Pyrolysis mass spectrometry and its applications in biotechnology

Royston Goodacre and Douglas B Kell

Pyrolysis mass spectrometry is a rapid and high-resolution method for the analysis of otherwise non-volatile material and has been widely applied for discriminating between closely related microbial strains. Recent advances in statistical and neural network methods based on supervised learning have now permitted exploitation of pyrolysis mass spectrometry in the quantitative analysis of many diverse samples of biotechnological interest; the technique may thus be regarded as an 'anything-sensor'.

Address
Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed, SY23 3DA, UK

Current Opinion in Biotechnology 1996, **7**:20–28

Abbreviations

ANN	artificial neural network
CVA	canonical variates analysis
ESI	electrospray ionization
HCA	hierarchical cluster analysis
MS	mass spectrometry
m/z	mass-to-charge
PCA	principal components analysis
PHB	poly(3-hydroxybutyrate)
PyGC/MS	pyrolysis gas chromatography/MS
PyMS	pyrolysis MS
RAPD	random amplified polymorphic DNA

Introduction

> "The ability [of pyrolysis mass spectrometry] to analyse small amounts of biological material with minimum sample preparations to obtain, in minutes, fingerprint data that can be used for identification and typing is unparalleled by other methods, including nucleic acid and fingerprinting methods."

AT Bull, M Goodfellow and JH Slater, *Annu Rev Microbiol* 1992, 46:219–252.

The above quote is from an excellent review on biodiversity as a source of innovation in biotechnology, highlighting the use of pyrolysis mass spectrometry (PyMS) as a technique for the rapid identification of micro-organisms. It is particularly noteworthy that at the time that review was published, PyMS had not been exploited for the quantitative analysis of biological material; this was soon to change with the application of artificial neural networks (ANNs). The following review discusses the wide range of applications now available to the biotechnologist.

The development of rapid and efficient methods for screening biologically active metabolites from large numbers of microbial cultures retains pre-eminence in drug discovery efforts [1,2••,3]. Such metabolites can also provide new structural templates for synthetic programmes using rational methods of drug design through chemical synthesis, including combinatorial methods [4••]. It is imperative, therefore, that the concentration of the fermentation product (i.e. the determinand) is assessed accurately, both so that the most high-yielding strains are selected and to assist the subsequent optimization of the bioprocess. The development of such monitoring methods is driven by economic and ecological needs and, more recently, by the requirements for better process documentation [5•,6•].

An ideal method for the rapid, precise and accurate analysis of the biochemical composition of fermenter broths, as well as of the characterization of the organisms that they contain, would permit the simultaneous estimation of multiple determinands, would have minimum sample preparation, would analyze samples directly (i.e. would not require reagents), and would be rapid, automated, accurate and (at least relatively) cheap. PyMS is a rapid, automated, instrument-based technique that permits the acquisition of spectroscopic data from 300 or more samples per working day.

Although on-line tandem MS has been used to analyze fermentation broth extracts for flavones [7], the majority of MS applications to fermentations has been either for the analysis of gases and volatiles in the bioreactor headspace [8] or for the analysis of volatile compounds dissolved in the broths via a membrane inlet probe [9]. It is obvious, however, that more worthwhile information could be gained by measuring the non-volatile components of fermentation broths.

The focus of this review is to highlight PyMS as an approach for the rapid analysis and identification of bacteria and fungi. We also reflect on the recent advances in the application of novel statistical methods based on supervised learning, such as ANNs, which have allowed us and others both to effect the rapid and quantitative analysis of a variety of fermentations and to permit the rapid screening of microbial cultures for the over-production of metabolites of interest. An overview of the PyMS

technique has been published on the World Wide Web (http://gepasi.dbs.aber.ac.uk/roy/pymshome.htm).

Pyrolysis mass spectrometry instrumentation

The first automated Curie-point pyrolysis mass spectrometer was built as far back as 1973 by Meuzelaar and Kistermaker (see [10]) specifically for the characterization of complex non-volatile biological samples such as micro-organisms. Even so, the only PyMS instrument presently available commercially is the RAPyD-400 manufactured by Horizon Instruments (Ghyll Industrial Estate, Heathfield, East Sussex, UK); it is arguably the availability of this instrument that has allowed this technology to become more accessible to biological chemists, and it is convenient to describe briefly its implementation of PyMS.

The first stage in the process is the preparation of samples. For microbial cultures, after incubation, bacteria or fungi are picked up carefully from the top of one or more well isolated colonies by means of disposable plastic loops, avoiding the plate surface, and smeared onto clean iron-nickel foils. Alternatively, if the sample to be analyzed is from a liquid culture, a few microlitres may be applied directly to the foil. The samples are then oven-dried or vacuum-desiccated. The rest of the analysis is carried out by the pyrolysis mass spectrometer under the control of a personal computer. The sample tubes are loaded sequentially and the pyrolysate generated in a vacuum by heating the sample. Heating is achieved by passing a radio-frequency current for 3 s through a pyrolysis coil that surrounds the sample-coated ferro-magnetic foil. The foil and sample heat rapidly, typically in <0.5 s, to the temperature corresponding to the Curie-point of the iron-nickel foil. (530°C is a common temperature, because this gives a balance between fragmentation from polysaccharides and protein fractions, both of which are abundant in bacteria). The pyrolysate then enters a gold-plated expansion chamber heated to 150°C, and diffuses down a molecular beam tube to the ionization chamber of the mass spectrometer. The pyrolysate is then bombarded with low-energy electrons (typically in the region of 25 eV) producing both molecular and fragment ions (because low energy is employed, the majority will carry only a single positive charge). Non-ionized molecules are deposited on a cold trap, which is cooled by liquid nitrogen. The ionized fragments are then focused by the electrostatic lens of a set of source electrodes, accelerated and directed into a quadrupole mass filter [11], separated on the basis of their mass-to-charge (m/z) ratio, and detected and amplified with an electron multiplier.

Almost all biological materials will produce pyrolytic degradation products, such as methane, ammonia, water, methanol and hydrogen sulphide, whose m/z ratio is <50; fragments with a m/z ratio >200 are rarely analytically important for bacterial discrimination [12]. The analytically useful data are thus constituted by a set of (150) normalized intensities versus m/z ratio in the range of 51 to 200 (see Fig. 1).

Identification using pyrolysis mass spectrometry and classic multivariate analyses

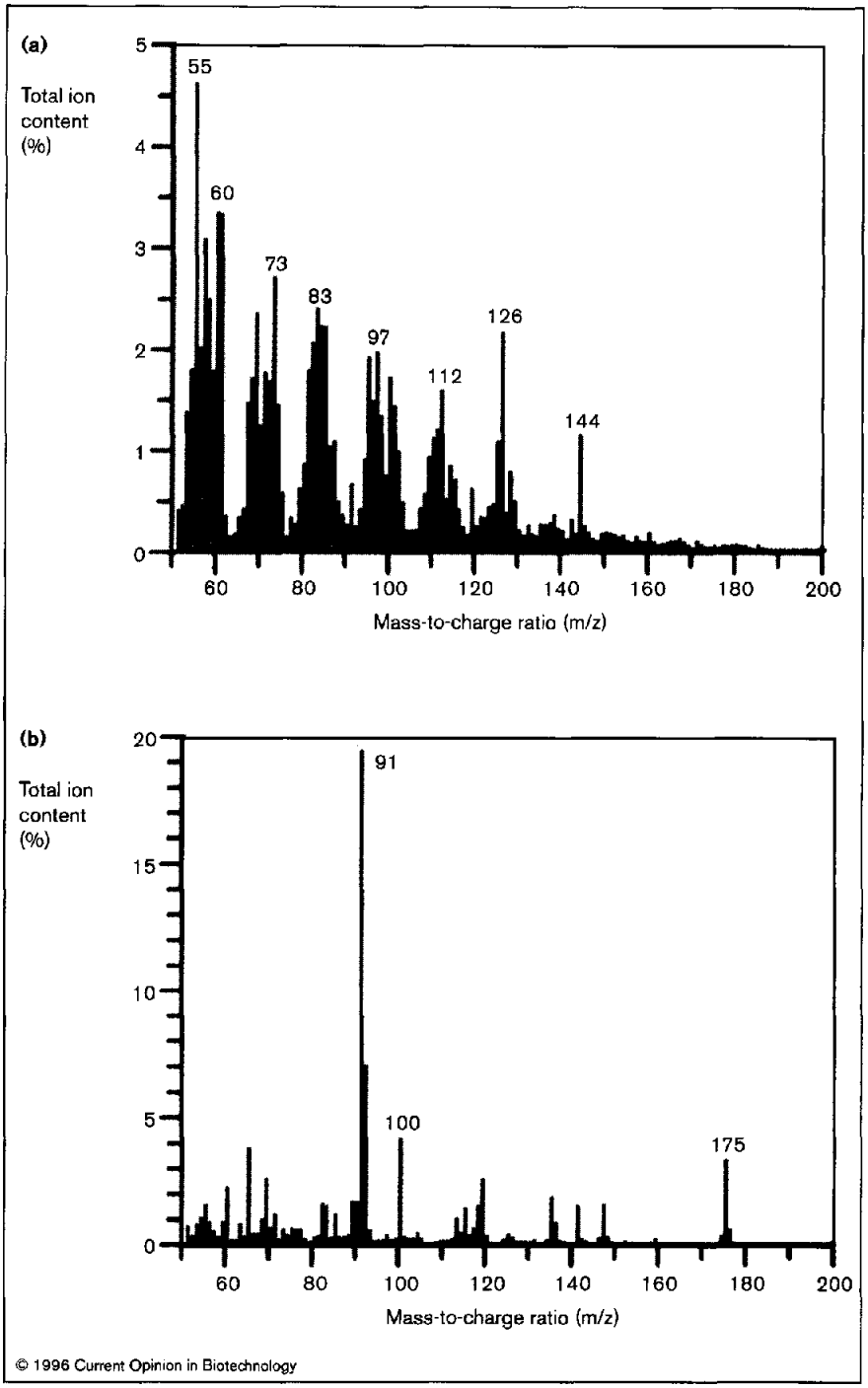
Conventionally, at least within microbiology and biotechnology, PyMS has been used as a taxonomic aid [13•,14, 15•,16•]. To this end, the reduction of the multivariate data generated by the PyMS system has normally been carried out using principal components analysis (PCA). PCA is a well known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance and so is an excellent technique for observing the natural relationships between samples. It neither takes account of any groupings in the data nor requires 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 enables the extraction of at least some chemically significant information.) The closely related canonical variates analysis (CVA) 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. The next stage involves the construction of a percentage similarity matrix by transforming the Mahalanobis' distance between *a priori* groups in CVA with the Gower similarity coefficient, S_G. Finally, hierarchical cluster analysis (HCA) may be employed to produce dendrograms, using average linkage clustering. Provided that the data set contains 'standards' (i.e. type strains or centro-strains), it is evident that one can establish the closeness of any unknown isolate to a known organism and thus effect the identification of the former [13•].

Within the clinical laboratory this approach has been used for the identification of potentially pathogenic isolates [17•,18,19•] and is now considered to be valuable for the rapid epidemiological typing of clinically significant pathogenic bacteria [14,16•]. Conventional typing systems are often slow and require specialized personnel; they also assay a rather restricted phenotype, often from only part of the bacterial cell (i.e. receptors located on the cell wall that bind to bacteriophage, antibodies or bacteriocins). PyMS, in contrast, analyzes the whole cell and so the pyrolysis mass spectrum contains all the relevant information. This technology has also been used for the analysis of causal agents of food spoilage [20•] and, within microbial ecology, for the characterization both of *Pseudomonas* spp. from estuarine environments [21] and of brown algae [22]. With regard to organisms of particular biotechnological interest, PyMS has been used for the detection of novel actinomycetes for pharmaceutical screening programs [23].

Because it is effectively studying the properties of a system in 150 dimensions (here, the m/z values range from 51 to 200) simultaneously, PyMS is a very high

Figure 1

The use of PyMS in the characterization and quantification of biotechnology-related bioprocesses. **(a)** Normalized pyrolysis mass spectrum of *Penicillium chrysogenum*; this complex 'fingerprint' can be used to type this organism. **(b)** Normalized pyrolysis mass spectrum of 200 µg pure penicillin G; this somewhat simpler 'biochemical profile' is one of many that can be obtained for the range of penicillins produced by *Penicillium chrysogenum*.



resolution technique and can easily discriminate between bacteria and fungi at the genus, species and subspecies levels. For example, this method has been used to analyze cell wall mutants of *Saccharomyces cerevisae* [24] and to characterize small changes in the lignin content of genetically engineered tobacco stems [25]. Furthermore, PyMS is not limited to whole cell analysis and can be used to differentiate bacteria on the basis of their fatty acid methyl ester distributions [26•]. Other volatilization methods have also been applied to the mass spectrometric analysis of microbial constituents in intact cells [27••,28•,29•,30,31].

Pyrolysis methods are extremely well suited for studies of the constitution of polymers. For instance, much interest is focused on poly-hydroxyalkanoates as potentially biodegradable plastics. Thus, the mechanisms for the thermal degradation of the bacterial polyester poly-hydroxybutyrates have been fervently discussed [32•,33•], and

particular attention has been paid to the complete and partial pyrolysis of poly(3-hydroxybutyrate) (PHB) [32•] as well as the effect that varying the sample thickness has on the thermolysis of PHB [34••].

With the recent advances in molecular biology, notably methods to study the similarity between bacterial strains on the basis of DNA and 16S-rRNA sequence homologies [35••], two schools of thought have developed as to how bacteria should be classified: some believe that the way forward is to study the bacterial genotype; others pursue phenetic classifications. The pyrolysis mass spectrum of a bacterium contains a fingerprint of its total biochemical make-up and, therefore, is a measure of the bacterial phenotype. Thus, it is necessary to compare PyMS, and indeed other measures of phenotype, with a microorganism's genotypic characteristics to get the overall microbial make-up.

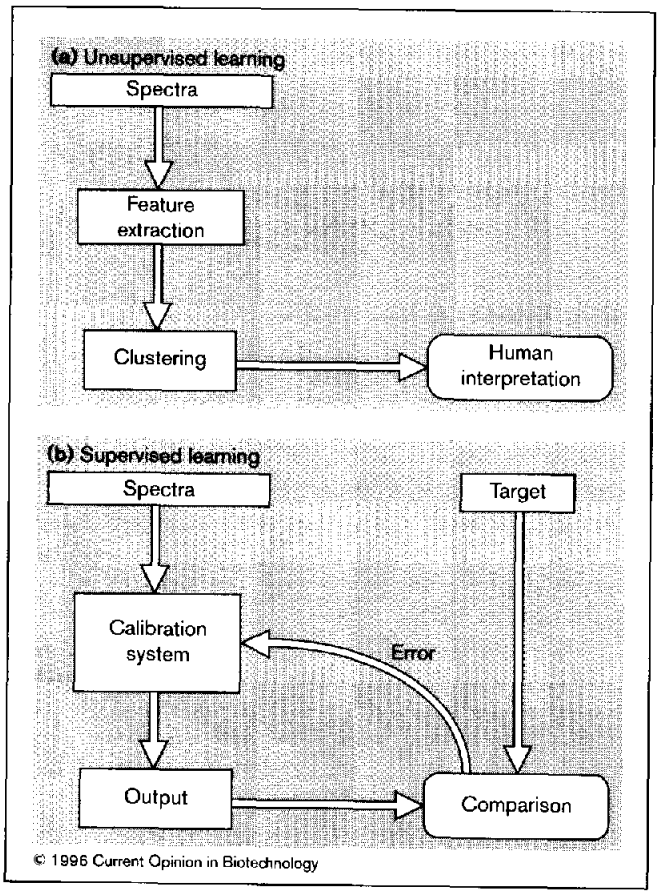
A recent study comparing the ability of random amplified polymorphic DNA (RAPD) method and PyMS in assessing the release of genetically engineered microorganisms in the environment investigated the survival of *Bradyrhizobium japonicum* in soil [36•]. Results indicated that both methods could be used to discriminate the rhizobial organism from other organisms in a similar fashion and were of value in studying the fate of the original inoculants. Congruence was also evident between the clusterings observed from PyMS and 16S rRNA sequence data in studies of the phenotypic and genotypic characteristics of several *Clostridium acetobutylicum* strains [37].

Given that any non-volatile biological material may be subjected to analysis by PyMS, it may be argued that this technique represents an 'anything sensor' for the analysis of any raw materials, intermediates or products of the bioprocess industries. Thus, the food industry has also exploited PyMS to confirm the authentication of orange juice [38] and scotch whisky [39•] as well as to detect the adulteration of virgin olive oil [40,41]; pyrolysis gas chromatography/MS (PyGC/MS) has been employed for the determination of the aspartame content in food products [42]. PyMS has also been exploited in the differentiation of beeswax products [43] and the characterization of Egyptian paint materials [44•].

Supervised versus unsupervised learning

The multivariate analyses used above (e.g. PCA, CVA and HCA) fall into the category of 'unsupervised learning', in which the relevant multivariate algorithms seek 'clusters' in the data [45]—although CVA may to some extent be considered a supervised method—thereby allowing the investigator to group objects together on the basis of their perceived closeness (Fig. 2a). Such methods, then, although in some sense quantitative, are better seen as qualitative because their chief purpose is merely to distinguish objects or populations. More recently, a

Figure 2



Flow diagrams representing the difference between unsupervised and supervised learning. **(a)** When learning is unsupervised, the system is shown a set of inputs (spectra) and then left to cluster the spectra into groups. For multivariate analysis, this optimization procedure is usually simplification or dimensionality reduction. This means that a large body of data (the spectral inputs) are summarized by means of a few parameters, with minimal loss of information. After clustering, the results must then be interpreted. **(b)** When the desired responses (targets) associated with each of the inputs (spectra) are known, the system may be supervised. The goal of supervised learning is to find a model that will correctly associate the inputs with the targets; this is usually achieved by minimizing the error between the target and the model's response (output).

variety of related, but much more powerful, methods, most often referred to within the framework of chemometrics [46••], have been applied to the 'supervised' analysis of multivariate PyMS data. In these methods, of which multiple linear regression, partial least squares regression and principal components regression are the most widely used, one seeks to relate the multivariate spectral inputs to the concentrations of target determinands (i.e. to generate a quantitative analysis, essentially via suitable types of multidimensional curve fitting or regression analysis [47]). A related approach [48,49••], which has been used to model and control bioprocesses [50•], is the use of ANNs.

ANNs are a well known means of uncovering complex, non-linear relationships in multivariate data, whilst still

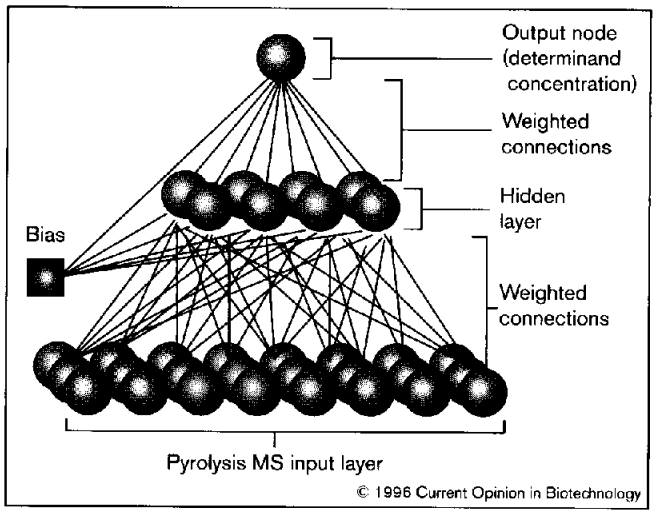
being able to map the linearities. ANNs can be considered as collections of very simple 'computational units' that take a numerical input and transform it (usually via summation) into an output [51–53,54••,55••,56•].

For a given analytical system, some patterns (e.g. mass spectra) have desired responses or values that are known (e.g. the concentration of target determinands). These two types of data form pairs that are called inputs and targets. The goal of supervised learning is to find a model or mapping that will correctly associate the inputs with the targets (Fig. 2b).

The relevant principle of 'supervised' learning in ANNs is thus that the ANNs take numerical inputs (the training data) and transform them into 'desired' (known or predetermined) outputs. The input and output nodes may be connected to the 'external world' and to other nodes within the network (for a diagrammatic representation, see Fig. 3). The way in which each node transforms its input depends on the so-called 'connection weights' (or 'connection strength') and 'bias' of the node, which are both modifiable. The output of each node, either to another node or to the external world, then depends both on its weight strength and bias and on the weighted sum of all its inputs, which are then transformed by a (normally non-linear) weighting function referred to as its activation, or squashing function. The great power of ANNs stems from the fact that they can be 'trained'. One can acquire sets of multivariate data (i.e. pyrolysis mass spectra) from standard materials of known identities and train ANNs using these identities as the desired outputs. Training is effected by continually presenting the networks with the 'known' inputs and outputs and modifying the connection weights between the individual nodes and the biases, typically according to some kind of back-propagation algorithm [51], until the output nodes of the network match the desired outputs to a stated degree of accuracy. The trained ANNs may then be exposed to unknown inputs (i.e. spectra) when they will immediately provide the globally optimal best fit to the outputs.

We [40,41] have provided the first demonstration of the ability of ANNs to discriminate successfully biological samples using pyrolysis mass spectra; in this work, which was performed double-blind such that the identities of the test set were not known to any of us, ANNs were trained with the spectra from 12 virgin olive oils and 12 adulterated oils, which at the output node were coded 1 and 0, respectively. This permitted the rapid and precise assessment of the adulteration of extra virgin olive oils with various seed oils, a task that previously was labour-intensive and very difficult. It was most significant that the traditional 'unsupervised' multivariate analyses of PCA, CVA and HCA failed to separate the oils according to their virginity or otherwise but rather discriminated them on the basis of their cultivar. Yet, partial least square regression and principal components

Figure 3



A back-propagation ANN consisting of 24 input nodes, eight hidden nodes, and one output node. Each node in the hidden layer is connected to all the nodes in the input and output layers. The actual number of nodes in the PyMS input layer was 150 (one for each mass from 51 to 200).

regression, which employ multivariate linear regression, were unable to effect the required distinction. This study thus demonstrated that 'supervised' learning methods that employ non-linear algorithms were needed for successful differentiation between virgin olive oils and adulterated olive oils. This combination of PyMS and ANNs has now been employed to effect the rapid identification of strains of *Escherichia* [57•], *Eubacterium* [58•], *Mycobacterium* [59•], *Propionibacterium* spp. [60••], and *Streptomyces* [61], whilst the time-dependent evolution of the PyMS has also been exploited to good effect [62•].

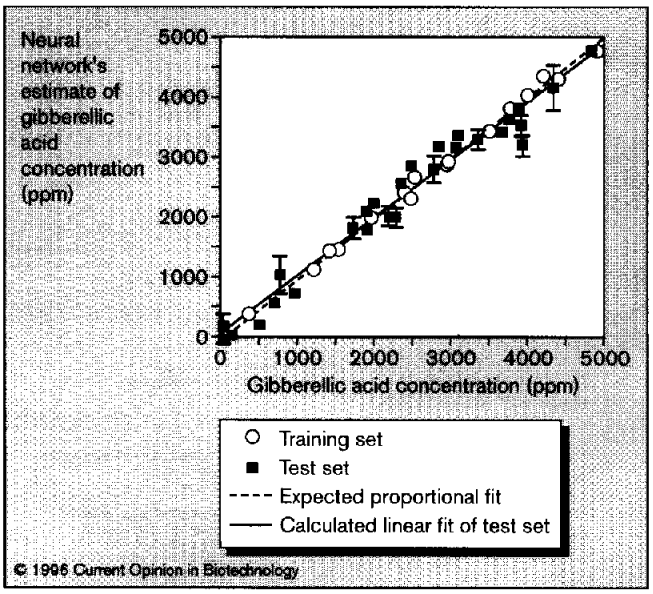
Quantification of biotechnological systems

Perhaps the most significant application of ANNs to the analysis of PyMS data in the review period is to gain accurate and precise quantitative information about the chemical constituents of microbial (and other) samples. For example, it has been shown that it is possible using this method to follow the production of indole in several strains of *Escherichia coli* grown on media incorporating various amounts of tryptophan [63], to quantify the (bio)chemical constituents of complex biochemical binary mixtures of proteins and nucleic acids in glycogen, and to measure the concentrations of tertiary mixtures of cells of the bacteria *Bacillus subtilis*, *E. coli* and *Staphylococcus aureus* [64,65•,66••]. We [66••] have also demonstrated that other supervised learning methods, such as partial least squares regression and principal components regression, could also be used to extract quantitative information from the spectra of the binary and tertiary mixtures.

The combination of PyMS and ANNs also has the potential for the screening and analysis of microbial

cultures producing recombinant proteins; for instance, this technique has permitted the amount of mammalian cytochrome b_5 expressed in *E. coli* to be accurately predicted [67•]. ANNs have also been applied to the quantitative analysis of the pyrolysis mass spectra of whole fermenter broths [68••]. Initially, a model system consisting of mixtures of the antibiotic ampicillin and either *E. coli* or *S. aureus* (to represent a variable biological background) was studied. It was especially interesting that ANNs trained to predict the amount of ampicillin in *E. coli*, having studied only mixtures of ampicillin and *E. coli*, were able to generalize so as to predict the concentration of ampicillin in a *S. aureus* background to ±~5% of the observed value, illustrating the very great robustness of ANNs to rather substantial variations in the biological background. Samples from fermentations of a single organism in a complex production medium were also analyzed quantitatively for a drug of commercial interest [68••]. It was found that the drug could also be quantified in a variety of mutant-producing strains cultivated in the same medium, thus effecting a rapid screen for the high-level production of desired substances [68••]. In a related study, *Penicillium chrysogenum* fermentation broths have been analyzed quantitatively for penicillins using PyMS and ANNs [69••]. Finally, we (R Goodacre, DAP Small, DB Kell, unpublished data) have also used this approach successfully to monitor *Gibberella fujikuroi* fermentations producing gibberellic acid (see Fig. 4).

Figure 4



A trained 150–8–1 ANN estimate of the amount of gibberellic acid in a range of Zeneca Bioproducts plc fermentations (samples supplied by DAP Small) compared with known values; the gibberellic acid titre was established using high-pressure liquid chromatography. ANNs were trained on 17 spectra with the standard back-propagation algorithm, until the error was optimal, as judged by test set cross-validation. Data points are the averages of the five separate trainings. Open circles represent spectra that were used to train the network and closed squares indicate 'unknown' spectra that were not in the training set. Error bars show standard deviation. The best linear fit is shown; the slope of this line is 0.957. The expected proportional fit is also shown.

Drift and reproducibility

The major problem with PyMS is that long-term reproducibility (>30 days) is poor and the mass spectral fingerprints of the same material analyzed at two different times are different; this lack of reproducibility largely results from instrumental drift in the mass spectrometer (and is not confined to PyMS). Therefore, within clinical microbiology, PyMS has really been limited to the typing of short-term outbreaks where all microorganisms are analyzed in a single batch [16•]. For PyMS to be used both for the routine identification of microorganisms and in combination with ANNs to quantify biological systems (e.g. metabolites of interest in fermenter broths), new spectra must be comparable with those previously collected.

We [P1••,70••] have found that ANNs can be used to correct for instrumental drift so that models created using old previously collected data can be employed to give accurate estimates of determinand concentration or bacterial identities from newly acquired spectra when calibrated with standards common to the two data sets. Calibration samples are run at the two times, and ANNs are set up in which the inputs are the 150 'new' calibration masses and the outputs are the 150 calibration masses from the 'old' spectra. Such autoassociative nets can thus be used as signal-processing elements to effect the transformation of data acquired one day to data that are acquired at a later date. Therefore, for the first time, PyMS can be used to acquire spectra that are comparable with those previously collected and held in a database.

Conclusions

Within biotechnology, PyMS is very useful for the discrimination of microorganisms at the genus, species and subspecies level. Compared with more conventional methods, PyMS offers the advantages of speed, sensitivity and the ability to analyze many hundreds of samples per day.

The exploitation of the novel multivariate analysis technique employing ANNs—and indeed the methods of partial least squares regression and principal components regression—that are based on supervised learning, rather than unsupervised methods, has permitted even better discrimination of industrially and medically important bacteria from their pyrolysis mass spectra and has allowed the rapid and quantitative analysis of microbial constituents. Within biotechnology, we may anticipate that the application of these powerful 'supervised' learning techniques, and the exploitation of other techniques of artificial intelligence [71••], will allow mass spectroscopists working in diverse fields to effect the rapid, sensitive and simultaneous analyses of the concentrations of many

substrates, metabolites and products in fermentation processes.

Acknowledgements

R Goodacre is currently funded as a Research Fellow by the Wellcome Trust (Grant number 042615/Z/94/Z). DB Kell thanks the Chemicals & Pharmaceuticals Directorate of the UK BBSRC for financial support under the terms of the LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, Neural Computer Sciences and Zeneca Bioproducts plc.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Porter N, Fox FM: **Diversity of microbial products – discovery and application.** *Pesticide Sci* 1993, **39**:161–168.

2. •• Bevan P, Ryder H, Shaw A: **Identifying small-molecule lead compounds – the screening approach to drug discovery.** *Trends Biotechnol* 1995, **13**:115–121.
The view from Xenova UK on untapped biodiversity, the rationale for high-throughput screening systems, and how they have been able to detect some 200 lead compounds (including 61 novel ones, mainly from fungal sources) from 10^6 cultures per year.

3. Tanaka Y, Mura S: **Agroactive compounds of microbial origin.** *Annu Rev Microbiol* 1993, **47**:57–87.

4. •• Gordon EM, Gallop MH, Campbell D, Holmes C, Bermak J, Look G, Murphy M, Needels M, Jacobs J, Sugarman J *et al.*: **Combinatorial organic synthesis – applications to drug discovery.** *Eur J Med Chem* 1995, **30(No SS)**:S337–S348.
The view from Affymax, the home of combinatorial chemistry, on solid-phase synthesis. An excellent entrée to the literature.

5. • Scheper T-H, Lammers F: **Fermentation monitoring and process control.** *Curr Opin Biotechnol* 1994, **5**:187–191.
In this review, improving the interface between the sensor and the bioprocess is highlighted as the major present hurdle to control. More on-line data, however, should lead to better process models.

6. • Kell DB, Sonnleitner B: **GMP – Good Modelling Practice.** *Trends Biotechnol* 1995, **13**:481–492.
Many bioprocess models, particularly those of a predictive nature, do not take into account some fundamental principles, and the somewhat arcane nature of the field has allowed their proponents to get away with it. In this paper, the authors provide a guide for assisting the development of these methods and their integration into mainstream biotechnology.

7. Lee MS, Hook DJ, Kerns EH, Volk KJ, Rosenberg IE: **Rapid screening of fermentation broths for flavones using tandem mass-spectrometry.** *Biol Mass Spectrom* 1993, **22**:84–88.

8. Heinzle E: **Present and potential applications of mass-spectrometry for bioprocess research and control.** *J Biotechnol* 1992, **25**:81–114.

9. Hansen KF, Lauritsen FR, Degn H: **An on-line sampling system for fermentation monitoring using membrane inlet mass-spectrometry (MIMS) – application to phenoxyacetic acid monitoring in penicillin fermentation.** *Biotechnol Bioeng* 1994, **44**:347–353.

10. Meuzelaar HLC, Haverkamp J, Hileman FD: *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials.* Amsterdam: Elsevier; 1982.

11. Chapman JR: *Practical Organic Mass Spectrometry.* New York: Wiley & Sons; 1993.

12. Berkeley RCW, Goodacre R, Helyer RJ, Kelley T: **Pyrolysis-mass spectrometry in the rapid identification of micro-organisms.** *Lab Practice* 1990, **39**:81–83.

13. • Goodacre R: **Characterisation and quantification of microbial systems using pyrolysis mass spectrometry: introducing neural networks to analytical pyrolysis.** *Microbiol Eur* 1994, **2**:16–22.
This short review, which is aimed at the bacterial taxonomist, gives an introduction to PyMS and illustrates how the application of ANNs can benefit systematics.

14. Magee JT: **Whole-organism fingerprinting.** In *Handbook of New Bacterial Systematics.* Edited by Goodfellow M, O'Donnell AG. London: Academic Press; 1993:383–427.

15. • Snyder AP, Smith PBW, Dworzanski JP, Meuzelaar HLC: **Pyrolysis-gas chromatography-mass spectrometry – detection of biological warfare agents.** *ACS Symp Ser* 1994, **541**:62–84.
In this paper, the US army attempts to develop analytical pyrolysis for extracting microbiological information from whole organisms in the field. The need for a simplified graphical representation, suitable for direct interpretation by non-microbiologists (as opposed to purely computerized 'black box' pattern recognition methods) is also discussed.

16. • Goodfellow M: **Inter-strain comparison of pathogenic micro-organisms by pyrolysis mass-spectrometry.** *Binary Comput Microbiol* 1995, **7**:54–60.
The nature of PyMS as a rapid screening tool for epidemiological studies is discussed along with its lack of long-term reproducibility. Therefore, the clinician still needs to seek stable bacterial markers for national records. This paper illustrates some of the pitfalls of using ANNs simply as a black box.

17. • Murdoch DA, Magee JT: **A numerical taxonomic study of the Gram-positive anaerobic cocci.** *J Med Microbiol* 1995, **43**:148–155.
The classifications derived from PyMS analysis compare favourably with those established using typing methods based on conventional test reaction patterns for species of *Peptostreptococcus.*

18. Magee JT, Philpot C, Yang J, Hosein IK: **Pyrolysis typing of isolates from a recurrence of systemic cryptococcosis.** *J Med Microbiol* 1994, **40**:165–169.

19. • Freeman R, Sisson PR, Jenkins DR, Ward AC, Lightfoot NF, O'Brien SJ: **Sporadic isolates of *Escherichia coli* O157:H7 investigated by pyrolysis mass-spectrometry.** *Epidemiol Infect* 1995, **114**:433–440.
This paper illustrates how the regional Public Health Laboratory in Newcastle is exploiting PyMS. As part of an ongoing series of such studies, PyMS is once again demonstrated to be a useful addition to existing typing methods in the investigation of the epidemiology of bacteria.

20. • Manchester LN, Toole A, Goodacre R: **Characterisation of *Carnobacterium* species by pyrolysis mass spectrometry.** *J Appl Bacteriol* 1995, **78**:88–96.
The typing of this group of food-spoilage organisms by PyMS compares favourably with genomic classification methods. Culture age had little effect on the PyMS spectra and short-term reproducibility over four weeks was excellent.

21. Helyer RJ, Bale SJ, Berkeley RCW: **The application of pyrolysis mass-spectrometry to microbial ecology – rapid characterization of bacteria isolated from an estuarine environment.** *J Anal Appl Pyrol* 1993, **25**:265–272.

22. Russell G: **Pyrolysis mass-spectrometry – a fresh approach to old problems in brown algal systematics.** *Mar Biol* 1995, **123**:153–157.

23. Sanglier JJ, Whitehead D, Saddler GS, Ferguson EV, Goodfellow M: **Pyrolysis mass-spectrometry as a method for the classification, identification and selection of actinomycetes.** *Gene* 1992, **115**:235–242.

24. De Nobel JG, Munnik T, Pureveen JBM, Eijkel GB, Mulder MM, Boon JJ, Van Den Ende H, Klis FM: **Analysis of cell-wall mutants of *Saccharomyces cerevisiae* by pyrolysis mass-spectrometry.** *Acta Bot Neerlandica* 1993, **42**:505–516.

25. Halpin C, Knight ME, Foxon GA, Campbell MM, Boudet AM, Boon JJ, Chabbert B, Tollier MT, Schuch W: **Manipulation of lignin quality by down-regulation of cinnamyl alcohol-dehydrogenase.** *Plant J* 1994, **6**:339–350.

26. • Basile F, Voorhees KJ, Hadfield TL: **Micro-organism Gram-type differentiation based on pyrolysis mass-spectrometry of bacterial fatty-acid methyl-ester extracts.** *Appl Environ Microbiol* 1995, **61**:1534–1539.
PyMS of a variety of organisms reveals an interesting correlation between Gram-type and fatty acid content, Gram-positive organisms containing more branched fatty acids. A correlation is also observed between fatty acid content and pathogenicity.

27. •• McClennen WH, Arnold NS, Meuzelaar HLC: **Field-portable hyphenated instrumentation – the birth of the tricorder.** *Trends Anal Chem*, 1994, **13**:286–293.
One-dimensional methods (e.g. MS) are now being replaced by two-dimensional or higher-dimensional techniques (e.g. GC/MS), also referred to as hyphenated techniques. Miniaturization is assuming increasing importance. With due reference to *Star Trek*, the authors postulate that 'the next generation' of field-portable analytical instruments will be based on analytical pyrolysis.

28. Smith PBW, Snyder AP, Harden CS: **Characterization of**
• **bacterial phospholipids by electrospray-ionization tandem mass-spectrometry.** *Anal Chem* 1995, **67**:1824–1830.
Electrospray ionization (ESI) tandem MS is applied to the profiling of glycerophospholipids present in the chloroform/methanol extracts of four different bacterial species. Differences in the fatty acid composition for each bacterial species are readily apparent from a visual examination of the data sets.

29. Black GE, Fox A, Fox K, Snyder AP, Smith PBW: **Electrospray**
• **tandem mass-spectrometry for analysis of native muramic acid in whole bacterial-cell hydrolysates.** *Anal Chem* 1994, **66**:4171–4176.
Muramic acid in hydrolysates from whole microbial cells are quantified using ESI tandem MS. Compared with the more conventional GC/MS, the sample preparation is simpler and the ESI tandem MS sample analysis much more rapid (2 min compared with 30–50 mins for GC/MS); the authors suggest (correctly) that this approach could be adapted for rapid bacterial identification.

30. Snyder AP: **Analyte detection in complex solid matrices with pyrolysis atmospheric-pressure chemical-ionization tandem mass-spectrometry.** *Trends Anal Chem*, 1993, **12**:296–303.

31. Smith PBW, Snyder AP: **Characterization of bacteria by oxidative and non-oxidative pyrolysis-gas chromatography ion trap mass-spectrometry.** *J Anal Appl Pyrol* 1993, **24**:199–210.

32. Lehrle R, Williams R, French C, Hammond T: **Thermolysis**
• **and methanolysis of poly(beta-hydroxybutyrate)—random scission assessed by statistical-analysis of molecular-weight distributions.** *Macromolecules* 1995, **28**:4408–4414.
Total thermolysis of PHB can be interpreted exclusively in terms of a random chain-scission mechanism. In this report, however, partial pyrolysis results were not consistent with random scission statistics; this implies that some kinetically favoured scissions are predominant (in fact, they occur near the ends of the molecules). The product distributions from methanolysis were also somewhat inconsistent with those expected on the basis of random scission, and the authors suggest that dimers hydrolyze more readily than do higher molecular weight species.

33. Abate R, Ballistreri A, Montaudo G, Impallomeni G: **Thermal**
• **degradation of microbial poly(4-hydroxybutyrate).** *Macromolecules* 1994, **27**:332–336.
The thermal degradation of poly(4-hydroxybutyrate) (P4HB) is investigated using thermogravimetry, direct PyMS and also by preparative pyrolysis and subsequent NMR analysis of the pyrolyzate. The authors' findings indicate that the thermal decomposition of P4HB yields a series of cyclic oligomers by an intramolecular exchange mechanism.

34. Lehrle RS, Williams RJ: **Thermal degradation of bacterial**
•• **poly(hydroxybutyric acid)—mechanisms from the dependence of pyrolysis yields on sample thickness.** *Macromolecules* 1994, **27**:3782–3789.
The authors use a novel approach, employing variation in the sample thickness, that provides indirect control of the residence time of primary products in the melt and thereby facilitates the detection of secondary reactions. From quantitative measurments of 'bonus yields' and 'deficit yields' of monomeric, dimeric, trimeric, and tetrameric products, further doubt is cast on the view that the pyrolysis products from PHB can be accounted for entirely in terms of random (β-elimination) scissions.

35. Amann RI, Ludwig W, Schleifer KH: **Phylogenetic identification**
•• **and *in-situ* detection of individual microbial cells without cultivation.** *Microbiol Rev* 1995, **59**:143–169.
A comprehensive account of the direct retrieval of ribosomal RNA sequences and of whole-cell oligonucleotide probing. The authors discuss how a combination of these approaches may be used both to detect specific rRNA sequences of uncultured bacteria in natural samples and to identify individual cells using microscopy or flow cytometry.

36. Kay HE, Coutinho HLC, Fattori M, Manfio GP, Goodacre R,
• Nut MP, Basaglia M, Beringer JE: **The identification of *Bradyrhizobium japonicum* strains isolated from Italian soils.** *Microbiology* 1994, **140**:2333–2339.
PyMS and the RAPD method are used to characterize the survival of rhizobial strains from fields in northern Italy. PyMS (but not RAPD) showed that the derivatives from one inoculant formed two distinct populations: one like the parent strain, the other altered phenotypically. This demonstrates that PyMS can be employed to assess the release of genetically engineered microorganisms in the environment.

37. Wilkinson SR, Young M, Goodacre R, Morris JG, Farrow JAE, Collins MD: **Phenotypic and genotypic differences between certain strains of *Clostridium acetobutylicum*.** *FEMS Microbiol Lett* 1995, **125**:199–204.

38. Aries RE, Gutteridge CS, Evans R: **Rapid characterization of orange juice by pyrolysis mass-spectrometry.** *J Food Sci* 1986, **51**:1183–1186.

39. Aylott RI, Clyne AH, Fox AP, Walker DA: **Analytical strategies**
• **to confirm scotch whisky authenticity.** *Analyst* 1994, **119**:1741–1746.
Characteristic analytical profiles for various higher-alcohol congeners provide a valuable method for checking the authenticity of scotch whisky. PyMS allows the rapid confirmation of conclusions from chromatographic analysis; furthermore, the authors speculate that PyMS has the potential to be a useful 'stand alone' technique for authenticity analysis.

40. Goodacre R, Kell DB, Bianchi G: **Neural networks and olive oil.** *Nature* 1992, **359**:594.

41. Goodacre R, Kell DB, Bianchi G: **Rapid assessment of olive oil adulteration using pyrolysis mass spectrometry and artificial neural networks.** *J Sci Food Agric* 1993, **63**:297–307.

42. Galletti GC, Chiavari G, Bocchini P: **Thermal-decomposition products of aspartame as determined by pyrolysis-gas chromatography mass-spectrometry.** *J Anal Appl Pyrol* 1995, **32**:137–151.

43. Beverly B, Kay PT, Voorhees KJ: **Principal component analysis of the pyrolysis-mass spectra from African, Africanized hybrid, and European beeswax.** *J Anal Appl Pyrol* 1995, **34**:251–263.

44. Chiavari G, Fabbri D, Galletti GC, Mazzeo R: **Use of**
• **analytical pyrolysis to characterize Egyptian painting layers.** *Chromatographia* 1995, **40**:594–600.
Archaeological objects are analyzed directly using PyGC/MS and Fourier-transform infrared (FT-IR) spectroscopy. PyGC/MS is shown to be a rapid approach for discriminating between wax-based binders and animal protein based binders and is able to recognize mixtures of the two products. FT-IR spectroscopy corroborates the pyrolysis findings, but sometimes provides less unequivocal results than PyGC/MS.

45. Everitt BS: *Cluster Analysis.* London: Edward Arnold; 1993.

46. Brown SD, Blank TB, Sum ST, Weyer LG: **Chemometrics.** *Anal*
•• *Chem* 1994, **66**:R315–R359.
A very comprehensive review in this biennial series, discussing the application of a wide range of statistical methods within analytical chemistry.

47. Martens H, Næs T: *Multivariate Calibration.* New York: John Wiley & Sons; 1989.

48. Næs T, Kvaal K, Isaksson T, Miller C: **Artificial neural networks in multivariate calibration.** *J Near Infrar Spectrosc* 1993, **1**:1–11.

49. Ripley BD: **Neural networks and related methods for**
•• **classification.** *J Roy Stats Soc Ser B* 1994, **56**:409–437.
Presents a general framework for classification within which methods from statistics, ANNs, pattern recognition and machine learning can be compared. ANNs emerge as one of a class of flexible non-linear regression methods that can be used to classify via supervised learning.

50. Montague G, Morris J: **Neural-network contributions in**
• **biotechnology.** *Trends Biotechnol* 1994, **12**:312–324.
A review of some of the applications of a variety of ANNs methodologies to bioprocess modelling, control and pattern recognition.

51. Rumelhart DE, McClelland JL, PDP Research Group: *Parallel Distributed Processing, Experiments in the Microstructure of Cognition*, vols 1 and 2. Cambridge, Massachusetts: MIT Press; 1986.

52. Wasserman PD: *Neural Computing: Theory and Practice.* New York: Van Nostrand Reinhold; 1989.

53. Zupan J, Gasteiger J: *Neural Networks for Chemists: an Introduction.* Weinheim: VCH Verlagsgeeselischaft; 1993.

54. Goodacre R, Neal MJ, Kell DB: **Quantitative analysis of**
•• **multivariate data using artificial neural networks: a tutorial review and applications to the deconvolution of pyrolysis mass spectra.** *Zbl Bakt* 1996, in press.
The implementation of ANNs for the analysis of multivariate data is reviewed, with particular reference to the analysis of pyrolysis mass spectra. The need for, and benefits of, multivariate data analysis are explained followed by a discussion of ANNs and their optimization.

55. Haykin S: *Neural Networks.* New York: Macmillan; 1994.
••
A very comprehensive book telling you all you ever wanted to know about ANNs. A must for both the beginner and the more experienced neural net enthusiast.

56. Dybowski R, Gant V: **Artificial neural networks in pathology and**
• **medical laboratories.** *Lancet* 1995, **346**:1203–1207.
A comprehensive account of the wide-ranging applications of ANNs in the clinical laboratory. ANNs are shown to extend the capabilities of the otherwise 'spacious electric filing cabinets' (computers) to the study of anatomi-

cal pathological system PAPNET (used to screen cervical smears), clinical chemistry, and general microbiology applications.

57. • Sisson PR, Freeman R, Law D, Ward AC, Lightfoot NF: **Rapid detection of verocytotoxin production status in *Escherichia coli* by artificial neural-network analysis of pyrolysis mass-spectra.** *J Anal Appl Pyrol* 1995, **32**:179–185.

A combination of ANNs and PyMS is exploited to detect verocytotoxin production in *E. coli* isolates. Although only two exemplars (mass spectra) were used to train the 150–8–1 ANN, satisfactory generalization was (perhaps surprisingly) found.

58. • Goodacre R, Hiom SJ, Cheeseman SL, Murdoch D, Weightman AJ, Wade WG: **Identification and discrimination of oral asaccharolytic *Eubacterium* spp. using pyrolysis mass spectrometry and artificial neural networks.** *Curr Microbiol* 1996, in press.

The need for numerical methods that allow easy direct interpretation of pyrolysis mass spectra for the identification of bacteria is illustrated well with this large set of data. ANNs can be encoded simply and the results read off in a tabulated format, whereas the methods still employed by most users of PyMS involve the more complex examination of three-dimensional ordination plots and dendrograms.

59. • Freeman R, Goodacre R, Sisson PR, Magee JG, Ward AC, Lightfoot NF: **Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of pyrolysis mass spectra.** *J Med Microbiol* 1994, **40**:170–173.

PyMS and ANNs are used to distinguish between *Mycobacterium tuberculosis* and *Mycobecterium bovis*, irrespective of their susceptibility to antituberculosis agents. This is significant because it is not presently possible using DNA probes to differentiate between species of the *M. tuberculosis* complex.

60. •• Goodacre R, Neal MJ, Kell DB, Greenham LW, Noble WC, Harvey RG: **Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs.** *J Appl Bacteriol* 1994, **76**:124–134.

The classic approach for the identification of bacteria from their pyrolysis mass spectra is often subjective because it relies on the interpretation of complicated ordination plots and dendrograms; however, ANNs are a more objective alternative for this purpose. Furthermore, this paper contains the first demonstration, within microbiology, of the use of self-organizing feature maps (an unsupervised neural clustering technique) for grouping pyrolysis mass spectra.

61. Chun J, Atalan E, Ward AC, Goodfellow M: **Artificial neural network analysis of pyrolysis mass-spectrometric data in the identification of *Streptomyces* strains.** *FEMS Microbiol Lett* 1993, **107**:321–325.

62. • Mavrovouniotis ML, Harper AM, Ifarraguerri AI: **Classification of pyrolysis mass-spectra of biological materials using convex cones.** *J Chemometrics* 1994, **8**:305–331.

A rather complex account of the use of convex cones to classify high-dimensional time-dependent pyrolysis mass spectra of biological samples; well worth a glance for the more mathematically inclined.

63. Goodacre R, Kell DB: **Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production.** *Anal Chim Acta* 1993, **279**:17–26.

64. Goodacre R, Edmonds AN, Kell DB: **Quantitative analysis of the pyrolysis-mass spectra of complex mixtures using artificial neural networks: application to casamino acids in glycogen.** *J Anal Appl Pyrol* 1993, **26**:93–114.

65. • Neal MJ, Goodacre R, Kell DB: **On the analysis of pyrolysis mass spectra using artificial neural networks. Individual input scaling leads to rapid learning.** In *Proceedings of the World Congress on Neural Nets, San Diego.* San Diego: Lawrence Erlbaum Associates Publishers; 1994:I-318–I-323.

Standard back-propagation ANNs often take a long time to train. In this paper, a 100-fold decrease in the training time, with no loss in precision of prediction of the test set, is achieved by scaling each input individually.

66. •• Goodacre R, Neal MJ, Kell DB: **Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks.** *Anal Chem* 1994, **66**:1070–1085.

A comprehensive comparison of ANNs and other supervised learning methods that employ multivariate linear regression, such as partial least squares and principal components regression, for the deconvolution of pyrolysis mass spectra obtained from binary mixtures of the protein lysozyme and nucleic acids in glycogen and tertiary mixtures of bacteria. It is noteworthy that all three strains in tertiary mixtures of bacteria could be quantified accurately and simultaneously.

67. • Goodacre R, Karim A, Kaderbhai MA, Kell DB: **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* 1994, **34**:185–193.

The combination of PyMS and ANNs is demonstrated to have potential for the screening and analysis of microbial cultures producing recombinant proteins.

68. •• Goodacre R, Trew S, Wrigley-Jones C, Neal MJ, Maddock J, Ottley TW, Porter N, Kell DB: **Rapid screening for metabolite overproduction in fermenter broths using pyrolysis mass spectrometry with multivariate calibration and artificial neural networks.** *Biotechnol Bioeng* 1994, **44**:1205–1216.

Levels of a drug of commercial interest are quantified accurately in fermentation samples of a single organism. Mutant producer strains are also analyzed, and PyMS/ANNs is demonstrated to be a rapid method for screening for the high-level production of desired substances. Furthermore, when predicting ampicillin concentration in model mixtures, ANNs are observed to be very robust to rather substantial variations in the biological background (either *E. coli* or *S. aureus*).

69. •• Goodacre R, Trew S, Wrigley-Jones C, Saunders G, Neal MJ, Porter N, Kell DB: **Rapid and quantitative analysis of metabolites in fermenter broths using pyrolysis mass spectrometry with supervised learning: application to the screening of *Penicillium chrysogenum* fermentations for the overproduction of penicillins.** *Anal Chim Acta* 1995, **313**:25–43.

As well as illustrating the capacity of PyMS/ANNs for the rapid screening of fermentation broths, this study compares the applicability of various squashing functions on the output nodes of otherwise identical ANNs. It is shown that ANNs employing linear output functions give more accurate estimates of determinand concentration near the edges of the concentration range than those using sigmoidal functions. In addition, these ANNs can be used successfully to extrapolate beyond the concentration range on which they have been trained.

70. •• Goodacre R, Kell DB: **On correcting mass spectral drift using neural networks.** *Anal Chem* 1996, in press.

Reports the utility of ANNs for correcting instrumental drift. ANN or other multivariate calibration models, created using previously collected data, can be used to give accurate estimates of determinand concentration or the nature of bacteria (and indeed other materials) from newly acquired pyrolysis mass spectra by running standards each time and using ANNs to transform 'old' spectra into 'new' spectra.

71. •• Michie D, Spiegelhalter DJ, Taylor CC: **Machine learning: neural and statistical classification.** In *Ellis Horwood Series in Artificial Intelligence.* Edited by Campbell J. Chichester: Ellis Horwood; 1994.

A useful comparison, based on a European Commission collaboration, of the main methods of artificial intelligence currently in use for classification. Which one is best depends rather on the data set used.

Patents

- • of special interest
- •• of outstanding interest

P1. •• Goodacre R, Kell DB: **Composition analysis.** 8/6/95, UK 9511619.0.

A description of how ANNs can be used successfully to correct for instrumental drift is disclosed, with particular reference to PyMS.