RAPID ANALYSIS OF MULTIPLE DETERMINANDS USING PYROLYSIS MASS SPECTROMETRY AND SUPERVISED LEARNING WITH ARTIFICIAL NEURAL NETWORKS

Douglas B. Kell, Royston Goodacre & Mark J. Neal

Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, U.K.

Curie-point pyrolysis mass spectrometry is a rapid, high-resolution technique which has enjoyed historical success in the operational fingerprinting and classification of microbial and other biological systems. More recently we have exploited the analytical power of this technique together with *supervised* learning, based on artificial neural networks and cognate chemometric methods, for the accurate and quantitative analysis of complex biological materials, including microorganisms, fermentor broths and agricultural products. When suitable standards are available, supervised learning is much more powerful than is unsupervised learning for the chemometric analysis of multivariate spectroscopic data.

Introduction

There is a continuing need for more rapid, precise and accurate analyses of the chemical composition of microbial systems, both within biotechnology and for the identification of potentially pathogenic organisms. An ideal method would permit the simultaneous estimation of multiple determinands, would have negligible reagent costs, and would run under the control of a PC, to allow flexible operation of the sample handling, instrument calibration, and data analysis and visualisation routines.

Pyrolysis is the thermal degradation of a material in an inert atmosphere, and leads to the production of volatile fragments from non-volatile material such as microorganisms[1]. Curie-point pyrolysis is a particularly reproducible and straightforward version of the technique, in which the sample, dried onto an appropriate metal is rapidly heated (0.6s is typical) to the Curie point of the metal, which may itself be chosen (358, 480, 510, 530, 610 and 770°C are common temperatures). The volatile fragments resulting from the Curie-point pyrolysis may then be separated and analysed in a mass spectrometer[2], and the combined technique is then known as Pyrolysis Mass Spectrometry or PyMS.

Almost all biological materials will produce pyrolytic degradation products such as methane, ammonia, water, methanol and H_2S, whose mass:charge (m/z) ratio < 50, and fragments with m/z > 200 are rarely analytically important[3]; in the commercially available instruments (cf.[4]) the analytically useful data are thus constituted by a set of (150) normalised intensities versus m/z in the range 51-200.

Conventionally (within microbiology and biotechnology), PyMS has been used as a taxonomic aid in the *identification* and *discrimination* of different microorganisms[5], such as members of the genera *Bacillus*[6], *Corynebacterium*[7] and *Legionella*[8]. In particular, PyMS, because of its high discriminatory ability, has been successfully applied to the inter-strain comparison of a wide range of medically important bacterial species and groups, including: *Corynebacterium*[2], *Escherichia coli*[9], *Legionella*[8], mycobacteria[10,11], salmonellae[12] and streptococci[13], highlighting the usefulness of this technique in the detection of small differences between microbial samples. More complex and powerful methods such as pyrolysis tandem-MS (e.g. [14,15]) and pyrolysis GC-MS (e.g. [4, 16-19]) have also demonstrated their great utility in the detection of the presence of bacteria *via* chemical biomarkers. One of the major advantages that PyMS has over other diagnostic methods, such as ELISA (e.g. [20-23]) and nucleic acid probing[24], however, is that it is rapid, both for a single sample and with respect to the (automated) throughput of samples, since the typical sample time is less than 2 min.

The reduction of the multivariate data generated by the PyMS system is normally carried out using Principal Components Analysis (PCA)[25-28], a well-known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance. Whilst PCA does not take account of any groupings in the data, neither does it require that the populations be normally distributed, i.e. it is a non-parametric method. (In addition, it permits the loadings of each of the m/z ratios on the principal components to be determined, and thus the extraction of at least *some* chemically significant information.) The closely-related Canonical Variates Analysis technique then separates the samples into groups on the basis of the principal components and some *a priori* knowledge of the appropriate number of groupings[29]. Provided that the data set contains "standards" (i.e. type or centro-strains) it is evident that one can establish the closeness of any unknown samples to a known organism, and thus effect the identification of the former, a procedure known as operational fingerpriting[2]. An excellent example of the discriminatory power of the approach is the demonstration[30] that one can use it to distinguish 4 strains of *E. coli* which differ only in the presence or absence of a single plasmid.

(Artificial) neural networks (ANNs) are, by now, a well-known means of uncovering complex, nonlinear relationships in multivariate data (see e.g.[31] and references therein for full details, [32-37] for excellent introductions). Independently of the numerous possible network architectures for so-called supervised learning, the principle of operation is that one can acquire sets of multivariate data (e.g. normalised intensities at 150 values of m/z) for mixtures of determinands *whose concentrations are known*, and train an ANN using the (known) concentrations as the desired outputs. Training is effected by continually presenting the networks with the training data and modifying ("updating") the weights between the individual neurons or processing elements, typically according to some kind of back-propagation algorithm[32,38], until the outputs of the network match the "true" (desired) outputs within a predetermined degree of accuracy. The network, the effectiveness of whose training is usually determined in terms of the root mean square error between the actual and the desired outputs, may then be exposed to unknown inputs (i.e. spectra) and will then "immediately" output the globally optimal best fit in terms of the relative concentrations of the desired determinands. We have therefore sought to exploit ANNs in the quantitative analysis of PyMS data from a variety of biological systems, a series of studies which we here review.

Experimental

The pyrolysis mass spectrometer used in our work was the Horizon Instruments PYMS-200X as described by Aries *et al.*[39]. The sample tube carrying the foil was heated, prior to pyrolysis, at

100°C for 5s. Curie-point pyrolysis was at 530°C for 3s, with a temperature rise time of 0.5s. This pyrolysis temperature was chosen because it has been shown[40,41] to give a balance between fragmentation from polysaccharides (carbohydrates) and protein fractions. The pyrolysate then entered a gold-plated expansion chamber heated to 150°C, whence it diffused down a molecular beam tube to the ionisation chamber of the mass spectrometer. To minimize secondary fragmentation of the pyrolysate the ionisation method used was low voltage electron impact ionisation (25eV). Non-ionised molecules were deposited on a cold trap, cooled by liquid nitrogen. The ionised fragments were focussed by the electrostatic lens of a set of source electrodes, accelerated and directed into a quadrupole mass filter. The ions were separated by the quadrupole, on the basis of their mass-to-charge ratio, and detected and amplified with an electron multiplier. The mass spectrometer scans the ionised pyrolysate 160 times at 0.2s intervals following pyrolysis. Data were collected over the m/z range 51 to 200, in one tenth of a mass-unit intervals. These were then integrated to give unit mass. Given that the charge of the fragment was unity the mass-to-charge ratio can be accepted as a measure of the mass of pyrolysate fragments. The IBM-compatible PC used to control the pyrolysis mass spectrometer, was also programmed (using software provided by the manufacturers) to record spectral information on ion count for the individual masses scanned and the total ion count for each sample analysed. This software also carried out PCA and CVA.

Normalised ion count data were preprocessed using Microsoft Excel, and imported into a Windows-based neural network simulation program, NeuralDesk (version 1.2) (Neural Computer Sciences, Lulworth Business Centre, Nutwood Way, Totton, Southampton, Hants, SO1 0JR, U.K.), which runs under Microsoft Windows 3.1 on an IBM-compatible PC. To ensure maximum speed, an accelerator board for the PC (NeuSprint) based on the AT&T DSP32C chip, which effects a speed enhancement of some 100-fold over a 33MHz 386-based PC, permitting the analysis (and updating) of some 400,000 weights per second, was used.

Results

Using the above system, we and our collaborators have shown that the combination of ANNs and PyMS may be used for the quantification of indole production in bacteria[42], of biopolymers in binary[43,44] and tertiary mixtures[45], of recombinant protein production in whole cells of *Escherichia coli*[46], of mixed microbial cultures[45], and in the rapid and quantitative screening of cultures and fermentor broths for metabolite overproduction[47]. *Inter alia*, we have also been the first to apply PyMS and ANNs to the successful identification of the adulteration of extra virgin olive oils[48,49] and in demonstrating that canine isolates of *Propionibacterium acnes* strains are the same as human wild-type strains[50], and have reviewed this material in more detail[51,52]. It is worth mentioning that the quantification of indole production in bacteria[42] represented the first occasion on which it was possible *quantitatively* to determine the concentration of a known biological molecule that lacked a unique ion in the pyrolysis mass spectrum of a biological system, and that one might expect that this principle could to be adopted for the study of a variety of other microbial problems of interest in the present context.

CONCLUSIONS

We would stress two findings in particular: (i) in our olive oil studies[49], although unsupervised methods indicated that the variance in the PyMS between *cultivars* of olives dominated that caused by adulteration, (nonlinear) supervised learning methods were easily able to distinguish oils based on the presence of adulterating seed oils, and (ii) the discrimatory power of these

chemometric methods is such that in one study[47] we were able to assay quantitatively for the concentrations of ampicillin in mixtures with the Gram-positive *S. aureus* when the training set consisted of mixtures of ampicillin and *an entirely different biological background, viz.* the Gram-negative *E. coli*. This shows, importantly, that chemometric methods of this type which are designed to effect the quantification of biomolecules in complex biological backgrounds may indeed be made highly resistant to changes in the background cooncentrations of metabolites and macromolecles. Pyrolysis mass spectrometry combined with ANNs thus represents a powerful and novel approach to the rapid characterization of complex biological material.

ACKNOWLEDGMENTS

We thank the Chemicals & Pharmaceuticals Directorate of the BBSRC for support.

REFERENCES

1. Irwin, W.J. (1982) *Analytical pyrolysis: a comprehensive guide*. New York, Marcel Dekker.

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

3. Berkeley, R.C.W., Goodacre, R., Helyer, R.J. & Kelley, T. (1990) Pyrolysis-MS in the identification of micro-organisms. *Lab. Pract.* **39**, 81-83.

4. Smith, P.B. & Snyder, A.P. (1992) Characterization of bacteria by quartz tube pyrolysis-gas chromatography ion trap mass-spectrometry. *J. Anal. Appl. Pyrol.* **24**, 23-38.

5. Gutteridge, C.S. (1987) Characterization of microorganisms by pyrolysis mass spectrometry. *Meth. Microbiol.* **19**, 227-272.

6. Shute, L.A., Gutteridge, C.S., Norris, J.R. & Berkeley, R.C.W. (1988) Reproducibility of pyrolysis mass spectrometry: effect of growth medium and instrument stability on the differentiation of selected *Bacillus* species. *J. Appl. Bacteriol.* **64**, 79-88.

7. Hindmarch, J.M., Magee, J.T., Hadfield, M.A. and Duerden, B.I. (1990) A pyrolysis-mass spectrometry study of *Corynebacterium* spp. J. Med Microbiol., 31, 137-149.

8. Kajioka R. and Tang, P.W. (1984) Curie-point pyrolysis mass-spectrometry of *Legionella* species. *J. Anal. Appl. Pyrol.* **6**, 59-68.

9. Goodacre, R., Beringer, J.E. & Berkeley, R.C.W. (1991) The use of pyrolysis-mass spectrometry to detect the fimbrial adhesive antigen F41 from *Escherichia coli* HB101(pSLM204). *J. Anal. Appl. Pyrol.* **22**, 19-28.

10. Wieten, G., Haverkamp, K., Meuzelaar, H.L.C., Engel H.B.W. & Berwald, L.G. (1981) Pyrolysis mass spectrometry: a new method to differentiate between the mycobacteria of the 'Tuberculosis' complex and other mycobacteria. *J. Gen. Microbiol.* **122**, 109-118.

11. Wieten, G., Haverkamp, K., Engel, H.B.W. & Berwald, L.G. (1981) Application of pyrolysis mass spectrometry to the classification and identification of mycobacteria. *Rev. Infectious*

Diseases **3**, 871-877.

12. Freeman, R., Goodfellow, M., Gould, F.K., Hudson, S.J. & Lightfoot, N.F (1990) Pyrolysis-mass spectrometry (Py-MS) for the rapid epidemiological typing of clinically significant bacterial pathogens. *J. Med. Microbiol.* **32**, 283-286.

13. Magee, J.T. Hindmarch, J.M. and Nicol, C.D. (1991) Typing of *Streptococcus pyogenes* by pyrolysis mass spectrometry. *J. Med. Microbiol.* **35**, 304-306.

14. Voorhees, K.J., DeLuca, S.J. & Noguerola, A. (1992) Identification of chemical biomarker compounds in bacteria and other biomaterials by pyrolysis-tandem mass spectrometry. *J. Anal. Appl. Pyrol.* **24**, 10-21.

15. Snyder, A.P. (1993) Analyte detection in complex solid matrices with pyrolysis atmospheric-pressure chemical-ionization tandem mass-spectrometry. *Trends Anal. Chem.* **12**, 296-303

16. Dworzanski, J.P. & Meuzelaar, H.L.C. (1991) Rapid detection and identification of microorganisms via pyrolytic on-line derivatization-gas chromatography-mass spectrometry. In *Modern techniques for rapid microbiological analysis* (W.H. Nelson, ed.) pp.19-42. VCH, Weinheim.

17. Morgan, S.L., Fox, A., Rogers, J.C. & Watt, B.E. (1991) Identification of chemical markers for microbial differentiation and detection by gas chromatography-mass spectrometry. In *Modern techniques for rapid microbiological analysis* (W.H. Nelson, ed.) pp.1-18. VCH, Weinheim.

18. Smith, P.B. & Snyder, A.P. (1993) Characterization of bacteria by oxidative and nonoxidative pyrolysis-gas chromatography ion trap mass-spectrometry. *J. Anal. Appl. Pyrol.* **24**, 199-210.

19. Snyder, A.P., Smith, P.B.W., Dworzanski, J.P. & Meuzelaar, H.L.C. (1994) Pyrolysis-gas chromatography-mass spectrometry. Detection of biological warfare agents. *ACS Symp Ser.* **541**, 62-84.

20. Phillips, A.P. & Martin, K.L. (1983) Immunofluorescence analysis of *Bacillus* spores and vegetative cells by flow cytometry. *Cytometry* **4**, 123-131.

21. Phillips, A.P., Martin, K.L. & Broster, M.G. (1983) Differentiation between spores of *Bacillus anthracis* and *Bacillus cereus* by a quantitative immunofluorescence technique. *J. Clin. Microbiol.* **17**, 41-47.

22. Phillips, A.P. & Martin, K.L. (1985) Dual parameter scatter-flow immunofluorescence analysis of *Bacillus* spores. *Cytometry* **6**, 124-129.

23. Chantler, S.M. and McIllmurray, M.B. (1987) Labelled-antibody methods for the detection and identification of microorganisms. *Meth. Microbiol.* **19**, 273-332.

24. Saano, A. and Lindström, K. (1990) Detection of rhizobia by DNA-DNA-hybridization from soil samples: problems and perspectives. *Symbiosis* **8**, 61-73.

25. Chatfield, C. & Collins, A.J. (1980) *Introduction to Multivariate Analysis*, pp 57-81. Chapman & Hall, London.

26. Causton, D.R. (1987) *A Biologist's Advanced Mathematics*, pp 48-72. London: Allen & Unwin.

27. Gutteridge, C.S., Vallis, L. & MacFie, H.J.H. (1985) Numerical methods in the classification of microorganisms by pyrolysis mass spectrometry. In: *Computer-assisted bacterial systematics*, eds. Goodfellow M, Jones D, Priest F, pp 369-401. Academic Press, London.

28. Flury, B. and Riedwyl, H. (1988) *Multivariate statistics: a practical approach*, pp 181-233. London: Chapman & Hall.

29. MacFie, H.J.H., Gutteridge, C.S. and Norris, J.R. (1978) Use of canonical variates in differentiation of bacteria by pyrolysis gas-liquid chromatography. *J. Gen. Microbiol.* **104**, 67-74.

30. Goodacre, R. and Berkeley, R.C.W. (1990) Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry. *FEMS Microbiol. Lett.* **71**, 133-138.

31. Wasserman, P.D. & Oetzel, R.M. (1989) *NeuralSource: the bibliographic guide to artificial neural networks*, Van Nostrand Reinhold, New York, 1014pp.

32. Rumelhart, D.E., McClelland J.L. and the PDP Research Group (*Parallel Distributed Processing. Experiments in the Microstructure of Cognition.* MIT Press, Cambridge, Mass.,1986.

33. McClelland, J.L. & Rumelhart, D.E. (1989) *Explorations in Parallel Distributed Processing.* MIT Press, Cambridge, Massachusetts.

34. Wasserman, P.D. (1989) *Neural Computing: Theory and Practice.* Van Nostrand Reinhold, New York.

35. Eberhart, R.C. & Dobbins, R.W. (1990) *Neural Network PC Tools.* Academic Press, London.

36. Simpson, PK (1990) *Artificial Neural Systems.* Pergamon Press, Oxford.

37. Hertz, J., Krogh, A. & Palmer, R.G. (1991) *Introduction to the Theory of Neural Computation.* Addison-Wesley, Redwood City.

38. Werbos, P.J. (1993) *The roots of back-propagation: from ordered derivatives to neural networks and political forecasting.* John Wiley & Sons, Chichester.

39. Aries, R.E., Gutteridge, C.S. & Ottley, T.W. (1986) Evaluation of a low-cost, automated pyrolysis-mass spectrometer. *J. Anal. Appl. Pyrol.* **9**, 81-98.

40. Windig, W., Kistemaker, P.G., Haverkamp, J. & Meuzelaar, H.L.C. (1980). Factor analysis of the influences of changes in experimental conditions in pyrolysis mass spectrometry. *J. Anal. Appl. Pyrol.* **2**, 7-18.

41. Goodacre, R. (1992) *The effects of genotypic and phenotypic changes on bacterial identification using pyrolysis mass spectrometry*, Ph.D. Thesis, University of Bristol.

42. Goodacre, R. & Kell, D.B. (1993) Rapid analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production. *Anal. Chim. Acta* **279**, 17-26.

43. Goodacre, R. Edmonds, A.N. & Kell, D.B. (1993). Quantitative analysis of the pyrolysis mass spectra of complex mixtures using artificial neural networks: application to amino acids in glycogen. *J. Anal. Appl. Pyrol.* **26**, 93-114.

44. Neal, M.J., Goodacre, R. & Kell, D.B. (1994) On the analysis of pyrolysis mass spectra using artificial neural networks. Individual input scaling leads to rapid learning. Proceedings of the World Congress on Neural Networks 1994, San Diego, I-318 - I-323.

45. Goodacre, R., Neal, M.J. & Kell, D.B. (1994) Rapid and quantitative deconvolution of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. *Anal. Chem.* **66**, 1070-1085.

46. Goodacre, R., Karim, A., Kaderbhai, M. & Kell, D.B. (1994) Rapid and quantitative analysis of recombinant protein expression using pyrolysis mass spectrometry and artificial neural networks: application to mammalian cytochrome b_5 in *Escherichia coli*. *J. Biotechnol.* **34**, 185-193.

47. Goodacre, R., Trew, S., Wrigley-Jones, C., Neal, M.J., Maddock, J., Ottley, T.W., Porter, N. & Kell, D.B. (1994) Rapid screening for metabolite overproduction in fermentor broths, using pyrolysis mass spectrometry with multivariate calibration and artificial neural networks. *Biotechnol. Bioeng.*, in the press.

48. Goodacre, R., Kell, D.B. & Bianchi, G. (1992) Neural networks and olive oil. *Nature* **359**, 594.

49. Goodacre, R., Kell, D.B. & Bianchi, G. (1993) Rapid assessment using pyrolysis mass spectrometry and artificial neural networks of the adulteration of virgin olive oils by other seed oils. *J. Sci. Food Agric.* **63**, 297-307.

50. Goodacre, R., Neal, M.J., Kell, D.B., Greenham, L.W., Noble, W.C. & Harvey, R.G. (1993) Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. *J. Appl. Bacteriol.* **76**, 124-134.

51. Goodacre, R. (1994) Characterization and quantification of microbial systems using pyrolysis mass spectrometry: introducing neural networks to analytical pyrolysis. *Microbiol. Eur.* **2** (2), 16-22.

52. Goodacre, R., Neal, M.J. & Kell, D.B. (1995) Quantitative analysis of multivariate data using artificial neural networks: a tutorial review and applications to the deconvolution of pyrolysis mass spectra. *Zentralblatt. für Bakteriologie*, in the press.