.

is being prepared for publication. The intensities at m/z 206 and 224 may point to comparable pyrolysis products of an unknown linkage type originating from the combination of rhamnose with hexose units or to acetylated rhamnose units. The presence of this type of structural units in DCI mass pyrograms illustrates the optimal pyrolysate transfer typical of this method.

Additional information in this spectrum can be found in the m/z 300–350 range. The $[M + NH_4 - H_2O]^+$ ions of dideoxyhexose, deoxyhexose-hexose and dihexose residues are visible in the spectrum at m/z 310, 326 and 342 respectively. Probably this provides valuable information on disaccharides, but it cannot be ruled out that these products arise due to ion–molecule reactions.

Principal component-discriminant analysis applied to selected signals of the mass pyrograms of the cell wall fractions reflects two main tendencies in the data: differences in the cell wall composition of the yeast and hyphal forms and longer-term reproducibility effects (Fig. 8). The first two discriminant functions obtained, showed $\lambda/(1 - \lambda)$ values of 198 and 3, representing 14% and 34%, respectively, of the variance contained in the selected variables.

The first effect is the most important one because it is represented by the first discriminant function, the ratio of the between-group and the within-group variance being much larger than the second function. Analysis of the factor spectrum in the 180° direction reveals the higher abundance of deoxyhexose residues in the yeast form as can be concluded from the m/z values 310 and 164, the $[M + NH_4 - H_2O]^+$ ions of the di- and monosaccharide units, respectively (Fig. 9). This observation is in agreement with reflection infrared measurements of these cell wall fractions, but in contrast
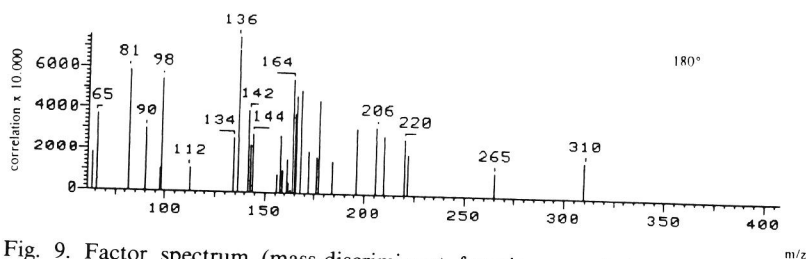
Fig. 9. Factor spectrum (mass-discriminant function correlation) after counter-clockwise rotation of the first discriminant function over 180 ° C (corresponding to Fig. 8).
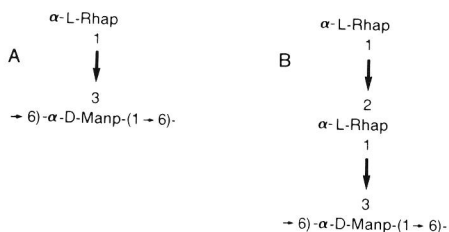

Fig. 10. Rhamnomannan biopolymer of the yeast (A) and hyphal (B) forms of *S. schenckii.*

to the results described for the dimorphic fungus *Sporothrix schenkii* in which a higher content of rhamnose units is detected in the hyphal form (Fig. 10).

The effects of the longer-term reproducibility are represented by the second discriminant function. This effect encloses biological as well as instrumental variations. No clear structural information can be inferred from the factor spectra in the 90 ° and 270 ° direction.

CONCLUSIONS

Py–DCIMS allows detection of a broad range of pyrolytic products from a variety of materials. Specificity tuning is possible by the choice of the proper reagent gas, the detection of positive or negative ions, the resolution of the mass spectrometer, or as shown in this study the application of MS/MS methods.

In the universal mode overall trends can be detected effectively as has been demonstrated in the study on the fungus *C. albicans.* Yeast and hyphal forms could be differentiated with pattern recognition techniques, even when cultured in the same medium. The application of these techniques also enabled the extraction of the relevant information producing leads for identification. Differences in polysaccharide composition were an important tool for this differentiation.

The pyrolysis method demonstrates its power by providing not only information on biopolymers, but also on changes of the sterol metabolic pattern effected by the SBI application. Furthermore, the presence of lombazole metabolites could be detected, despite their low concentration in the medium. The latter points to an accumulation of the SBI and a metabolite in the cell or in, either on, the cell envelope. Also with *O. ulmi* information on the polysaccharides present in cell wall fractions was clearly exhibited by ammonia DCI–pyrolysis and played a crucial role in the differentiation of yeast and hyphal forms.

ACKNOWLEDGEMENT

REFERENCES

1 H.L.C. Meuzelaar and P.G. Kistemaker, Anal. Chem., 45 (1973) 587.
2 H.L.C. Meuzelaar, P.G. Kistemaker, W. Eshuis and A.J.H. Boerboom, Adv. Mass Spectrom., 7B (1978) 1452.
3 F.J. Vastola, A.J. Pirone and B.E. Knox, Proceedings of the 14th Annual Conference on Mass Spectrometry and Allied Topics, ASTM Commitee, E-14, Philadelphia, 1966, p. 78.
4 J.L. Wiebers, Anal. Biochem., 51 (1973) 542.
5 H.-R. Schulten, U. Bahr and W.J. Görtz, J. Anal. Appl. Pyrolysis, 3 (1981) 229.
6 H.-R. Schulten and W. Görtz, Anal. Chem., 50 (1978) 428.
7 H.-R. Schulten, U. Bahr, H. Wagner and H. Hermann, Biomed. Mass Spectrom., 9 (1982) 115.
8 A.C. Tas, J. van der Greef, J. de Waart, J. Bouwman and M.C. ten Noever de Brauw, J. Anal. Appl. Pyrolysis, 7 (1985) 249.
9 A.C. Tas, J. de Waart, J. Bouwman, M.C. ten Noever de Brauw and J. van der Greef, J. Anal. Appl. Pyrolysis, 11 (1987) 329.
10 A.C. Tas, G. Wieten, J. de Waart, L. Berwald and J. van der Greef, J. Microbiol. Methods, 8 (1988) 333.
11 J. van der Greef, A.C. Tas, J. Bouwman and M.C. ten Noever de Brauw, Adv. Mass Spectrom., 10 (1986) 1227.
12 D. Barug, R.A. Samson and A. Kerkenaar, Arzneim.-Forsch. 33 (1983) 528.
13 D. Barug, H.B. Bastiaanse, J.M. van Rossum and A. Kerkenaar, Antimicrob. Agents Chemother., 30 (1986) 238.
14 W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81.
15 A. Kerkenaar, R.J. Scheffer, H.B. Nilander, A.C. Brakenhoff and D.M. Elgersma, Abstract 6th International Congress Pesticide Chem. IUPAC, 10–15 August, 1986, Ottawa.
16 R.K. Kulkarni and K.W. Nickerson, Exp. Mycol., 5 (1981) 148.
17 P.A.J. Gorin and J.F.T. Spencer, Carbohydr. Res., 13 (1970) 339.
18 L.R. Travassos and K.O. Lloyd, Microbiol. Rev., 44 (1980) 683.

19 G. Strobel, N. van Alfen, K.D. Hapner, M. McNeil and P. Albersheim, Biochim. Biophys. Acta, 538 (1978) 60.

20 Pattern recognition package ARTHUR 81, Infometrix, Inc., Seattle, U.S.A.

21 K.R. Gabriel, Biometrika, 58 (1971) 453.

22 D. Barug and H.B. Bastiaanse, Arzneim.-Forsch., 33 (1983) 538.

23 M.A. Ghannoum, G. Janini, L. Khamis and S.S. Radwan, J. Gen. Microbiol., 132 (1986) 2367.

24 R. Svaldi and D.M. Elgersma, Eur. J. Forest. Pathol., 12 (1982) 29.

25 M.A. Dewerchin and A.J. van Laere, Trans Br. Mycol. Soc., (1985) 167.

26 R.J. Scheffer, Dutch elm disease, aspects of pathogenesis and control, Thesis, University of Amsterdam, 1984.
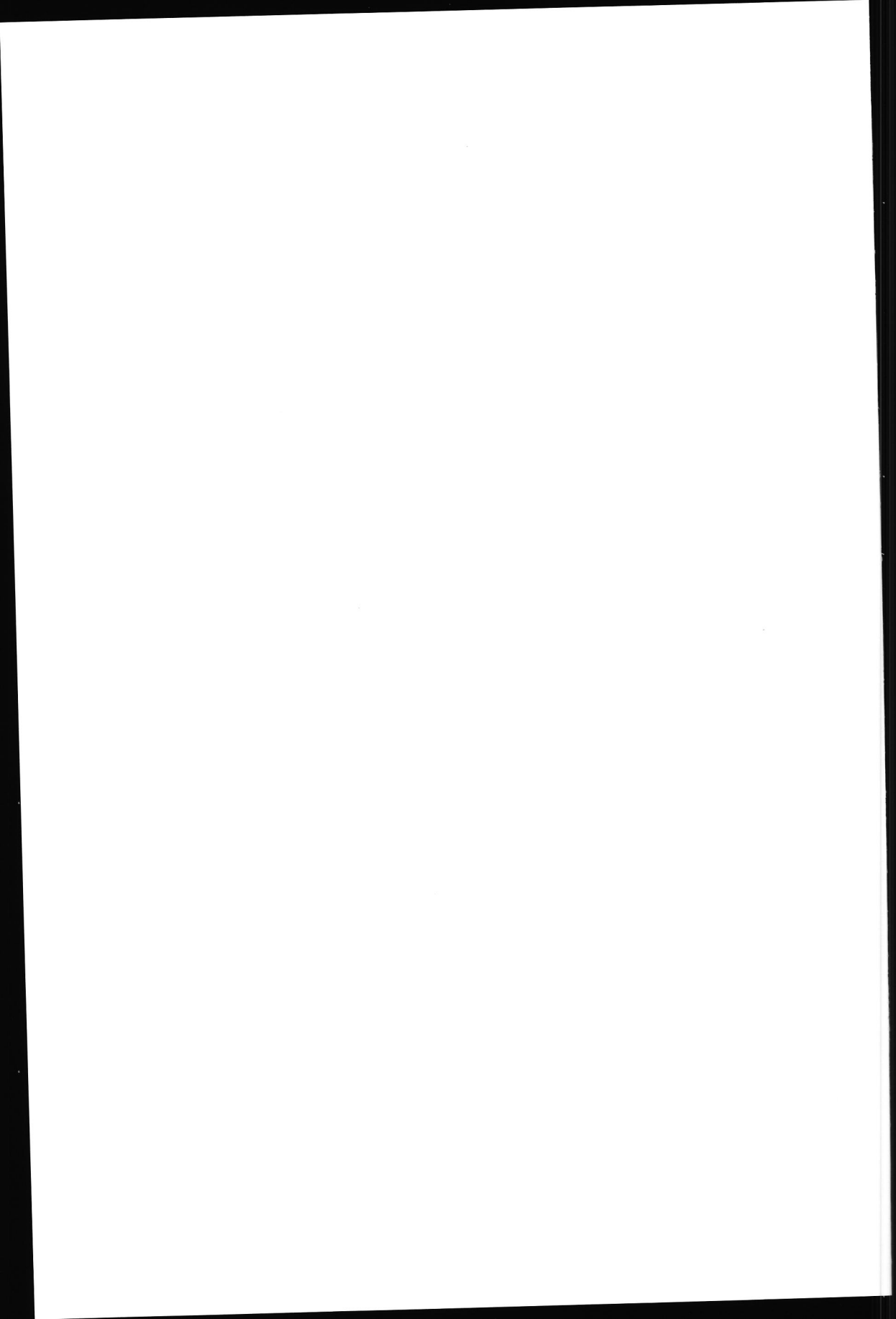
Chapter 9

PYROLYSIS–DIRECT CHEMICAL IONIZATION MASS SPECTROMETRY OF SOME BIOPOLYMERS IN THE POSITIVE AND NEGATIVE IONIZATION MODE

A.C. TAS *

TNO-CIVO Food Analysis Institute, P.O. Box 360, 3700 AJ Zeist (The Netherlands)

A. KERKENAAR

TNO Institute of Applied Chemistry, P.O. Box 108, 3700 AC Zeist (The Netherlands)

G.F. LaVOS and J. VAN DER GREEF

TNO-CIVO Food Analysis Institute, P.O. Box 360, 3700 AC Zeist (The Netherlands)

ABSTRACT

Pyrolysis–direct chemical ionization mass spectrometry experiments were carried out under positive and negative ionization conditions on the polysaccharides amylose and dextran, $(1 \rightarrow 4)$-α-glucan and $(1 \rightarrow 6)$-α-glucan. Ions were observed in the higher mass range, corresponding to hexose fragments up to four hexose units. Some of these specific fragments were investigated by tandem mass spectrometry.

The applicability of positive and negative ionization conditions with respect to the specific detection of compounds in complex pyrolysates from a *Penicillium italicum* strain, cultured with and without the fungicide imazilil, were compared. Different constituents of the complex pyrolysates were more or less emphasised by the ionization conditions applied. Variation of these conditions provided information on both pyrolytic and ion chemistry processes.

Amylose; dextran; direct chemical ionization; glucan; mass spectrometry; negative ionization; *Penicillium italicum*; polysaccharides; positive ionization; pyrolysis; tandem mass spectrometry.

INTRODUCTION

Pyrolysis–mass spectrometry has a great potential for the investigation of complex biopolymers. Biopolymers with different physical and chemical properties can be investigated directly. The great variety of chemical compounds to be detected in pyrolysates is very demanding with respect to the choice of instrumental conditions [1–3]. The use of versatile instruments enabling the variation of these conditions provides better opportunities for research in this area.

Tuning of the resolution, the use of tandem mass spectrometry (MS/MS), extension of the mass range for pyrolysate detection and the application of different ionization conditions are important tools for identification of unknown compounds or for specific detection of target compounds.

The recently developed pyrolysis–direct chemical ionization mass spectrometry (Py–DCIMS) [4] opens new possibilities for this kind of biopolymer research. Pyrolysis in the ion source guarantees an optimum pyrolysate transfer of larger and polar molecules of low volatility to the ionization region [4,5]. The combination with relatively soft ionization results in the detection of ions over an extended mass range. In addition, chemical ionization conditions can be adjusted to provide more specific detection of biopolymers of interest. Such compounds reflect substructures of larger biopolymers or may be present as such in biopolymer mixtures.

The combination of various ionization conditions, high resolution [6] and, especially, MS/MS measurements provide extra specificity or can be used for identification purposes.

When ammonia is applied as a reactant gas in the positive ionization mode a great variety of compounds present in pyrolysates are ionized, the main reactions being proton transfer, association reactions and charge exchange reactions [7]. In the negative ionization mode a specific subset of compounds can be detected in the pyrolysate. The presence of high concentrations of thermalized electrons and the strong base NH_2^- gives rise to mainly $M^{-\cdot}$ and $[M - H]^-$ ions [7,8].

In this study Py–DCI experiments were carried out under positive and negative ionization conditions on two polysaccharides, amylose and dextran, $(1 \rightarrow 4)$-α-glucan and $(1 \rightarrow 6)$-α-glucan respectively. Some ions, originating from pyrolytic fragmentations of these biopolymers, were investigated by MS/MS [5].

To compare the scopes of different ionization conditions for the detection of compounds in pyrolysates obtained from complex mixtures, *Penicillium italicum* strains [9–11] were subjected to DCI pyrolysis in the positive and negative mode. The results of the comparison of the strain (W-5), cultured with and without the fungicide imazalil, are presented.

EXPERIMENTAL

Pyrolysis–mass spectrometry

DCI pyrolysis was performed on a Finnigan-MAT 8230 mass spectrometer coupled to a Finnigan-MAT SS300 data system. Ammonia was used as a reactant gas at an indicated source pressure of 53 Pa; the temperature of the ion source was kept at 240°C. During heating of the DCI wire by linear current programming (8 mA/s) spectra were acquired over the mass range

m/z 60–1000 in the positive mode and m/z 42–900 in the negative mode, at a scan speed of 0.7 s/decade. Spectra obtained during pyrolysate formation were summed to one profile.

Pyrolysis–tandem mass spectrometry

Pyrolysis–DCIMS/MS was carried out on a Finnigan-MAT HSQ-30 mass spectrometer coupled to a Finnigan-MAT SS300 datasystem. Except for the scanning parameters, the same conditions as described above were maintained in the sector field mass spectrometer which was used as the first mass analyzer. The collision energy was set at approximately 45 V. Argon was used as collision gas.

Samples

$(1 \rightarrow 4)$-α-Glucan (amylose) was obtained from Avebe and $(1 \rightarrow 6)$-α-glucan (dextran T-70) was purchased from Pharmacia Fine Chemicals.

Samples of *Penicillium italicum* were made available by Dr. M.A. de Waard [9]. *P. italicum* W-5 is an isolate sensitive to various sterol biosynthesis inhibitors [10]. Cultural conditions for growth of mycelium in liquid medium and the effect of imazalil have been described previously [10,11]. Mycelium (0.5 g dry weight) harvested from 8 hour cultures with and without imazalil (0.1 mg/ml) was used. Differences in sterol patterns in untreated and treated mycelium have been clearly detected using different analytical technique.

RESULTS

Py–DCIMS of amylose and dextran (positive mode)

Comparison of the DCI spectra of amylose and dextran obtained in the positive mode reveals substantial differences between the profiles in the lower as well as in the higher mass range (Figs. 1 and 2). The observation of the greater complexity of the pyrogram of dextran is in accordance with pyrolysis–field ionization (Py–FIMS) experiments with glycogen and dextran [6].

Although in both pyrograms m/z 180 is the base peak, the relative intensity of this ion is higher for amylose. This higher relative intensity of the m/z 180 ($[M + NH_4]^+$ ion) in the spectrum of amylose (Fig. 1) can be well explained by the ease of formation of levoglucosan from this $(1 \rightarrow 4)$-α-polymer [6,12,13]. Curie-point pyrolysis–gas chromatographic analysis (Py–GC) of amylose under neutral (suspended in methanol) conditions revealed this compound to be the major component in the pyrolysate [12].
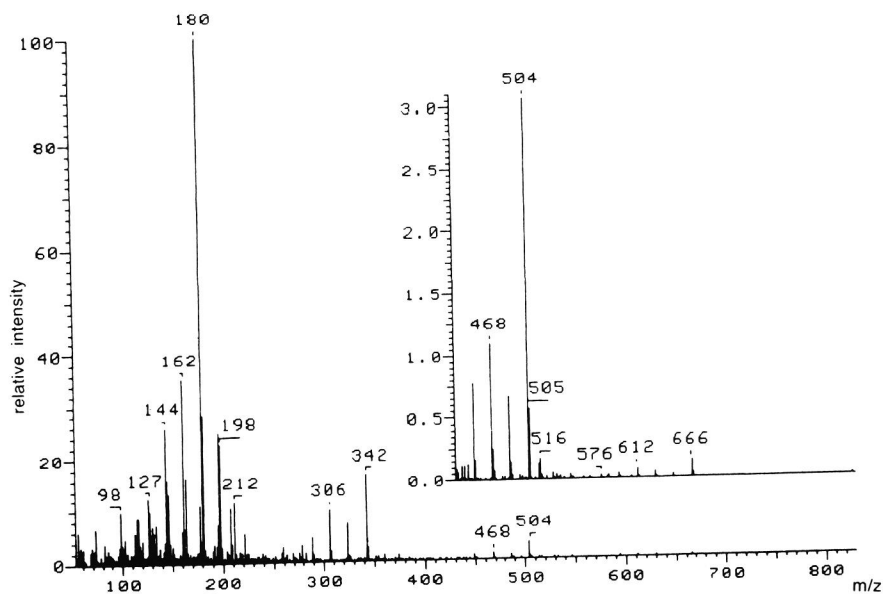
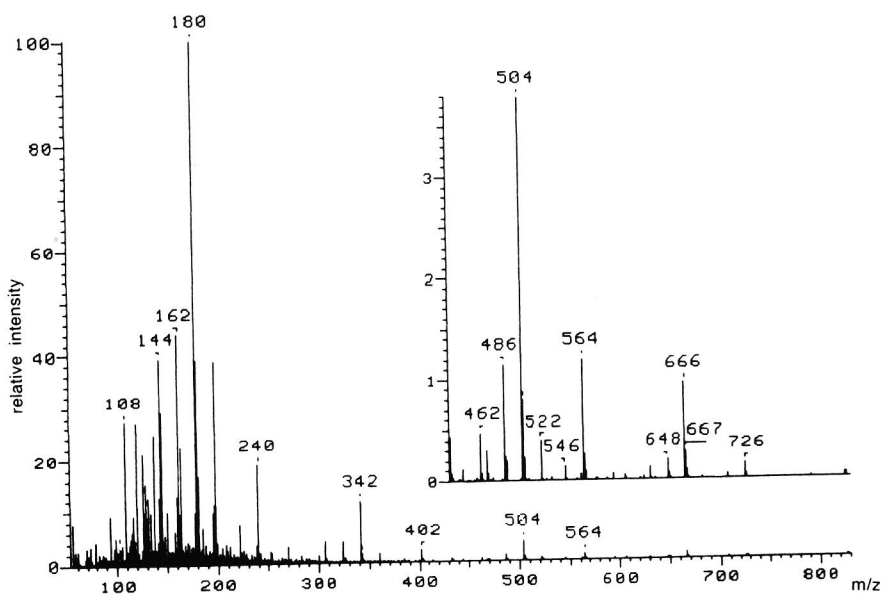Fig. 1. Pyrogram of amylose; positive DCI conditions.

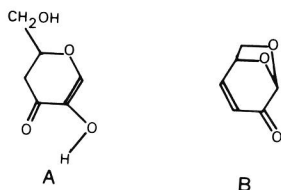Fig. 2. Pyrogram of dextran; positive DCI conditions.

Fig. 3. (A) 1,4-Dideoxy-D-glycero-hex-1-enopyranos-3-ulose. (B) 1,6-Anhydro-3,4-dideoxy-D-glycero-hex-3-enopyranos-2-ulose (levoglucosenone).

The ion with the next highest intensity in the amylose pyrogram, m/z 162, probably corresponds to the $[M + NH_4]^+$ ion of 1,4-dideoxy-D-glycero-hex-1-enopyranos-3-ulose (Fig. 3, A) which has been found to be an important pyrolysis product in Py–GC [12] and Py–MS analysis [14] of amylose under neutral conditions. In addition, a contribution to this ion of other anhydrohexose compounds can be expected. The ion at m/z 144 may be composed of the $[M + NH_4]^+$ ions of a number of dehydration products such as 5-(hydroxymethyl)-2-furaldehyde and 1,6-anhydro-3,4-dideoxy-D-glycero-hex-3-enopyranos-2-ulose (levoglucosenone, Fig. 3, B). The latter compound may give rise preferentially to the $[M + H]^+$ ion at m/z 127 due to the favourable position of the proton simultaneously bound to two geminal oxygen atoms.

A remarkable phenomenon in the medium mass range of this pyrogram is the relatively high intensity of the m/z 197 ion which exceeds that of the m/z 198 ion. A probable origin of this ion can be the levoglucosan $[M + 2(NH_3) + H]^+$ ion, because of the high concentration of this compound in the pyrolysate and a relatively high abundance of $[2(NH_3) + H]^+$ ions in the reactant gas.

The same mass (m/z 197) is also observed in comparable intensities in the dextran pyrogram, but the ion m/z 198 is more abundant, originating from the $[M + NH_4]^+$ ion of a complete monosaccharide unit. In the dextran pyrogram the m/z 180 ion is also the base peak, although at a lower relative intensity. In the lower mass range a number of masses, for example m/z 164, 162, 146, 144, 127, 116 and 98, are present in the pyrograms of both biopolymers. However, an additional group of masses, such as m/z 150, 138, 120 and 108 have a higher intensity in dextran. The masses m/z 150 and 120 can be explained from retrograde aldolization (RA) reactions with losses of formaldehyde and hydroxyacetaldehyde, respectively, from the m/z 180 ion [6]; m/z 138 and 108 may originate from m/z 198, the $[M + NH_4]^+$ ion of a monosaccharide unit, also by a RA reaction sequence, giving rise to losses of hydroxyacetaldehyde and formaldehyde, respectively. Apparently, the RA reaction channels are more important in the pyrolysis of dextran, leading to the formation of additional pyrolysis products.
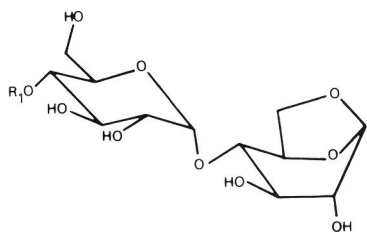
Fig. 4. Levoglucosan series (m/z 342, 504, 666); R_1 = H or one or two hexose units.

In the higher mass range of the amylose pyrogram, ion series can be observed corresponding to disaccharide, trisaccharide and tetrasaccharide units. The ions at m/z 342, 504 and 666 correspond to the respective $[M + NH_4 - H_2O]^+$ ions of these compounds. Probably these ions consist of a levoglucosan unit, connected to a mono-, di- and trisaccharide unit respectively (Fig. 4). Also a 1-deoxy-D-glycero-hexanopyranos-2-ulose unit, arising from 1,2-*trans*-elimination and analogously attached, may contribute to this ion series. In all series identical fragmentation patterns are observed, which can be explained from one or more losses of water, for example in the mass range corresponding to trisaccharide units, the ions at m/z 486, 468 and 450. Corresponding ions of the di- and tetrasaccharide units can be found at a distance of 162 mass units.

In the higher mass range of the dextran pyrogram comparable mass series of di-, tri- and tetrasaccharide units can be observed (Fig. 2). However, an additional series of ions is present, 60 mass units higher than the corresponding levoglucosan series. The ions at m/z 240, 402, 564 and 726 are probably formed by RA fragmentations of the adjacent hexose rings. An acetaldehyde unit coupled to mono-, di-, tri- and tetrasaccharide units, respectively, can well explain the masses observed (Fig. 5). The correspond-
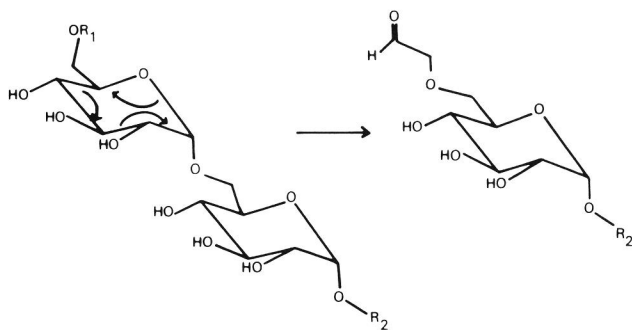
Fig. 5. Retrograde aldolization reaction; R_1 and R_2 = polysaccharide chains.

ing ion series, 18 mass units lower, can also be observed in the dextran pyrogram. Two attachments sites (1- or 6-position) are possible of which only the 6-position is indicated. One of the ions in this series, m/z 222, is also visible in the amylose pyrogram; the masses at m/z 240, 402, 564 and 726 are absent. Signs of the presence of similar but further degraded compounds, containing more than six carbon atoms were observed in Py–FIMS spectra [6] of dextran. However, it appeared from Py–DCIMS of cellulose [5] that these compounds are also formed in this $(1 \rightarrow 4)$-β-polysaccharide, although to a relatively lesser extent, and cannot be considered as exclusive $(1 \rightarrow 6)$-linkage markers.

Py–DCIMS / MS of some amylose and dextran pyrolysis products

Direct measurements of compounds in pyrolysates with Py–DCIMS/MS can provide important structural information, especially on higher molecular weight substances. Identification of structural units repeatedly occurring in pyrograms, for example connected to mono- and oligosaccharide units, may give a better insight into the structure of series of pyrolysis products and into pyrolysis mechanisms.

Investigation of the parent ion at m/z 240, occurring in a relatively high abundance in the pyrogram of dextran and correlated to compounds at m/z 402, 564 and 726, revealed a number of specific fragments, almost all of which being odd-numbered ions (Fig. 6). Ions with a relatively high intensity
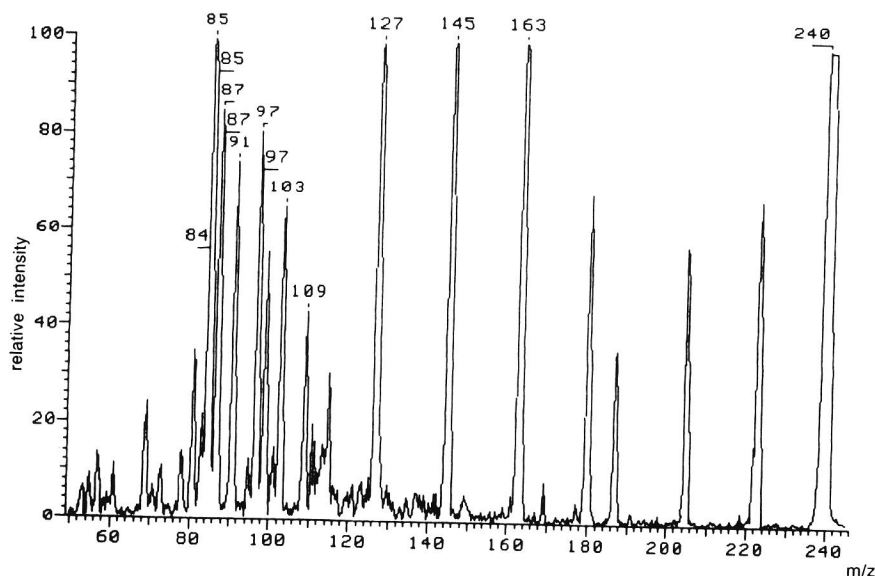
Fig. 6. Py–DCIMS/MS spectrum of m/z 240 of dextran.

104
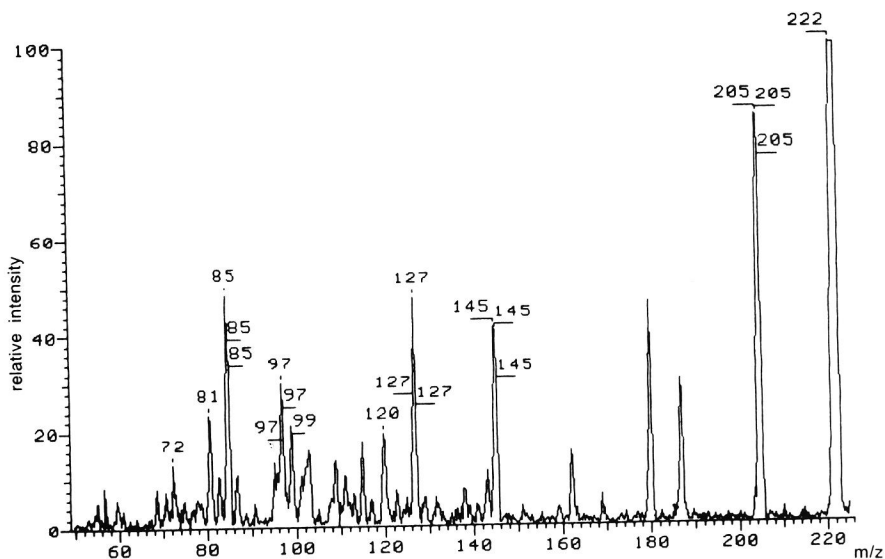
Fig. 7. Py–DCIMS/MS spectrum of m/z 222 of dextran.

are present at m/z 163, 145 and 127. The same ions are observed in isobutane CI spectra of levoglucosan and in other hexoses, such as amylose [12]. They correspond to protonated anhydro-, dianhydro- and tri-anhydrohexose units.

In the higher mass range the ion at m/z 223 has been formed from m/z 240 by loss of NH_3 and can be ascribed to the protonated coupling product of acetaldehyde and hexose. Consecutive losses of water result in the formation of the ions at m/z 205, 187 and 169. One of the scarce even ions in the spectrum, m/z 180, probably arises from loss of ketene and water from the parent ion m/z 240. The most likely structure of the ion at m/z 180 is an ammoniated anhydrohexose unit.

The parent ion at m/z 222 originating from dextran, probably an anhydro form of the compound m/z 240, has a similar fragmentation pattern (Fig. 7). Again loss of ammonia is observed with formation of m/z 205 from the parent ion. Repeated losses of water result in the formation of the ions at m/z 187 and 169. The intensities at m/z 163, 145 and 127 are also observed in this spectrum, emphasizing the hexose nature of the parent ion. The formation of the even ion at m/z 180 can be explained by ketene loss from the parent ion.

A comparable ion at m/z 222, present in low abundance in the pyrogram of amylose, shows a similar fragmentation pattern in the higher mass range (Fig. 8). Loss of ammonia as well as losses of water can also be observed. However, in the lower mass range some significant differences compared to the fragmentation pattern of the dextran parent ion m/z 222 are found. The
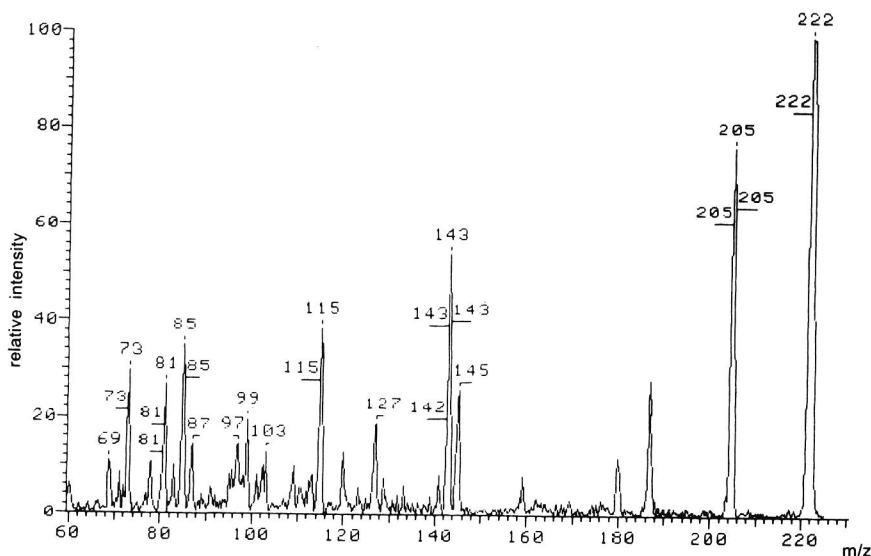
Fig. 8. Py–DCIMS/MS spectrum of m/z 222 of amylose.

ions at m/z 180 and 163 are less abundant and a higher intensity is observed for the ions at m/z 143 and 115. The structure of the $[M + NH_4]^+$ ion at m/z 222, which presumably arises from levoglucosan, coupled to an acetaldehyde unit at the 4-position, is prone to loss of ammonia, rearrangement and a 3,6 dehydration step, resulting in the formation of the protonated ion at m/z 187 (Fig. 9). Subsequent loss of acetaldehyde might be favoured for this compound, resulting in a higher intensity of the ion at m/z 143. The resulting ketone can undergo loss of carbon monoxide (Fig. 9). The observed differences in intensity of the ions at m/z 143 and 115 might be an indication of the attachment site of the acetaldehyde group: when coupled to the ring system, loss of the acetaldehyde group is promoted, due to the formation of the oxonium ion.

DCIMS/MS comparison of the ions at m/z 180 of dextran and amylose revealed the strong similarity between the fragmentation patterns (Figs. 10
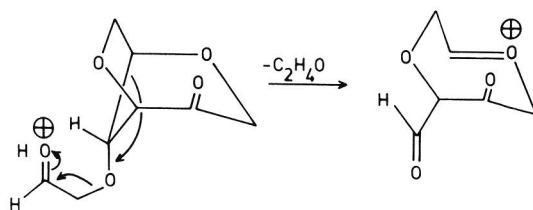
Fig. 9. Loss of acetaldehyde from an amylose fragment (m/z 187).
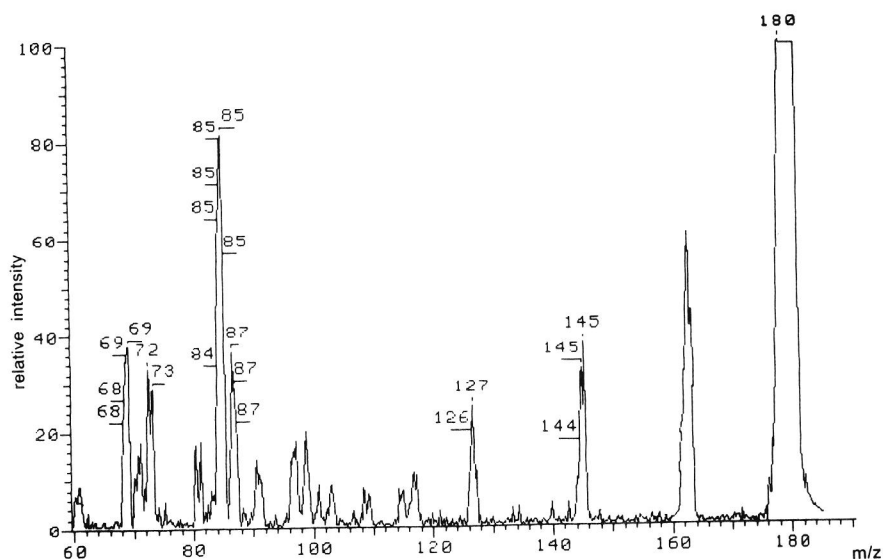
Fig. 10. Py–DCIMS/MS spectrum of *m/z* 180 of dextran.

and 11). After loss of ammonia (*m/z* 163), fragmentation by repeated losses of water can be observed providing the ions at *m/z* 145 and 127. Also in the lower mass range the fragmentation patterns are very similar, taking into
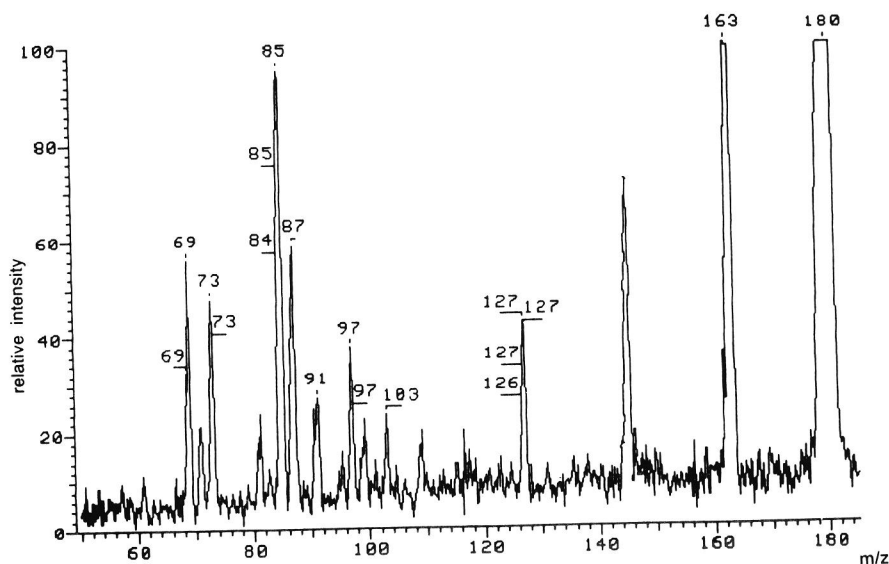
Fig. 11. Py–DCIMS/MS spectrum of *m/z* 180 of amylose.

account the diminished reproducibility of the method in comparison to normal mass spectrometry. These data strongly indicate the identity of the m/z 180 ions of both biopolymers.

Py–DCIMS of amylose and dextran (negative mode)

When ammonia is used under negative ionization conditions, a number of different ionization processes occur [7]. In the ion source NH_2^- ions are generated and a great density of thermalized ions exists, which initiate several processes. Anion formation via resonance capture takes place only for a specific group of compounds for which resonance stabilization for instance via a π system is possible [8]. For example, α-diketones can give intense molecular anions under negative electron impact conditions. Also electron transfer can occur for compounds with a strong electron affinity. $[M - H]^-$ ions can be produced by proton abstraction of both thermalized ions and NH_2^- ions [7]. Due to the specificity of the reactions, it may be expected that different groups of compounds in the pyrolysate are accentuated, compared to positive ionization conditions.

In the pyrogram of amylose this effect is clearly demonstrated in the occurrence of the ion series at m/z 207, 369, 531 and 693, with 162 mass units interval, which may arise from proton abstraction of compounds with m/z 208, 370, 532 and 694. The latter compounds are probably generated by negative ion-induced RA fragmentation of a hexose ring, which specifically occurs in amylose (Figs. 12 and 13). These intensities can be ascribed to formyl esters of the mono-, di-, tri- and tetrahexoses, respectively, 28 mass units higher than the $[M - H]^-$ ions of the corresponding hexoses (Fig. 12) of which ions can be found at m/z 179, 341 and 503. The ion at m/z 45 may indicate a formyl anion. However, it is unclear why such compounds are strongly favoured in the ionization.

An additional series of ions, 18 mass units lower than the previously mentioned one and displayed at m/z 161, 323, 485 and 647, arise from proton abstraction from anhydrohexoses, corresponding to the $[M + NH_4]^+$ ions at m/z 180, 342, 504 and 666 found in the positive ion spectrum and ascribed to levoglucosan and higher analogous compounds. Two main series of even ions are detected at m/z 144, 306, 468 and 630, and at m/z 126,
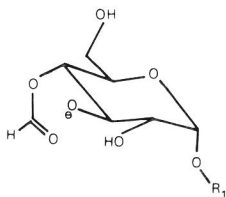

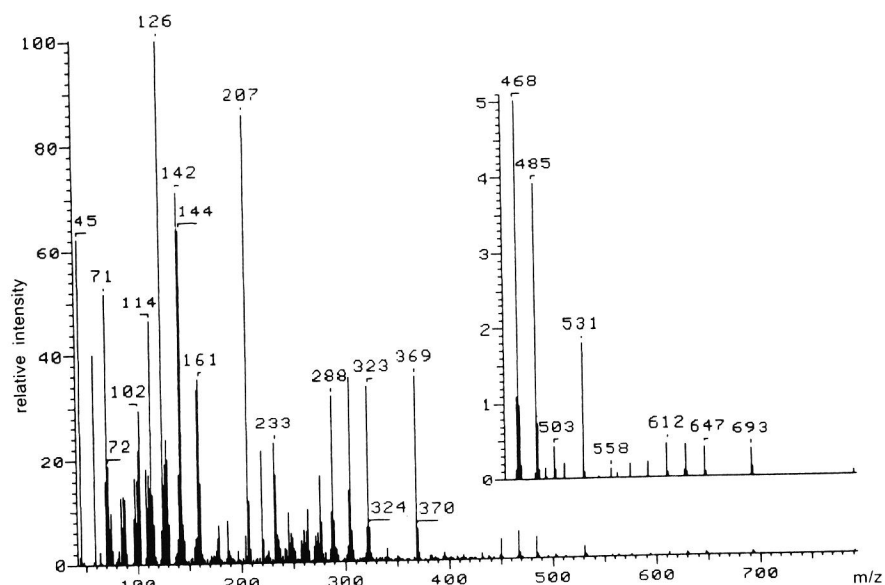Fig. 12. Tentative structure of an hexose formiate anion; $R_1 = H$ or one to three hexose units.

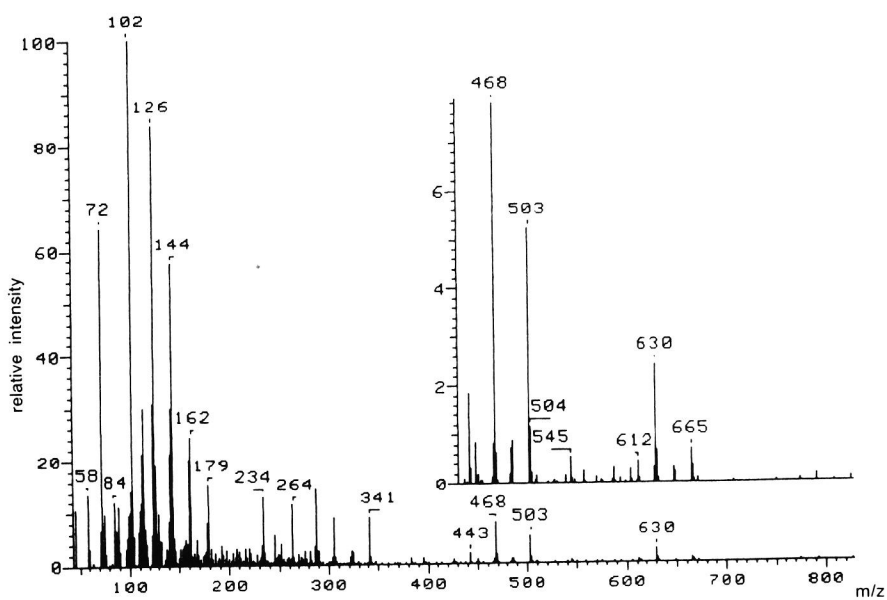Fig. 13. Pyrogram of amylose; negative DCI conditions.

Fig. 14. Pyrogram of dextran; negative DCI conditions.

288, 450 and 612. They may be formed by resonance capture of enones, which probably can accommodate an electron by resonance stabilization via a π system. For example, compounds such as 1,4-dideoxy-D-glycero-1-hexenopyranos-3-ulose and 1,6-anhydro-3,4-dideoxy-D-glycero-3-hexeno-pyranos-2-ulose (levoglucosenone) can be detected in the pyrograms under negative ionization conditions.

In the pyrogram of dextran a comparable series of odd ions at m/z 179, 341, 503 and 665 can be observed; the series at m/z 45, 207, 369, 531 and 693 is completely absent (Fig. 14). Both even ion series, starting at m/z 144 and 126, respectively, are also found in this pyrogram. The base peak at m/z 102 in the pyrogram, which is small in amylose, can be correlated to the $[M + NH_4]^+$ ion at m/z 120, a specific ion in the positive ion pyrogram of dextran. The ion which probably originates from a RA reaction, can well be ascribed to 4-hydroxy-2-oxobutanal [6]. It can also be found as a molecular unit in the ion series m/z 264, 426 and 588, connected to one to three hexose units. Under negative CI conditions the ion series corresponding to the positive ion series at m/z 240, 402, 564 and 726 could not be detected. This is an indication that the latter RA series are induced by ion chemistry processes.
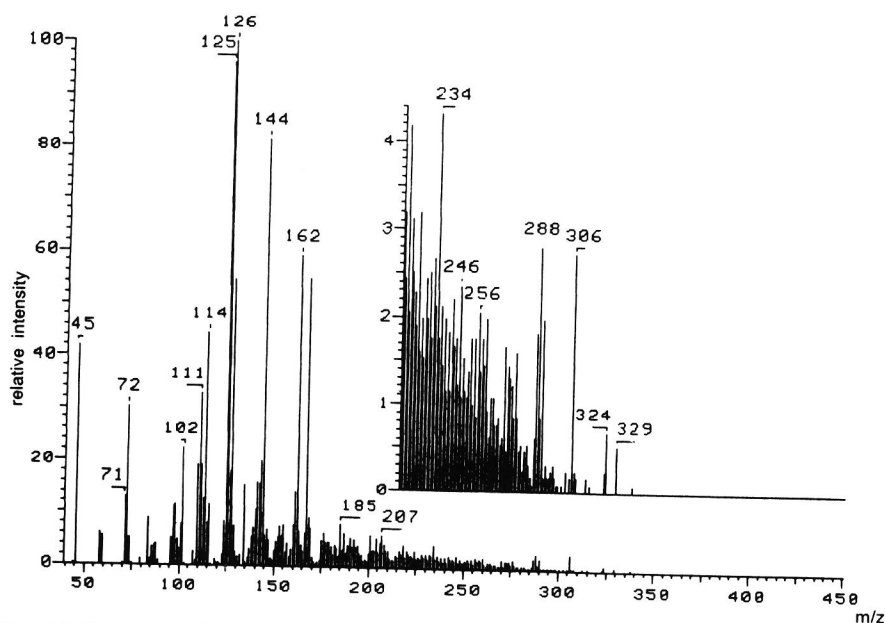
Fig. 15. Pyrogram of *P. italicum*, cultured without imazalil; negative DCI conditions.

Application of Py–DCIMS under positive and negative ionization conditions to a P. italicum strain cultured with and without a fungicide

DCI pyrolysis was performed on cells of a non-resistant *P. italicum* strain (W-5) cultured with and without the fungicide imazalil. Differences in cell composition can be expected for these organisms due to alterations in the sterol biosynthesis [10,11] which may lead to differences in cell wall composition.

Comparison of the pyrograms of the strains measured under positive and negative ionization conditions shows the greater specificity in detection when negative ionization conditions are used (Figs. 15–18), simpler pyrograms being obtained under these conditions. Amongst the ions observed, those related to polysaccharides, for example at m/z 126, 144, 162, 288 and 306, are more dominant. In addition, an ion series at m/z 167, 185, 187 and 201 can be clearly distinguished from the background. These ions originate from *N*-acetylglucosamine residues, present in higher abundances in the strain cultured in the presence of the fungicide imazalil. The corresponding ions in the positive DCI pyrograms at m/z 221, 204, 186 and 168 also show the higher abundance of this compound in the latter strain. However, these signals will easily disappear under noise at lower abundances of this com-
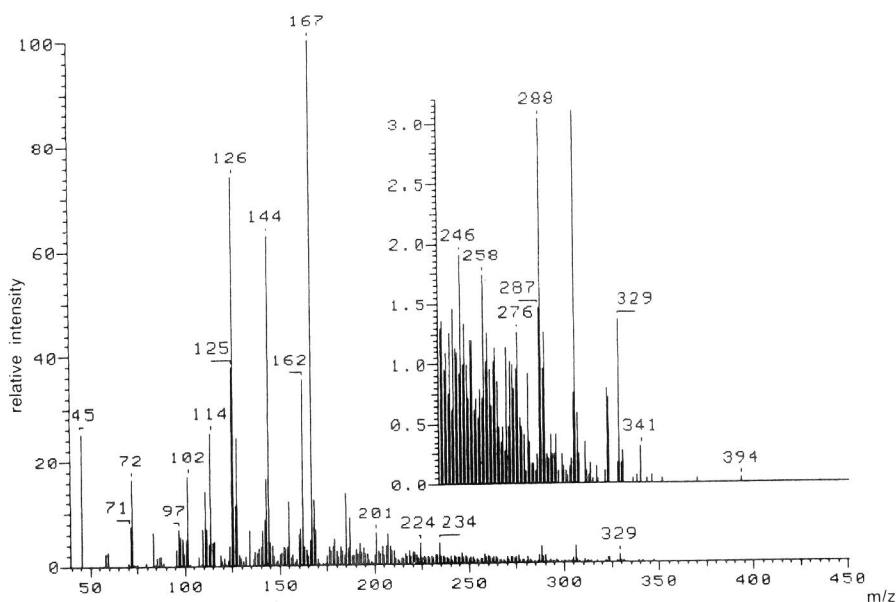

Fig. 16. Pyrogram of *P. italicum*, cultured with imazalil; negative DCI conditions.
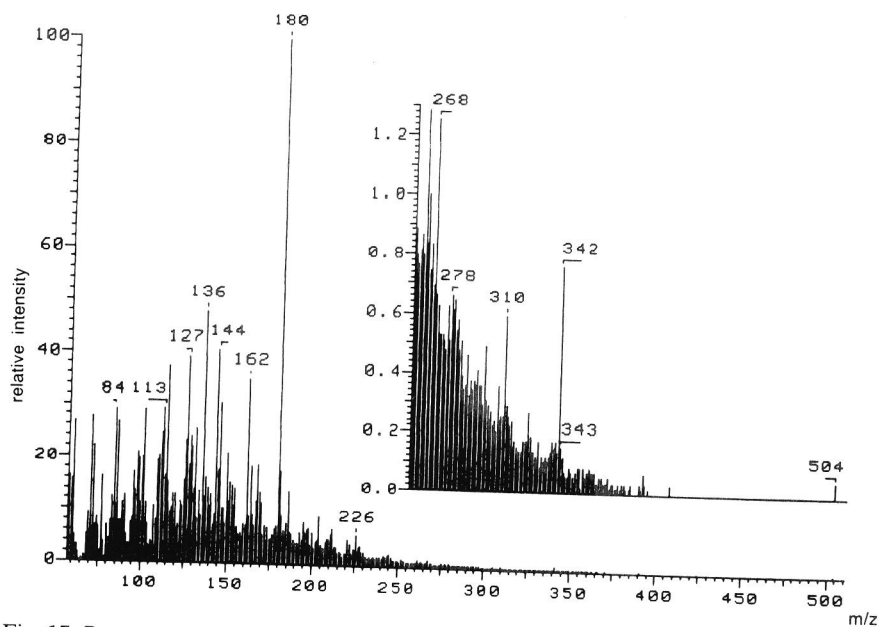
Fig. 17. Pyrogram of *P. italicum*, cultured without imazalil; positive DCI conditions.
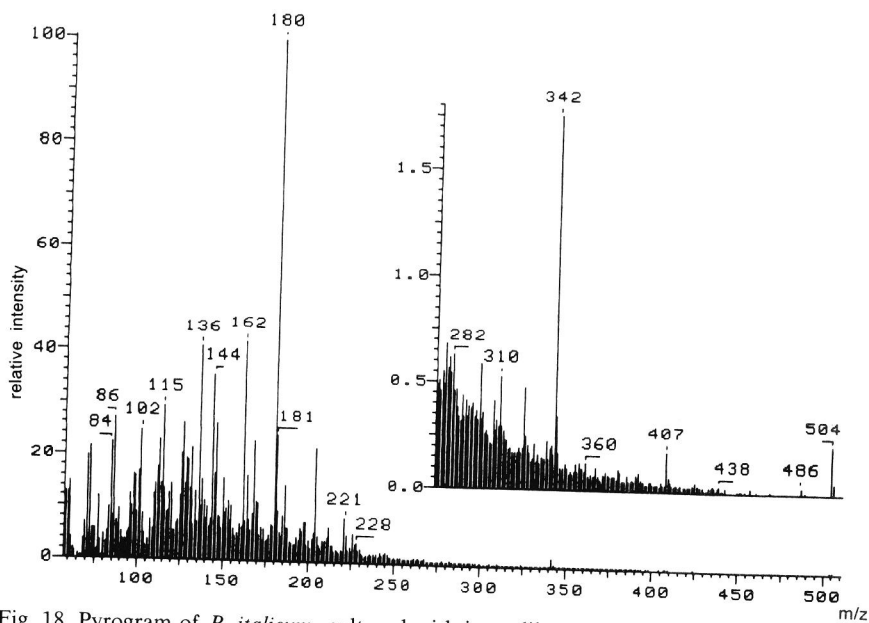

Fig. 18. Pyrogram of *P. italicum*, cultured with imazalil; positive DCI conditions.

pound due to the less specific detection under positive CI conditions, as is shown in the pyrogram of the strain cultured without fungicide (Fig. 17).

CONCLUSIONS

The application of negative CI conditions for pyrolysate detection, using ammonia as a reactant gas was often more compound selective. Proper tuning of ionization conditions, resulting in the suppression of the bulk of compounds present in pyrolysates, is a means of improving the detection of compounds of interest as was shown for polysaccharides and chitin in *P. italicum*.

Several ion series observed in the positive and negative spectra of amylose and dextran could be correlated. Application of the different ionization conditions enabled differentiation between pyrolytic and ion chemistry processes.

ACKNOWLEDGEMENT

REFERENCES

1 S.M. Huff, H.L.C. Meuzelaar, D.L. Pope and C.R. Kjeldsberg, J. Anal. Appl. Pyrolysis, 3 (1981) 95.
2 H.L.C. Meuzelaar, J. Haverkamp and F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials, Elsevier, Amsterdam, 1982.
3 H.-R. Schulten and R.P. Lattimer, Mass Spectrom. Rev., 3 (1984) 231.
4 A.C. Tas, J. de Waart, J. Bouwman, M.C. ten Noever de Brauw and J. van der Greef, J. Anal. Appl. Pyrolysis, 11 (1987) 329.
5 A.C. Tas, H.B. Bastiaanse, J. van der Greef and A. Kerkenaar, J. Anal. Appl. Pyrolysis, 14 (1989) 309.
6 H.-R. Schulten and W. Görtz, Anal. Chem., 50 (1978) 428.
7 J.B. Westmore and M.M. Alauddin, Mass Spectrom. Rev., 5 (1986) 381.
8 H. Budzikiewicz, Angew. Chem. Int. Ed. Engl., 20 (1981) 624.
9 M.A. de Waard, H. Groeneweg and J.G.M. van Nistelrooy, Neth. J. Plant Pathol., 88 (1982) 99.
10 M.A. de Waard and J.G.M. van Nistelrooy, Pest. Sci., 22 (1988) 371.
11 J. Guan, A. Kerkenaar and M.A. de Waard, unpublished results.
12 A. van der Kaaden, J. Haverkamp, J.J. Boon and J.W. de Leeuw, J. Anal. Appl. Pyrolysis, 5 (1983) 199.
13 F. Shafizadeh and Y.L. Fu, Carbohydr. Res., 29 (1973) 113.
14 A. van der Kaaden, J.J. Boon and J. Haverkamp, Biomed. Mass Spectrom., 11 (1984) 486.

Chapter 10

CHARACTERIZATION OF ALGAE BY PYROLYSIS–DIRECT CHEMICAL IONIZATION (TANDEM) MASS SPECTROMETRY

J. de Waart, A.C. Tas , G.F. La Vos and J. van der Greef

TNO-CIVO Food Analysis Institute, P.O. Box 360, 3700 AJ Zeist (The Netherlands)

ABSTRACT

Several marine algae, belonging to the classes of brown, red and green algae were investigated by pyrolysis–direct chemical ionization mass spectrometry. Factor-discriminant analysis techniques applied to the pyrograms enabled differentiation between these classes. Pyrograms indicated relatively high levels of mannitol in brown algae and relatively high levels of methylhexose in red algae as was confirmed by tandem mass spectrometry. *Codium fragile*, a green algal strain, was characterized by higher intensities of peptide markers in its pyrogram. Analytical data obtained from these peptide markers were compared with those from *Spirulina madras*, a micro-algal species.

INTRODUCTION

The exploitation of algae for the production of commercially important products is increasingly receiving attention. In addition to the production of alginic acid, agar-agar, furcellaran and carrageenan, recent interest has focussed on the isolation of bioactive components [1], such as antitumour compounds [2], vermicides and vitamins [3, 4].

For selection of productive strains and determination of optimum culturing conditions for the production of desired compounds, sophisticated analytical techniques are essential and rapid screening techniques are advantageous.

Pyrolysis—mass spectrometry (Py—MS) is a versatile technique for this type of analytical monitoring [5, 6]. The technique is fast, requires only minimum sample work-up, and enables the simultaneous detection of a wide range of chemical compounds. The scope of this technique has been further extended with the introduction of soft ionization detection techniques such as field ionization (FI) [6] and direct chemical ionization mass spectrometry (Py—DCI/MS) [7, 8]. Owing to the use of a temperature ramp, in-source heating of complex (bio)polymeric materials results in desorption as well as pyrolytic processes [9, 10]. A wide range of chemical components, varying with respect to polarity and molecular mass, can be monitored this way.

After its initial introduction as a pyrolytic technique in the analysis of starch [7] the potential of Py—DCI/MS was investigated in the characterization of organisms, such as bacteria [8, 10] and fungi [11, 12], and of virus-infected cell cultures [13], in the differentiation of healthy and diseased women suffering from premenstrual syndrome [14] and in the pyrolysis of polysaccharides [11]. Recently, a further extension of its potential was accomplished through the application of tandem mass spectrometry (Py—DCIMS/MS) [11, 15] to the structural analysis of pyrolytic products from polyhexoses of different linkage types. Intact and more or less dehydrated oligosaccharide units originating from the polymeric chain were detected in the pyrolysate. The presence of oligomeric products in pyrolysates provides a key for obtaining sequence-related information from biopolymers by MS/MS. Series of such oligomeric pyrolysis products have also been detected in DCI pyrolysates of cellulose [16] and citrus and apple pectins [17].

In this report the profiling by Py—DCI/MS of a number of red and brown algae and one green alga will be described. Principal component-discriminant (PC-D) analysis [18] was applied for display and for tracing the mass intensities of compounds relevant to strain differentiation. A number of these masses were investigated by tandem mass spectrometry (Py—DCIMS/MS).

EXPERIMENTAL

Samples

The red algae *Porphyra umbilicalis*, *Gigartina stellata*, *Chondrus crispus* and *Pterocladia*, the brown algae *Fucus serratus*, *F.spiralis*, *F.vesiculosus*, *Ascophyllum nodosum*, *Sargassum muticum* (1985 and 1986) and the green alga *Codium fragile* were dried under a 150 W Elstein infrared lamp at 40 - 50°C for a period of about 48 h. The micro-algal strain *Spirulina madras* was obtained from Dr. W. Becker, Universität Tübingen, Germany.

Before to analysis samples were milled in an ultracentrifugal Retch mill (type ZM-1) and suspended in methanol.

The reference compounds mannitol and bovine serum albumin (BSA) were obtained from Sigma Chemical Company, St. Louis, MO.

Pyrolysis—mass spectrometry

Py—DCI/MS was performed on a Finnigan-MAT 8230 mass spectrometer coupled to a Finnigan-MAT SS300 data system. Ammonia was used as reactant gas at an indicated source pressure of 50 Pa; the temperature of the ion source was kept at 240°C. Sample amounts of approximately 100 μg were applied to the wire. During heating of the DCI wire by linear current programming (8 mA/s) spectra were acquired over the mass range *m/z* 60-1000, at a scan speed of 0.7 s/decade. After pyrolysis the wire was cleaned by heating to a maximum temperature of approximately 1200 °C. Spectra obtained during pyrolysate formation were averaged to one profile. All strains were analysed in triplicate.

High-resolution Py—DCI/MS was conducted on the same apparatus at a resolution of 10000.

Pyrolysis—tandem mass spectrometry

Py—DCIMS/MS was carried out on a Finnigan-MAT HSQ-30 mass spectrometer (BEQQ geometry) coupled to a Finnigan-MAT SS300 data system. Except for the scanning parameters, the same conditions as described above were maintained in the sector field mass spectrometer which was used as the first mass analyser. Collision activation was carried out in the RF-only quadrupole; the collision energy was set at approximately 45 V. Argon was used as collision gas.

Data analysis

Data analysis was performed with the pattern recognition package ARTHUR. This package has been extended with PC-D analysis and graphic rotation routines developed by the FOM Institute, Amsterdam, as well as routines developed at TNO for conversion of spectral data into ARTHUR format, for conversion of factor spectral data into SS300 spectrum format and for construction of comparable plots of objects and variables (BIPLOT routines) [12]. Factor spectra represent the correlation of the original masses with the discriminant axis, indicating the relevance of the masses to the differentiation observed.

Each spectrum was reduced to a pattern of 77 mass intensities. Fisher weights were used as selection criterion; triplicate measurements were defined as categories. Differences in sample size were corrected for by normalization to 100% total ion current, excluding masses with the highest intensities. The data set was transformed by autoscaling (zero mean and unit standard deviation) to perform PCA on a correlation about the mean matrix. Subsequently, PCs (autoscaled score vectors) were orthogonally rotated to discriminant functions (PC-D analysis).

RESULTS AND DISCUSSION

A characteristic of the Py–DCI/MS technique, the capacity to detect polar and higher molecular mass pyrolysis products [7-8], is attractive for the monitoring of chemical components in algae. For example, in the pyrogram of *Fucus vesiculosus*, a brown alga (Fig. 1), the highly polar polyol compound mannitol (the $[M + NH_4]^+$ and $[M + NH_4-H_2O]^+$ ions at m/z 200 and 182) features prominently. The intensities at m/z 164 and 146 are indicative of the presence of deoxyhexose residues ($[M + NH_4-H_2O]^+$ and $M + NH_4-2H_2O]^+$) such as fucose, although mannitol residues subject to further dehydration may also contribute to these ions. The ability to detect highly polar compounds is illustrated by the occurrence of mass intensities at m/z 180 and 342 from monomeric and oligomeric hexoses [9]. In addition, the pyrogram reveals the presence of mono- and diglyceride residues probably originating from phospholipids, as reflected by some of the peaks in the mass range m/z 300-400 and m/z 550-650. For example, the intensities at m/z 348 and 374 indicate the $[M + NH_4]^+$ ions of 16:0 and 18:1 monoglycerides; the intensities at m/z 584 and 612 point to 16:0-16:1 and 16:0-18:1 diglycerides, respectively. The alcohol *trans*-3,7,11,15-tetramethylhexadec-2-en-1-ol (phytol), which frequently occurs in algae either in a free form or as the alcohol moiety of chlorophyll, gives rise to m/z 296 as the $[M + NH_4-H_2O]^+$ ion. A number of these indicated structures were substantiated by Py–MS/MS measurements, as will be
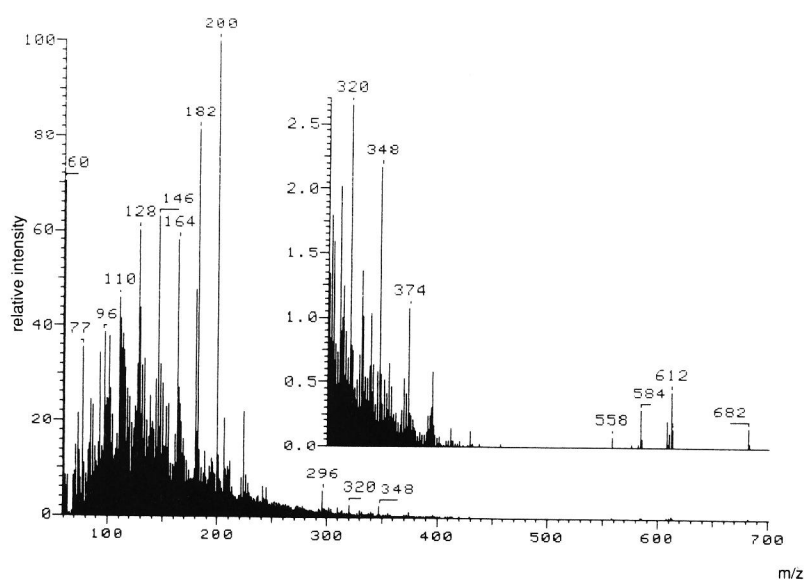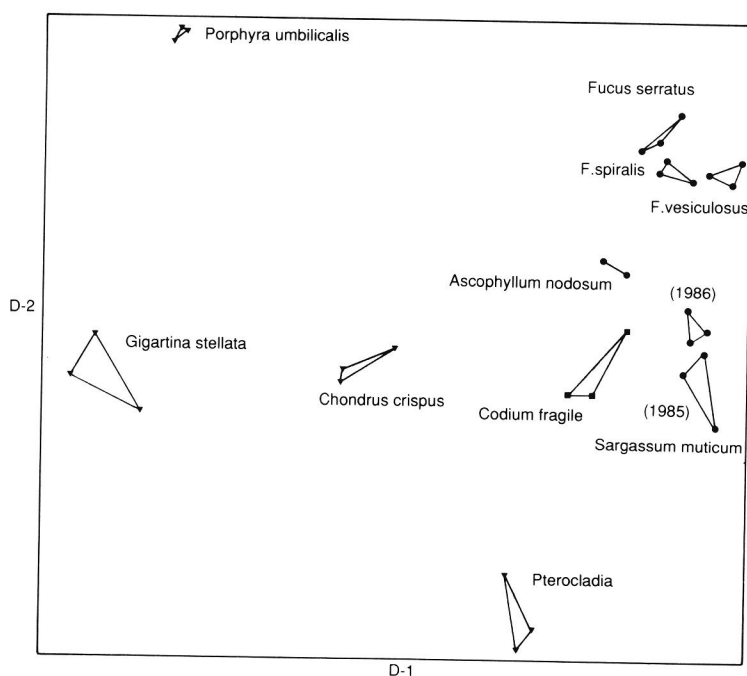
Fig. 1   DCI pyrogram of *Fucus vesiculosus*

Fig. 2   Discriminant plot (D_1 v/s D_2) of algal strains. Brown algae (●), Red algae (▼) and Green alga (■)
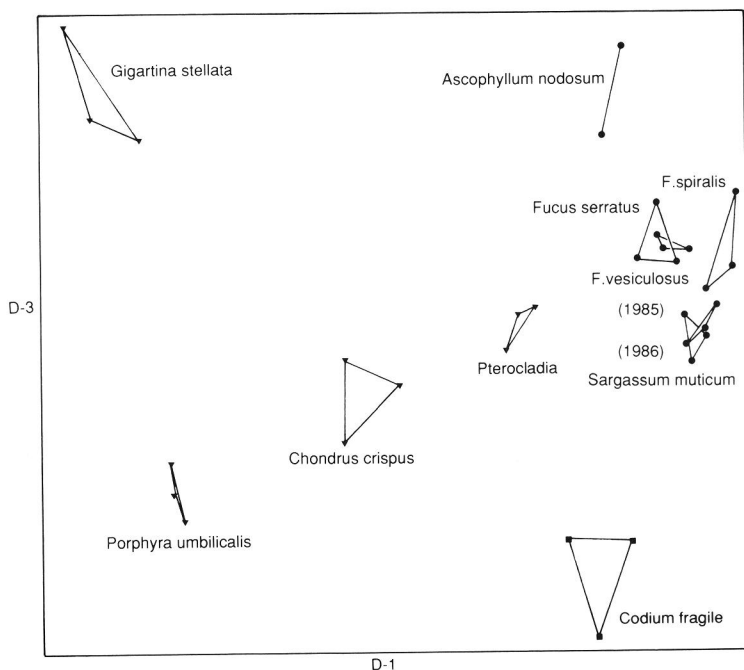
Fig. 3   Discriminant plot (D_1 v/s D_3) of algal strains. Brown algae (●), Red algae (▼) and Green alga (■)
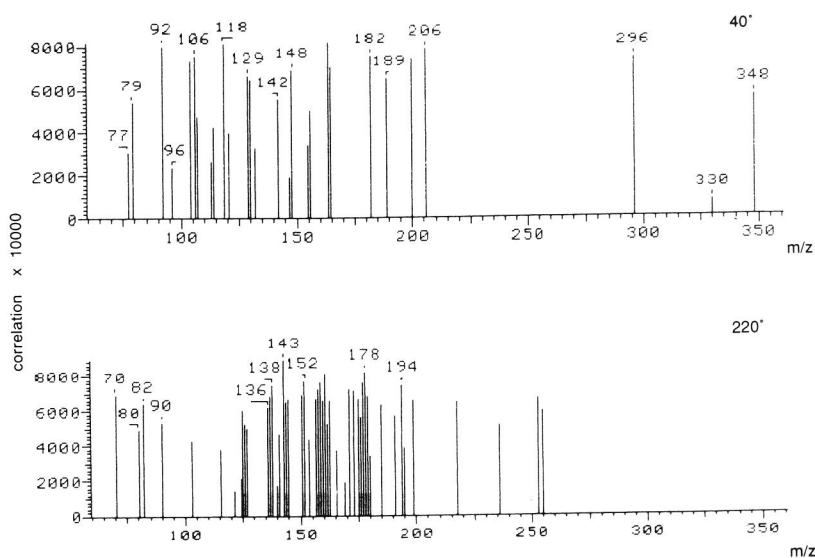

Fig. 4   Factor spectra (mass-discriminant function correlation) after counterclockwise rotation of the first discriminant function over 40° and 220°.

discussed in connection with the factor spectra obtained after discriminant analysis.

In the discriminant plot red and brown algae can clearly be distinguished (Fig. 2). Within the division of brown algae the *Fucus* species and *Sargassum muticum* (1985 and 1986) tend to cluster. *Codium fragile*, the only green alga strain investigated, can be discriminated from brown and red algae in the D-plot 1 vs. 3 (Fig. 3). The discriminant analysis applied was of an unsupervised nature because groups were defined consisting of triplicate measurements and no further information concerning the algal divisions was used in the analysis. The fact that brown and red algae are located in different areas of the plot indicates that different (bio)chemical factors are present which are typical of these divisions.

Factor spectra in the 40° and 220° direction, in the D-plot 1 vs. 2 (Fig. 4), reflect the mass series responsible for differentiation between the red and brown algae. For example, the masses m/z 182 and 200 are positively correlated with the 40° function, indicating the relatively high levels of mannitol in brown algae. The daughter spectra of these ions (Fig. 5 and 6) were indentical to the spectra of the reference compound. In the daughter ion spectrum of m/z 200, formation of m/z 183 is prominent owing to loss of NH_3. Subsequently, repeated water losses from the latter fragment are observed giving rise to the fragments at m/z 165, 147, 129 and 111. Fragments at m/z 99 and 69 are formed from m/z 129 by ring fragmentation with losses of formaldehyde. Comparable fragmentation processes starting at the ions m/z 147 may generate the ions at m/z 87 and 57. The highly comparable daughter ion spectrum of m/z 182 indicates that this is probably the $[M+NH_4-H_2O]^+$ ion of mannitol.

In the higher mass range of the 40° factor spectrum positive correlations are found for the intensities at m/z 206, 330, 348 and for m/z 296. The daughter spectrum of m/z 206 (Fig. 7), showing a ketene loss from m/z 189 to 147, points to an acetylated deoxyhexose unit. Most of the fragments observed, such as m/z 147, 129, 111, 85, 81 and 69, are common in deoxyhexose spectra. The intensities at m/z 330 and 348 probably originate from the $[M+NH_4-H_2O]^+$ and $[M+NH_4]^+$ ions of 16:0 monoglyceride. This structure was confirmed by the daughter spectrum of m/z 348 (Fig. 8). In this spectrum the fragment at m/z 239 points to a 16:0 acyl cation, while fragments of the alkyl chain are observed at m/z 57, 71, 85, 99 and 113. For the ion at m/z 296, the daughter ion spectrum points to the presence of a longer alkenyl chain (Fig. 9). The fragment series starting at m/z 55 and 57, with doublets up to m/z 197, is indicative. The ion at m/z 197, probably formed by the combined loss of NH_3 and hexadiene (C_6H_{10}), indicates that the ion at m/z 296 may well be the $[M-H_2O+NH_4]^+$ ion of phytol ($C_{20}H_{39}OH$).

The factor spectrum in the 220° direction contains mass intensities of compounds present in relatively high levels in red algae (Fig. 4). In the higher mass range, intensities at m/z 253 and 255 may point to the $[M+H]^+$ ions of the related
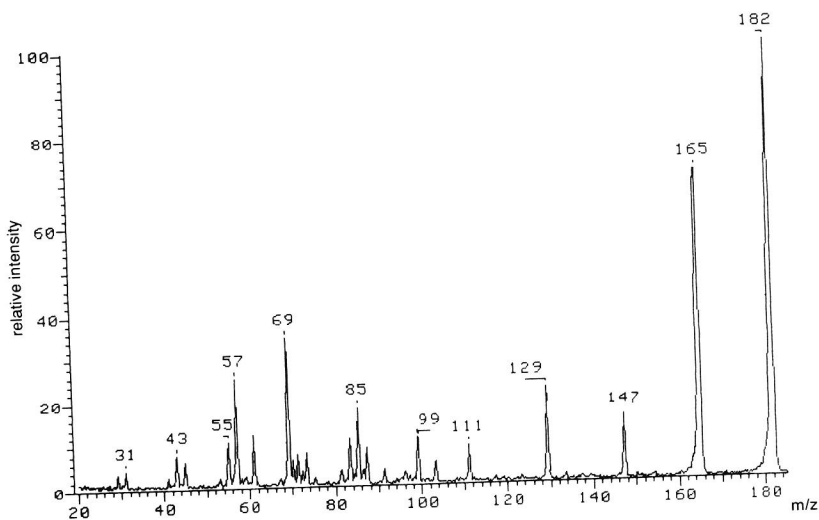
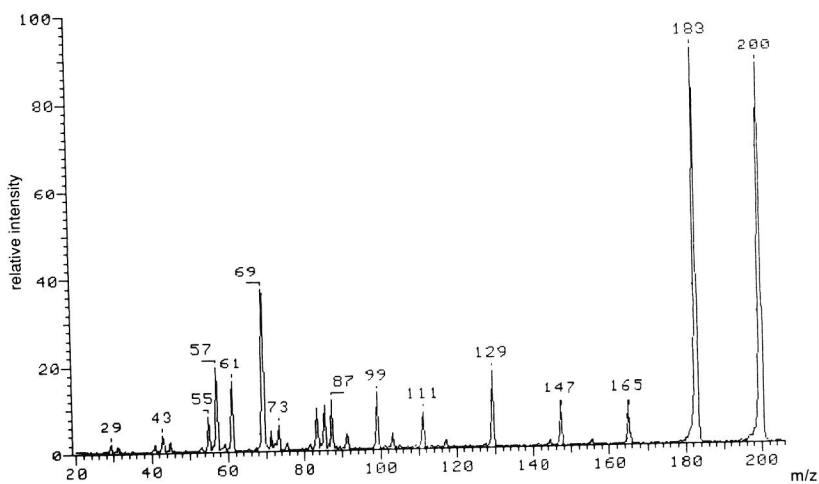Fig. 5   Py–DCIMS/MS spectrum of *m/z* 182 of *Fucus vesiculosus*

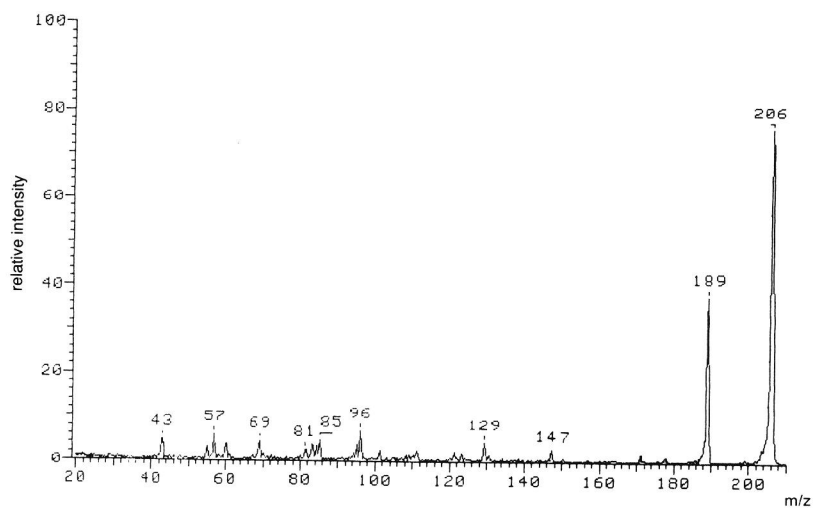Fig. 6   Py–DCIMS/MS spectrum of *m/z* 200 of *Fucus vesiculosus*

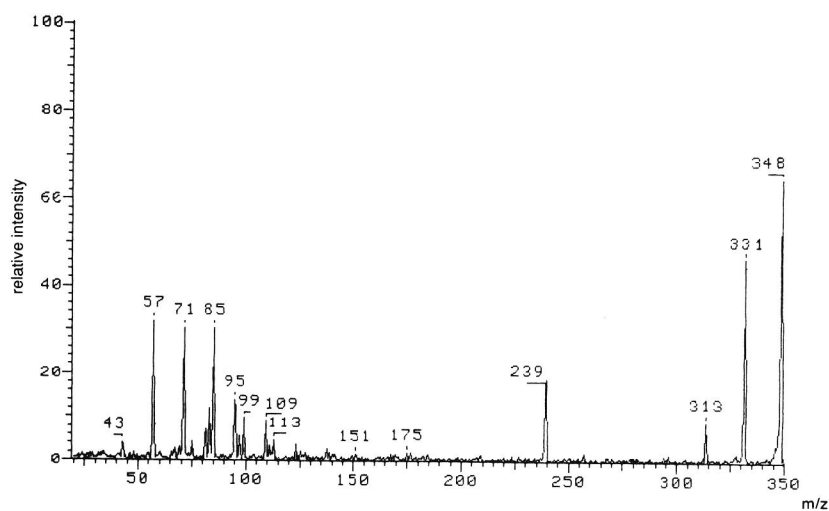Fig. 7   Py–DCIMS/MS spectrum of *m/z* 206 of *Fucus vesiculosus*

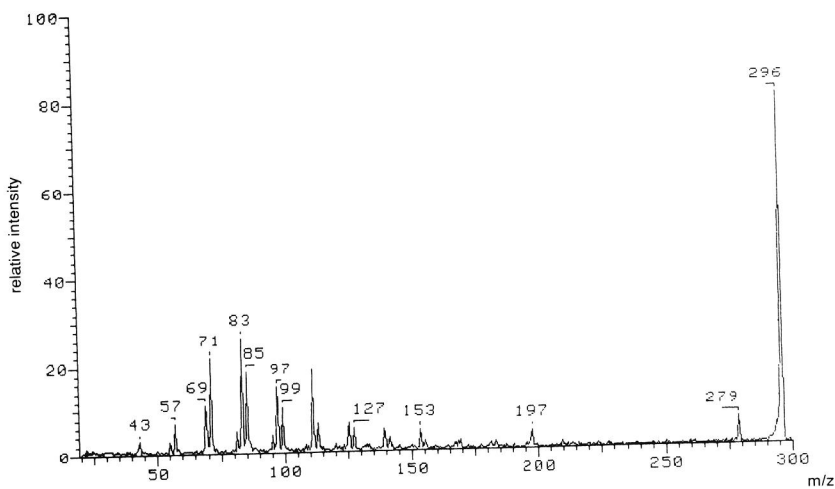Fig. 8   Py–DCIMS/MS spectrum of *m/z* 348 of *Fucus vesiculosus*

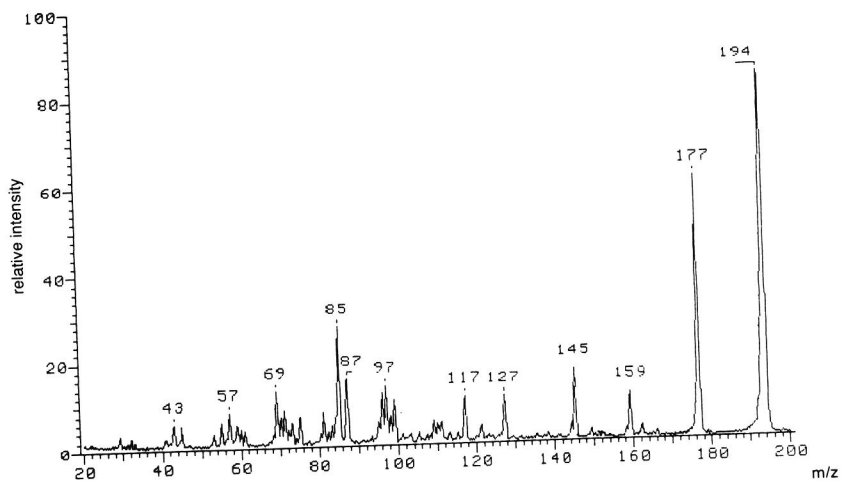Fig. 9   Py–DCIMS/MS spectrum of *m/z* 296 of *Fucus vesiculosus*

Fig. 10   Py–DCIMS/MS spectrum of *m/z* 194 of *Gigartina stellata*

compounds dehydrofloridoside and floridoside, respectively. In the original spectrum of *Gigartina stellata* the $[M + NH_4]^+$ ions at m/z 270 and 272 may also support this suggestion. However, daughter ion spectra of these compounds did not provide conclusive evidence for this structure.

In the lower mass range m/z 194 could either be ascribed to a uronic acid or to a methylated hexose. The daughter ion spectrum (Fig. 10) points strongly to the latter structure as can be inferred from the observed loss of 32 amu (CH_4O) from m/z 177 and from the dominant hexose nature of the fragments at m/z 145, 127, 97, 87 and 85 [9].

The structures of compounds represented by clusters of mass intensities in the lower mass range (Fig. 4) can be indicated only tentatively or are not yet known. m/z 162 and 180 are present at higher intensities in the pyrograms of the red algae investigated and can be attributed to hexose residues; m/z 136 and 152 may point to the $[M + H]^+$ ions of adenine and guanine, respectively. Additional Py–MS/MS investigations are needed for the confirmation of these assignments and for additional identification of unknown compounds.

As mentioned above, *Codium fragile* is differentiated from the brown and red algae along the third D-function (Fig. 3). The corresponding factor spectrum reveals the positive correlation of the ions m/z 141, 155 and 169 with this function, which reflects their relatively high intensities in the pyrogram of this strain. The daughter spectra (Figs. 11-13) of the ions at m/z 141, 155 and 169 indicate close structural similarity. The fact that these ions were also observed in the pyrogram of *Spirulina madras* as part of a series of homologous compounds at m/z 99, 113, 127, 141, 155, 169, 183, 197 and 211 (Fig. 14), prompted us to compare analytical data from these strains [19]. The daughter spectra from m/z 169 of both strains were found to be similar. Moreover, high-resolution (HR/MS) measurements yielded a mass of 169.134 for both strains, most likely corresponding to an elemental composition of $C_9H_{17}N_2O$. Extending this result to the lower homologues the elemental compositions of m/z 155 and 141 are probably $C_8H_{15}N_2O$ and $C_7H_{13}N_2O$, respectively. The possibility that one of the nitrogen atoms originated from ammonia adduct ion formation was ruled out by the observation of the same homologous ion series in the pyrogram of the *Spirulina* strain under methane chemical ionization conditions. Daughter spectra and HR/MS results provide an indication that these pyrolytic products originate from algal proteins. The proposed compounds and the possible route for their formation during pyrolysis are indicated in Fig. 15.

After ring closure, cationization of the amide nitrogen of the intermediate imidazolidon compounds (II) can promote the expulsion of the amide side chain. Metal cations as well as protons may act as acidic species. The other side chain, connected to the ring nitrogen by a weak allylic bond (induced by two $C = O$ groups and a $C = N$ group), can be split off either by a homolytic cleavage as indicated, or
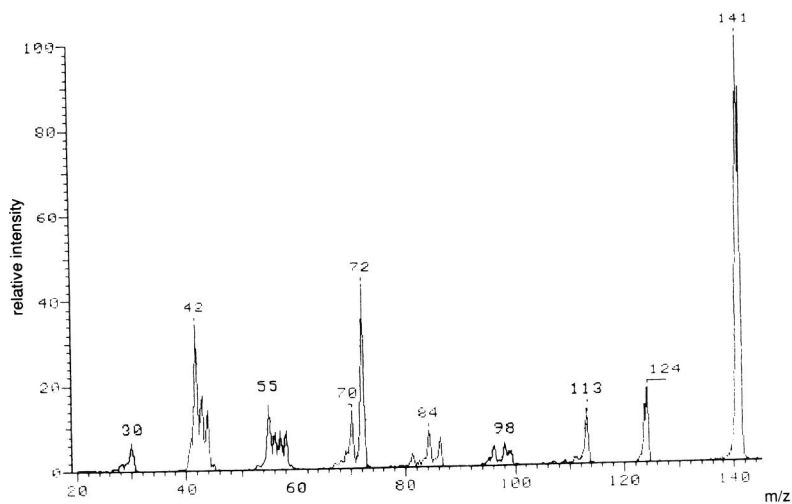
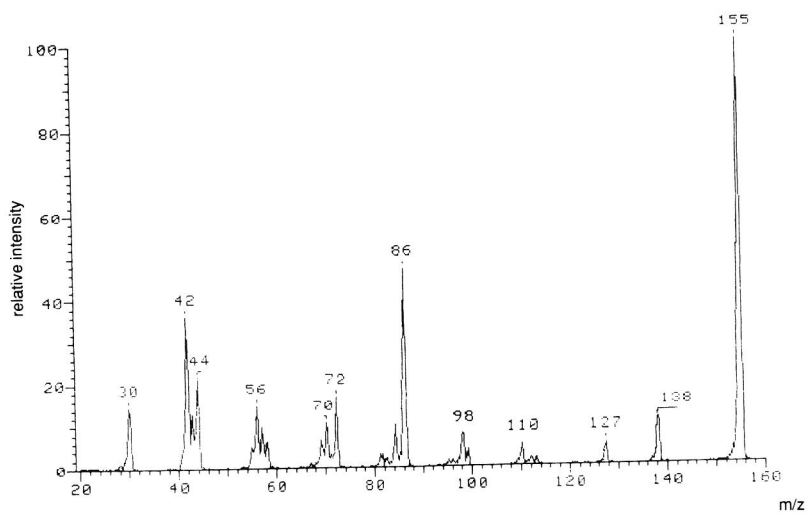Fig. 11   Py–DCIMS/MS spectrum of *m/z* 141 of *Codium fragile*

Fig. 12   Py–DCIMS/MS spectrum of *m/z* 155 of *Codium fragile*
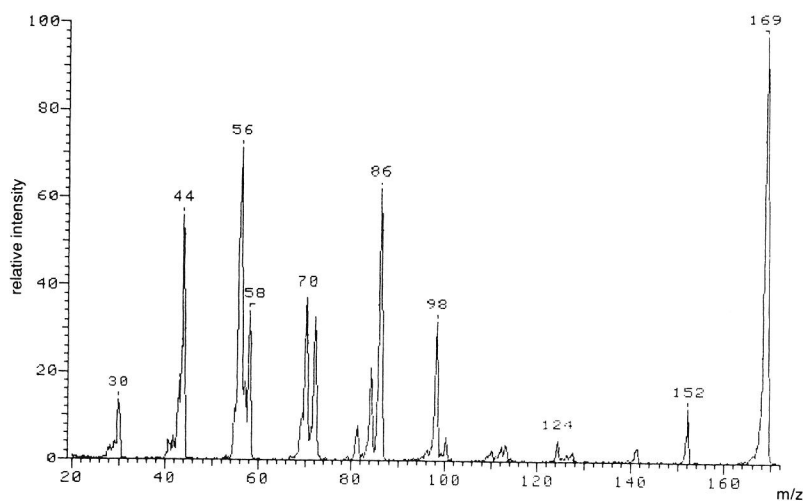
Fig. 13   Py–DCIMS/MS spectrum of *m/z* 169 of *Codium fragile*
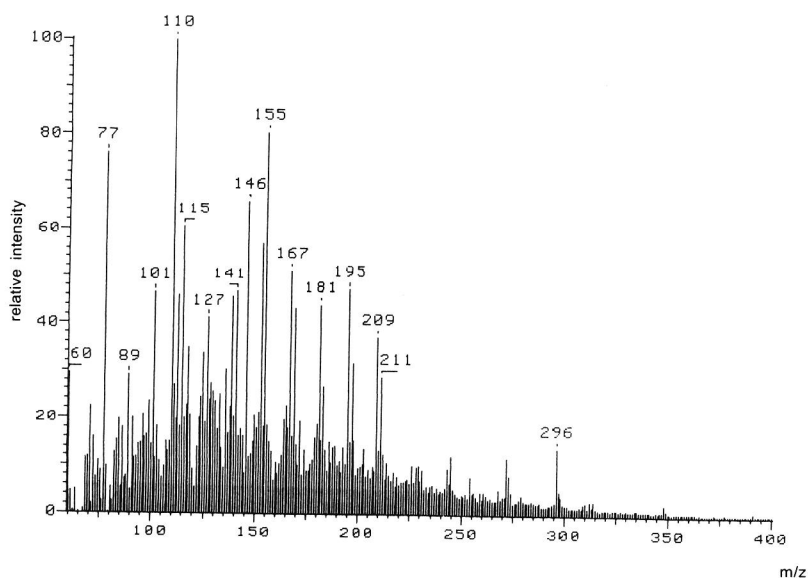

Fig. 14   DCI pyrogram from *Spirulina madras*

by heterolytic elimination. The resulting alkylidene compounds (III) can be reduced under chemical ionization conditions to type IV compounds [20]. The ring systems of III and IV, containing two nitrogen atoms and one oxygen atom, will have various side-chains, depending on the (two) amino acids engaged. The ion series at m/z 99, 113, 127, 141, 155, 169, 183, 197 and 211 (Fig. 14) can be completely explained through compounds (IV) by the combinations of Gly, Ala, Val, Leu and/or Ile. For example, m/z 99: Gly-Gly (R1 = R2 = H); m/z 127: Ala-Ala ($R_1 = R_2 = CH_3$); m/z 141: Gly-Val ($R_1 = H$, $R_2 = i\text{-}C_3H_7$); m/z 155: Ala-Val ($R_1 = CH_3$, $R_2 = i\text{-}C_3H_7$); m/z 169: Ala-Leu ($R_1 = CH_3$, $R_2 = i\text{-}C_4H_9$); m/z 183: Val-Val ($R_1 = R_2 = i\text{-}C_3H_7$); m/z 197: Val-Leu ($R_1 = i\text{-}C_3H_7$, $R_2 = i\text{-}C_4H_9$); m/z 211: Leu-Leu ($R_1 = R_2 = i\text{-}C_4H_9$). For most of the ions a variety of isomers can exist. Firstly, the combination of different amino acids can result in isomers with the same molecular mass, and secondly, interchange of the sequence of two amino acids engaged in the ring formation is possible (interchange of the substituents on R_1 and R_2).

In the pyrogram of *S. madras* the ion series corresponding to the proposed unsaturated compounds III, at m/z 111, 125, 139, 153, 167, 181, 195 and 209, is also prominent (Fig. 14). High-resolution measurements of some of the ions of this series, m/z 167, 195 and 209, revealed that the elemental compositions are $C_9H_{15}N_2O$, $C_{11}H_{19}N_2O$ and $C_{12}H_{21}N_2O$, respectively. One of these ions, m/z 195, was also found in the factor spectrum of D-3 in the direction of *C. fragile*.

In addition to ions originating from alifatic amino acids only, markers for combinations of phenylalanine with different alifatic amino acids are also present in this pyrogram. For example, the intensities at m/z 203, 231, 245 may indicate Ala-Phe, Val-Phe, Ile-Phe and Leu-Phe combinations of type IV.

The assignment of the ion series of type III and IV to protein pyrolysis products was also substantiated by the observation of such series in the pyrogram of BSA.

It is interesting to note that several compounds with molecular weights of m/z 195 and 209 have been detected in Py–GC/MS analyses of proteinaceous materials of different origins, using electron impact as well as chemical ionization detection [21]. These compounds were found to originate from combinations of valine, leucine and isoleucine. However, a different elemental composition ($C_{12}H_{19}NO_2$) was found for the ion at m/z 209.

As stated before, most ions from compounds III and IV can be expected to consist of a mixture of isomers. This may complicate an unequivocal interpretation of the daughter spectra. Indications for the structure of the daughter ions can be inferred. In the daughter ion spectra of m/z 169 and 155 a series of even ions is present at m/z 98, 86, 84, 72, 70, 56, 44, 42 and 30, which points to protonated nitrogen-containing compounds. Because of the isomeric mixtures contained in the parent ions it is not apparent from these daughter spectra if fragments are formed directly or consecutively. Possible elemental compositions for the main ions in the
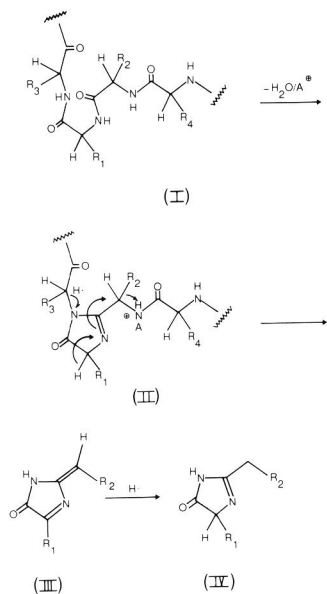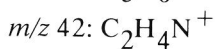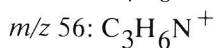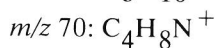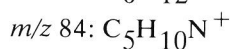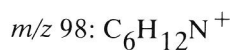
Fig. 15   Proposed route for the formation of 2-alkylidene-imidazole-4-one (III) and 2-alkyl-imidazole-4-one (IV) compounds

Alkyl cyanides

m/z 98: $C_6H_{12}N^+$
m/z 84: $C_5H_{10}N^+$
m/z 70: $C_4H_8N^+$
m/z 56: $C_3H_6N^+$
m/z 42: $C_2H_4N^+$

Alkyl imines

m/z 100: $C_6H_{14}N^+$
m/z 86: $C_5H_{12}N^+$
m/z 72: $C_4H_{10}N^+$
m/z 58: $C_3H_8N^+$
m/z 44: $C_2H_6N^+$
m/z 30: CH_4N^+
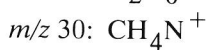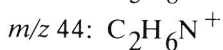
Fig. 16   Proposed daughter ions from the protonated pyrolysis compounds m/z 169, 155 and 141

128

spectra are indicated in Fig. 16. The fragments at m/z 98, 84, 70, 56 and 42 may point to protonated alkylcyanides, originating from leucine or isoleucine, valine or alanine residues. From the proposed structure IV it can be inferred that each amino acid may provide two homologous alkyl cyanide ions with 14 amu difference, depending on the position of the amino acid res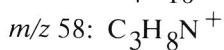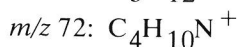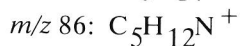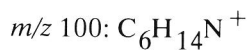idue in the ring (same alkyl group on R_1 or R_2). The fragments at m/z 86, 72, 58 and 30 can be explained by analogy from protonated alkyl imines. In the higher mass range of the spectra small intensities of the M-17 ions may be due to loss of ammonia from the parent ions, occurring to a limited extent. Also losses of $C=O$ can be observed, either directly from the parent ion or from the M-17 fragment ion.

Several investigators have already pointed out the complexity of protein pyrolysates [21-23]. Factors that add to this complexity are the large number of possible isomers arising from the structural differences of amino acids, the varying stabilities of the peptide bonds and the differences in three-dimensional structures of proteins.

CONCLUSIONS

Py–DCI/MS profiling followed by MS/MS characterization of relevant pyrolysate components offers an improved analytical approach to the characterization of complex mixtures. Mixtures of lower, medium and high molecular weight compounds, as present in cells and cell walls, can be analysed in more detail.

In general, an extended storage capacity of pyrolytic profiles is required to accommodate the large amount of chemical information that arises from the complexity of biological systems. It is therefore essential to obtain profiles over an extended mass range and to impose limitations on fragmentation processes, either induced by pyrolysis or by ionization. Py–DCI/MS largely complies with these conditions.

ACKNOWLEDGEMENT

REFERENCES

1. M.A. Ragan and C.J. Bird (Editors), Developments in Hydrobiology, 12th International Seaweed Symposium, Dr. W. Junk Publishers, Dordrecht, Boston, New York, 1987.

2. J. de Waart and J. van Putten, in S. Paoletti and G. Blunden (Editors), Proceedings of a Workshop on Phycocolloids and Fine Chemicals, COST 48, Triest, Commission of the European Communities, Brussels, 1987, p. 254.

3. H.A Hoppe, T. Levring and Y. Tanaka (Editors), Marine Algae in Pharmaceutical Science, Walter de Gruyter (1979) Berlin, New York.

4. H. van den Berg, P.C. Dagnelic and W.A. van Staveren, Lancet, 8679 (1988) 242.

5. H.L.C. Meuzelaar, J. Haverkamp and F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials, Elsevier, Amsterdam (1982).

6. H.-R. Schulten, N. Simmleit and R. Müller, Anal. Chem., 59 (1987) 2903.

7. J. Metzger, Fresenius Z. Anal. Chem., 308 (1981) 29.

8. A.C. Tas, J. van der Greef, J. de Waart, J. Bouwman and M.C. ten Noever de Brauw, J. Anal. Appl. Pyrolysis, 7 (1985) 249.

9. R. Stoll and F.W. Roellgen, Org. Mass Spectrom., 16 (1981) 72.

10. A.C. Tas, J. de Waart, J. Bouwman, M.C. ten Noever de Brauw and J. van der Greef, J. Anal. Appl. Pyrolysis, 11 (1987) 329.

11. A.C. Tas, A. Kerkenaar, G.F. La Vos and J. van der Greef, J. Anal. Appl. Pyrolysis, 15 (1989) 55.

12. A.C. Tas, H.B. Bastiaanse, J. van der Greef and A. Kerkenaar, J. Anal. Appl. Pyrolysis, 14 (1989) 309.

13. A.C. Tas, J. Odink, J. van der Greef, M.D. Ferrari, L. van Ekdom, A.C.B. Peters and W. Boogerd, Biomed. and Environ. Mass Spectrometry, 18 (1989) 757.

14. A.C. Tas, H. van den Berg, J. Odink, H. Korthals, J.T.N.M. Thissen and J. van der Greef, J. Pharm. Biomed. Analysis, 7 (1989) 1239.

15. R.J. Helleur, R. Guevremont, J. Anal. Appl. Pyrolysis, 15 (1989) 85.

16. A.D. Pouwels, G.B. Eijkel, P.W. Arisz and J.J. Boon, J. Anal, Appl. Pyrolysis, 15 (1989), 71.

17. J.J. Boon, M. Nip, G.B. Eijkel, B. Brandt-de Boer and M. van Amsterdam, Proceedings of the 11th Int. Mass Spectrometry Conf., Bordeaux, (1988).

18. W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81.

19. J. de Waart, A.C. Tas and J. van der Greef, unpublished results.

20. K.P. Madhusudanan, V.S. Murthy and D. Fraisse, Org. Mass Spectrom., 22 (1987) 665.

21. J.J. Boon and J.W. de Leeuw, J. Anal. Appl. Pyrolysis, 11 (1987) 313.

22. T.O. Munson and D.D. Fetterolf, J. Anal. Appl. Pyrolysis, 11 (1987) 15.

23. A. Ballistrieri, M. Guiffrida and G. Montaudo, J. Polym. Sci., 23 (1985) 1731.
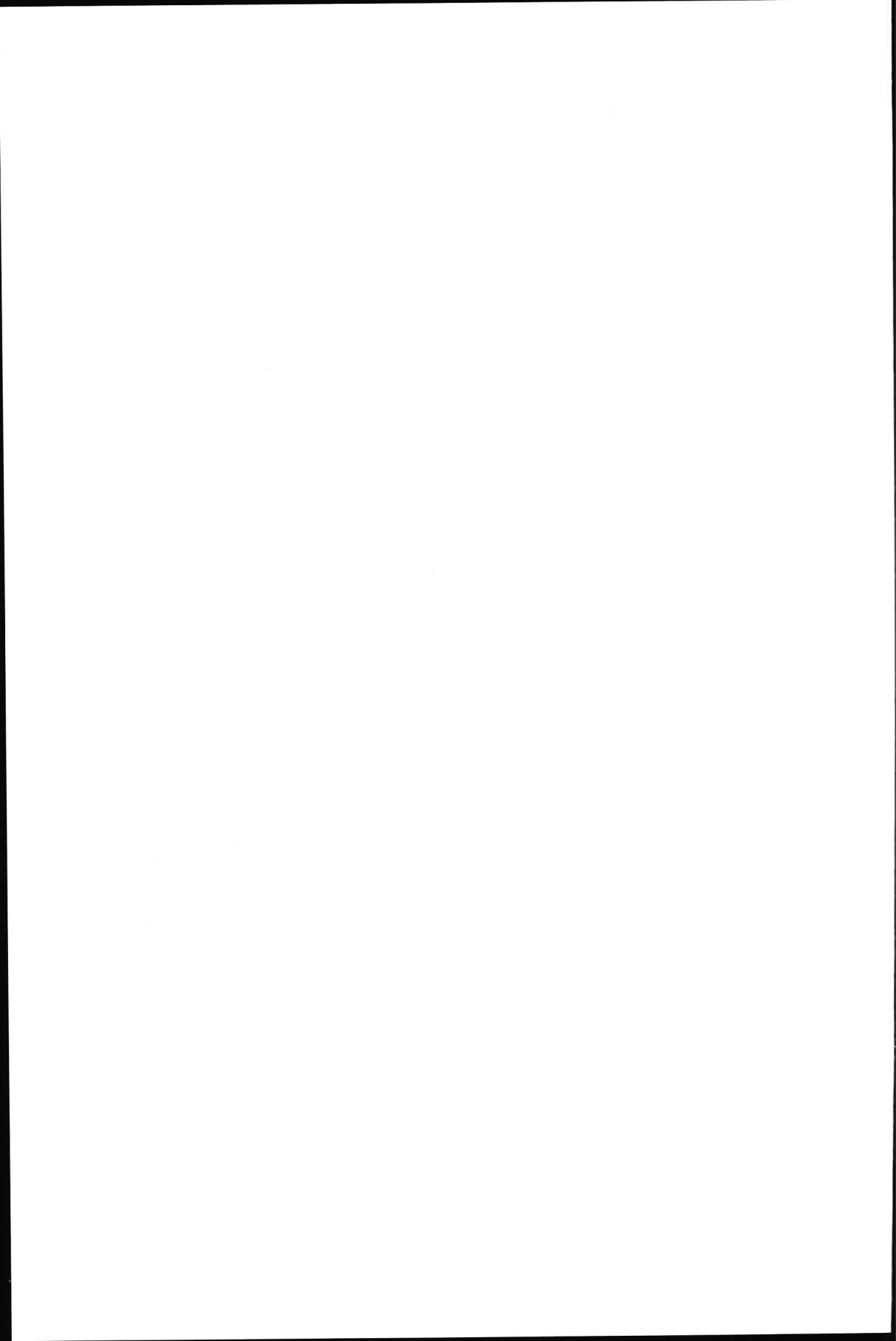
PART III
MATHEMATICAL INTRODUCTION
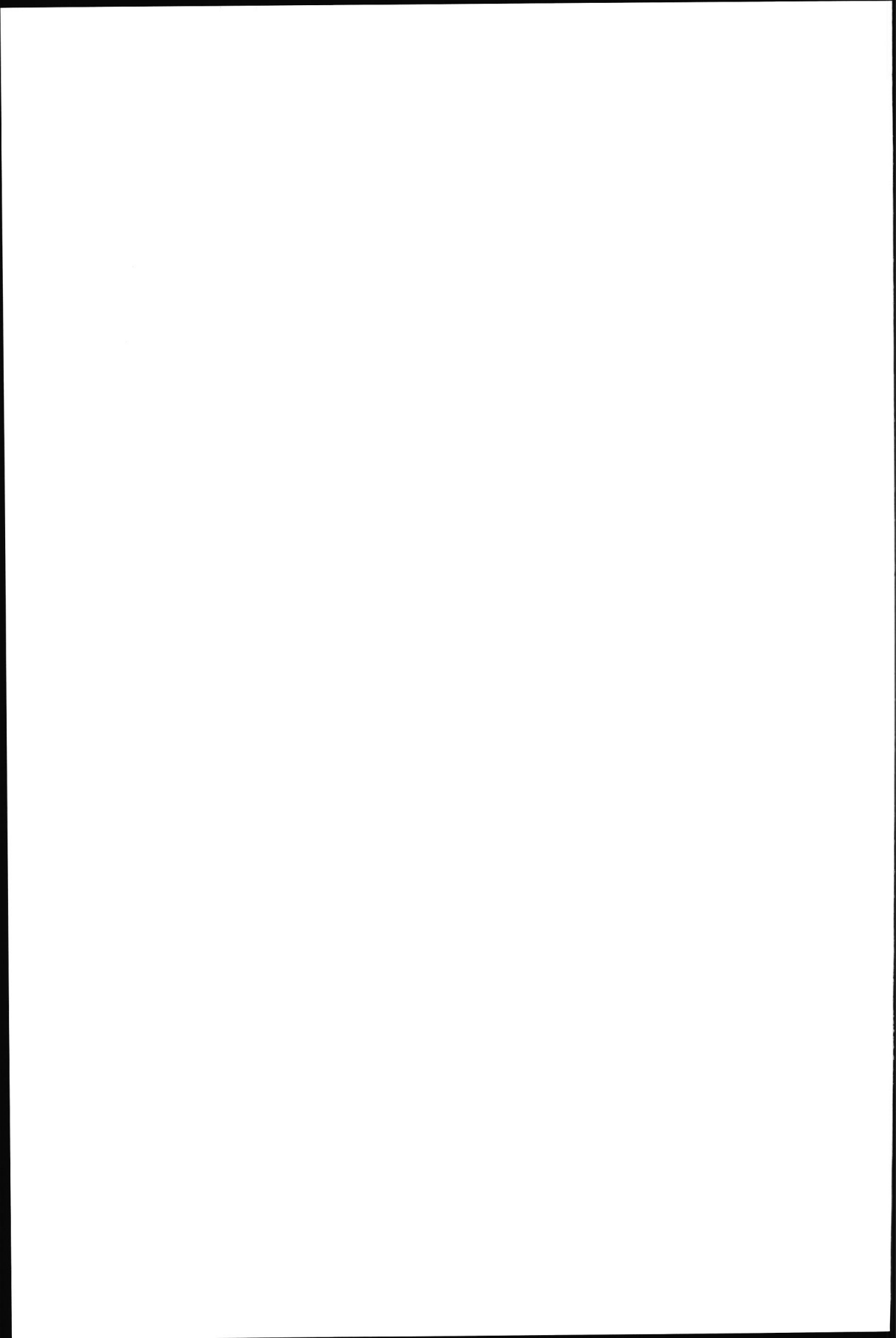
Chapter 11

MATHEMATICAL INTRODUCTION TO PRINCIPAL COMPONENTS AND DISCRIMINANT ANALYSIS

1. PRINCIPAL COMPONENTS ANALYSIS

1.1 Mathematical representation

Principal components analysis (PCA) is a method of converting a $(n \times q)$ data matrix $\mathbf{X}$, consisting of q column vectors $\mathbf{x}$, into a set of transformed orthogonal vectors $\mathbf{t}_h$, so-called factor scores vectors or PCs. These vectors represent the variance contained in the data set more efficiently in that the vectors $\mathbf{t}_h$ are ranked according to decreasing variance. In most cases a smaller number of these vectors than the original variables is needed to reproduce or approximate the original data set $\mathbf{X}$.

A basic idea in PCA is that a data matrix $\mathbf{X}$ $(n \times q)$ can be written as a sum of n or q rank 1 $(n \times q)$ matrices $\mathbf{A}_h$, depending on the rank of matrix $\mathbf{X}$ [1, 2]. In this chapter it is assumed that matrix $\mathbf{X}$ is of rank k or rank q $(n > q > k)$. Matrices $\mathbf{A}_h$ are products of the factor scores vector $\mathbf{t}_h$ and the factor loadings vector $\mathbf{p}^T_h$, the transpose of the eigenvector $\mathbf{p}_h$:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}^T_1 + \mathbf{t}_2\mathbf{p}^T_2 + ... + \mathbf{t}_q\mathbf{p}^T_q \tag{1}$$
(Fig. 1)

$$\mathbf{X} = \mathbf{A}_1 + \mathbf{A}_2 + ... + \mathbf{A}_q \tag{2}$$
(Fig. 2)

If the data matrix $\mathbf{X}$ is of reduced rank k $(k < q)$ the matrix is also completely reproduced by the sum of products of the first k scores- and loadings vectors. Because the variance of the last $q - k$ eigenvectors is equal to zero, it is clear that the n profile vectors are confined to a k-dimensional subspace of the q-dimensional space. However, in most practical situations where PCA is applied, measurement error is present in the data set causing a larger number of eigenvectors than is required to span the real subspace, making the rank of the matrix more or less diffuse.

In that case the last $q - k$ eigenvectors will represent error. By deleting these error eigenvectors [1] data matrix $\mathbf{X}$ can be approximated in a least-squares sense [2] by the sum of the products of the first k scores and loadings vectors:

$$\hat{\mathbf{X}} = \mathbf{t}_1\mathbf{p}^T_1 + \mathbf{t}_2\mathbf{p}^T_2 + \ldots + \mathbf{t}_k\mathbf{p}^T_k \tag{3}$$

In doing so, part of the error of $\mathbf{X}$ (extracted error, (XE) is removed from the data set, resulting in an improved data matrix $\hat{\mathbf{X}}$ upon reconstruction [1]. This will be further discussed in section 1.2 of this chapter.

The set of eigenvectors required to span the space is calculated as follows. The first eigenvector $\mathbf{p}_1$ indicates a linear combination of the q original variables (direction cosines in the q-dimensional space), such that the variance of

$$\mathbf{t}_1 = \mathbf{X}\,\mathbf{p}_1 \tag{4}$$
(Fig. 3)

is a maximum over all possible linear combinations, subject to $\mathbf{p}^T_1\mathbf{p}_1 = 1$. The latter constraint prevents the arbitrary increase of the variance of $\mathbf{X}\,\mathbf{p}$ by making the elements of $\mathbf{p}$ large, i.e. multiplying the vector with a constant. If the length of this vector is constant only its direction determines the maximum of the product $\mathbf{X}\,\mathbf{p}$. The variance of $\mathbf{t}$ is equal to $\mathbf{t}^T\mathbf{t}$. The second eigenvector $\mathbf{p}_2$, orthogonal to the first one, is calculated from the matrix of residuals $\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}^T_1$, such that the variance of $\mathbf{t}_2 = \mathbf{E}_1\mathbf{p}_2$ is a maximum, subject to $\mathbf{p}^T_2\mathbf{p}_2 = 1$. The third eigenvector, orthogonal to the first two, is calculated in the same way from $\mathbf{E}_2 = \mathbf{E}_1 - \mathbf{t}_2\mathbf{p}^T_2$, etc. Consequently, eigenvector matrix $\mathbf{P}$ containing k or q eigenvectors as columns is orthonormal; the factor scores matrix $\mathbf{T}$ containing the k or q factor scores vectors $\mathbf{t}$ as columns is orthogonal.

$$\mathbf{P}^T\mathbf{P} = \mathbf{I} \tag{5}$$

$$\mathbf{T}^T\mathbf{T} = \Lambda \tag{6}$$

where $\mathbf{I}$ is an identity matrix and Λ is the diagonal matrix of eigenvalues, the variances of the scores vectors $\mathbf{t}_h$. The factor scores matrix $\mathbf{T}$ is calculated by right multiplication of data matrix $\mathbf{X}$ by the eigenvector matrix $\mathbf{P}$.

$$\mathbf{T} = \mathbf{X}\,\mathbf{P} \tag{7}$$
(Fig. 4)

Fig. 1 $\mathbf{X} = \mathbf{t}_1\mathbf{p}^T{}_1 + \mathbf{t}_2\mathbf{p}^T{}_2 + \ldots + \mathbf{t}_q\mathbf{p}^T{}_q$

Fig. 2 $\mathbf{X} = \mathbf{A}_1 + \mathbf{A}_2 + \ldots + \mathbf{A}_q$

Fig. 3 $\mathbf{t}_1 = \mathbf{X}\,\mathbf{p}_1$

Fig. 4 $\mathbf{T} = \mathbf{X\,P}$

Fig. 5 $\mathbf{X} = \mathbf{T\,P}^{\mathrm{T}}$
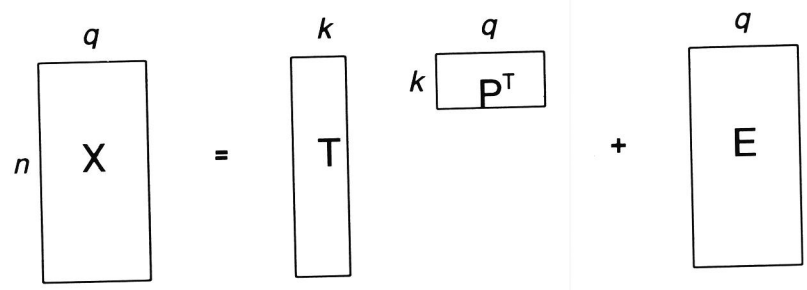
Fig. 6 $\mathbf{X} = \mathbf{T\,P}^{\mathrm{T}} + \mathbf{E}$

Matrix **T** will be of full rank q or of reduced rank k, depending on the number of eigenvectors used in matrix **P**.

The mathematical problem of solving the eigenvalues and eigenvectors is treated in many textbooks [3, 4] and will only be summarized here. The problem is to find a q-element vector **p** and a scalar λ (eigenvalue) such that

$$\mathbf{R}\,\mathbf{p} = \lambda\,\mathbf{p} \tag{8}$$

where the solution $\mathbf{p} = 0$ is excluded. Geometrically, an eigenvector **p** is a vector which is transformed by a matrix **R**, with PCA a symmetrical matrix $\mathbf{R} = \mathbf{X}^T\mathbf{X}$ or $\mathbf{X}\mathbf{X}^T$, into a vector with the same direction but lengthened or shortened by the scalar λ. The matrix **R** might be a covariance about the origin, covariance about the mean, correlation about the origin or correlation about the mean matrix, depending on the transformation applied to the original data matrix [5].

Several methods for calculating eigenvalues and eigenvectors have been described [6]. Two well known methods for calculating eigenvectors are the power method [7] and non-linear iterative partial least squares (NIPALS) [8, 9], which have been shown mathematically to be equivalent [6, 10].

According to the matrix notation introduced in this section, the decomposition of a data matrix **X** into a sum of the q product terms of the factor scores vectors $\mathbf{t}_h$ and the transposed eigenvectors $\mathbf{p}_h$, as indicated in Eqn. 1, is summarized:

$$\mathbf{X} = \mathbf{T}\,\mathbf{P}^T \tag{9}$$
(Fig. 5)

where **X** is an $n \times q$ matrix, **T** is an $n \times q$ matrix of factor scores (columns) and $\mathbf{P}^T$ is a $q \times q$ matrix of eigenvectors (rows). If **X** $(n \times q)$ is approximated $(\hat{\mathbf{X}})$ by the first k product terms (Eqn. 3), **T** is an $n \times k$ matrix and $\mathbf{P}^T$ a $k \times q$ matrix, while **E** is an $n \times q$ matrix containing the difference terms (e.g. error) between $\hat{\mathbf{X}}$ and **X**.

$$\mathbf{X} = \mathbf{T}\,\mathbf{P}^T + \mathbf{E} \tag{10}$$
(Fig. 6)

Equation 10 is closely related to the singular value decomposition (SVD) [2]. The singular value decomposition theorem states that an $n \times q$ matrix **X** can be written as

$$\mathbf{X} = \mathbf{V}\,\Phi\,\mathbf{P}^T \tag{11}$$
(Fig. 7)

where $\mathbf{V}$ is an $n \times q$ column matrix of eigenvectors of $\mathbf{X}\mathbf{X}^T$, connected to the n rows (objects), Φ is a $q \times q$ diagonal matrix containing the square roots of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ or $\mathbf{X}\mathbf{X}^T$, and $\mathbf{P}^T$ is the $q \times q$ row matrix of eigenvectors of $\mathbf{X}^T\mathbf{X}$, connected to the q columns (variables). The q vectors of both $\mathbf{V}$ and $\mathbf{P}$ are orthogonal and of unit length ($\mathbf{V}^T\mathbf{V} = \mathbf{I}$ and $\mathbf{P}^T\mathbf{P} = \mathbf{I}$). It is an essential property that the eigenvalues of the square matrix $\mathbf{X}^T\mathbf{X}$ and the square matrix $\mathbf{X}\mathbf{X}^T$ are identical. Therefore matrix $\mathbf{V}$ can be calculated from $\mathbf{P}$, and *vice versa*:

$$\mathbf{V} = \mathbf{X}\mathbf{P}\Phi^{-1} \tag{12}$$

$$\mathbf{P} = \mathbf{X}^T\mathbf{V}\Phi^{-1} \tag{13}$$
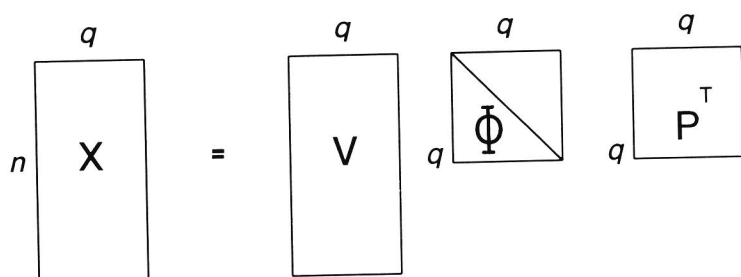
Fig. 7 $\mathbf{X} = \mathbf{V}\Phi\mathbf{P}^T$

These properties enable the choice of the smallest symmetrical matrix $\mathbf{X}^T\mathbf{X}$ or $\mathbf{X}\mathbf{X}^T$ for calculation of the eigenvectors, which may provide important computational advantages. Two principal component solutions are to be distinguished: right multiplication of $\mathbf{V}$ with Φ results in the orthogonal scores matrix $\mathbf{T} = \mathbf{V}\Phi$ ($\mathbf{T}^T\mathbf{T} = \Lambda$, $\mathbf{P}^T\mathbf{P} = \mathbf{I}$), or right multiplication of $\mathbf{P}$ with Φ, provides the orthogonal loadings matrix $\mathbf{L} = \mathbf{P}\Phi$ ($\mathbf{V}^T\mathbf{V} = \mathbf{I}$, $\mathbf{L}^T\mathbf{L} = \Lambda$). The latter solution has the advantage of providing a good survey of the importance of feature contributions to the PCs. If, for example, the matrix $\mathbf{X}$ consists of autoscaled data divided by the square root of n (autoscaled over the columns: zero mean and unit standard deviation), linear combinations of the features with the PCs are equal to the correlation coefficients. For both solutions the original data matrix $\mathbf{X}$ can be reconstructed by Eqn. 9 as well as by Eqn. 14:

$$\mathbf{X} = \mathbf{V}\mathbf{L}^T \tag{14}$$

1.2   Determination of the number real eigenvectors of a data matrix

A number of methods have been developed to determine the number of real eigen-vectors or, in other words, to distinguish real eigenvectors carrying systematic information from error eigenvectors representing noise. In fact, this is equivalent to determining the intrinsic rank of a data matrix.

The effects of experimental error have been studied systematically [1]. An essential idea in error theory is the formulation of a Pythagorean relationship between real error (RE), imbedded error (IE) and extracted error (XE):

$$(RE)^2 = (IE)^2 + (XE)^2 \tag{15}$$

where RE is defined as the difference between the pure data and the raw experimental data and is identical to the residual standard deviation (RSD). The relation between RE, IE and XE is given by:

$$IE = (k/q)^{\frac{1}{2}} \ RE \tag{16}$$

$$XE = [(q-k)/q]^{\frac{1}{2}} \ RE \tag{17}$$

$$(RE)^2 = (1/nq) \sum_{i=1}^{n} \sum_{h=1}^{q} e_{ih}^{2} \tag{18}$$

The scalars n and q refer to the number of rows and columns respectively in the data matrix X; k refers to the number of selected eigenvectors (Eqn. 3) and e represents experimental error. IE is the particular part of RE that mixes into the first k eigenvectors which span the k-dimensional sub-space containing systematic information. This part of the error cannot be removed from the data upon reconstruction of X by $\hat{X}$. The extracted error (XE) is removed by deleting the last $q-k$ error eigenvectors. Therefore

$$X - \hat{X} = XE \tag{19}$$

where X is the original data matrix and $\hat{X}$ is the reconstructed data matrix by using the first k terms (Eqn. 3).

If no knowledge exists of RE, the size of this error can be estimated from XE, using Eqn. 17. For the calculation of XE the number of real eigenvectors (k) has to be determined. One of the methods for this determination consists of detecting a minimum in the IE value as a function of the (increasing) number of eigenvectors employed. It has been demonstrated that if error eigenvectors are included the IE values often start to rise [1, 11]. However, this behaviour usually fails if errors are

140

not fairly uniform throughout the data set, not random, or if errors are sporadic or systematic. This causes enhanced differences in size of error eigenvalues rather than the levelling off of these eigenvalues which occurs if we deal with true random error. The latter has the geometrical effect of a spherical distribution of object vectors in error space resulting in a set of similar error eigenvalues. In recent papers the effect of different error distributions, such as the Gaussian distribution, on the factor structure of the error eigenvectors has been reported [11]. A second criterion, the indicator function (IND), is also based on RE.

$$IND = RE/(q-k)^2 \tag{20}$$

This empirical function also indicates a minimum before error eigenvectors are selected in addition to the k real eigenvectors. Although the IND function seems to be less sensitive to non-uniformly distributed error as compared to the IE function it should also be applied with caution.

In one of the recent papers the distribution of error eigenvalues is considered for uniformly distributed errors in the data set [11]. For this kind of error distribution the error eigenvalues have equal so-called, reduced eigenvalues (REV) which can again be related to RE:

$$REV = l(RE)^2 \tag{21}$$

where $l = n(q-k)/\sum_{h=k+1}^{q} (n-h+1)(q-h+1)$.

The terms $(n-h+1)$ and $(q-h+1)$ refer to the degrees of freedom involved in the determination of λ_h. The scalars n and q refer to the number of objects and variables, respectively; k refers to the number of real eigenvectors.

REV_h, the h-th reduced eigenvalue, is related to the h-th eigenvalue by:

$$REV_h = \lambda_h/(n-h+1)(q-h+1) \tag{22}$$

The distribution of error eigenvalues can be predicted from Eqn. 22. Real eigenvectors are recognized by higher than predicted REVs and by a sharper levelling off to almost constant REV values when error eigenvectors are involved. In a number of examples the detected number of real eigenvectors was in agreement with the IND function and with the well known cross-validation criterion [12-14].

Cross-validation is another well known method for the determination of the number of real eigenvectors. Cross validation estimates the number of real eigenvectors by deleting a small part of the data set and comparing the predicted values of this deleted set based on k eigenvectors with the original values. This procedure (prediction residual sum of squares, PRESS) is repeated until all data vectors have

been classified once. Including error eigenvectors results in a decreased prediction ability.

1.3 Additional properties of eigenvectors and eigenvalues

We will state a number of useful properties of eigenvalues, eigenvectors and factor scores vectors in relation to the data matrix from which they are derived. The Gramian symmetrical matrix $\mathbf{R}$ plays a central role in factor analysis. A Gramian matrix is obtained by left or right multiplication of a data matrix with its transpose: $\mathbf{X}^T\mathbf{X}$ or $\mathbf{X}\mathbf{X}^T$. Covariance and correlation matrices are Gramian matrices. Important properties of Gramian matrices are:
- they can always be factored, such that $\mathbf{R} = \mathbf{F}^T\mathbf{F}$
- their rank is equal to the rank of $\mathbf{X}$.
- the eigenvectors are orthogonal
- none of the eigenvalues is negative
- if $\mathbf{X}$ is of full rank q, $\mathbf{R}$ is a non-singular matrix (it has an inverse) and all eigenvalues are greater than zero.
- if $\mathbf{X}$ is of rank k ($k < q$), $\mathbf{R}$ is a singular matrix (it has no inverse). There are k non-zero eigenvalues and $q - k$ zero eigenvalues.

A number of additional properties of Gramian matrices are given in various textbooks [3, 15].

Starting from $\mathbf{X} = \mathbf{T}\mathbf{P}^T$ (Eqn. 9) a symmetrical matrix $\mathbf{R}$ is reconstructed by left multiplication of Λ by its eigenvector matrix $\mathbf{P}$ and right multiplication by $\mathbf{P}^T$ (cf. Eqn. 6).

$$\mathbf{R} = \mathbf{X}^T\mathbf{X} = \mathbf{P}\mathbf{T}^T\mathbf{T}\mathbf{P}^T = \mathbf{P}\Lambda\mathbf{P}^T \tag{23}$$

Conversely, the diagonal matrix of eigenvalues Λ is obtained from the symmetrical matrix $\mathbf{R}$ by left and right multiplication by the transpose of its eigenvector matrix ($\mathbf{P}^T$) and its eigenvector matrix ($\mathbf{P}$) (cf. Eqn. 5 and 23).

$$\mathbf{P}^T\mathbf{R}\mathbf{P} = \mathbf{P}^T\mathbf{P}\Lambda\mathbf{P}^T\mathbf{P} = \Lambda \tag{24}$$

In fact, these multiplications diagonalize the matrix $\mathbf{R}$, keeping the sum of the diagonal elements (the trace) which represents the variances of the matrix $\mathbf{X}$ equal to the sum of the eigenvalues, if all q eigenvalues are calculated. The eigenvalues are ranked from higher to lower on the diagonal of Λ.

$$\text{trace } \mathbf{R} = \lambda_1 + \lambda_2 + ... + \lambda_q \tag{25}$$

Multiplication of matrix $\mathbf{R}$ by an arbitrary scalar c results in eigenvalues $c\Lambda$ and the same eigenvectors $\mathbf{P}$:

$$c\mathbf{R}\,\mathbf{P} = \mathbf{P}\,c\Lambda \tag{26}$$

The power of the matrix $\mathbf{R}$ yields the same eigenvectors $\mathbf{P}$ with Λ^a as eigenvalues:

$$\mathbf{R}^a\,\mathbf{P} = \mathbf{P}\,\Lambda^a \tag{27}$$

This theorem also holds for $a = -1$, the inverse of the matrix $\mathbf{R}$:

$$\mathbf{R}^{-1}\mathbf{P} = \mathbf{P}\,\Lambda^{-1} \tag{28}$$

2. DISCRIMINANT ANALYSIS

2.1 Mathematical representation

If data sets consist of categorized data, i.e. objects can be arranged into groups, the variance present in the data set can be split into within-group variance ($\mathbf{W}$) and between-group variance ($\mathbf{B}$). Addition of these variances yields the total variance ($\mathbf{V_t}$).

$$\mathbf{V_t} = \mathbf{B} + \mathbf{W} \tag{29}$$

$\mathbf{W}$, a $q \times q$ symmetrical matrix, is calculated by adding the individual within-group sum-of-squares matrices ($\mathbf{W}_k$) equal to (for group k):

$$\mathbf{W}_k = (\mathbf{X}_k - \mathbf{M}_k)^T (\mathbf{X}_k - \mathbf{M}_k) \tag{30}$$

Matrix $\mathbf{X}_k$ is a $n_k \times q$ matrix of n_k objects belonging to group k; $\mathbf{M}_k$ is a matrix of n_k rows with q mean values of group k. Summing of the within-group sum of squares yields

$$\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2 + \dots + \mathbf{W}_k \text{ (for } k \text{ groups)} \tag{31}$$

The between-group sum-of-squares matrix $\mathbf{B}$ is a $q \times q$ matrix, calculated from $\mathbf{M}$ ($n \times q$) of which the first n_1 rows represent the group means on q variables for the first group, the next n_2 rows the group means for the second group, and so on,

until the last n_k rows representing group k, and in addition an $n \times q$ matrix $\mathbf{G}$, containing the grand mean values of $\mathbf{X}$:

$$\mathbf{B} = (\mathbf{M} - \mathbf{G})^T (\mathbf{M} - \mathbf{G}) \tag{32}$$

Now the calculation of the discriminant functions can be formulated as: find a linear combination of the original variables that maximizes:

$$\lambda = \mathbf{e}^T \mathbf{W}^{-1} \mathbf{B} \, \mathbf{e} \tag{33}$$

This is performed by taking the eigenvectors $\mathbf{e}$ of $\mathbf{W}^{-1}\mathbf{B}$, which are the linear discriminant functions (LDFs). In discriminant analysis the first eigenvector has the largest discriminating power (λ) of all possible linear combinations of the original variables, the next eigenvectors are ranked in descending order of magnitude of their discriminant criterion λ. Matrix $\mathbf{B}$ has a rank equal to the minimum of $k - 1$ or q. For spectral data with many variables mostly $q > k - 1$. Therefore, the rank of $\mathbf{W}^{-1}\mathbf{B}$ is mostly $k - 1$, and consequently $k - 1$ discriminant functions are calculated. For the calculation of the inverse of matrix $\mathbf{W}$, data of full rank are required, because a singular matrix cannot yield an inverse [3]. This implies that $n - k > q$, a condition rarely fulfilled for spectral data, such as pyrolysis mass spectra. Because the matrix $\mathbf{W}^{-1}\mathbf{B}$ is not symmetric the eigenvectors $\mathbf{e}$ will be oblique [4, 15]. In contrast to PCA, D functions are not obtained by orthogonal rotation, but by oblique rotation.

To enable calculation of $\mathbf{W}^{-1}$ in the case of lower-rank data $(n - k < q)$ feature selection is frequently applied, which inevitably leads to loss of information. A very useful alternative to feature selection is PCA, which rather compresses the information present in the data set into a limited number of latent variables. By using the real factor scores vectors contained in $\mathbf{T}$ for discriminant analysis [16, 17] the systematic variation is retained in the transformed data set, while part of the error $(\mathbf{XE})$ is removed. The total variance matrix $\mathbf{V}_t$, the between-group variance matrix $\mathbf{B}$ and the within-group variance matrix $\mathbf{W}$ are now calculated from $\mathbf{T}$ $(\mathbf{V}_t = \mathbf{T}^T\mathbf{T})$. An important modification is attained by autoscaling the factor scores yielding $\mathbf{V}_t = n\mathbf{I}$. Because $\mathbf{V}_t = \mathbf{B} + \mathbf{W}$ and $\mathbf{B} = n\mathbf{I} - \mathbf{W}$, $\mathbf{W}^{-1}\mathbf{B}$ now reduces to:

$$\mathbf{W}^{-1}\mathbf{B} = \mathbf{W}^{-1}(n\mathbf{I} - \mathbf{W}) = n\mathbf{W}^{-1} - \mathbf{I} \tag{34}$$

This matrix is symmetrical and, consequently, has orthonormal eigenvectors as LDFs $(\mathbf{U}^T\mathbf{U} = \mathbf{I})$ [4, 17].

$$\mathbf{U}^T(n\mathbf{W}^{-1} - \mathbf{I})\mathbf{U} = \Lambda \tag{35}$$

From Eqns. 34 and 35 it can be concluded that calculation of the axis of maximum variance from **B** also maximizes $\mathbf{W}^{-1}\mathbf{B}$. Therefore the eigenvectors can directly be calculated from

$$\mathbf{U}^T\mathbf{B}\,\mathbf{U} = \Lambda \tag{36}$$

Matrix **U** contains as columns the eigenvectors (LDFs) which are linear combinations of the factor scores vectors **T**. Therefore **U** is a rotation matrix which conducts a orthogonal rotation of the vector scores vectors of **T** to orthogonal discriminant scores **D** [18].

$$\mathbf{D} = \mathbf{T}\,\mathbf{U} \tag{37}$$

(Fig. 8)

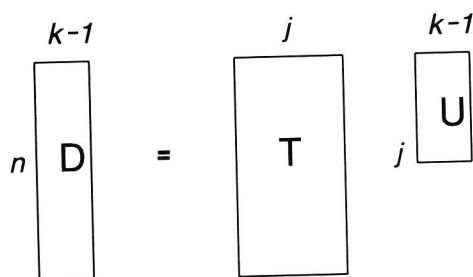Fig. 8   **D = T U**

This leads to the important conclusion that when standardized factors scores vectors (matrix **T**) are used instead of the original data, orthogonal LDFs are obtained. If the original factors scores matrix **T** is used Eqn. 34 is converted into $\mathbf{W}^{-1}\Lambda - \mathbf{I}$. This matrix is not symmetrical and, consequently, yields oblique eigenvectors. It has to be noted that, when the autoscaled factor scores **T** are rotated, the discriminant scores vectors **D** are of the same length:

$$\mathbf{T}^T\mathbf{T} = \mathbf{D}^T\mathbf{D} = n\mathbf{I} \tag{38}$$

From Eqn. 37 it can be concluded that the number of eigenvectors in **U** determines the number of discriminant vectors in **D**. The rank of matrix **B**, which is j or $k-1$,

determines the number of eigenvectors in U: if the number of groups (k) exceeds the number (j) of factor score vectors, a $j \times j$ matrix U is obtained; otherwise a $j \times (k - 1)$ matrix U will be calculated.

The discriminant loadings contained in matrix C, the correlation coefficients between the original variables and a discriminant function may be obtained in two different ways: by determination of the correlation coefficient between the original variables X and the discriminant scores matrix D or by right multiplication of the factor loading matrix L (cf. Eqn 14) by the rotation matrix U:

$$L\,U = C \tag{39}$$

It is important to note that the discriminant loadings vectors contained in matrix C are oblique.

The calculation process of the discriminant scores D and loadings matrices C from the factor scores T and loadings matrices L can be summarized upon reconstructing the data matrix $\hat{X}$ (cf. Eqn 14), the approximation of the autoscaled data matrix X:

$$\hat{X} = T\,L^{T} \tag{40}$$

T is the autoscaled factor scores matrix $(T^{T}T = nI)$; L is the factor loadings matrix, which contains the correlation coefficients of the original variables with the factor scores vectors. After discriminant analysis by rotation of T and L, $\hat{X}$ can be reconstructed by combining Eqns. 37 and 39:

$$\hat{X} = T\,U\,U^{T}L^{T} = D\,C^{T} \tag{41}$$

REFERENCES

1. E.R. Malinowski and D.G. Howery, Factor Analysis in Chemistry, John Wiley & Sons, New York, 1980.

2. K.G. Jöreskog, J.E. Klovan and R.A. Reyment, Geological Factor Analysis, Elsevier Scientific Publishing Company, Amsterdam, 1976.

3. R.J. Rummel, Applied Factor Analysis, Northwestern University Press, Evanston, Ill., 1970.

4. W.R. Dillon and M. Goldstein, Multivariate Analysis, Methods and Applications, John Wiley & Sons, New York, 1984.

5. R.W. Rozett and E.M. Petersen, Anal. Chem., 47 (1975) 1301-1308

6. A. Lorber, L.E. Wangen and B.R. Kowalski, J. of Chemometrics, 1 (1987) 19-31.

146

7.   G. Strang, Introduction to Applied Mathematics, Wellesley Cambridge Press, Wellesley, MA, 1986.

8.   S. Wold in, H. Martens and H. Russwurm (Eds.), Food Research and Data Analysis, Applied Science Publishers, London, 1983.

9.   B.M.G. Vandeginste, C. Sielhorst and M. Gerritsen, Trends in Anal. Chemistry, 7 (1988) 286-287.

10. P. Geladi and B.R. Kowalski, Anal. Chim. Acta, 185 (1986) 1-17.

11. E.R. Malinowski, J. of Chemometrics, 1 (1987) 33-40.

12. S. Wold, Technometrics, 20 (1978) 397-406.

13. G. Scarponi, I. Moret, G. Capodaglio and M. Romanazzi, J. of Chemometrics, 4 (1990) 217-240.

14. H.T. Eastman and W.J. Krzanowski, Technometrics, 24 (1982) 73-79.

15. M.M. Tatsuoka, Multivariate Analysis, John Wiley & Sons, New York, 1971.

16. R. Hoogerbrugge, S.J. Willig and P.G. Kistemaker, Anal. Chem., 55 (1983) 1711-1712.

17. W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81-88.

18. D. Nierop, Department of Data Theory, University of Leiden, unpublished results, 1989.

SUMMARY

In this thesis the potential of Py–DCI/MS is demonstrated for characterization of complex macromolecular mixtures and for structural investigations of biopolymers. Exploratory data analysis techniques were used for differentiation of and trend detection in pyrograms of Gram-negative micro-organisms, algae, fungi, supernatants of virus-infected cell cultures and urines of patients suffering from premenstrual syndrome. These techniques were also applied to the selection of relevant variables for structural determination with HR/MS and MS/MS.

In Chapter 3, Py–DCI/MS is compared with direct probe pyrolysis under electron impact (EI) and chemical ionization (CI) conditions. Py–DCI/MS was found to have a much higher discriminative power and a better reproducibility. Furthermore, sample loading was more convenient and analysis time shorter than for the direct probe technique.

Chapter 4 describes the differentiation of a number of epidemiologically impor-tant *Salmonella* strains from bacterial strains which often interfere during microbiological identification procedures. An unsupervised approach to dis-criminant analysis, defining triplicates as groups, revealed that profiles of almost all *Salmonella* strains tend to cluster. This indicates that these profiles contain characteristic ratios of compounds. Ion series in the higher mass range of the pyrograms (m/z 540-650) were identified as diglycerides, the characteristic build-ing blocks of phospholipids. These masses were particularly important as discriminating variables. This clearly illustrated that compounds of higher molecular weight in pyrolysates often embody characteristic information on the original cell wall components.

The longer-term reproducibility, evaluated over periods of 2 and 4 months, was found to be excellent for the shorter period, but diminished after 4 months. A decrease of ion intensities in the higher mass range, due to ion source contamina-tion, was responsible for the loss of characteristic information.

In Chapter 5, diglyceride information (DCI/MS) and fatty acid methyl ester (FAME) data (GC) from *Salmonella* strains and potentially interfering strains were compared. The data for both techniques were in good agreement, as was shown by the similar object plots of the FAME and diglyceride profiles. For most fatty acids a clear relation was found between FAME and diglyceride data.

Chapter 6 shows an improved approach to trend analysis in the investigation of DCI/MS urine profiles from PMS patients, which were compared with profiles of controls. Because of the great variety of compounds detected, these profiles may provide new leads to the understanding of the biochemical basis of PMS. Differentiation was based on different trends in both groups, observed for profiles of urines collected at day 11 and day 25 of the menstrual cycle. These trends were reflected in the so-called quotient profiles. Quotient profiles enable detection of

group-related changes in individual profiles with conditions, despite large within-group variances.

In Chapter 7, profiles of supernatants of Vero cell cultures after infection with a herpes simplex virus and a poliovirus were compared to study the potential value of Py-DCI/MS in the early diagnosis of viral meningoencephalitis. Blind herpes profiles were correctly distinguished from polio and blank profiles, while discrimination between blind polio and blank profiles was unreliable. However, none of these profiles were misclassified as herpes samples.

In Chapter 8, a study of the fungi *Candida albicans* and *Ophiostoma ulmi* is described. For *C. albicans* dimorphism-related differences, and effects on cell wall composition and sterol metabolism induced by the sterol biosynthesis inhibitor (SBI) lombazole were reflected in the pyrograms. Inhibition of the formation of ergosterol and concomitant higher levels for higher methylated sterols, due to the blocking of a demethylation process in the ergosterol biosynthesis, were established. Moreover, metabolites from lombazole were detected despite its low concentration in the medium.

Also for *O. ulmi* differences were observed between the yeast and hyphal forms which could be attributed to higher levels of deoxyhexose residues in the yeast form. This result was in agreement with reflection infrared measurements.

Chapter 9 reports the DCI pyrolysis of amylose and dextran, $(1-4)$-α-glucan and $(1-6)$-α-glucan respectively, under positive and negative ion detection using ammonia as a reactant gas. Both biopolymers yielded characteristic fragments from ring fragmentations, as well as monomeric and oligomeric products with up to 4 saccharide units originating from glycosidic cleavages. The occurrence of oligomeric pyrolysis products is of potential importance in that it may enable in-pyrolysate characterization of linkage types in biopolymeric samples by MS/MS.

Under positive ion detection mostly $[M + NH_4]^+$ ions were formed. Higher intensities of ring fragments connected to monomeric and oligomeric saccharide units were observed in the pyrogram of dextran.

Differences between the pyrograms of these glucans were even more pronounced in negative ion detection. Odd ions, formed by proton abstraction processes, were detected from non-dehydrated and mono-dehydrated pyrolysis products. Even ions were observed for further dehydrated compounds, probably formed by resonance capture of enones and diketones. Several monomeric and oligomeric pyrolysis products were characterized by MS/MS.

Negative ion detection, applied to the pyrolysate detection of the fungus *Penicillium italicum*, cultured with and without a fungicide, showed a greater specificity for certain cell wall components, such as poly-N-acetylglucosamine (chitin). Under these conditions clear differences were found in the composition of the cell wall, due to alterations in sterol biosynthesis.

Chapter 10 describes the investigation of red and brown algae, and a green algal strain with Py–DCI/MS. Structural characterization of a number of ions of discriminating compounds was performed by MS/MS and HR/MS. In brown algae, among other things, higher levels of mannitol and phytol were found, as was substantiated by MS/MS. In the pyrogram of the green algal strain *Codium fragile* some ions were observed, which had been detected before in *Spirulina* strains (micro-algae) as part of an extended series of homologous ions. Comparison of HR/MS and MS/MS data of a number of these ions revealed the similarity of these compounds from these algal strains. These data provided indications that these compounds were pyrolysis products from proteins. This assignment was further substantiated by the observation of the same ion series in the pyrogram of bovine serum albumin. Based on these data, two basic structures for these compounds are proposed, which explained the homologous ion series from combinations of aliphatic amino acids. Ions originating from combinations of aromatic and aliphatic amino acids were also detected.
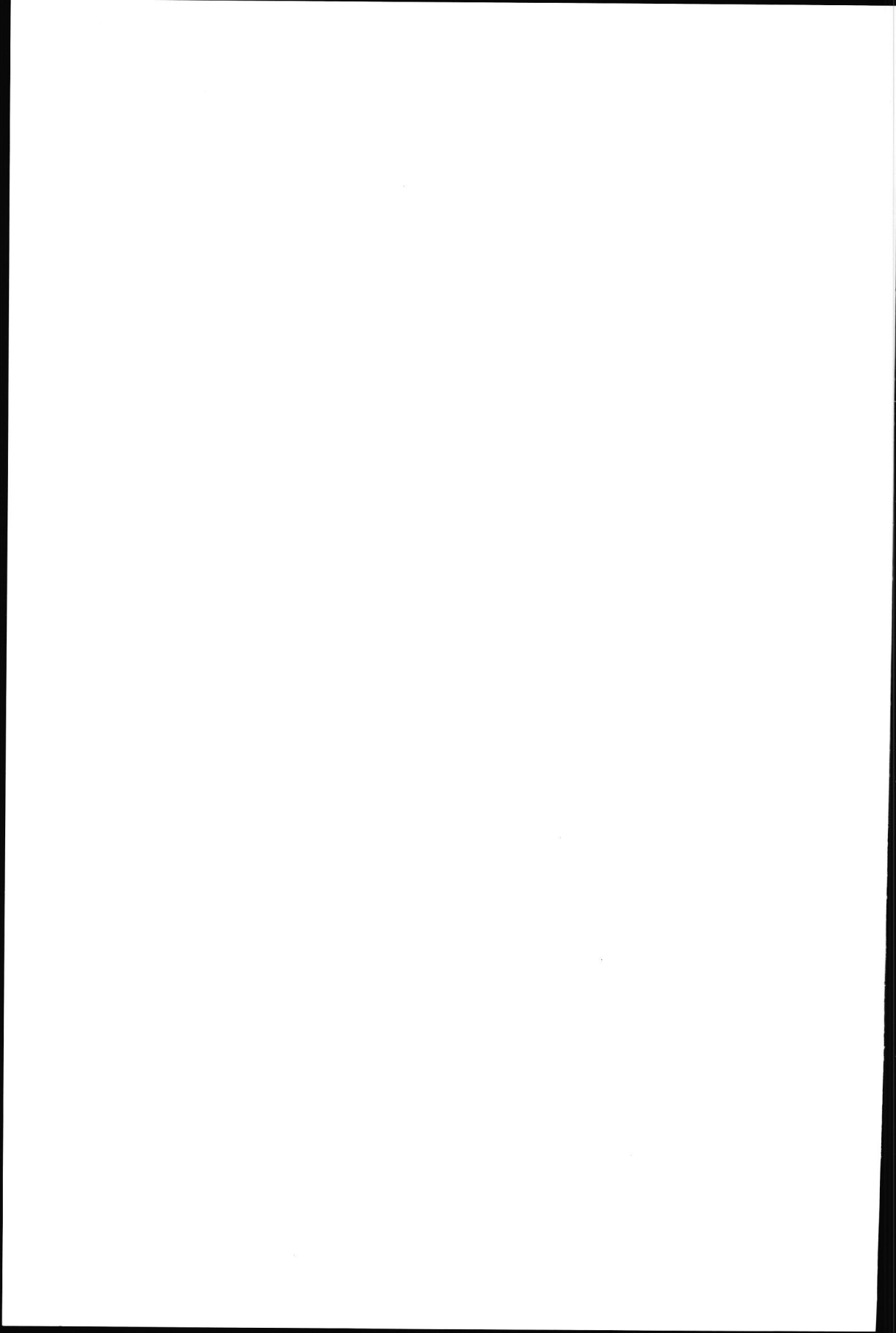
SAMENVATTING

In dit proefschrift zijn de mogelijkheden beschreven van pyrolyse-directe chemische ionisatie massaspectrometrie (Py—DCI/MS) voor de karakterisering van complexe macromoleculaire mengsels en voor structuuronderzoek van biopolymeren. Data-analysetechnieken met een verkennend karakter werden gebruikt voor de differentiatie en detectie van trends in de complexe pyrogrammen van Gram-negatieve micro-organismen, algen, fungi, supernatanten van viraal geïnfecteerde celcultures en urine-extracten van premenstrueel syndroom-patiënten. Deze technieken werden bovendien toegepast op de selectie van variabelen voor verder structuuronderzoek met HR/MS en MS/MS.

Hoofdstuk 3 beschrijft de vergelijking van Py—DCI/MS met "direct probe"-pyrolyse onder "electron impact" en chemische-ionisatie condities. Py—DCI/MS bleek de beste techniek voor wat betreft het onderscheidend vermogen en de reproduceerbaarheid. Bovendien waren de analysetijden korter en was de monsterintroductie eenvoudiger.

In hoofdstuk 4 wordt de differentiatie beschreven van een aantal epidemiologisch belangrijke *Salmonella*-bacteriën en bacteriën die vaak storen in de microbiologische identificatieprocedures. Na discriminantanalyse, waarbij de replica-analyses als groepen werden gedefinieerd, bleken vrijwel alle *Salmonella*-bacteriën in een cluster te worden geprojecteerd. Dit is het gevolg van karakteristieke verhoudingen van bepaalde verbindingen bij deze groep van bacteriën. Van groot belang voor de differentiatie was de serie ionen in het hogere massagebied van het pyrogram, die afkomstig waren van diglyceriden, de bouwstenen van fosfolipiden. Dit is een voorbeeld van wat vaak geldt: fragmenten met een hoger molecuul gewicht bevatten zeer specifieke structuurinformatie over de oorspronkelijke verbindingen.

De reproduceerbaarheid op de langere termijn, die werd bepaald over perioden van 2 en 4 maanden aan de hand van het voorspellend vermogen van een dis-criminantfunctie, was uitstekend over de eerste periode maar verminderde na 4 maanden. De oorzaak hiervan was het verlies aan intensiteit van de ionen in het hogere massagebied en daarmee het verlies aan karakteristieke informatie.

Hoofdstuk 5 is gewijd aan de vergelijking van DCI-pyrogrammen, voor wat betreft de diglyceride-informatie daarin, en gaschromatografische gegevens van vetzure methylesters van *Salmonella*-bacteriën en potentieel interfererende bacteriën. Er werd een duidelijk verband gevonden tussen de gegevens van beide methoden wat betreft het discriminerend vermogen van bepaalde vetzuren.

Hoofdstuk 6 behandelt de analyse met DCI/MS van urines van patiënten met premenstrueel syndroom. Trends in de profielen van urines verzameld op dag 11 en dag 25 zijn vergeleken met die van een controlegroep, waarbij gebruik werd gemaakt van zogenaamde quotiëntprofielen. Met quotiëntprofielen kunnen

groepsgerelateerde verschillen in trends worden gevonden bij individuele profielen, ondanks grote binnen-groepsvarianties.

In hoofstuk 7 wordt de profilering met DCI/MS beschreven van supernatanten van Vero-celcultures na infectie met een herpes simplex-virus en een poliovirus. Het onderzoek was gericht op de potentiële bruikbaarheid van deze methode voor een vroegtijdige diagnose van virale meningoencephalitis. Profielen van een aantal testmonsters werden geklasseerd op basis van de berekende discriminantfunctie. Herpes-profielen konden worden onderscheiden van polio-monsters en blanco monsters, terwijl het onderscheid tussen polio-monsters en blanco monsters on-betrouwbaar bleek te zijn. Geen van deze profielen werd echter als herpes geclasseerd.

In hoofstuk 8 wordt het onderzoek met Py–DCI/MS beschreven van de fungi *Candida albicans* en *Ophiostoma ulmi*. Effecten op de samenstelling van de cel-wand en op het sterolmetabolisme door de sterolbiosyntheseremmer lombazole werden teruggevonden in de profielen, alsmede de met dimorfie gerelateerde verschillen tussen gist- en myceliumvorm. De remming van de sterolbiosynthese was zichtbaar in de lagere gehalten aan ergosterol en de hogere gehalten aan hoger gemethyleerde sterolen bij toepassing van lombazole in het kweekmedium. Tevens werden enkele metabolieten gevonden van dit fungicide, hoewel het fungicide slechts in lage concentraties in het medium aanwezig was.

In de profielen van *O. ulmi* waren eveneens verschillen zichtbaar tussen gist en myceliumvorm. Hogere gehalten aan desoxyhexose-bevattende polymeren werden gevonden in de gistvorm, in overeenstemming met reflectie-infraroodmetingen.

Hoofdstuk 9 beschrijft de DCI-pyrolyse van amylose en dextran, $(1-4)$-α-glucaan and $(1-6)$-α-glucaan, onder positieve en negatieve ionendetectie, waarbij ammonia als reactiegas werd gebruikt. In het pyrolysaat van beide biopolymeren werden karakteristieke fragmenten afkomstig van de ring, alsmede monomere en oligomere produkten aangetroffen, gevormd door fragmentering van de glycosidische bindingen. Het detecteren van oligomeren is belangrijk, omdat dit de mogelijkheid biedt voor de rechtstreekse typering in het pyrolysaat van monomere en oligomere suikers en van de glycosidische bindingen, bij voorbeeld met MS/MS.

Onder positieve ionendetectie werden over het algemeen $[M + NH_4]^+$-ionen gevormd. Het pyrogram van dextran kon duidelijk van dat van amylose worden onderscheiden, met name door de hogere intensiteiten van fragmenten uit de ring, al dan niet gebonden aan monomere of oligomere hexosen.

Bij negatieve ionendetectie waren de verschillen tussen pyrogrammen van amylose en dextran nog groter. In de regel werden van niet-gedehydrateerde en mono-gedehydrateerde pyrolyseprodukten de oneven ionen gevonden. Deze zijn gevormd door protonabstractie. Even ionen kwamen vooral voor bij meer gedehydrateerde verbindingen. De vorming hiervan vindt waarschijnlijk plaats door "resonance capture" van enonen en diketonen. Bij de karakterisering van een

aantal monomere en oligomere pyrolyseprodukten in het pyrolysaat werd gebruik gemaakt van MS/MS.

Bij de negatieve ionendetectie van de pyrolyseprodukten van de fungus *Penicillium italicum*, gekweekt met en zonder een fungicide, bleek een grotere specificiteit voor bepaalde verbindingen uit de celwand in vergelijking tot positieve ionendetectie. Dit leidde tot eenvoudiger profielen van de fungus en tot een betere herkenbaarheid van de optredende verschillen. Poly-N-acetylglucosamine bleek een belangrijke verschilcomponent te zijn, waarvan het ion met een hoge intensiteit in het negatief ionenprofiel aanwezig was, maar in het positief ionenprofiel nauwelijks opviel.

In hoofdstuk 10 wordt het onderzoek naar componenten in de pyrolysaten van rode en bruine algen alsmede van een groene alg beschreven. Data-analyse werd gebruikt bij de selectie van discriminerende variabelen voor verder onderzoek met HR/MS en MS/MS. In bruine algen werden, met name, hogere gehalten aan mannitol en fytol aangetroffen. In het pyrogram van de groene alg *Codium fragile* kwamen een aantal ionen voor die tevens waren gevonden in *Spirulina*-algen (micro-algen), waar deze deel uitmaakten van een uitgebreide serie van homologe ionen. De vergelijking van de HR/MS- en MS/MS-gegevens toonde de overeenkomst in structuur van deze verbindingen uit beide algensoorten aan. Deze analyses leverden sterke aanwijzingen op dat het hier om pyrolyseprodukten van proteïnen ging. Op basis hiervan werden twee basisstructuren voorgesteld die deze homologe series goed kunnen verklaren. De gesignaleerde ionen konden worden toegeschreven aan alifatische dipeptiden en alifatische-aromatische dipeptiden.
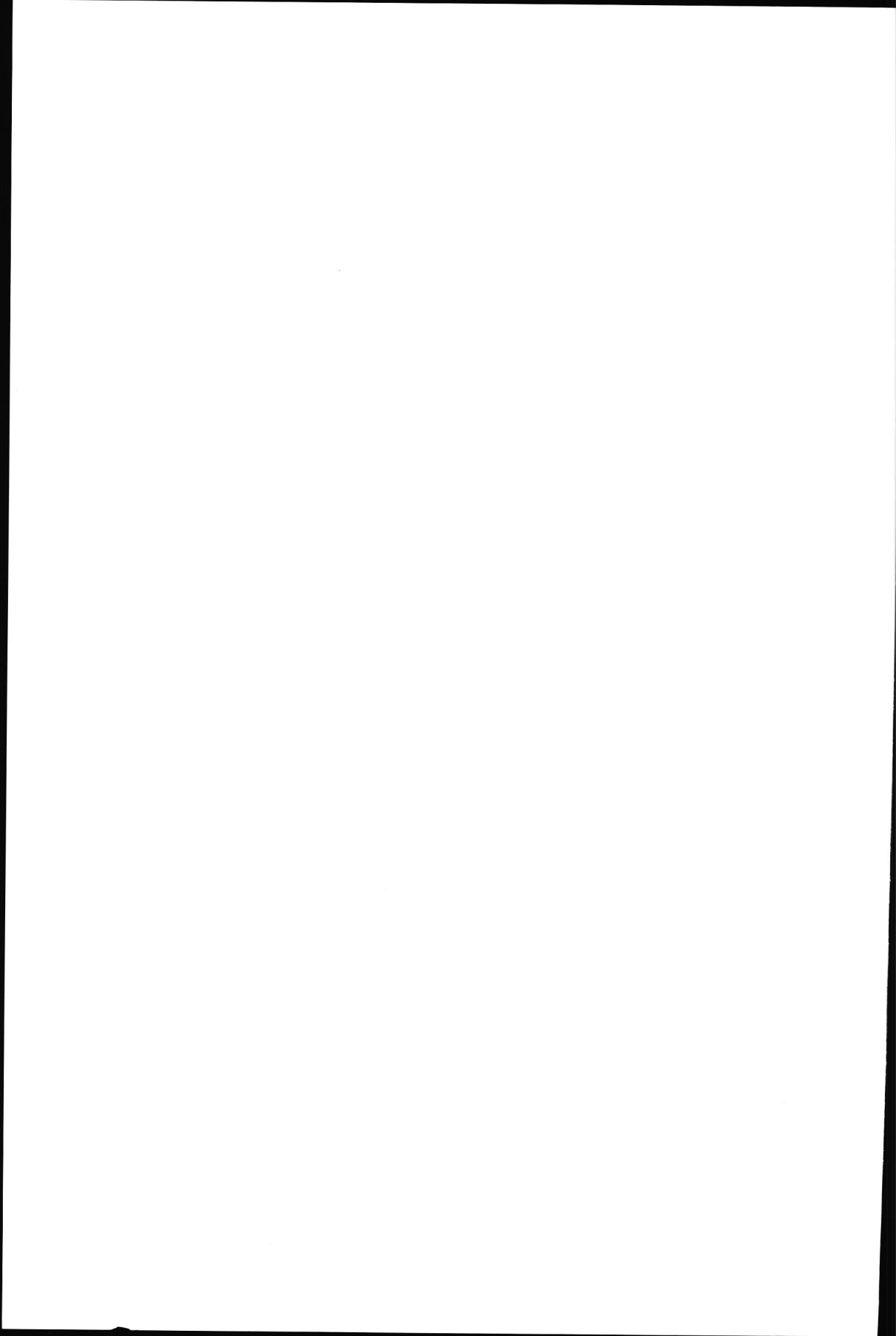
NAWOORD

Het in dit proefschrift beschreven onderzoek is grotendeels uitgevoerd in het laboratorium van de Sectie Instrumentele Analyse van het Instituut CIVO-Analyse TNO. Velen hebben bijgedragen aan de totstandkoming.

In de eerste plaats mijn promotor, die op een uiterst plezierige manier dit onderzoek heeft begeleid. Ik denk met genoegen terug aan de discussies die wij hadden, die naar mijn gevoel doorgaans onmiddellijk leidden naar de kern van de zaak. Ze waren kernachtig, ontspannen door de goede en directe communicatie die er altijd was en heel belangrijk vanwege de stimulans voor het onderzoek die ervan uitging.

Van groot belang was ook de stimulerende rol van ir. M.C. ten Noever de Brauw. Vanuit een filosofisch perspectief voorzag hij al lange tijd een ontwikkeling die zou leiden tot dit wetenschappelijke werk. Op de voor hem karakteristieke wijze heeft hij dit werk altijd ondersteund. Hij slaagde er steeds in een klimaat te scheppen waarin onderzoek goed kan gedijen.

Ook andere medewerkers van de sectie hebben een bijdrage geleverd. Mw. G.F. La Vos ben ik dankbaar voor haar inzet bij de MSMS-metingen. Haar kritische houding ten aanzien van de verkregen meetresultaten zorgde ervoor dat haar metingen garant stonden voor kwaliteit. Gedegenheid en accuratesse zijn belangrijke kenmerken van de software die door J. Bouwman is ontwikkeld en geïmplementeerd voor de patroonherkenning. Deze werkzaamheden hebben een belangrijke bijdrage aan dit onderzoek geleverd.

Met genoegen denk ik terug aan de samenwerking met dr. J. de Waart, die vaak bij nacht en ontij in de weer was om de juiste kast-en veldstammen op tijd in blakende conditie te brengen. Onvergetelijk waren de discussies over patroonherkenning, afgewisseld met kwinkslagen, waarbij ter illustratie van deze "duistere technieken" zelfs vogelsilhouetten en schoenzoolprofielen zijn gebruikt.

Zeer stimulerend waren voor mij de discussies met mijn referent over mechanistische aspecten van pyrolytische en massaspectrometrische fragmentaties, met name bij Py-DCI/MSMS.

Onderzoeksprojecten hebben de meeste kans van slagen wanneer de deelnemers worden gedreven door gemeenschappelijke wetenschappelijke interesse. Vanuit vaak zeer verschillende invalshoeken en disciplines maakt die interesse inzet mogelijk en samenwerking boeiend. Voor de collega-onderzoekers die in de loop van het onderzoek mijn pad kruisten, mede-auteurs met een belangrijke inbreng in de in dit proefschrift gebundelde artikelen, en artikelen die nog op stapel staan, was dit een belangrijk gemeenschappelijk kenmerk. Zonder anderen tekort te willen doen in waardering wil ik in dit verband graag noemen: dr. G. Wieten en L. Berwald (Rijksinstituut voor Volksgezondheid en Milieuhygiëne), dr. J. Odink en dr. H. van den Berg (Instituut CIVO-Toxicologie en Voeding TNO), dr. M.D. Ferrari en drs. A. Nierop (Rijksuniversiteit Leiden), dr. A. Kerkenaar (Instituut

CURRICULUM VITAE

Albert Cornelis Tas werd in 1941 geboren te Aalsmeer. Na het behalen van het MULO-diploma in 1958 volgde hij de opleiding voor leerling-analist (Amsterdam, 1958-1960). Na militaire dienst (1961-1963) sloot hij zijn opleiding voor wetenschappelijk chemisch analist in 1966 af met het behalen van het diploma 2-B. In 1968 en 1971 legde hij de tentamens U.T.K. voor Organische Scheikunde en Hogere Wiskunde af. De tentamens U.T.K. werden ingesteld door de toenmalige Commissie Uitgebreid Theoretische Kennis van de KNCV.

In 1960-1967, onderbroken door militaire dienst, heeft hij gewerkt op het onderzoekslaboratorium van Brocades waar hij werkzaam was in de organische synthese van geneesmiddelen, zowel op laboratoriumschaal als op semi-technische schaal.

In 1967 trad hij in dienst bij CIVO-TNO, in de functie van researchmedewerker. Later kreeg hij daar de functie van wetenschappelijk medewerker. Het accent van zijn werkzaamheden lag op synthese en het gebruik van instrumenteel-analytische technieken als infraroodspectrometrie, NMR en massaspectrometrie. In deze periode waren zijn werkzaamheden vooral gericht op het conformeren met synthese van -op basis van spectroscopische technieken verkregen-veronderstelde structuren van onbekende verbindingen, zoals smaakcomponenten en verbindingen uit modelsystemen, alsmede contaminanten en hun metabolieten in voedingsmiddelen.

Vanaf 1976 is het accent van de werkzaamheden verder verlegd naar de massaspectrometrie, in de sectie Instrumentele Analyse onder leiding van ir. M.C. ten Noever de Brauw. Aanvankelijk werd veel aandacht besteed aan identificatie van verbindingen met GC/MS en "direct probe" - massaspectrometrie en de kwantitatieve bepaling van verbindingen in voedingsmiddelen en lichaamsvloeistoffen.

De laatste jaren is in deze sectie gewerkt aan het ontwikkelen van massaspectrometrische technieken voor de profilering van complexe biochemische mengsels, in samenhang met het gebruik en de ontwikkeling van wiskundige technieken voor data-verwerking voor de directe analyse van complexe mengsels, zoals lichaamsvloeistoffen, micro-organismen, celwanden en biopolymeren. Onder leiding van professor dr. J. van der Greef werd daar het onderzoek verricht dat in dit proefschrift wordt beschreven.
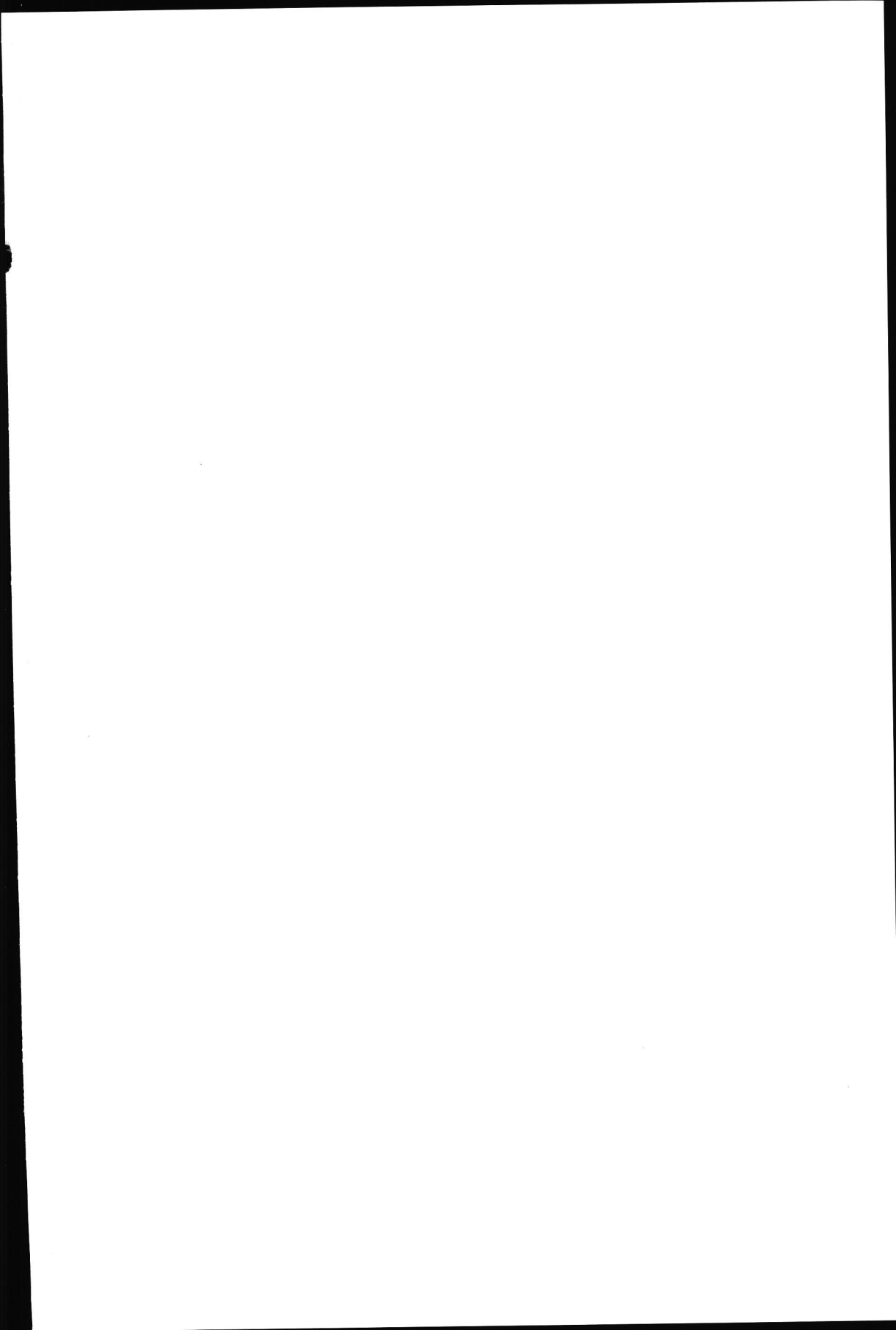
STELLINGEN

1. Beperking van de door ionisatie geïnduceerde fragmentatie vergroot de mogelijkheden van detectie van interrelaties tussen verbindingen in complexe mengsels.

2. Het voorkomen van logische, bij voorbeeld sequentiële, trends in plots van objecten na 'unsupervised' analyse moet als zeer significant worden beschouwd.

3. Bij het vaststellen van het gebruik van niet toegestane middelen in de (pluim)veeteelt wordt te weinig gebruik gemaakt van de mogelijkheid naast de eigenlijke verbinding tevens één of meer metabolieten te meten.

4. Bij derivatisering van verbindingen in zeer lage concentraties verminderen de reactiesnelheden zodanig dat hiervoor vaak in onvoldoende mate kan worden gecompenseerd door verhoging van de concentratie van het reagens.

5. Bij het gebruik van principale componentenanalyse voor de weergave van systematische informatie met minder componenten dan oorspronkelijke variabelen verdient het aanbeveling te spreken van data-compressie in plaats van data-reductie.

6. De invloed van drie-dimensionale structuren op de distributie van pyrolyse-componenten verdient meer onderzoek.

7. De veronderstelling dat peptidepyrolyseprodukten met verschillende retentie-tijden maar met bijna dezelfde massaspectra "dezelfde of zeer overeenkomstige structuren hebben" is onjuist.
    T.O. Munson and D.D. Fetterolf, J. Anal. Appl. Pyrolysis, 11 (1987) 15-24.

8. De mate waarin secundaire reacties optreden bij pyrolyse wordt behalve door fysische condities bepaald door de aard van de verbindingen die hierbij ontstaan.

9. Met de keuze van een eiwit (BSA) en polyhexosen (glycogeen en dextraan) voor kalibratie en predictie van de verhoudingen van deze verbindingen uit de Py—MS-spectra van mengsels van bekende samenstelling, is gekozen voor een model waarbij intensieve secundaire reacties kunnen worden verwacht.

   L.V. Vallis, H.J.H. MacFie en C.S. Gutteridge, Anal. Chem., 57 (1985) 704-709.

10. Pyrogrammen gemeten over grotere massagebieden hebben een grotere opslagcapaciteit voor chemische informatie, vooral wanneer de ionisatie-geïnduceerde fragmentatie wordt beperkt door het gebruik van zachte ionisatiemethoden.

    Dit proefschrift.

11. Het linksom draaien van monumentale windmolens is alleen voor korenmolens afdoende verklaard door de handmolentheorie van J. den Besten. Het feit dat ook later ontstane molens met andere functies, zoals water-, zaag- en oliemolens, meestal linksomdraaien kan worden verklaard uit de voorbeeldfunctie van korenmolens en op cultuurhistorische gronden.

    E. Smit, Het Gilde van Vrijwillige Molenaars, Informatie-XVI.

12. Zekerheid is in zekere zin een illusie.

13. Het tijdelijke blijkt vaak zeer lang te duren, terwijl ideeën of ideologieën die voor de eeuwigheid zijn geproclameerd, veelal een kort leven zijn beschoren.

14. Popmuziek verveelt of irriteert snel doordat wordt gestreefd naar een van het begin tot het einde volgehouden hoogspanning zonder gebruik te maken van compositorische technieken van opbouw van spanning gevolgd door ontspanning.