Cytotechnology **21:** 231–241, 1996.

Rapid authentication of animal cell lines using pyrolysis mass spectrometry and auto-associative artificial neural networks

Royston Goodacre,* Deborah J. Rischert, Peter M. Evans & Douglas B. Kell
Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed, SY23 3DA, Wales, U.K.

Received 21 November 1995; accepted in final form 9 April 1996

Key words: authentication, auto-associative neural networks, chemometrics, feature extraction, pyrolysis mass spectrometry, cell line

Summary

Pyrolysis mass spectrometry (PyMS) was used to produce biochemical fingerprints from replicate frozen cell cultures of mouse macrophage hybridoma 2C11–12, human leukaemia K562, baby hamster kidney BHK 21/C13, and mouse tumour BW-O, and a fresh culture of Chinese hamster ovary CHO cells. The dimensionality of these data was reduced by the unsupervised feature extraction pattern recognition technique of auto-associative neural networks. The clusters observed were compared with the groups obtained from the more conventional statistical approaches of hierarchical cluster analysis. It was observed that frozen and fresh cell line cultures gave very different pyrolysis mass spectra. When only the frozen animal cells were analysed by PyMS, auto-associative artificial neural networks (ANNs) were employed to discriminate between them successfully. Furthermore, very similar classifications were observed when the same spectral data were analysed using hierarchical cluster analysis. We demonstrate that this approach can detect the contamination of cell lines with low numbers of bacteria and fungi; this approach could plausibly be extended for the rapid detection of mycoplasma infection in animal cell lines. The major advantages that PyMS offers over more conventional methods used to type cell lines and to screen for microbial infection, such as DNA fingerprinting, are its speed, sensitivity and the ability to analyse hundreds of samples per day. We conclude that the combination of PyMS and ANNs can provide a rapid and accurate discriminatory technique for the authentication of animal cell line cultures.

Introduction

Within animal cell technology it is of paramount importance to have a reliable source of pure authenticated cell cultures (Hay, 1988; Mowles & Doyle, 1990; Stacey *et al.*, 1992a). Research conducted on either mixed cell lines or cultures that have been contaminated with mycoplasma (Uphoff *et al.*, 1992a; Uphoff *et al.*, 1992b), bacteria, or virions (Nicklas *et al.*, 1993) can be deceptive and therefore invalid. Moreover, if contaminated cultures were employed within the pharmaceutical industry for the production of vaccines or other therapeutics including the production of recombinant DNA-derived proteins (Werner & Noe, 1993), or for the testing of novel pharmacophores, then there would be obvious and major concern about both the safety of the products and the validity of the *in vitro* testing procedures (Werner *et al.*, 1992).

Cell lines have typically been characterised by a variety of different methods, one of the common procedures used being isoenzyme analysis (Steube *et al.*, 1995); this method exploits the facts that isoenzymes from different cell lines have distinct molecular structures, and that the isoenzymes can thus be separated by electrophoresis and used to type the cell line. Alternatively, cells can be discriminated by karyotyping (Flores & Donis, 1995); this cytogenetic analysis procedure involves the visual examination of the

* Author for correspondence.

chromosomes, and if present, aberrations in chromosome number and/or morphology may then be detected. Finally, DNA fingerprinting has been exploited to identify specific cell lines (Stacey *et al.*, 1992a; Stacey *et al.*, 1992b; Park *et al.*, 1995). This technique uses multilocus probes (Stacey *et al.*, 1992b) allowing a wide range of repetitive DNA sequences to be detected in many animal species. Isoenzyme analysis and karyotyping are slow and the inter-laboratory reproducibility poor; indeed the quality of the data from cytogenetic analysis are dependent on the experience of the operator (Stacey *et al.*, 1992a). DNA fingerprinting, although relatively rapid after the extraction of the DNA, also has some serious pitfalls. The reproducibility of data acquired are often dependent on tissue culture conditions and the effects of storing cells for long periods of time; however, provided the experimental conditions are stringently controlled the data obtained are considered to be acceptable (Stacey *et al.*, 1992a). A particular advantage of DNA fingerprinting cell cultures is that the banding pattern observed can be digitised and stored in a computer database; unfortunately, there has been limited success in the automatic screening of such databases to effect the identification of unknown cell lines.

In addition to the characterisation of cell lines the methods outlined above have also been used to detect the infection of cell cultures with microorganisms. Perhaps the most significant and common group of microbes that have been found to contaminate animal cell cultures are the mycoplasmas. It is probable that these bacteria, which lack cell walls and thus are favoured by the prescence of penicillin and other antibiotics in the standard tissue culture media, might change cell growth (Werner & Noe, 1993). Rather more specialised methods which exploit the polymerase chain reaction (PCR) have been developed to detect mycoplasma contamination in cell lines (Hopert *et al.*, 1993a; Hopert *et al.*, 1993b; Dussurget *et al.*, 1994).

The ideal method for the rapid, precise and accurate analysis characterization of cell lines, would permit the simultaneous estimation of cell line authentication and/or microbial infection, would have minimum sample preparation, would analyse samples directly (i.e. would not require reagents), and would be rapid, automated, accurate and (at least relatively) cheap. Pyrolysis mass spectrometry (PyMS) is a rapid, automated, instrument-based technique which permits the acquisition of spectroscopic data from 300 or more samples per working day.

Pyrolysis is the thermal degradation of complex molecules in a vacuum causing their cleavage to smaller, volatile fragments called pyrolysate (Irwin, 1982) separable on the basis of their mass-to-charge ratio (m/z) so as to produce a pyrolysis mass spectrum, which can then be used as a "chemical profile" or fingerprint of the complex material analysed (Meuzelaar *et al.*, 1982; Magee, 1993; Goodacre, 1994). PyMS has been applied to the characterisation and identification of a variety of micro-organisms and their products (Magee, 1993; Goodacre, 1994; Snyder *et al.*, 1994; Goodacre & Kell, 1996a) and, because of its high discriminatory ability (Goodacre & Berkeley, 1990), presents a powerful fingerprinting technique, which is applicable to any biological material.

Recently there has been a number of important advances in the field of pyrolysis mass spectrometry (PyMS), in particular the fact that the relatively new numerical techniques of artificial neural networks (ANNs) (see refs. Rumelhart *et al.*, 1986; Kohonen, 1989; Wasserman, 1989; Beale & Jackson, 1990; Hecht-Nielsen, 1990; Simpson, 1990; Hertz *et al.*, 1991; Zupan & Gasteiger, 1993; Haykin, 1994; Ripley, 1994; Werbos, 1994; Chauvin & Rumelhart, 1995; Goodacre *et al.*, 1996c for excellent introductions) have been applied to pyrolysis mass spectra to gain *quantitative*, as well as qualitative, information about the chemical constituents of microbial (and other) samples analysed. The first exploitation for discriminating biological samples from their pyrolysis mass spectra was the successful demonstration (Goodacre *et al.*, 1992; Goodacre *et al.*, 1993) of the assessment of the presence of lower-grade seed oils as adulterants in extra virgin olive oils. This combination of PyMS and ANNs has now been employed to effect the rapid identification of strains of *Escherichia* (Sisson *et al.*, 1995), *Eubacterium* (Goodacre *et al.*, 1996a), *Mycobacterium* (Freeman *et al.*, 1994), *Propionibacterium* spp. (Goodacre *et al.*, 1994b), and *Streptomyces* (Chun *et al.*, 1993).

Recently, we have also investigated the ability of self-organising feature maps (Kohonen, 1989; Hecht-Nielsen, 1990; Hertz *et al.*, 1991) to carry out unsupervised learning and hence classify the pyrolysis mass spectra of canine *Propionibacterium acnes* isolates (Goodacre *et al.*, 1994b) and *P. acnes* isolated from man (Goodacre *et al.*, 1996b). Another neural network-based method for unsupervised feature extraction called auto-associative neural networks (Kramer, 1991; Kramer, 1992; Leonard & Kramer, 1993; Kuespert & McAvoy, 1994) has been used

to reduce the dimensionality of the infrared spectra of polysaccharides and hence extract features due to polysaccharides (Jacobsson, 1994), to detect plasmid instability using on-line measurements from an industrial fermentation producing a recombinant protein expressed by *Escherichia coli* (Montague & Morris, 1994), and to classify plant seeds from their pyrolysis mass spectra (Goodacre *et al.*, 1996d).

The aim of this study was therefore to use PyMS to determine whether one could type and discriminate animal cell cultures. We chose to examine three replicate cultures of four frozen animal cell lines and one fresh culture of Chinese hamster ovary cells. Whilst we recognise that the number of objects analysed in the present study was relatively small, they served more than adequately to illustrate the principles of our approach. Once data were collected auto-associative neural networks were employed to cluster the mass spectral data; the results obtained showed that PyMS could indeed be used to effect the rapid typing of cell cultures. The groups from the neural network-based methods were compared with the conventional multivariate statistical approaches of hierarchical cluster analysis and the clusters produced by both methods were very similar. Finally, the replicate cultures from the frozen mouse tumour BW-O cell line were found to give different pyrolysis mass spectra, rather than this being due to the effects of prolonged freezer storage it was found that this cell culture had become infected with bacteria and fungi.

Materials and methods

Animal cell lines and sample preparation

Details of the animal cell lines that were used in this study are shown in Table 1. Baby hamster kidney cells BHK21/C13 were originally obtained from Flow Laboratories (Scotland), and K562cl.6 human leukaemia cells from the European Collection of Animal Cell Cultures (ECACC, Salisbury). The BW-O tumour line is a non-invasive derivative of a mouse lymphosarcoma cell line (De Baetselier *et al.*, 1984b); and the 2C11–12 macrophage line is a mouse x mouse fusion product resulting from somatic hybridisation of a sarcoma cell line with isolated macrophages (De Baetselier *et al.*, 1984a). The murine cell lines were originally supplied by courtesy of Prof. Patrick de Baetselier, Brussels Free University, Belgium.

All cell cultures had been expanded in our laboratory prior to freezing and stored for up to 6 years under N_2. Triplicate samples of each of the four cell lines were taken from separate vials and a fresh sample of Chinese hamster ovary (CHO) cells was also included for comparison. Each cell line was harvested aseptical-

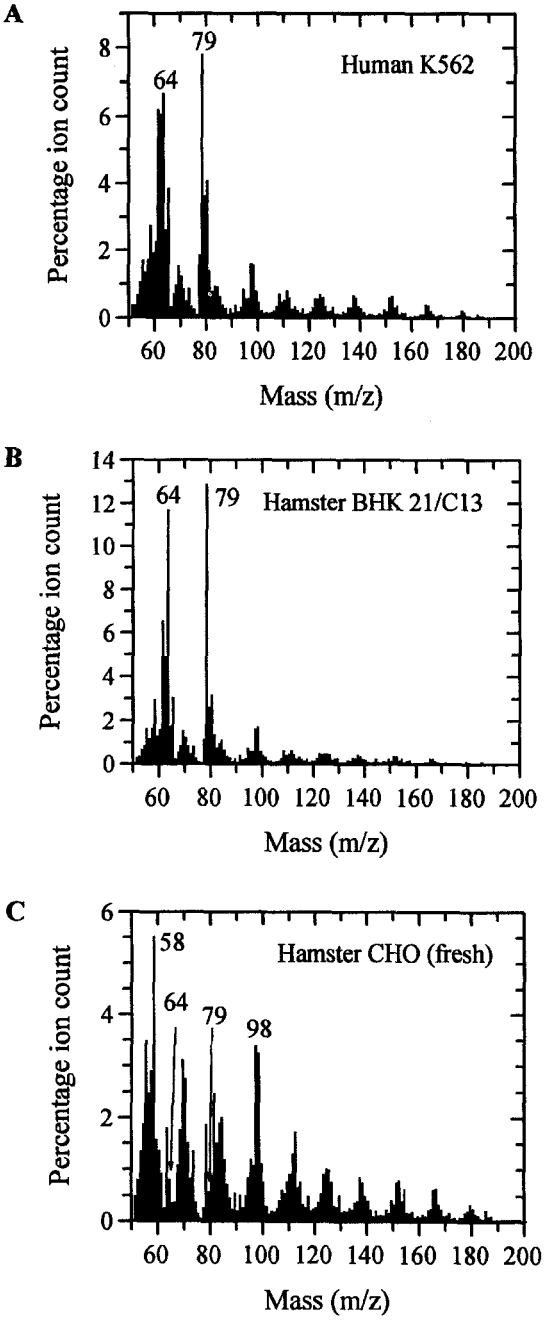

Figure 1. Normalised and averaged pyrolysis mass spectra of human leukaemia K562 (A), baby hamster kidney BHK 21/C13 (B), and Chinese hamster ovary CHO (C).

Table 1. Details of the animal cell lines used in this study

Type of cell line	Cell line identifier	Date frozen	Identifier used in multivariate analyses
Mouse macrophage	2C11–12	13 September, 1990	A1
Mouse macrophage	2C11–12	13 September, 1990	A2
Mouse macrophage	2C11–12	13 September, 1990	A3
Human leukaemia	K562	14 October, 1989	B1
Human leukaemia	K562	4 December, 1989	B2
Human leukaemia	K562	4 December, 1989	B3
Baby hamster kidney	BHK 21/C13	27 January, 1987	C1
Baby hamster kidney	BHK 21/C13	3 February, 1988	C2
Baby hamster kidney	BHK 21/C13	27 January, 1987	C3
Mouse tumour	BW-O	15 December, 1989	D1
Mouse tumour	BW-O	15 December, 1989	D2
Mouse tumour	BW-O	15 December, 1989	D3
Chinese hamster ovary	CHO	Fresh	E

ly by centrifugation and washed in phosphate buffered saline (PBS). All cell lines were then resuspended to approximately 1.5×10^6 cells per ml using PBS. The cells were then ready for analysis by PyMS.

Pyrolysis Mass Spectrometry (PyMS)

5 μl of the washed cells were evenly applied onto iron-nickel foils to give a thin uniform surface coating. Prior to pyrolysis the samples were oven-dried at 50 °C for 30 min. Each sample was analysed in triplicate. The pyrolysis mass spectrometer used was the Horizon Instruments PYMS–200X; for full operational procedures see Goodacre *et al.* (Goodacre, 1994; Goodacre *et al.*, 1994a; Goodacre *et al.*, 1994b; Goodacre *et al.*, 1995). The sample tube carrying the foil was heated, prior to pyrolysis, at 100 °C for 5 sec. Curie-point pyrolysis was at 530 °C for 3 sec, with a temperature rise time of 0.5 sec. The data from PyMS were collected over the m/z range 51 to 200 and may be displayed as quantitative pyrolysis mass spectra (e.g. as in Fig. 1). The abscissa represents the m/z ratio whilst the ordinate contains information on the ion count for any particular m/z value ranging from 51–200. Data were normalised as a percentage of total ion count to remove the influence of sample size *per se*.

Hierarchical cluster analysis

The normalised data were processed with the GENSTAT package (Nelder, 1979) running under Microsoft DOS 6.22 on an IBM-compatible PC. The method has been previously described (Gutteridge *et al.*, 1985; Goodacre, 1994; Goodacre *et al.*, 1996d). In essence, the first stage was the reduction of the data by principal components analysis (Jolliffe, 1986; Causton, 1987; Flury & Riedwyl, 1988; Martens & Næs, 1989; Everitt, 1993) keeping only those principal components whose eigenvalues accounted for more than 0.1% of the total variance. Canonical variates analysis then separated the samples into groups on the basis of the retained principal components and some *a priori* knowledge of the appropriate number of groupings (MacFie *et al.*, 1978; Windig *et al.*, 1983). The next stage was the construction of a percentage similarity matrix by transforming the Mahalanobis' distance between a priori groups in canonical variates analysis with the Gower similarity coefficient S_G (Gower, 1971). Finally, hierarchical cluster analysis was employed to produce a dendrogram, using average linkage clustering (Gutteridge *et al.*, 1985).

Auto-associative artificial neural networks

All artificial neural network (ANN) analyses were carried out under Microsoft Windows NT on an IBM-compatible PC. Data were normalised prior to analysis using the Microsoft Excel 4.0 spreadsheet. The back propagation neural network simulation program employed was WinNN version 0.96 (Dr Yaron Danon, 14 Beman Lane, Troy New York 12180, U.S.A. The program is available via ftp: ftp://sunsite.doc.ic.ac.uk/packages/windows3/programr/, the most recent file name to down load is winnn97.zip).

For in-depth descriptions of the analysis of PyMS data using back propagation ANNs the reader is referred to Goodacre *et al.* (Goodacre *et al.*, 1994b; Goodacre *et al.*, 1995; Goodacre *et al.*, 1996c; Goodacre *et al.*, 1996d).

The structure of the ANN used in this study to analyse pyrolysis mass spectra consisted of 5 layers containing processing nodes (neurons or units) made up of the 150 input nodes (normalised averaged pyrolysis mass spectra), 150 output nodes (normalised average pyrolysis mass spectra), and three "hidden" layers containing 8, 3 and 8 nodes respectively; this may be represented as a 150–8–3–8–150 architecture (Fig. 2). This ANN can be referred to as a fully interconnected feedforward multilayer perceptron where each of the layers of nodes was connected to the next (hidden) layer using abstract interconnections (connections or synapses). Connections each have an associated real value, termed the weight, that scale signals passing through them. Nodes in the hidden layers and output layer sum the signals feeding to them and output this sum to each driven connection scaled by a "squashing" function (f) with a sigmoidal shape. Typically the function $f = 1/(1 + e^{-x})$, where $x = \Sigma$inputs.

These signals are then passed to the next layer which sums them and then they are in turn squashed by the sigmoidal activation function (Fig. 2); the product of the final layer of nodes was then fed to the "outside world".

Before training commenced the values applied to the input and output nodes were normalised across the whole mass range such that the lowest ion count was set to 0 and the highest to 1. Finally, the connection weights were set to small random values (typically between –0.0001 and +0.0001).

The algorithm used to train the neural network was the standard back-propagation (BP) (Rumelhart *et al.*, 1986; Werbos, 1994; Chauvin & Rumelhart, 1995). For the training of the ANN each input (i.e. normalised averaged pyrolysis mass spectrum) is paired with a desired output (i.e., the same pyrolysis mass spectrum); together these are called a training pair (or training pattern). An ANN is trained over a number of training pairs; this group is collectively called the training set. The input is applied to the network, which is allowed to run until an output is produced at each output node. The differences between the actual and the desired output, taken over the entire training set are fed back through the network in the reverse direction to signal flow (hence back-propagation) modifying the weights as they go. This process is repeated until a suitable level of error is achieved. In the present work, we used a learning rate of 0.05 and a momentum of 0.9.

Each epoch represents the connection weight updatings and a recalculation of the root mean squared (RMS) error between the true and desired outputs (mass spectra) over the entire training set. During training a plot of the error versus the number of epochs represents the "learning curve", and may be used to estimate the extent of training. Training may be said to have finished when the network has found the lowest error. Provided the network has not become stuck in a local minimum, this point is referred to as the global minimum on the error surface. In the present experiment we trained the auto-associative neural networks until the RMS error was 0.005; this was chosen because at this low RMS error point the output layer would be very similar to the mass spectra applied to the neural networks' input nodes.

After training, each of the pyrolysis mass spectra were applied in turn to the input layer and the overall activation on the three nodes in the 'bottle-neck' layer calculated. The compression of the 150 inputs through only three nodes in the middle layer allows non-linear principal components analysis to be performed; plots of the activations of the nodes in the 'bottle-neck' layer therefore allow 'clusters' to be found in the data. For a more detailed account of this data compression through the 'bottle-neck' layer please refer to Kramer (1991; 1992).

Results and discussion

After collection of the pyrolysis mass spectra the first stage was to normalise them to total ion count, to average the three replicate spectra, and then to inspect them visually; Fig. 1 displays the normalised and averaged pyrolysis mass spectra of human leukaemia K562 (Fig. 1a), baby hamster kidney BHK 21/C13 (Fig. 1b), and Chinese hamster ovary CHO (Fig. 1c). It was seen that the spectra from the frozen and fresh cell lines were very different; the frozen samples all have intense peaks at 64 and 79 m/z which contribute between 15 to 25% of the total ion count, whilst these peaks are very small (<2% of the total ion count) in the fresh sample. It is possible that this observation is due to the freezing process causing a permeabilty change in the cell walls and thus leakage of cellular components when washed. If true then provided the same number of cells were analysed the total ion count might be expected to be reduced in the frozen samples, this was

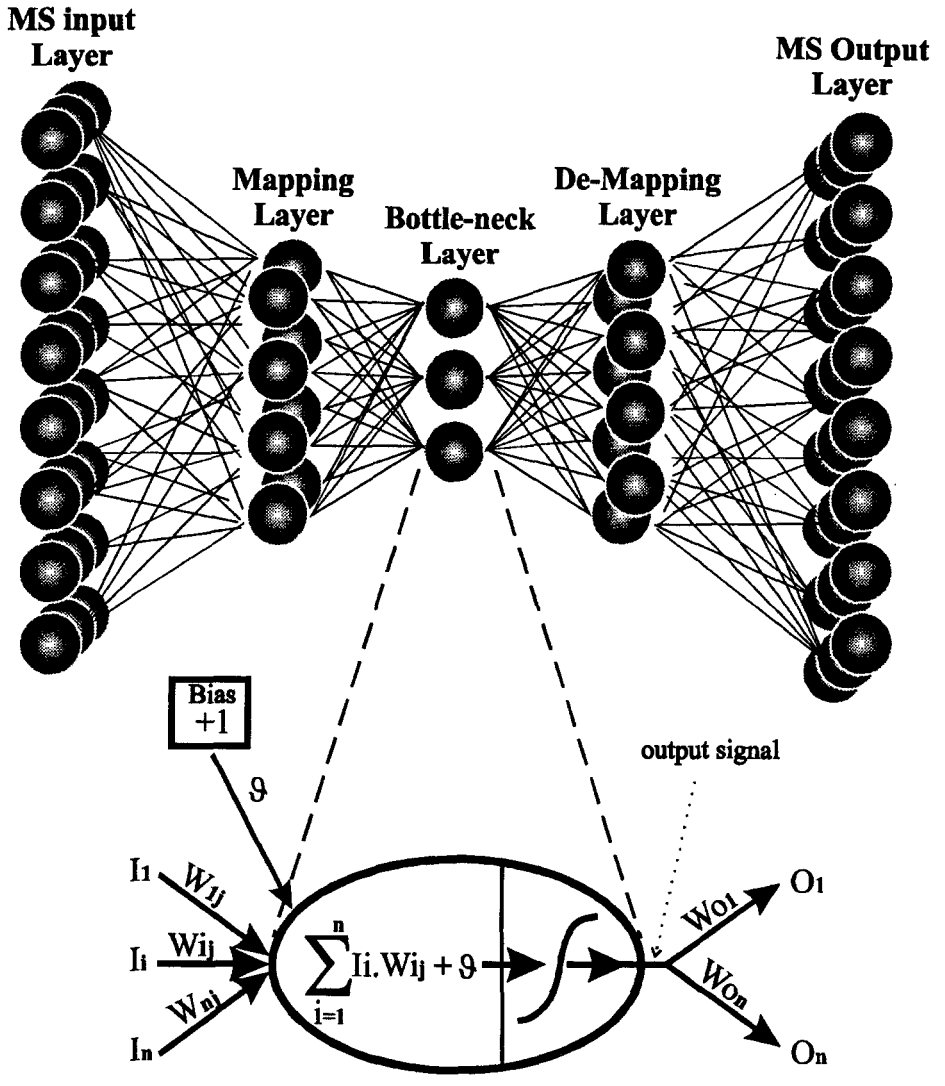

Figure 2. Architecture of an auto-associative neural network consisting of 5 layers. In the architecture shown, adjacent layers of the network are fully interconnected. The input and output layer are presented with identical PyMS data (in this figure there are 24 nodes in these layers; in the present work the number of nodes was actually 150 inputs/masses). A key feature of the auto-associative network is the data compression in the middle (third) bottle-neck layer of 3 nodes. The second and fourth layers each consisted of 8 nodes and these map and de-map the mass spectra allowing feature extraction in the bottle neck layer; this is equivalent to non-linear principal components analysis. Also displayed is the information processing by a node in one of the hidden layers or output layer. An individual node sums its input (the Σ function) from nodes in the previous layer, including the bias (ϑ), transforms them via a "sigmoidal" squashing function, and outputs them to the next node to which it is linked via a connection weight. The bias has an activation which was always set to +1 and is applied to the 3 hidden layers and the output layer.

not found to be the case. Indeed the method of freezing was to reduce the temperature very slowly so as to avoid the formation of ice crystals which are normally thought to be to blame for the lysis occuring in frozen cell samples; moreover, the typical viability of animal cell lines frozen in our laboratory is >90%. A more detailed analysis of the intense peaks found at 64 and 79 m/z in the pyrolysates of the frozen cell lines might have been effected using either pyrolysis tandem mass spectrometry or pyrolysis gas chromatography mass spectrometry; however, these facilities were not available to us.

The next stage was to perform unsupervised learning using auto-associative ANNs to effect the classification of the 13 cell lines. The architecture of the multi-layer perceptrons employed was 150–8–3–8–150 (illustrated in Fig. 2), the training pairs consisted of the *same* normalised pyrolysis mass spectrum, and the 13 averaged pyrolysis mass spectra were then applied in turn to the 150 input and 150 output nodes. These ANNs were trained as described above until the RMS error was 0.005, this took approximately 1×10^4 epochs.

After training to this point each of the averaged pyrolysis mass spectra were applied to the input layer of the auto-associative ANN and the activation on the three nodes in the 'bottle-neck' layer calculated. All the frozen cell lines gave activations of between 0 and 0.000001 on all three nodes and the fresh cell line Chinese hamster ovary gave activations of 0.978947, 0.978739, 0.97874 at each of the three nodes. It was therefore obvious that the major feature extracted reflected the freezing process rather than the different types of cell; since visual inspection alone could differentiate between frozen and fresh cell lines this result was not surprising. Therefore other auto-associative ANNs were used to perform unsupervised learning on the pyrolysis mass spectra from the 12 frozen cell lines.

After training this ANNs until the RMS error was 0.005, which took approximately 5×10^5 epochs, each of the 12 averaged pyrolysis mass spectra were applied to the input layer of the auto-associative ANN and the activation on the three nodes in the 'bottle-neck' layer calculated and plotted as a pseudo-three dimensional graph (Fig. 3). In Fig. 3 it can be seen that for the cell lines mouse macrophage 2C11–12, human leukaemia K562, and baby hamster kidney BHK 21/C13 each of the three replicate samples from different vials group together and that the three different cell lines cluster separately. However, the replicate samples from the mouse tumour cell line BW-O, although frozen at the same time (Table 1), were recovered separately.

The samples from human leukaemia K562, baby hamster kidney BHK 21/C13, and mouse tumour BW-O had all been stored for over four years and had been frozen between January 1987 and December 1989 (Table 1). That the three replicate samples from the human leukaemia K562 and the triplicate samples from the baby hamster kidney BHK 21/C13 cell line cluster together shows that keeping the cell lines frozen does not significantly alter their pyrolysis mass spectra, therefore PyMS could be employed to characterize them. However, the three mass spectra from the mouse tumour BW-O cell line were different as judged by auto-associative ANNs (Fig. 3). Since neural computation methods are relatively new to PyMS the question therefore arises as to whether these three mass spectra really are different, either due to the freezing process or some other phenomenon, or whether it is a failing in the feature extraction technique.

The 12 spectra were therefore analysed using the standard statistical procedure of hierarchical cluster analysis, as detailed above. The resulting dendrogram (Fig. 4) indeed shows congruence with the NLPCA plot; the replicates from the mouse macrophage 2C11–12, human leukaemia K562, and baby hamster kidney BHK 21/C13 cell lines group together and away from one another, and the replicate samples from the mouse tumour BW-O cell line are recovered separately. In particular, replicates D2 and D3 are 80% similar and only 62% similar with D1. This result shows conclusively that auto-associative ANNs give very similar results to the classical approaches used to classify samples on their pyrolysis mass spectra, moreover it shows that the three replicate samples from the mouse tumour cell line are definitely different from each other. Since the three vials from this cell line were all frozen at the same time (15 December, 1989) the question therefore arose as to whether the cells had become contaminated.

50 μl samples from each of the 12 vials were lawned aseptically on to nutrient agar plates and incubated at 28 °C for 48 hrs. No bacteria or fungi grew on the samples taken from mouse macrophage 2C11–12, human leukaemia K562, and baby hamster kidney BHK 21/C13 cell lines. However a mixture of unidentified bacteria and fungi (approximately 1×10^4 colony forming units ml^{-1}) were cultivated on the agar plates from the mouse tumour BW-O cell line, and replicate D1 was particularly contaminated with fungi. It was therefore evident that the differences observed in the pyrolysis mass spectra of these three replicate cultures of the same cell line was due to contamination with mixed microbial populations rather than an irreproducible change in the biochemical profile during storage under N_2. That PyMS was able to detect this very low level of contamination, 1×10^4 ml^{-1}, in a relative large background population of 1.5×10^6 cells ml^{-1} is encouraging and highlights the exquisite sensitivity of this technique. Because it is effectively studying the properties of a system in 150 dimensions (here the m/z values from 51–200) simultaneously, PyMS is a very high-resolution technique and it has been shown that one can even use PyMS to detect the prescence or absence of antibiotic plasmids in *Escherichia coli*

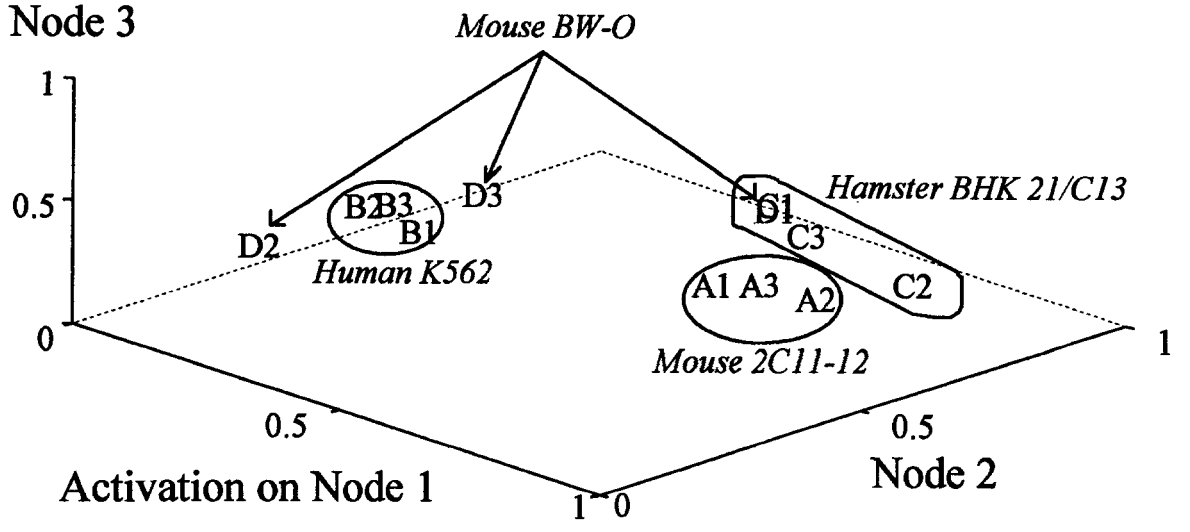

Figure 3. Pseudo three-dimensional non-linear principal components plot based on PyMS data analysed by the 150–8–3–8–150 auto-associative neural network showing the relationship between the four cell lines. The activations of the three nodes in the bottle-neck layer are shown. ANNs were trained using the standard-back propagation algorithm, to a RMS error of 0.005 which typically took 5×10^5 epochs.

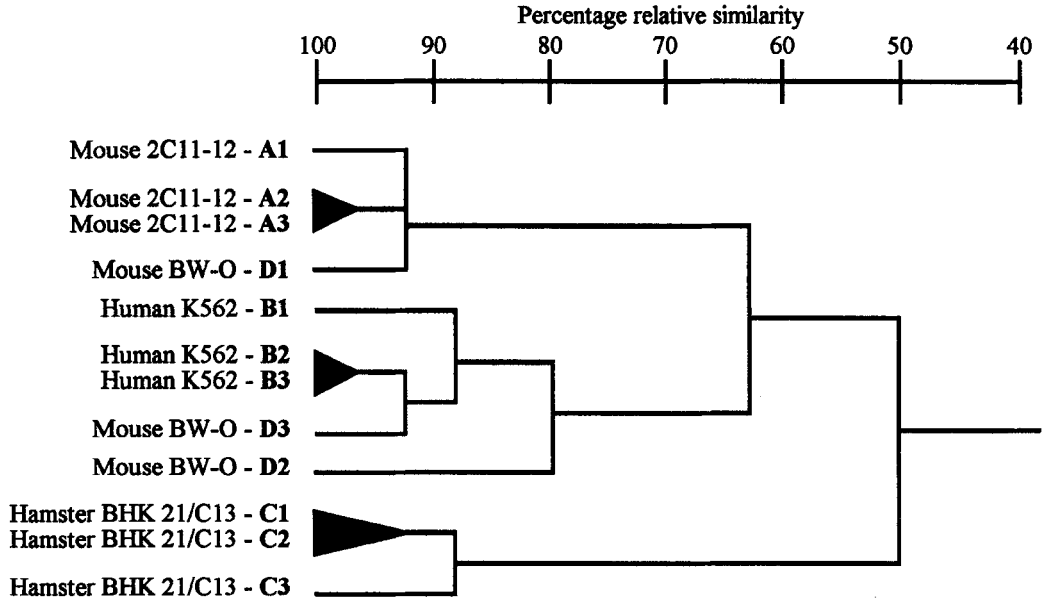

Figure 4. Dendrogram representing the relationships between the four frozen animal cell lines based on PyMS data analysed by GENSTAT.

(Goodacre & Berkeley, 1990). It is therefore likely that the sensitivity of this technique is such that one could exploit it for the detection of mycoplasma infections in animal cell cultures.

Concluding remarks

These results show clearly that PyMS and auto-associative neural networks, which carry out unsupervised learning, can be employed to discriminate between animal cell lines, and that very similar classification was observed when the same spectral data

were analysed using established multivariate statistical procedures, *viz.* hierarchical cluster analysis.

We also demonstrated that this approach can detect the contamination of cell lines with bacteria and fungi, and believe that this approach could be extended for the rapid detection of mycoplasma (and indeed other microbial) infection in animal cell lines. Moreover, we have previously shown that PyMS and ANNs trained using supervised learning can be used to measure the concentrations of tertiary mixtures of cells of the bacteria *Bacillus subtilis*, *Escherichia coli* and *Staphylococcus aureus* (Goodacre *et al.*, 1994a); such an approach could also be exploited to quantify the level of contamination in infected animal cell lines.

The major problem with PyMS is that long-term reproducibility (>30 days) is poor and the mass spectral fingerprints of the same material analysed at two different times may be different; this lack of reproducibility is largely due to 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 micro-organisms are analysed in a single batch (Magee, 1993; Goodfellow, 1995). For PyMS to be used (a) for the routine identification of cell lines, new spectra must be able to be compared to those previously collected. We have recently found that neural networks can be used successfully 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 (Goodacre and Kell, 1995; 1996b). Calibration samples were run at the two times, up to 2 years apart, and ANNs set up in which the inputs were the 150 'new' calibration masses and the outputs were the 150 calibration masses from the 'old' spectra. These neural networks could thus be used as signal-processing elements to effect the transformation of data acquired one day to those which would have been acquired on a later date. Therefore for the first time PyMS can be used to acquire spectra which could be compared to those previously collected and held in a library.

The major advantages that PyMS offers over more conventional methods used to type cell lines and to screen for microbial infection, such as DNA fingerprinting, are its speed, sensitivity and the ability to analyse many hundreds of samples per day. We conclude that the combination of PyMS and neural networks can provide a rapid and accurate discriminatory technique for the authentication of animal cell lines.

Acknowledgments

R.G. is funded as a research fellow by the Wellcome Trust grant number 042615/Z/94/Z. D.B.K. thanks the Chemicals & Pharmaceuticals Directorate of the UK BBSRC for financial support.

References

Beale R & Jackson T (1990) Neural Computing: An Introduction. Adam Hilger, Bristol.

Causton DR (1987) A Biologist's Advanced Mathematics. Allen and Unwin, London.

Chauvin Y & Rumelhart DE (ed.) (1995) Backpropagation: Theory, Architectures, and Applications. Erlbaum, Hove, UK.

Chun J, Atalan E, Kim SB, Kim HJ, Hamid ME, Trujillo ME, Magee JG, Manfio GP, Ward AC & Goodfellow M (1993) Rapid identification of *Streptomycetes* by artificial neural network analysis of pyrolysis mass spectra, FEMS Microbiol Lett 114: 115–119.

De Baetselier P, Brys L, Vercauteren E, Mussche L, Hamers R & Schram E (1984a) Generation of macrophage-hybridomas for the study of macrophage chemiluminescence In: Analytical Application of Bioluminescence and Chemiluminescence. Academic Press, New York.

De Baetselier P, Roos E, Brys L, Remels L & Feldman M (1984b) Generation of invasive and metastatic variants of a non-metastatic T-cell lymphoma by *in vivo* fusion with normal host cells, Int J Cancer 34: 731–738.

Dussurget O, Henry A, Lemercier B & Roullanddussoix D (1994) Polymerase chain reaction based diagnosis of mollicute infection of commercial animal sera, J Microbiol Meth 20: 125–135.

Everitt BS (1993) Cluster Analysis. Edward Arnold, London.

Flores EF & Donis RO (1995) Isolation of a mutant MDBK cell line resistant to bovine viral diarrhea virus infection due to a block in viral entry, Virology 208: 565–575.

Flury B & Riedwyl H (1988) Multivariate Statistics: A Practical Approach. Chapman and Hall, London.

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

Goodacre R (1994) Characterisation and quantification of microbial systems using pyrolysis mass spectrometry: Introducing neural networks to analytical pyrolysis, Microbiol Eur 2: 16–22.

Goodacre R & Berkeley RCW (1990) Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry, FEMS Microbiol Lett 71: 133–137.

Goodacre R, Hiom SJ, Cheeseman SL, Murdoch D, Weightman AJ & Wade WG (1996a) Identification and discrimination of oral asaccharolytic *Eubacterium* spp. using pyrolysis mass spectrometry and artificial neural networks, Cur Microbiol 32: 77–84.

Goodacre R, Howell SA, Noble WC & Neal MJ (1996b) Subspecies discrimination using pyrolysis mass spectrometry and self-organising neural networks of *Propionibacterium acnes* isolated from normal human skin, Zbl Bakt: in the press.

Goodacre R & Kell DB (1995) Composition analysis, UK Patent #9511619.0 of June 8th, 1995.

Goodacre R & Kell DB (1996a) Pyrolysis mass spectrometry and its applications in biotechnology, Cur Opin Biotechnol 7: 20–28.

Goodacre R & Kell DB (1996b) Correction of mass spectral drift using artificial neural networks, Anal Chem 68: 271–280.

Goodacre R, Kell DB & Bianchi G (1992) Neural networks and olive oil, Nature 359: 594–594.

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

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

Goodacre R, Neal MJ & Kell DB (1996c) Quantitative analysis of multivariate data using artificial neural networks: a tutorial review and applications to the deconvolution of pyrolysis mass spectra, Zbl Bakt: in the press.

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

Goodacre R, Pygall J & Kell DB (1996d) Plant seed classification using pyrolysis mass spectrometry with unsupervised learning; the application of auto-associative and Kohonen artificial neural networks, Chemometr Intell Lab Sys 34: 69–83.

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

Goodfellow M (1995) Inter-strain comparison of pathogenic microorganisms by pyrolysis mass spectrometry, Binary – Comp Microbiol 7: 54–60.

Gower JC (1971) A general coefficient of similarity and some of its properties, Biomet 27: 857–874.

Gutteridge CS, Vallis L & MacFie HJH (1985) Numerical methods in the classification of microorganisms by pyrolysis mass spectrometry. In: Goodfellow M, Jones D & Priest F (ed.) Computer-assisted Bacterial Systematics. (pp. 369–401) Academic Press, London.

Hay RJ (1988) The seed stock concept and quality control for cell lines, Analytical Biochemistry 171: 225–237.

Haykin S (1994) Neural Networks. Macmillan, New York.

Hecht-Nielsen R (1990) Neurocomputing. Addison-Wesley, Massachusetts.

Hertz J, Krogh A & Palmer RG (1991) Introduction to the Theory of Neural Computation. Addison-Wesley, California.

Hopert A, Uphoff CC, Wirth M, Hauser H & Drexler HG (1993a) Mycoplasma detection by PCR analysis, In Vitro Cellular & Developmental Biology – Animal 29A: 819–821.

Hopert A, Uphoff CC, Wirth M, Hauser H & Drexler HG (1993b) Specificity and sensitivity of polymerase chain reaction (PCR) in comparison with other methods for the detection of mycoplasma contamination in cell lines, J Immunol Meth 164: 91–100.

Irwin WJ (1982) Analytical Pyrolysis: A Comprehensive Guide. Marcel Dekker, New York.

Jacobsson SP (1994) Feature extraction of polysaccharides by low-dimensional internal representation neural networks and infrared spectroscopy, Anal Chim Acta 291: 19–27.

Jolliffe IT (1986) Principal Component Analysis. Springer-Verlag, New York.

Kohonen T (1989) Self-Organization and Associative Memory. Springer-Verlag, Berlin.

Kramer MA (1991) Non linear principal components analysis using auto-associative neural networks, AIChE J 37: 233–243.

Kramer MA (1992) Autoassociative neural networks, Comput Chem Eng 16: 313–328.

Kuespert DR & McAvoy TJ (1994) Knowledge extraction in chemical process control, Chem Eng Comm 130: 251–264.

Leonard JA & Kramer MA (1993) Diagnosing dynamic faults using modular neural nets, IEEE Expert 8: 44–53.

MacFie HJH, Gutteridge CS & Norris JR (1978) Use of canonical variates in differentiation of bacteria by pyrolysis gas-liquid chromatography, J Gen Microbiol 104: 67–74.

Magee JT (1993) Whole-organism fingerprinting. In: Goodfellow M & O'Donnell AG (ed.) Handbook of New Bacterial Systematics. (pp. 383–427) Academic Press, London.

Martens H & Næs T (1989) Multivariate Calibration. John Wiley and Sons, New York.

Meuzelaar HLC, Haverkamp J & Hileman FD (1982) Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials. Elsevier, Amsterdam.

Montague G & Morris J (1994) Neural network contributions in biotechnology, Trends Biotechnol 12: 312–324.

Mowles JM & Doyle A (1990) Cell culture standards – time for a rethink, Cytotechnology 3: 107–108.

Nelder JA (1979) Genstat Reference Manual. Scientific and Social Service Program Library, University of Edinburgh.

Nicklas W, Kraft V & Meyer B (1993) Contamination of transplantable tumors, cell lines, and monoclonal antibodies with rodent viruses, Lab Animal Sci 43: 296–300.

Park JG, Lee JH, Kang MS, Park KJ, Jeon YM, Lee HJ, Kwon HS, Park HS, Yeo KS, Lee KU, Kim ST, Chung JK, Hwang YJ, Lee HS, Kim CY, Lee YI, Chen TR, Hay RJ, Song SY, Kim WH, Kim CW & Kim YI (1995) Characterization of cell lines established from human hepatocellular carcinoma, Int J Cancer 62: 276–282.

Ripley BD (1994) Neural networks and related methods for classification, J Roy Stats Soc Series B-Methodological 56: 409–437.

Rumelhart DE, McClelland JL & The PDP Research Group (1986) Parallel Distributed Processing, Experiments in the Microstructure of Cognition. MIT Press, Cambridge, Mass.

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

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

Snyder AP, Smith PBW, Dworzanski JP & Meuzelaar HLC (1994) Pyrolysis-gas chromatography-mass spectrometry – detection of biological warfare agents, ACS Symp Ser 541: 62–84.

Stacey GN, Bolton BJ & Doyle A (1992a) DNA fingerprinting transforms the art of cell authentication, Nature 357: 261–262.

Stacey GN, Bolton BJ, Morgan D, Clark SA & Doyle A (1992b) Multilocus DNA fingerprint analysis of cellbanks – stability studies and culture identification in human B-lymphoblastoid and mammalian cell lines, Cytotechnology 8: 13–20.

Steube KG, Grunicke D & Drexler HG (1995) Isoenzyme analysis as a rapid method for the examination of the species identity of cell cultures, In Vitro Cellular & Developmental Biology – Animal 31: 115–119.

Uphoff CC, Gignac SM & Drexler HG (1992a) Mycoplasma contamination in human leukemia cell lines. 1. Comparison of various detection methods, J Immunol Meth 149: 43–53.

Uphoff CC, Gignac SM & Drexler HG (1992b) Mycoplasma contamination in human leukemia cell lines. 2. Elimination with various antibiotics, J Immunol Meth 149: 55–62.

Wasserman PD (1989) Neural Computing: Theory and Practice. Van Nostrand Reinhold, New York.

Werbos PJ (1994) The Roots of Back-Propagation: From Ordered Derivatives to Neural Networks and Political Forecasting. John Wiley and Sons, Chichester.

Werner RG & Noe W (1993) Mammalian cell cultures. 1. Characterization, morphology and metabolism, Arzneimittel-Forschung/Drug Research 43: 1134–1139.

Werner RG, Walz F, Noe W & Konrad A (1992) Safety and economic aspects of continuous mammalian cell culture, J Biotechnol 22: 51–68.

Windig W, Haverkamp J & Kistemaker PG (1983) Interpretation of sets of pyrolysis mass spectra by discriminant analysis and graphical rotation, Anal Chem 55: 81–88.

Zupan J & Gasteiger J (1993) Neural Networks for Chemists: An Introduction. VCH Verlagsgesellschaft, Weinheim.

Address for correspondence: Royston Goodacre, Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed, SY23 3DA, Wales, U.K.