ELSEVIER

Analytica Chimica Acta 313 (1995) 25–43

ANALYTICA CHIMICA ACTA

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 ☆

Royston Goodacre [a,*], Sally Trew [b], Carys Wrigley-Jones [b,1], Gunter Saunders [c], Mark J. Neal [a], Neil Porter [b], Douglas B. Kell [a]

[a] *Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, UK*
[b] *Xenova Ltd., 545 Ipswich Road, Slough, Berkshire SL1 4EQ, UK*
[c] *Department of Biological and Health Sciences, University of Westminster, 115 New Cavendish Street, London W1M 8JS, UK*

Received 6 December 1994; revised 14 March 1995; accepted 15 March 1995

Abstract

The combination of pyrolysis mass spectrometry (PyMS) and artificial neural networks (ANNs) can be used to quantify levels of penicillins in strains of *Penicillium chrysogenum* and ampicillin in spiked samples of *Escherichia coli*. Four *P. chrysogenum* strains (NRRL 1951, Wis Q176, P1, and P2) were grown in submerged culture to produce penicillins, and fermentation samples were taken aseptically and subjected to PyMS. To deconvolute the pyrolysis mass spectra so as to obtain quantitative information on the titre of penicillins, fully-interconnected feedforward artificial neural networks (ANNs) were studied; the weights were modified using the standard back-propagation algorithm, and the nodes used a sigmoidal squashing function. In addition the multivariate linear regression techniques of partial least squares regression (PLS), principal components regression (PCR) and multiple linear regression (MLR) were applied. The ANNs could be trained to give excellent estimates for the penicillin titre, not only from the spectra that had been used to train the ANN but more importantly from previously unseen pyrolysis mass spectra. All the linear regression methods failed to give accurate predictions, because of the very variable biological backgrounds (the four different strains) in which penicillin was produced and also of the inability of models using linear regression accurately to map non-linearities. Comparisons of squashing functions on the output nodes of identical 150-8-1 neural networks revealed that networks employing linear functions gave more accurate estimates of ampicillin in *E. coli* near the edges of the concentration range than did those using sigmoidal functions. It was also shown that these neural networks could be successfully used to extrapolate beyond the concentration range on which they had been trained. PyMS with the multivariate clustering technique of principal components analysis was able to differentiate between four strains of *P. chrysogenum* studied, and was also able to detect phenotypic differences at

☆ Invited paper on Analytical Biotechnology.

* Corresponding author.

[1] Present address: Department of Biological and Health Sciences, University of Westminster, 115 New Cavendish Street, London W1M 8JS, UK.

SSDI 0003-2670(95)00170-0

five, seven, nine or 11 days growth. A crude sampling procedure consisting of homogenised agar plugs proved applicable for rapid analysis of a large number of samples.

Keywords: Mass spectrometry; Artificial neural networks; Fermentor broths; Regression analysis; Chemometrics; Biotechnology; Pyrolysis mass spectrometry

1. Introduction

Within medicine and biotechnology there is a continuing need to find new pharmaceuticals, and hence for the development of rapid and efficient methods for the screening of large numbers of microbial cultures for the production of biologically active metabolites (e.g., see [1–9]). Such metabolites can provide new structural templates for drug design and development through chemical synthesis. Screening for such metabolites usually involves the modulation of a particular biochemical step implicated in a particular disease process. Very often metabolites showing activity during screening are produced only in very small amounts by the organism, and therefore increasing the titre of the metabolite is essential in order to provide sufficient material for chemical characterization and further biological evaluation. Metabolite titres can be significantly improved through the generation and isolation of overproducing mutants derived from the original wild-type organism [5,10–24]. Over-producing mutants arise typically at frequencies around 10^{-4} [25], and therefore tens of thousands of cultures need to be screened in the search for an improved, overproducing strain.

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. The method typically involves the thermal degradation of complex material in a vacuum by Curie-point pyrolysis; this causes molecules to cleave at their weakest points to produce smaller, volatile fragments called pyrolysate [26]. A mass spectrometer can then be used to separate the components of the pyrolysate on the basis of their mass-to-charge ratio (m/z) to produce a pyrolysis mass spectrum, which can then be used as a chemical signature (fingerprint) of the complex material analysed.

PyMS has been applied to the characterisation and identification of a variety of microbial and biotechnological systems over a number of years [26–34] and, because of its high discriminatory ability [35], represents a powerful fingerprinting technique, which is applicable to any organic material.

Our own aims have been to extend the PyMS technique for the *quantitative* analysis of the chemical constituents of microbial and other samples [36]. To this end, we have sought to apply fully-interconnected feedforward artificial neural networks (ANNs) (see [37–49] for introductory surveys), and the multivariate linear regression techniques of partial least squares regression (PLS) and principal components regression (PCR) (see [50–60] for first-rate texts) to the deconvolution and interpretation of pyrolysis mass spectra. Thus, we have been able to follow the production of indole in a number of strains of *Escherichia coli* grown on media incorporating various amounts of tryptophan [61], to estimate the amount of casamino acids in mixtures with glycogen [62,63], to deconvolute the pyrolysis mass spectra of complex biochemical and microbiological mixtures [64], and for the analysis of recombinant cytochrome b_5 expression in *E. coli* [65].

With regard to discrimination of materials from their pyrolysis mass spectra we have also exploited ANNs for the rapid and accurate assessment of the presence of lower-grade seed oils as adulterants in extra virgin olive oils [66,67], and for the identification of strains of *Propionibacterium* spp. [68,69]. Chun et al. [70] and Freeman et al. [71] have also used the combination of PyMS and ANNs for the differentiation of strains of *Streptomyces* and mycobacteria respectively.

Industry exploits the biosynthetic capabilities of microorganisms to produce pharmaceuticals and other products through fermentation. It is imperative therefore that the concentration of the product (the determinand) is assessed accurately so as to optimise control of the fermentation process. Whilst on-line tandem mass spectrometry (MS/MS) has been used

to analyse fermentation broth extracts for flavones [72], the majority of mass spectrometry (MS) applications during fermentations have been for the analysis of gases and volatiles produced over the reactor [73,74], or by employing a membrane inlet probe for volatile compounds dissolved in the broths [75–80]. Although, Hansen et al. [81], using membrane-inlet MS, have shown that compounds of *relatively* low volatility may be monitored in fermentors, it is obvious, that more worthwhile information could be gained by measuring the non-volatile components of fermentation broths. Indeed, Heinzle et al. [82] were able to characterise the states of fermentations using off-line PyMS, and this technique was extended to on-line analysis [83]. These authors, however, were not very satisfied with their system and although they have continued to use mass spectrometry for the analysis of volatiles produced during fermentation [74,84] the analysis of non-volatiles by PyMS seems to have been less than fully exploited.

More recently, therefore, we have also applied ANNs to the quantitative analysis of the pyrolysis mass spectra of fermentor broths, thus effecting a rapid screening for the high-level production of desired substances [85]. A particularly noteworthy feature of that study [85] was that ANNs trained to predict the amount of ampicillin in *E. coli* were able to generalise so as to predict the concentration of ampicillin in a *Staphylococcus aureus* background, illustrating the great robustness of ANNs to rather substantial variations in the biological background. In other words, whilst the pyrolysis mass spectra measure the presence of *all* molecules simultaneously they contain sufficient spectral information from the target molecules of interest to allow their quantification.

The major aim of the present study was to demonstrate that PyMS, with multivariate calibration and ANNs, can be used to effect the rapid prediction of the amount of penicillins in fermentor broths containing *Penicillium chrysogenum*. The β-lactam antibiotic penicillin was chosen because it is a well-characterised, industrially important system [4,5,86–91], and there is access to over-producing mutants. It is worth pointing out that the penicillin production of Fleming's *Penicillium notatum* isolate was about 2 International Units ml^{-1}, whilst today's processes yield a penicillin titre of at least 85,000 units ml^{-1} [5]. This represents a substantial increase from 1.2 mg l^{-1} to some 50 g l^{-1} and well illustrates the value and power of strain selection (although the efficiency of glucose conversion to penicillin is still less than 10% [91]).

The first noteworthy improvement in yield came in 1943 with the isolation of *P. chrysogenum* strain NRRL 1951; this organism was chosen because it was better suited to submerged production than Fleming's isolate. Strain NRRL 1951 was subsequently subjected to mutagenesis using X-ray and ultraviolet radiation treatment to produce Wis Q176, the original strain in the famous Wisconsin culture line [92]. Strain Q176 has been used as the starting strain for many improvement programs, one of which by Panlabs Inc. produced a number of high yielding derivatives including P1 and P2 [93]. The wild type *P. chrysogenum* NRRL 1951, and its higher yielding derivatives Q176, P1 and P2 were used in this study; in terms of their genealogy P1 and P2 are distantly related to NRRL 1951 and Q176 which are evolutionary closer.

When screening mutants for the overproduction of metabolites of interest there is a need to go beyond the realm of the "knowledge base", that is, there is a desire to extrapolate beyond the concentration range of the metabolite already observed [85]. We have previously reported [62] that ANNs should not expected to give wholly correct estimates near the edges of or outside their training sets because they cannot be expected to extend beyond the range of the nonlinear squashing function on the output node(s) (see later), a phenomenon also reported by other workers [94–97].

One way around this is simply to spike samples with the determinand, so that the knowledge base is effectively extended to encompass the higher concentrations desired [85]. (However, spiking with product does not take into account underlying biochemical changes which might be needed to give increased titre.) An alternative and arguably more attractive approach (in that it requires no a priori knowledge of the range attainable), is to alter the existing models so that they can then be used to extrapolate. When the individuals in a population of samples are (more-or-less) linearly separable methods such as multiple linear regression (MLR), PLS and PCR should be able to extrapolate. Therefore we

used these methods on the pyrolysis mass spectra of binary mixtures of ampicillin (0–5000 μg ml^{-1} in steps of 250 μg ml^{-1}) with *Escherchia coli* (as a very simple model of a fermentor previously studied [85]) to show here that these multivariate linear regression models could indeed extrapolate beyond the knowledge base.

In a previous study we showed that ANNs could not extrapolate [62] when we were employing a sigmoidal squashing function on the output node and we indicated in that report that future work would involve the assessment of the use of a linear activation function to ascertain if it may improve the ability of ANNs to generalise better from pyrolysis mass spectral data. We also therefore compared sigmoidal versus linear squashing functions on the output nodes of identical neural networks and found that (1) more accurate estimates near the edges of the concentration range could be gained using linear scalars on the output nodes and (2) that neural networks employing linear squashing function could be used to extrapolate successfully in producing predictions of concentration significantly greater than those used in the training set.

2. Experimental

2.1. Preparation of Penicillium chrysogenum fermentation broths and agar plugs

The strains of *Penicillium chrysogenum* used were the wild type NRRL 1951 and three higher penicillin producing strains Wis Q176, P1 and P2 and were obtained from Dr. Saunders' laboratory at the University of Westminster. Details of how all four strains are related can be found in Hersbach et al. [98]. All strains were maintained on slants of GM agar (containing: glycerol, 7.5 g; molasses, 2.5 g; yeast extract, 1 g; $MgSO_4$, 0.05 g; KH_2PO_4, 0.06 g, Bacto peptone, 5 g; NaCl, 10 g, $FeSO_4 \cdot 7H_2O$, 3 mg; $CuSO_4 \cdot 5H_2O$, 1 mg; per litre water). Slant cultures were incubated at 25°C until well sporulated.

Spores were harvested and suspended in sterile 20% glycerol +0.1% Tween 80 before inoculation into sterile 10 ml Panlab seed medium [99] held in a large test tube fitted with a loose lid. The seed cultures were incubated at an angle of 60° at 25°C and with agitation at 240 rpm for 2 days. 10% v/v seed culture was used to inoculate 10 ml PFM medium [99] (with cotton seed flour instead of Pharmamedia), again held in a large test tube. Incubation was conducted under the same conditions for up to 10 days. Samples of whole broth were removed aseptically and stored at −20°C until required for PyMS analysis. Samples for antibiotic bioassay were centrifuged for 5 min in a microfuge. The clarified broth was also stored at −20°C prior to bioassay.

In order to confine growth and impose nutrient-limiting conditions, solid phase cultures were grown as mini cultures in 24-well sterile tissue culture plate. PFM agar was prepared as above with 1.5% LabM Agar No. 2 and 1 ml aliquots of agar were dispensed into each well. Spores of each *P. chrysogenum* strain were spotted onto the centre of each agar-filled well and incubated at 25°C for between 3 and 11 days. Samples were removed on days 3, 5, 7, 9 and 11 by punching a 7 mm diameter cork borer into the culture, and stored at −20°C until required for bioassay and PyMS; before PyMS analysis the agar plug and biomass were ground using a plastic loop to form a smooth paste.

Penicillin production was measured by bioassay against *Bacillus subtilis* var *niger* using the agar diffusion assay technique. $\log_{10}$ dose response curves were constructed using pure penicillin G (obtained as the sodium salt from Sigma). Penicillin G was chosen because it is considered to be the most abundantly produced penicillin by *P. chrysogenum*. Samples of clarified broth were analysed by loading 10 μl aliquots onto sterile bioassay discs. Agar plugs were bioassayed by placing directly onto the seed agar. The standard curve was constructed by casting a known concentration of pure penicillin G in molten agar and removing the plug in the same way as culture samples were taken. All bioassay determinations were conducted in triplicate with the error in these values being typically 20–30%.

2.2. Preparation of the ampicillin mixture with Escherichia coli

The bacterium used was *E. coli* W3110 [61]; this is ampicillin-sensitive, indicating that any spectral features observed are not due for instance to β-lactamase activity. The mixtures were prepared as

previously used [85]: the strain was grown in 4 l liquid medium (glucose (BDH), 10.0 g; peptone (LabM), 5.0 g; beef extract (LabM), 3.0 g; per litre water) for 16 h at 37°C in a shaker. After growth the cultures were harvested by centrifugation and washed in phosphate buffered saline (PBS: $NaH_2PO_4 \cdot 2H_2O$, 2.652 g; $Na_2HPO_4 \cdot 12H_2O$, 65.6055 g; per litre water, then adjust pH to 7.8). The dry weights of the cells were estimated gravimetrically and used to adjust the weight of the final slurries using PBS to approximately 40 mg/ml. Ampicillin (desiccated D[−]-α-aminobenzylpenicillin sodium salt, 98% (titration), Sigma) was prepared in each of the bacterial slurries to give a concentration range of from 0 to 5000 μg/ml in 250 μg/ml steps.

2.3. Pyrolysis mass spectrometry

Clean iron-nickel foils (Horizon Instruments, Heathfield, UK) were inserted, using clean forceps, into clean pyrolysis tubes (Horizon Instruments), so that 6 mm was protruding from the mouth of the tube. 5 μl aliquots of the above materials were evenly applied to the protruding foils. The samples were oven dried at 50°C for 30 min, then the foils were pushed into the tube using a stainless steel depth gauge so as to lie 10 mm from the mouth of the tube. Finally, viton 'O'-rings (Horizon Instruments) were placed on the tubes. Samples were run in triplicate.

The pyrolysis mass spectrometer used in this study was the Horizon Instruments PYMS-200X, as initially described by Aries et al. [100]. The sample tube carrying the foil was heated, prior to pyrolysis, at 100°C for 5 s. Curie-point pyrolysis was at 530°C for 3 s, with a temperature rise time of 0.5 s. This pyrolysis temperature was chosen because it has been shown [101,102] to give a balance between fragmentation from polysaccharides (carbohydrates) and protein fractions. The pyrolysate then entered a gold-plated expansion chamber heated to 150°C, whence it diffused down a molecular beam tube to the ionisation chamber of the mass spectrometer. To minimize secondary fragmentation of the pyrolysate the ionisation method used was low voltage electron impact ionisation (25 eV). These conditions were employed because it has been found that the stated expansion chamber temperature gives the most reproducible spectra [103], whilst the spectra from samples ionised at 25 eV are much more robust to small changes in ionisation voltage than are those [28] obtained at lower ionisation voltages, whilst much higher ionisation voltages lead to excessive fragmentation. Non-ionised molecules were deposited on a cold trap, cooled by liquid nitrogen. The ionised fragments were focused by the electrostatic lens of a set of source electrodes, accelerated and directed into a quadrupole mass filter. The ions were separated by the quadrupole, on the basis of their mass-to-charge ratio, detected and amplified with an electron multiplier [104]. The mass spectrometer scans the ionised pyrolysate 160 times at 0.2 s intervals following pyrolysis. Data were collected over the m/z range 51 to 200, in one tenth of a mass-unit intervals. These were then integrated to give unit mass. Given that the charge of the fragment was unity the mass-to-charge ratio can be accepted as a measure of the mass of pyrolysate fragments. The IBM-compatible PC used to control the PYMS-200X, was also programmed (using software provided by the manufacturers) to record spectral information on ion count for the individual masses scanned and the total ion count for each sample analysed.

Prior to any analysis the mass spectrometer was calibrated using the chemical standard perfluorokerosene (Aldrich), such that the abundance of m/z 181 was one tenth of that of m/z 69.

The data from PyMS 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.

2.4. Multivariate data analysis

The normalised data were then processed with the GENSTAT package [105] which runs under Microsoft DOS 6.2 on an IBM-compatible PC. This method has been previously described by MacFie and Gutteridge [106] and Gutteridge et al. [107]. In essence, the first stage was the reduction of the data by principal components analysis (PCA) [29,54,108–112], which is a well-known technique for reducing the dimensionality of multivariate data whilst pre-

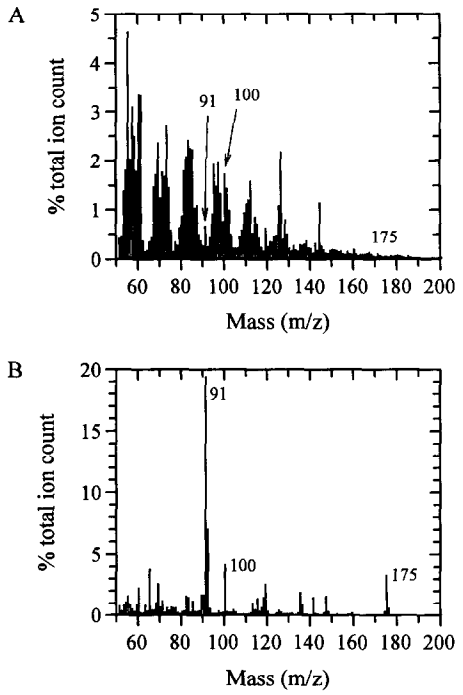

Fig. 1. Pyrolysis mass spectra of (A) *Penicillium chrysogenum* strain P1 grown for six days in shaken submerged culture producing 112 μg ml^{-1} penicillin and of (B) 200 μg pure penicillin G (sodium salt).

serving most of the variance. Data were reduced by keeping only those principal components (PCs) whose eigenvalues accounted for more than 0.1% of the total variance. Canonical variates analysis (CVA) then separated the samples into groups on the basis of the retained PCs and the a priori knowledge of the appropriate number of groupings [113,114]. The next stage was 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 [115]. Finally, hierarchical cluster analysis (HCA) was employed to produce a dendrogram, using average linkage clustering [107].

2.5. *Artificial neural networks*

All ANN analyses were carried out under Microsoft Windows 3.1 or WindowsNT on an IBM-compatible PC. Data were normalised prior to analysis using the Microsoft Excel 4.0 spreadsheet. Three back propagation neural network simulation programs were employed. The first was NeuralDesk version 1.2 (Neural Computer Sciences, Southampton, UK) with an accelerator board for the PC (NeuSprint) based on the AT&T DSP32C chip; the advantage bestowed by using this package was that simulations could be carried out relatively quickly (typically 20 min). The other two programs that were used were WinNN version 0.93 (Yaron Danon, Troy, NY) and Thinks version 1.02c (Logical Designs Consulting, La Jolla, CA). These did not use a co-processor board and so were relatively slow (typically a tenth of the speed compared with NeuralDesk using NeuSprint); however, the major advantage that these packages offered was that, among other topological and processing options, the "squashing" function of the output node (and indeed of the nodes on the hidden layer) could be varied. NeuralDesk 1.2 permitted only sigmoidal squashing functions.

The structure of the ANN used in this study to analyse pyrolysis mass spectra consisted of 3 layers containing 159 processing nodes (neurons or units) made up of the 150 input nodes (normalised pyrolysis mass spectra), 1 output node (amount of determinand), and one "hidden" layer containing 8 nodes (i.e., a 150-8-1 architecture). Each of the 150 input nodes was connected to the 8 nodes of the 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 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, the function $f = 1/(1 + e^{-x})$, where $x = \Sigma$inputs. These signals are then passed to the output node which sums them and in turn squashed either by the previously mentioned sigmoidal activation function or a linear scalar; the product of this node was then feed to the "outside world".

In addition, the hidden layer and output node were connected to a bias (whose activation was always set to $+1$), making a total of 1217 connections, whose weights will be altered during training. Before training commenced the values applied to the input and output nodes were normalised between 0 and $+1$; the input layer was scaled globally, i.e., 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 −0.005 and +0.005).

The algorithm used to train the neural network was the standard back-propagation (BP) [37,116]. For the training of the ANN each input (i.e., normalised pyrolysis mass spectrum) is paired with a desired output (i.e., the amount of antibiotic, the determinand); 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 0.1 and a momentum of 0.9.

Each epoch represented 1217 connection weight updatings and a recalculation of the error between the true and desired outputs 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.

It is known [37,41,44,61,64] that neural networks can become over-trained. An over-trained neural network has usually learnt perfectly the stimulus patterns it has seen but cannot give accurate predictions for unseen stimuli, i.e., it is no longer able to generalise. For ANNs accurately to learn and predict the concentrations of determinands in biological systems networks must obviously be trained to the correct point. Therefore for the various neural network runs the pyrolysis mass spectral data were usually split into three sets: (1) data used to train the ANN; (2) data employed to cross-validate the model; (3) spectra whose determinand concentration is "unknown" and used to test the "calibrated" system. During training the network was interrogated with the cross validation set and the %error between the output seen and that expected was calculated, thus allowing a second learning curve for the cross-validation set to be drawn. Training was stopped when the error on the cross-validation data was lowest. Once trained to the best generalisation point, the neural network was challenged with stimuli (i.e., pyrolysis mass spectra) whose determinand concentrations were "unknown".

2.6. Principal components and partial least squares regression

All PCR and PLS analyses were carried out using the program Unscrambler II Version 4.0 (CAMO, Trondheim, Norway) (and see [54]) which runs under Microsoft MS-DOS 6.2 on an IBM-compatible PC. Data were also processed prior to analysis using the Microsoft Excel 4.0 spreadsheet, run under Microsoft Windows NT on an IBM-compatible PC.

The first stage was the preparation of the data. This was achieved by presenting the "training set" as two data matrices to the program; X-, which contains the normalised triplicate pyrolysis mass spectra, and Y- which represents the concentration of the determinand. Unscrambler II also allows the addition of "start noise" (i.e., noise to the X-data); this option was not used. Finally, the X-data were mean centred and scaled in proportion to the reciprocal of their standard deviations.

The next stage is the generation of the calibration model; this first requires the user to specify the appropriate algorithm. The Unscrambler II program has one PCR algorithm and two PLS algorithms: PLS1 which handles only one Y-variable at a time, and PLS2 which will model several Y-variables simultaneously [54]. Since we wanted to predict only one Y-variable the PCR and PLS1 algorithms were used.

The method of validation used was full cross-validation, via the leave-one-out method. This technique sequentially omits one sample from the calibration; the PCR or PLS model is then re-determined on the basis of this reduced sample set. The concentration (μg/ml) of the omitted sample is then predicted with the use of the model. This method is required to determine the optimal size of the calibration model, so as to obtain good estimates of the precision of the multivariate calibration method (i.e.,

neither to under- nor over-fit predictions of unseen data) [52,54,117,118]. Unscrambler also has reasonably sophisticated outlier detection methods; although these were employed we did not find it necessary to delete any of the objects from the calibration models formed.

Cross-validation can indicate the optimal number of principal components (PCs) or PLS factors to use in predictions after the model is calibrated. To establish the accuracy of the suggestions produced by Unscrambler we therefore calculated the RMS error between the true and desired concentrations over the entire calibration model, for the known training set, cross-validation set and unknown mass spectra, and plotted these RMS errors vs. the number of latent variables (factors) used in predictions. Using this approach, after calibration, to choose the optimal number of PCs or PLS factors to use in the prediction, all pyrolysis mass spectra were used as the "unknown" inputs (test data); the model then gave its prediction in terms of the concentration of determinand.

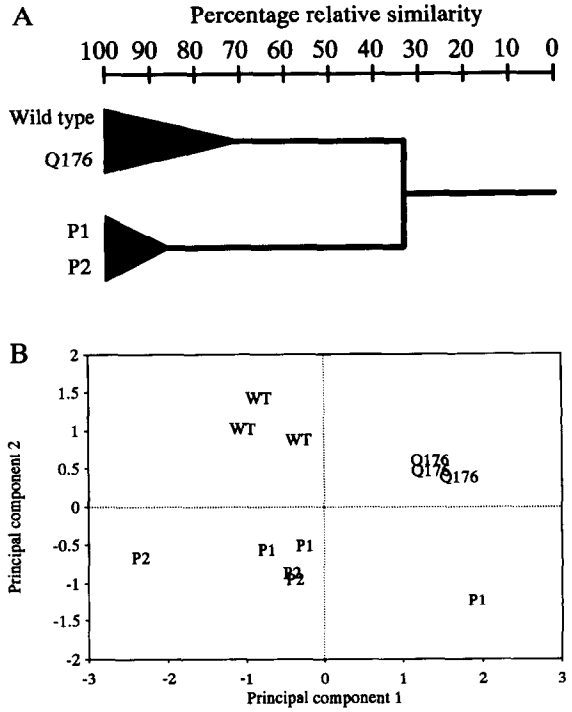

Fig. 2. Abridged dendrogram (A) and principal components analysis plot (B) representing the relationships between the *Penicillium chrysogenum* wild type strain and three mutants (Q176, P1 and P2) grown on solid media, based on PyMS data of the biomass and the agar plug analysed by GENSTAT. In the ordination diagram the first two principal components are displayed which respectively account for 55.8% and 26.3% of the total variance.

3. Results and discussion

3.1. Analysis of the agar plugs

After collection of the pyrolysis mass spectra from the *P. chrysogenum*, grown on solid media and homogenised with the agar plug to form a paste; the first stage of the experiment was to perform multivariate statistical analysis using the GENSTAT package to establish the relationships between the *Penicillium* strains. Each of the four species grown for three days, represented by the three replicate spectra, were coded to give four groups; the resulting dendrogram and principal components analysis plot are shown in Fig. 2. In the abridged dendrogram (Fig. 2a) it can be seen that at 33% relative similarity the strains cluster into two groups: the first of the wild type NRRL 1951 and Wis Q176, having a relative similarity of 70% (although 70% may seem rather small this is due to the size of the universe analysed and hence why the word *relative* is used, that is to say if a different *Penicillium* species were also analysed then these two strains would be close to 100% similar); the second cluster contains P1 and P2 which were 85% similar. PCA is the best method for reducing the dimensionality of multivariate data whilst preserving most of the variance; in our pyrolysis mass spectral data this reduction was from the 150 m/z values to 2 principal components (PCs). Plots of the first two PCs of the variance in the PyMS for these four strains (Fig. 2b) show that the majority of the variation was preserved in the first two PCs and was 55.8% and 26.3% (82.1% total) of the variance respectively; with the foresight of the hierarchical cluster analysis (Fig. 2a) the same two clusters may be observed. Similar results (data not shown) were found when the same organisms were analysed when sampled after five, seven, nine or 11 days of growth. Furthermore, although only four strains were studied the degree of separation ob-

served between them appears to reflect their different genealogical history.

The method of analysing these samples is novel in that the organism and the agar matrix, containing diffused antibiotic, were subjected to PyMS together. PCA displays the *natural* relationships between multivariate data, i.e., there is no a priori knowledge of the triplicate sampling, it is reassuring therefore that the three spectra for the wild type strain and the three for Q176 cluster together indicating that good reproducible data could be obtained from materials prepared using this rather 'crude' sampling technique. It is likely that this method of sample preparation could be exploited for the accurate discrimination of other organisms which grow into the solid agar media on which they are being cultured.

The next stage was to use PCA to observe whether there were any differences between the same *P. chrysogenum* strain cultured over 11 days. Fig. 3 displays the PCA plots based on the PyMS data from the homogenised biomass/agar plugs of the wild type (Fig. 3a) and mutant P2 (Fig. 3b) grown for three, five, seven, nine and 11 days; the first two principal components are displayed which for NRRL 1951 account for 72.0% and 13.1% (85.1% total) of the total variance respectively, and for P2 82.3% and 8.6% (90.9% total). The reproducibility of the three replicate samples for each strain on different days was again good.

The most obvious feature seen in both of the PCA plots (Fig. 3) was that the first PC roughly accounted for the time the culture had been grown, indicating that PyMS was detecting a phenotypic difference in the *P. chrysogenum* strains (This was also observed in the cultures of Q176 and P1 (data not shown)). The differences observed also proceed on from one another and followed a roughly linear trend. That this trend ceased after seven days indicates that a shift in their metabolism may have occurred; if this is true then PyMS might be exploited to highlight such changes.

3.2. Quantification of the penicillins produced by *Penicillium chrysogenum*

After the analysis of the relationships between the cultures the next stage was to assess whether the penicillin titre could be accurately assessed using PyMS. The results of the bioassay measurements of the penicillin titre (μg ml^{-1}) produced by the four

Table 1
The amount of antibiotic activity (expressed as the titre of all penicillins, μg ml^{-1}) produced in the various samples taken from fermentation broths of *Penicillium chrysogenum* wild type and mutants

Strain	Incubation period (days)	penicillin titre (μg ml^{-1})	Use in neural network training	Strain	Incubation period (days)	Penicillin titre (μg ml^{-1})	Use in neural network training
WT	2	0.079	Train	Q176	2	3.1	Test
WT	10	1.5	Cross-Val	Q176	5	5.9	Test
WT	6	3.4	Train	Q176	9	1.4	Train
WT	5	5.5	Train	Q176	8	2.6	Test
WT	3	8.7	Cross-Val	Q176	10	3.6	Test
WT [a]	9	10.3	Train	Q176	3	4.1	Test
WT [a]	4	12.5	Cross-Val	Q176	7	3.4	Test
WT [a]	5	15.5	Train	P1	10	39	Train
WT [a]	3	18.7	Cross-Val	P1	8	69	Test
WT [a]	7	21.4	Train	P1	6	112	Train
WT [a]	6	23.4	Cross-Val	P1	5	81	Test
WT [a]	10	31.5	Cross-Val	P1	3	50.1	Test
WT [a]	7	41.4	Train	P1	2	61.6	Train
WT [a]	3	48.0	Cross-Val	P2	10	22	Train
WT [a]	5	55.5	Cross-Val	P2	9	43	Test
WT [a]	6	63.4	Train	P2	7	91	Train
WT [a]	6	73.4	Cross-Val	P2	2	43	Train
WT [a]	9	80.3	Train	P2	4	229	Test

[a] Sample was spiked with pure penicillin G (Sigma) to titre shown.

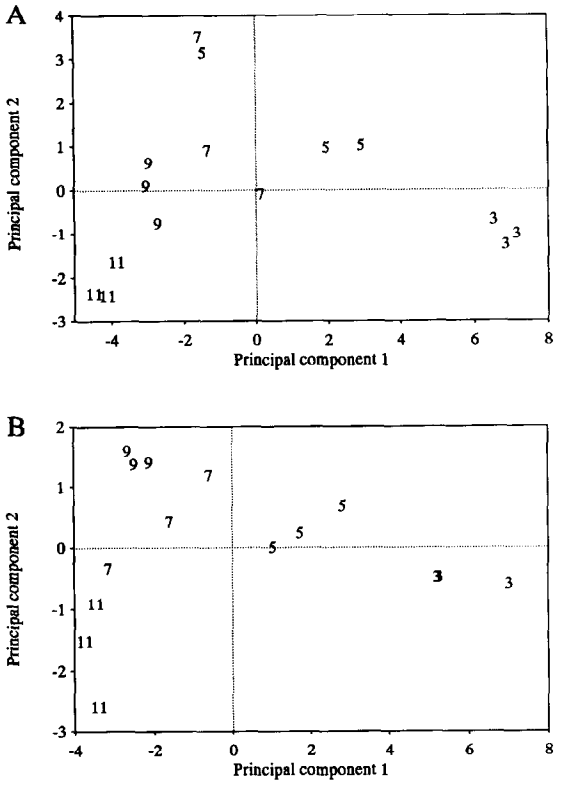

Fig. 3. Principal components analysis plots showing the relationships between (A) the *Penicillium chrysogenum* wild type strain and (B) the *P. chrysogenum* mutant P2 grown on solid media for either 3, 5, 7, 9 and 11 days, based on PyMS data of the biomass and the agar plug analysed by GENSTAT. The first two principal components are displayed which for A account for 72.0% and 13.1% of the total variance, and for B 82.3% and 8.6%, respectively.

Penicillium chrysogenum strains from fermentation liquors are given in Table 1. The error of the bioassay was typically 20–30%.

Pyrolysis mass spectral fingerprints of *P. chrysogenum* P1 grown for 6 days producing a penicillin titre of 112 μg ml^{-1}, and of the pure β-lactam (Penicillin G sodium salt), are shown in Fig. 1. The pyrolysis mass spectrum of the *P. chrysogenum* fermentation producing penicillin is quite complex and difficult to interpret (Fig. 1a); however the chemical fingerprint of the pure antibiotic (Fig. 1b) is relatively simple and contains intense (characteristic) peaks at m/z 91, 100 and 175, and the m/z 91 peak represents almost 20% of the total ion count. We have previously observed the m/z 100 peak in the pyrolysis mass spectra of ampicillin [85] and this pyrolysis fragment has also been observed in penicillins analysed under somewhat different mass spectral conditions by Meuzelaar et al. [28]. The exact structure(s) of the ions described by this peak and of m/z 91 and m/z 175 have yet to be elucidated. If the origin of these peaks were determined by either pyrolysis MS/MS or pyrolysis GC/MS the complicated fragmentation of penicillins by pyrolysis may become known, but for present purposes the detailed interpretation of the individual mass spectra is not important.

The peaks m/z 91, 100 and 175 which are characteristic of penicillin G (Fig. 1b) are relatively minor in the mass spectra of *P. chrysogenum* producing penicillins (Fig. 1a). If these masses can be considered characteristic for the antibiotic, the intensities should alter linearly depending on the relative titre of penicillin produced by *P. chrysogenum*. A plot of the average intensities of the mass m/z 91 vs. the penicillin titre from the fermentation liquors, with standard error bars and the best linear fit (which has a slope of 0.0004 and intercept at 0.53), is shown in Fig. 4. The intensity of the m/z 91 peak was evidently not well correlated with increasing penicillin titre. Similar results were observed when m/z 100 and m/z 175 were plotted against the amount of

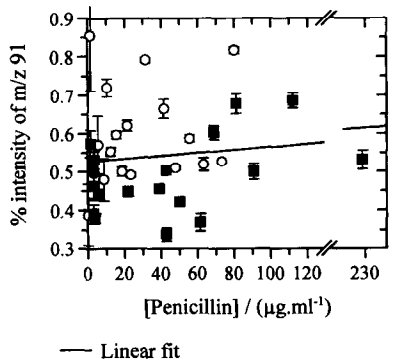

Fig. 4. Effect of the concentration of penicillins on the percentage intensity of m/z 91 in the pyrolysis mass spectra of the *Penicillium chrysogenum* wild-type strains producing penicillin(s) and the same strains spiked with penicillin G (open circles) and three mutant strains (closed squares). Error bars show standard deviation. The linear fit is shown and has a slope of 0.0004 and intercept at 0.53.

antibiotic (data not shown). Thus, changes in single ions could not be used to estimate the amount of penicillin produced in the *P. chrysogenum* fermentations.

We therefore trained 150-8-1 ANNs, using the standard back-propagation algorithm, with normalised ion intensities from the averaged pyrolysis mass spectra from 16 fermentation liquors (see Table 1 for details) as inputs and the penicillin titres as outputs; the latter covered the range from 0.079 μg ml^{-1} to 112 μg ml^{-1} and were scaled to lie between 0 and 1. To ensure good generalisation it is normally important to have a training set that fills the concentration range [62], and therefore some of the fermentation liquors from the wild type *Penicillium* were spiked with pure penicillin G (this penicillin was chosen because it is the most abundantly produced by *P. chrysogenum*). The effectiveness of training was expressed in terms of the percentage error between the actual and desired network outputs; this "learning curve" is shown in Fig. 5a (open circles). During training the network was interrogated with a second collection of 9 pyrolysis mass spectra (these were the remaining spectra from wild type liquors; see Table 1 for details), termed the cross validation set, the %error between the seen and desired outputs calculated and a second learning curve was also plotted in Fig. 5a (open squares). It can be seen that whereas the learning curve of the training set continues to decrease during training the cross validation set's learning curve initially decreases for approximately 1900 epochs and then increases. This indicates that the ANN was being over-trained, and it is important not to over-train ANNs since (by definition) the network will not then generalise well [37,44,45,61,119]. This over-training appears even more marked when the %error of the cross validation set is plotted against the %error of the training set (Fig. 5b); the minimum RMS error in the cross validation was reached (13.19%) when the RMS error of the training set was 6.55% and optimal training had occurred.

The method of using a cross validation set during the training of neural networks is invaluable in achieving the best generalisation. In the present experiment we were confident that the neural network training was terminated at the correct point because subsequent neural networks were also trained and the

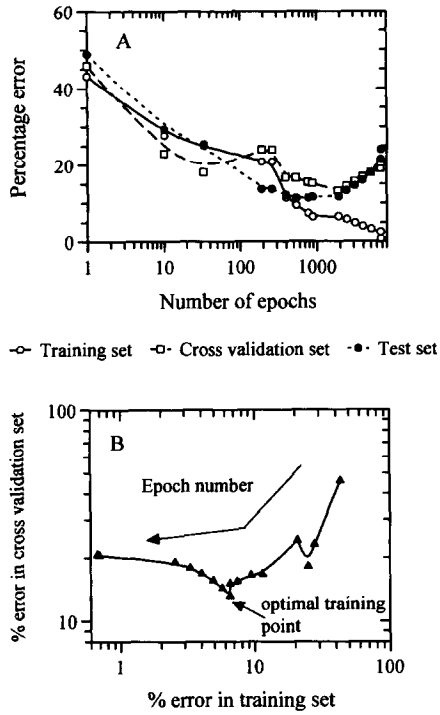

Fig. 5. (A) Typical learning curves for the ANN, using the standard back propagation algorithm and with one hidden layer consisting of eight nodes, trained to estimate the amount of penicillin (μg ml^{-1}) production. The open circles represent the percentage error of the data used to train the neural network (the training set), the open squares from the cross validation data set, and the closed circles the data from the test set. A plot of the percentage RMS error of the test set versus the percentage error of the cross validation set (B) shows that optimal training (to produce a network which generalised well) occurred at 6.55% error in the training set; the number of epochs (and hence extent of training) increases from right to left.

%error between the actual and desired outputs for the test set calculated and plotted versus the number of epochs (Fig. 5a; closed circles). It can be seen that the point at which the %error of the results from the test sets error is lowest is at a very similar point to that when the error in the cross validation set is minimal.

At the optimal point indicated by the cross validation set, the ANN was then interrogated with all the PyMS data, including the 11 spectra from the test set (refer to Table 1 for details), and a plot of the network's estimate versus the true titre of penicillins (Fig. 6) gave an approximately linear fit; the slope of

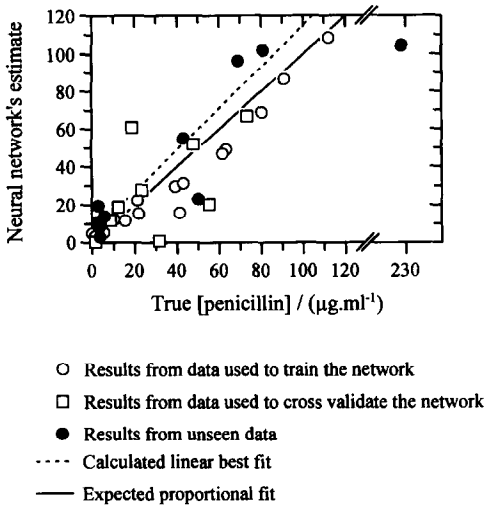

Fig. 6. The estimates of trained 150-8-1 neural networks versus the true amount of penicillin produced by *Penicillium chrysogenum*; refer to Table 1 for details of training, cross validation and test sets. Networks were trained using the standard back propagation algorithm employing a sigmoidal squashing function on the output layer to the point at which best generalisation occurred (this is the point indicated by Fig. 5 and was at 6.55% error in the training set); open circles represent spectra that were used to train the network, open squares the cross validation set and closed circles indicate "unknown" spectra which were not in the training or validation sets. The calculated linear best fit and the expected proportional fit are shown.

the best fit line for the test data was 1.102 and the intercept was 4.9 (Table 2). Given that the error in bioassays for the true penicillin titre was between 20–30% and that these values were used to train the neural networks, it was indeed very encouraging that the error in the test set was "only" 11.71%; it was therefore evident that the network's estimate of the quantity of penicillin in the fermentor liquors was similar to the true quantity, both for spectra that were used to train (open circles) and cross validate (open squares) the neural network and, most importantly, for the "unknown" pyrolysis mass spectra (closed circles). It is particularly noteworthy that those "unseen" cultures actually containing the 3 highest concentrations of penicillin were indeed predicted to have the 3 highest concentrations. The prediction of the penicillin titre for *P. chrysogenum* P2 grown for 4 days producing 229 μg ml^{-1} of the penicillin was, however, rather inaccurately assessed to be 104.4 μg ml^{-1} (Fig. 6). This was unsurprising since it is generally accepted [61,94,95] that neural networks which perform non-linear mappings cannot be expected to extrapolate beyond their knowledge base, that is to say the range on which they had been trained. That the value given for P2 was near the maximum output (112 μg ml^{-1}) is encouraging because this would indicate to the investigator that

Table 2
Comparison of artificial neural network calibration with partial least squares, principal components regression and multiple linear regression in the deconvolution of pyrolysis mass spectra for determining the concentration of penicillin (μg ml^{-1}) produced in fermentations by *Penicillium chrysogenum*

Percentage error between true values and estimates of the concentration of penicillin (μg.ml^{-1}):				
	ANNs [a]	PLS [b]	PCR [c]	MLR [d]
Training data	6.55	2.64	15.04	0.00
Cross validation	13.19	18.77	18.84	23.07
Test data	11.71	33.96	51.57	50.90
Results for the test data:				
Intercept	4.9	48.9	−63.6	80.4
Slope	1.102	0.410	2.002	−0.170
Correlation coefficient	0.922	0.336	0.704	−0.098

[a] Neural network trained to optimal point (1900 epochs) with the standard back propagation algorithm exploiting a *sigmoidal* squashing function on the output layer.
[b] Results using partial least squares (PLS) were from the optimal calibration models which were formed using 11 latent variables.
[c] Results using principal components regression (PCR) were from the optimal calibration models which were formed using 8 principal components.
[d] Multiple linear regression (MLR).

fermentation liquors with higher titres of penicillin than those previously seen should lie near the maximum and that these samples would then be targeted for further study.

As outlined above, PCR and PLS were also used to create calibration models, using the same data that were used to train ANNs, in order to predict the titre of penicillin produced by the *P. chrysogenum* strains. Table 2 gives the percentage error on the predictions produced by PCR and PLS on the training, cross validation and test sets, and these are compared with results from ANNs and MLR. It can be seen that whereas ANNs were able to predict the test set reasonably well (11.7% error in the test set) these linear regression methods were really very poor at interpolation and the %test set error was from 34.0% to 51.6%. In addition, the number of latent variables used to obtain optimal calibration models were 11 factors for PLS and 8 principal components for PCR, although one might have presumed that using more than 3 factors would cause overfitting [118], that is to say inaccurate predictions on the test data. That optimal calibration occurred using > 3 latent variables, a phenomenon that has however been seen previously [64,85], usually implies that there are non-linear relationships within the pyrolysis mass spectral data [54]. The neural network architecture used in this study was 150-8-1 and employed non-linear squashing functions on the 8 nodes in the hidden layer; networks of this topology permit the accurate mapping of *non-linear* features within data [120,121]. It is therefore not surprising that the ANNs used in this study predicted the penicillin titre more accurately than either PCR, PLS or MLR, techniques which rely only on *linear* regression. Furthermore, in the present study penicillin (the determinand) was produced in four *different* biological backgrounds and this may also explain why the linear regression techniques failed to predict the penicillin titre.

3.3. Investigation into extrapolation using neural networks and linear regression techniques

We have previously reported in a study to quantify the amount of casamino acids in glycogen [62], that when ANNs were trained with PyMS data, even though the network's estimate was linear, the edges of the data range were nearly always sigmoidal. This curvature in the graphs of neural networks estimates versus determinand concentration has also been observed by other workers using infrared spectroscopy [96,97]. Since this phenomenon has been observed when analysing IR and MS data it is not likely to be an effect of the data capture technique but due to the use of the sigmoidal squashing function on the output node(s). Neural networks should not therefore be expected to give wholly correct estimates near the edges of or outside their training sets because they cannot be expected to extend beyond the range of the non-linear activation function. To test this hypothesis it was desirable to compare sigmoidal versus linear squashing functions on the output nodes of topologically identical 150-8-1 neural networks to see if more accurate estimates near the edges of the concentration range could be gained.

We therefore prepared a mixture of 40 mg ml^{-1} *E. coli* containing from 0 to 5000 $\mu g\ ml^{-1}$ ampicillin, in 250 $\mu g\ ml^{-1}$ steps, and analysed these using PyMS. Next ANNs were trained, using the standard back-propagation algorithm, with normalised ion intensities from the averaged triplicate pyrolysis mass spectra from 0–5000 $\mu g\ ml^{-1}$ in 500 $\mu g\ ml^{-1}$ steps as inputs and the ampicillin concentration as outputs. The output node was scaled to lie between 0 and 1 and employed either a sigmoidal squashing function ($f = 1/(1 + e^{-x})$, where $x = \Sigma$inputs) or a linear scalar ($f = x$). To ensure good generalisation the spectra from samples containing 250 $\mu g\ ml^{-1}$–4750 $\mu g\ ml^{-1}$ ampicillin in 500 $\mu g\ ml^{-1}$ steps were used to cross validate the model. Training was stopped when the cross validation set error was 1.89% and 2.75% for ANNs with sigmoidal and linear squashing functions respectively; this was after approximately 5000 and 1500 epochs, respectively (the error in the training data were 2.47% and 1.61%). The network's estimate was then plotted versus the true ampicillin concentration (Fig. 7). It can be seen that the estimates from networks employing sigmoidal activation function on the output node (bold line) did indeed have a slight curvature at the edges of the concentration range; that is from 0 $\mu g\ ml^{-1}$ to 1000 $\mu g\ ml^{-1}$ and from 4000 $\mu g\ ml^{-1}$ to 5000 $\mu g\ ml^{-1}$, which is a reflection of the sigmoidal nature of the non-linear squashing function. However, the estimates near the edges from networks employing linear squashing functions (bold

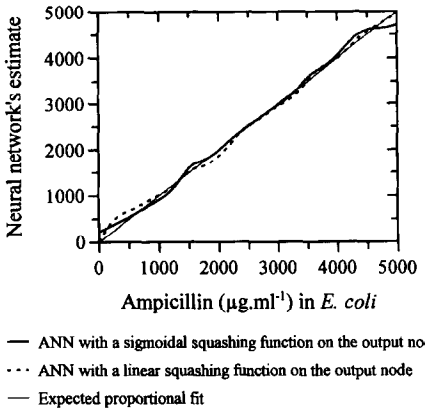

Fig. 7. The estimates of trained 150-8-1 neural networks vs. the true amount of ampicillin (0–5000 μg ml^{-1} in steps of 250 μg ml^{-1}) mixed in 40 mg ml^{-1} *E. coli*. Networks were trained on 0–5000 μg ml^{-1} in steps of 500 μg ml^{-1} using the standard back propagation algorithm employing either a *sigmoidal* (bold line) or a *linear* (bold broken line) squashing function on the output layer (node). The expected proportional fit is also shown.

broken line) were more accurate and closer to the expected proportional fit (thin line). It is concluded that the latter neural networks can be used to attain better results. It is noteworthy that although the error in the test set of networks employing only sigmoidal squashing functions was slightly better, these networks did take (3 times) longer to train. Furthermore, during the early stages of training the sigmoidal effect in plots of estimates vs. real values was even more pronounced; this phenomenon did not occur at any point in the training of networks using a linear scalar on the output node (data not shown).

Since neural networks using linear functions on the output node gave more accurate predictions at the edges of the concentration range it was worth considering whether they could be used to extrapolate beyond their knowledge base, that is can they extend beyond the range of the squashing function on the output node(s)?

ANNs were therefore trained, using the standard back-propagation algorithm, with normalised ion intensities from the averaged triplicate pyrolysis mass spectra from 0–2500 μg ml^{-1} in 500 μg ml^{-1} steps as inputs and the ampicillin concentration as outputs. The output node was scaled to lie between 0 and 1 and employed either a sigmoidal squashing function or a linear scalar. To ensure good generalisation the

Table 3
Comparison of artificial neural network calibration with partial least squares, principal components regression and multiple linear regression in the deconvolution of pyrolysis mass spectra from ampicillin mixed in *Escherichia coli*. These analyses were aimed at assessing the ability of these techniques to extrapolate above the range used to train the neural networks or to calibrate the linear regression models

Percentage error between true values and estimates of the amount of ampicillin (μg ml^{-1}) in 40 mg ml^{-1} *E. coli*:					
	ANNs [a]	ANNs [b]	PLS [c]	PCR [d]	MLR [e]
Training data [f]	2.08	1.21	1.35	3.81	0.00
Cross validation [g]	2.44	2.28	2.43	2.84	6.47
Extrapolation data set [h]	28.38	3.44	4.26	7.01	7.34
Results for the extrapolation data set:					
Intercept	2240	−503	−15	31	−894
Slope	0.06	1.16	0.96	0.90	1.30
Correlation coefficient	0.899	0.966	0.968	0.982	0.938

[a] Neural network trained to optimal point (2500 epochs) with the standard back propagation algorithm exploiting a *sigmoidal* squashing function on the output layer (node).
[b] Neural network trained to optimal point (750 epochs) with the standard back propagation algorithm exploiting a *linear* squashing function on the output layer (node).
[c] Results using partial least squares (PLS) were from the optimal calibration models which were formed using 2 latent variables.
[d] Results using principal components regression (PCR) were from the optimal calibration models which were formed using 2 principal components.
[e] Multiple linear regression (MLR).
[f] Pyrolysis mass spectral data from 0–2500 μg ml^{-1} ampicillin in 500 μg ml^{-1} steps (6 patterns).
[g] Pyrolysis mass spectral data from 250–2250 μg ml^{-1} ampicillin in 500 μg ml^{-1} steps (5 patterns).
[h] Pyrolysis mass spectral data from 2750–5000 μg ml^{-1} ampicillin in 250 μg ml^{-1} steps (10 patterns).

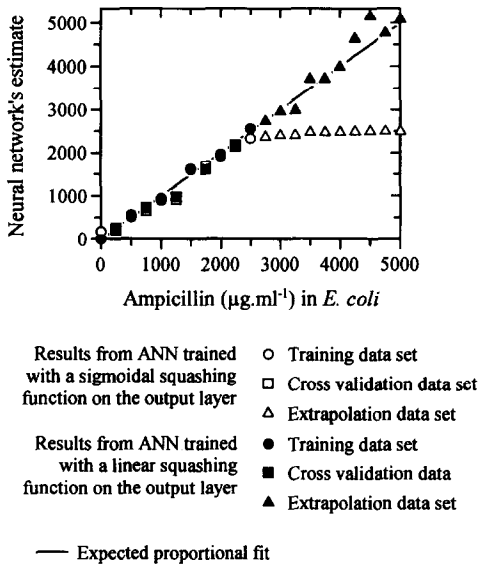

Fig. 8. The estimates of trained 150-8-1 neural networks vs. the true amount of ampicillin (0–5000 μg ml^{-1} in steps of 250 μg ml^{-1}) mixed in 40 mg ml^{-1} *E. coli*. (open symbols) Networks were trained using the standard back propagation algorithm employing a *sigmoidal* squashing function on the output layer; open circles represent spectra that were used to train the network, open squares represent the cross validation set and open triangles indicate "unknown" spectra which were *beyond* the training range to test the network's ability to extrapolate. In contrast (closed symbols) show data from neural networks set up as above except that they used a *linear* squashing function on the output node; again closed circles represent spectra that were used to train the network, closed squares the cross validation set and the closed triangles indicate the "unknown" extrapolation data set. The expected proportional fit is shown.

spectra from samples containing 250 μg ml^{-1}–2250 μg ml^{-1} ampicillin in 500 μg ml^{-1} steps were used to cross validate the model. Training was stopped when the cross validation error was minimal and was 2.44% and 2.28% for ANNs with sigmoidal and linear squashing functions, respectively; this was after approximately 2500 and 750 epochs, respectively (Table 3). The remaining mass spectra from 2750 μg ml^{-1}–5000 μg ml^{-1} ampicillin in 250 μg ml^{-1} steps were then passed through the trained networks and the results plotted against the true ampicillin concentration (Fig. 8). It can be seen that as expected the network using sigmoidal squashing functions was unable to extrapolate and all the estimates were near the maximum value seen, i.e., 2500 μg ml^{-1}. In contrast, neural networks employing a linear scalar on the output node gave accurate predictions of the concentration of ampicillin showing that these networks could indeed be used to extrapolate.

The linear regression methods of MLR, PCR and PLS were also used to create calibration models, using the same data that were used to train and cross validate ANNs, in order to assess their ability to extrapolate. Table 3 gives the percentage error on the predictions produced by MLR, PCR and PLS on the training, cross validation and test sets and are compared with results from the two ANNs; the intercept, slope and correlation coefficient were also calculated for these five analysis techniques. It can be seen that MLR, PCR and PLS can also be used to form models which will allow one to extrapolate to obtain the ampicillin concentration of samples analysed by PyMS at levels beyond those in the calibration model, and that PLS gave the best predictions with the error in the test set being 4.26%. The ANN using the linear squashing function did however still extrapolate best and the test set %error between estimated and real ampicillin concentration was only 3.44%. It is likely that the reason ANNs gave better predictions was because ANNs are considered to be robust to noisy data [120,121]. Furthermore, we have previously observed a small amount of noise in pyrolysis mass spectral data [64] (particularly in m/z values of low intensity) to which ANNs were robust but which PLS and PCR incorporated into their calibration models to give less accurate predictions for the determinand in binary mixtures [64].

In conclusion, ANNs (with linear squashing functions on the output node) and the linear regression methods can be used accurately to predict the concentration of ampicillin mixed with *E. coli* when this is beyond the concentration range used in the training or cross-validations sets; in the present study the range of the test set was 100% above that of the training and cross-validation datasets. It is likely that the reason this extrapolation is possible is because the spectra are linearly orderable in the high dimensional weight space; since linear regression methods were also able to extrapolate using PyMS data.

Although the PyMS of ampicillin mixed in *E. coli* cannot necessarily be expressed solely in terms of linear superpositions of subpatterns of spectra describing the pure components of the mixtures since

intermolecular reactions take place in the pyrolysate [85], because PLS could be used to quantify accurately the amount of ampicillin in *E. coli* from their spectra, the increasing additivity of ampicillin to *E. coli*, in terms of the features (variables) used by PLS to create calibration models from their pyrolysis mass spectra, can be considered linear. However, when the amount of determinand (e.g., ampicillin) becomes large, compared to the biological background (e.g., *E. coli*), saturation effects on the intense peaks in the spectrum will occur. This is when the smaller peaks increase more in size relative to larger peaks and is because the electron multiplier in the mass spectrometer is being overloaded with ions. Saturation will therefore lead to a deterioration in the linear superposition of the pyrolysis mass spectra and the spectra will no longer be linearly orderable in the high dimensional weight space, thus there is a limit beyond which ANNs and the linear regression methods will no longer be able to extrapolate accurately.

The last stage of this study was to see if ANNs and the linear regression techniques could be trained or calibrated with the pyrolysis mass spectra of *P. chrysogenum* strains producing penicillin to predict the titre of the antibiotic beyond the training data range. The spectra from the four mutant cultures producing the most penicillin (81, 91, 112 and 229 $\mu g\ ml^{-1}$) were used as an extrapolation test set and the remaining 32 spectra were used to train ANNs or calibrate regression models; the maximum value in the training set was 80.3 $\mu g\ ml^{-1}$. Table 4 gives the predictions produced by ANNs (employing either a sigmoidal or linear squashing function), MLR, PCR and PLS. It can be seen that ANNs using a sigmoidal squashing function could not be used to extrapolate and estimates produced were all near the maximum seen. All the linear regression techniques gave accurate estimates for cultures producing 112 $\mu g\ ml^{-1}$ penicillin, the ANN employing a linear squashing function slightly over estimated this to be 119.2 $\mu g\ ml^{-1}$. The estimates for 81 and 91 $\mu g\ ml^{-1}$ penicillin were less accurate and were between 102.7 and 123.5, and 60.9 and 67.5, respectively, for all methods. Finally, although the estimates for the culture producing *three* times as much penicillin (229 $\mu g\ ml^{-1}$) as the maximum value in the training set were low, typically 162 $\mu g\ ml^{-1}$ (MLR, PCR and PLS) and even lower (100 $\mu g\ ml^{-1}$) for the ANN, it was encouraging that these estimates were beyond the maximum value seen in the training set. Although these estimates were not always accurate, this does demonstrate that ANNs and linear regression techniques can be applied to the PyMS spectra of mutants in a screening program to give an indication which cultures are over-producing the desired

Table 4
Comparison of artificial neural network calibration with partial least squares, principal components regression and multiple linear regression in the deconvolution of pyrolysis mass spectra for determining the concentration of penicillin ($\mu g\ ml^{-1}$) produced in fermentations by *Penicillium chrysogenum*. These analyses were aimed at assessing the ability of these methods to extrapolate above the range used to train the neural networks or to calibrate the linear regression models.

The true values and estimates of the concentration of penicillin ($\mu g\ ml^{-1}$) produced in *Penicillium chrysogenum* fermentations					
True value	ANNs [a]	ANNs [b]	PLS [c]	PCR [d]	MLR [e]
81.0	71.2	123.5	106.2	103.2	102.7
91.0	78.7	67.5	60.9	61.7	61.0
112.0	71.6	119.2	115.1	112.1	111.7
229.0	73.8	99.5	163.4	161.9	160.7

[a] Neural network trained until the error in the training set was 2.5% (approximately 10,000 epochs) with the standard back propagation algorithm exploiting a *sigmoidal* squashing function on the output layer.
[b] Neural network trained until the error in the training set was 2.5% (approximately 40,000 epochs) with the standard back propagation algorithm exploiting a *linear* squashing function on the output layer.
[c] Results using partial least squares (PLS) were from calibration models which Unscrambler II suggested would give best results and were formed using 24 latent variables.
[d] Results using principal components regression (PCR) were from calibration models which Unscrambler II suggested would give best results and were formed using 31 principal components.
[e] Multiple linear regression (MLR).

metabolite(s) to the greatest extent, thus highlighting strains for further study.

4. Conclusions

PyMS was able to differentiate between the four strains of *Penicillium chrysogenum* grown on solid media and the clustering of these organisms was in agreement with the genealogical history of these strains. The method of analysing these samples was novel in that the organism and the agar matrix, containing diffused penicillin, were subjected to PyMS and it is possible that this method of sample preparation could be exploited for the accurate discrimination of other organisms which grow into the solid agar media on which they are being cultured, and for which existing sample preparation techniques are extremely cumbersome.

PyMS was able to detect phenotypic difference in the *P. chysogenum* strains grown for either five, seven, nine and 11 days growth. These differences may have been a reflection of their metabolic states, information which might be exploited in screening programs if correlated with the onset of secondary metabolism.

P. chrysogenum fermentation broths could be analysed quantitatively for penicillin (as judged by antimicrobial activity of culture to *B. subtilis* var *niger*) using PyMS and ANNs; the linear regression methods of PLS, PCR and MLR, using the same pyrolysis mass spectra, could not be used to predict the titre of penicillins (Table 2). It is likely that the latter linear regression methods could not form good calibration models because of (1) the very variable biological background (the four different strains) in which the penicillin was produced and (2) their inability accurately to map non-linearities.

Sigmoidal versus linear squashing functions on the output nodes of identical 150-8-1 neural networks were compared and it was found that networks employing linear functions gave more accurate estimates of ampicillin in *E. coli* near the edges of the concentration range than did networks using sigmoidal functions. Finally, it was shown that these neural networks could be used to extrapolate successfully for ampicillin mixed in *E. coli*, and to a lesser degree for the analysis of the fermentation broths of *P. chrysogenum* producing penicillin.

We conclude that the combination of PyMS and ANNs constitutes a rapid and convenient method for exploitation in microbial fermentation development programmes generally. PyMS is rapid (the typical sample time is less than 2 min) and automated; the present system allows 300 samples to be analysed in 8 h 45 min. Thus in a working day of two shifts (and allowing for two days down-time per month) one might expect to be able to analyse some 12,000 isolates per month.

Acknowledgements

R.G., M.J.N. and D.B.K. are supported by the Chemicals and Pharmaceuticals Directorate of the UK BBSRC, under the terms of the LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, Neural Computer Sciences and Zeneca Bioproducts plc. Ravi Manohar is thanked for his help in the preparation and assay of *P. chrysogenum* samples.

References

[1] E.J. Vandamme, in E.J. Vandamme (Ed.), Biotechnology in Industrial Antibiotics, Marcel Dekker, New York, 1984, pp. 3–31.
[2] M.S. Verrall, Discovery and Isolation of Microbial Products, Ellis Horwood, Chichester, 1985.
[3] S. Ōmura, Microb. Rev., 50 (1986) 259–279.
[4] R.P. Elander, in J. Bu'lock and B. Kristiansen (Eds.), Basic Biotechnology, Academic Press, London, 1987, pp. 217–251.
[5] W. Crueger and A. Crueger, Biotechnology: A textbook of Industrial Microbiology, Sinauer, Sunderland, MA, 1989.
[6] P.D. Crittenden and N. Porter, Trends Biotechnol., 9 (1991) 409–414.
[7] S. Ōmura, Gene, 115 (1992) 141–149.
[8] N. Porter and F. Fox, Pestic. Sci., 39 (1993) 161–168.
[9] Y. Tanaka and S. Ōmura, Ann. Rev. Microbiol., 47 (1993) 57–87.
[10] G.-H. An, J. Bielich, R. Auerbach and E.A. Johnson, E.A., Bio/Technol., 9 (1991) 70–73.
[11] T. Azuma, G.I. Harrison and A.L. Demain, Appl. Microbiol. Biotechnol., 38 (1992) 173–178.
[12] J.E. Bailey, Science, 252 (1991) 1668–1675.
[13] J. Betz, BTF-Biotech. Forum, 2 (1985) 74–80.
[14] D.B. Finkelstein, J.A. Rambosek, J. Leach, R.E. Wilson,

A.E. Larson, P.C. McAda, C.L. Soliday and C. Ball, in M. Alacevic et al. (Eds.), Proc. of the Fifth Int. Symp. on the Genetics of Industrial Microorganisms, Zagreb, 1987, pp. 101–110.
[15] D.V. Goeddel, D.G. Kleid, F. Bolivar, H.L. Heyneker, D.G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura and A.D. Riggs, Proc. Natl. Acad. Sci. USA, 76 (1979) 106–110.
[16] D.A. Hopwood, F. Malpartida, H.M. Kieser, H. Ikeda, J. Duncan, I. Fujii, B.A.M. Rudd, H.G. Floss and S. Ōmura, Nature, 314 (1985) 642–644.
[17] R. Hütter, in H.J. Rehm and G. Reed (Eds.), Biotechnology, Vol 4. VCH, New York, 1986, pp. 3–17.
[18] T. Ichikawa, M. Date, T. Ishikura and A. Ozaki, Folia Microbiol., 16 (1971) 218–224.
[19] J.F. Martin, J.A. Gil, G. Naharro, P. Liras and J.R. Villanueva, in O.K. Sebek and A.I. Laskin (Eds.), Genetics of Industrial Microorganisms (GIM78), American Soc. Microbiol., Washington, DC, 1979, pp. 205–209.
[20] J.B. McAlpine, J.S. Tuan, D.P. Brown, K.D. Grebnes, D.N. Whittern, A. Buko and L. Katz, J. Antibiot., 40 (1987) 1115–1122.
[21] G. Saunders and G. Holt, in J.F. Peberdy (Ed.), Penicillin and Acremonium, Plenum Press, New York, 1987, pp. 73–92.
[22] V.A. Saunders and J.R. Saunders, Microbial Genetics Applied to Biotechnology. Principles and Techniques of Gene Transfer and Manipulation, Croom Helm, London, 1987.
[23] G.C. Walker, Microbiol. Rev., 48 (1984) 60–93.
[24] K.D. Wittrup and J.E. Bailey, Cytometry, 9 (1988) 394–404.
[25] R.T. Rowlands, Enz. Microb. Technol., 6 (1984) 3–10.
[26] W.J. Irwin, Analytical Pyrolysis: A Comprehensive Guide, Marcel Dekker, New York, 1982.
[27] D.B. Drucker, Meth. Microbiol., 9 (1976) 51–125.
[28] H.L.C. Meuzelaar, J. Haverkamp and F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials, Elsevier, Amsterdam, 1982.
[29] C.S. Gutteridge, Meth. Microbiol., 19 (1987) 227–272.
[30] R.C.W. Berkeley, R. Goodacre, R.J. Helyer and T. Kelley, Lab. Pract., 39 (1990) 81–83.
[31] S.L. Morgan, A. Fox, J.C. Rogers and B.E. Watt, in W.H. Nelson (Ed.), Modern Techniques for Rapid Microbiological Analysis, VCH, Weinheim, 1991, pp. 1–18.
[32] J.J. Sanglier, D. Whitehead, G.S. Saddler, E.V. Ferguson and M. Goodfellow, Gene, 115 (1992) 235–242
[33] R. Goodacre, Microbiol. Eur., 2:2 (1994) 16–22.
[34] A.P. Snyder, P.B.W. Smith, J.P. Dworzanski and H.L.C. Meuzelaar, ACS Symp. Ser., 541 (1994) 62–84.
[35] R. Goodacre and R.C.W. Berkeley, FEMS Microbiol. Lett., 71 (1990) 133–138.
[36] D.B. Kell, M.J. Neal and R. Goodacre, Trends Biotechnol., 12 (1994) 434–435.
[37] D.E. Rumelhart, J.L. McClelland and the PDP Research Group, Parallel Distributed Processing, Experiments in the Microstructure of Cognition, Vols. I and II, MIT Press, Cambridge, MA, 1986.
[38] R. Beale and T. Jackson, Neural Computing: An Introduction, Adam Hilger, Bristol, 1990.
[39] R.C. Eberhart and R.W. Dobbins, Neural Network PC Tools, Academic Press, London, 1990.
[40] Y.-H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, Reading, MA, 1989.
[41] P.D. Wasserman, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.
[42] P.D. Wasserman and R.M. Oetzel, NeuralSource: the Bibliographic Guide to Artificial Neural Networks, Van Nostrand Reinhold, New York, 1989.
[43] P.K. Simpson, Artificial Neural Systems, Pergamon Press, Oxford, 1990.
[44] R. Hecht-Nielsen, Neurocomputing, Addison-Wesley, Massachusetts, 1990.
[45] J. Hertz, A. Krogh and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City, 1991.
[46] P. Peretto, An Introduction to the Modelling of Neural Networks, Cambridge University Press, Cambridge, 1992.
[47] S.I. Gallant, Neural Network Learning, MIT Press, Cambridge, MA, 1993.
[48] J. Zupan and J. Gasteiger, Neural Networks for Chemists: An Introduction, VCH, Weinheim, 1993.
[49] R. Goodacre, M.J. Neal and D.B. Kell, Zbl. Bakt, in press.
[50] K. Joreskog and H. Wold, Systems under Direct Observation, North Holland, Amsterdam, 1982.
[51] I.A. Cowe and J.W. McNicol, Appl. Spectrosc., 39 (1985) 257–266.
[52] D.M. Haaland and E.V. Thomas, Anal. Chem., 60 (1988) 1193–1202.
[53] R. Manne, Chemom. Intell. Lab. Syst., 2 (1987) 187–197.
[54] H. Martens and T. Næs, Multivariate Calibration, Wiley, New York, 1989.
[55] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991.
[56] R.G. Brereton, Multivariate Pattern Recognition in Chemometrics, Elsevier, Amsterdam, 1992.
[57] T.J. McAvoy, H.T. Su, N.S. Wang and M. He, Biotechnol. Bioeng., 40 (1992) 53–62.
[58] K.A. Martin, Appl. Spectrosc. Rev., 27 (1992) 325–383.
[59] I.E. Frank and J.H. Friedman, Technometrics, 35 (1993) 109–135.
[60] Y.-Z. Liang, O.M. Kvalheim and R. Manne, Chemom. Intell. Lab. Syst., 18 (1993) 235–250.
[61] R. Goodacre and D.B. Kell, Anal. Chim. Acta, 279 (1993) 17–26.
[62] R. Goodacre, A.N. Edmonds and D.B. Kell, J. Anal. Applied Pyrol., 26 (1993) 93–114.
[63] M.J. Neal, R. Goodacre and D.B. Kell, Proc. WCNN, San Diego, (1994) I-318–I-323.
[64] R. Goodacre, M.J. Neal and D.B. Kell, Anal. Chem., 66 (1994) 1070–1085.
[65] R. Goodacre, A. Karim, M.A. Kaderbhai and D.B. Kell, J. Biotechnol. 34 (1994) 185–193.
[66] R. Goodacre, D.B. Kell and G. Bianchi, Nature, 359 (1992) 594.
[67] R. Goodacre, D.B. Kell and G. Bianchi, J. Sci. Food Agric., 63 (1993) 297–307.

[68] R. Goodacre, M.J. Neal, D.B. Kell, L.W. Greenham, W.C. Noble and R.G. Harvey, J. Appl. Bacteriol., 76 (1994) 124–134.
[69] R. Goodacre, S.A. Howell, W.C. Noble and M.J. Neal, Zbl. Bakt, in press.
[70] J. Chun, E. Atalan, A.C. Ward and M. Goodfellow, FEMS Microbiol. Lett., 107 (1993) 321–325.
[71] R. Freeman, R. Goodacre, P.R. Sisson, J.G., Magee, A.C. Ward and N.F. Lightfoot, J. Med. Microbiol., 40 (1994) 170–173.
[72] M.S. Lee, D.J. Hook, E.H. Kerns, K.J. Volk and I.E. Rosenberg, Biolog. Mass Spectrom., 22 (1993) 84–88.
[73] E. Heinzle, J. Moes, M., Griot, H. Kramer, I.J. Dunn and J.R. Bourne, Anal. Chim. Acta, 163 (1984) 219–229.
[74] E. Heinzle, A. Oeggerli and B. Dettwiler, Anal. Chim. Acta, 238 (1990) 101–115.
[75] J.C. Weaver, in J.S. Cohen (Ed.), Noninvasive Probes of Tissue Metabolism, Wiley, New York, 1982.
[76] S. Bohatka, G. langer, J. Szilagyi and I. Berecz, Int. J. Mass Spectrom., 48 (1983) 277–280.
[77] E. Heinzle, H. Kramer and I.J. Dunn, Biotechnol. Bioeng., 27 (1985) 238–246.
[78] F.R. Lauritsen, T.K. Choudhury, L.E. Dejarme and R.G. Cooks, Anal. Chim. Acta, 266 (1992) 1–12.
[79] F.R. Lauritsen, L.T. Nielsen, H. Degn, D. Lloyd and S. Bohatka, Biolog. Mass Spectrom., 20 (1991) 253–258.
[80] D. Lloyd, J.E. Ellis, K. Hillman and A.G. Williams, J. Appl. Bacteriol., 73 (1992) 155–163.
[81] K.F. Hansen, F.R. Lauritsen and H Degn, Biotechnol. Bioeng., 44 (1994) 347–353.
[82] E. Heinzle, H. Kramer and I.J. Dunn, in A. Johnson (Ed.), Modelling and Control of Biotechnological processes, 1st IFAC Symposium. Pergamon Press, Oxford, 1985.
[83] E.P. Sandmeier, J. Keller, E. Heinzle, I.J. Dunn and J.R. Bourne, in E. Hienzle and M. Reuss (Eds.), Mass Spectrometry in Biotechnological Process Analysis and Control, Plenum Press, New York, 1988, pp. 209–215.
[84] E. Heinzle, J. Biotechnol., 25 (1992) 81–114.
[85] R. Goodacre, S. Trew, C. Wrigley-Jones, M.J. Neal, J. Maddock, T.W. Ottley, N. Porter and D.B. Kell, Biotechnol. Bioeng., 44 (1994) 1205–1216.
[86] C. Ball and M.P. McGonagle, J. Appl. Bacteriol., 45 (1987) 67–74.
[87] J.F. Martin, Ann. NY Acad. Sci., 646 (1991) 193–201.
[88] T. Barenschee, T. Scheper and K. Schugerl, J. Biotechnol., 26 (1992) 143–154.
[89] J. Barriosgonzalez, T.E. Castillo and A. Mejia, Biotechnol. Adv., 11 (1993) 525–537.
[90] T. Vichitsoonthonkul, Y.W. Chu, H. Sodhi and G. Saunders, in Y. Murooka (Ed.), Recombinant Microbes, Marcel Dekker, New York, 1994, pp. 119–135.
[91] D.B. Kell, Process. Biochem., 15 (1980) 18–23.
[92] M.P. Bakus and J.F. Stauffer, Mycologia, 47 (1995) 429–463.
[93] R.W. Swartz, Ann. Rep. Ferm. Proc., 3 (1979) 75–110.
[94] P.J. Haley and D. Soloway, Proc. Int. Joint Conf. Neural Networks, 4 (1992) paper 25.
[95] D.R. Hush and B.G. Horne, IEEE Sign. Process. Mag., (1993) 8–39.
[96] J.R. Long, V.G. Gregoriou and P.J. Gemperline, Anal. Chem., 62 (1990) 1791–1797.
[97] S.P. Jacobsson and A. Hagmann, Anal. Chim. Acta, 284 (1993) 137–147.
[98] G.J.M. Hersbach, C.P. van der Beek and P.W.M. van Dijck, in E.J. Vandamme (Ed.), Biotechnology of Industrial Antibiotics, Marcel Dekker, New York, 1984, pp. 45–140.
[99] J. Lein, in Z. Vanek and Z. Hostalek (Eds.), Overproduction of Microbial Metabolites, Butterworths, London, 1983, pp. 105–139.
[100] R.E. Aries, C.S. Gutteridge and T.W. Ottley, J. Anal. Applied Pyrol., 9 (1986) 81–98.
[101] W. Windig, P.G. Kistemaker, J. Haverkamp and H.L.C. Meuzelaar, J. Anal. Applied Pyrol., 2 (1980) 7–18.
[102] R. Goodacre, Ph.D. thesis, University of Bristol, UK, 1992.
[103] W. Windig, P.G. Kistemaker, J. Haverkamp and H.L.C. Meuzelaar, J. Anal. Applied Pyrol., 1 (1979) 39–52.
[104] J.R. Chapman, Practical Organic Mass Spectrometry, Wiley, New York, 1993.
[105] J.A. Nelder, Genstat Reference Manual, University of Edinburgh, Scientific and Social Service Program Library, 1979.
[106] H.J.H. MacFie and C.S. Gutteridge, J. Anal. Applied Pyrol., 4 (1982) 175–204
[107] C.S. Gutteridge, L. Vallis and H.J.H. MacFie, in M. Goodfellow et al. (Eds.), Computer-assisted Bacterial Systematics, Academic Press, London, 1985, pp. 369–401.
[108] C. Chatfield and A.J. Collins, Introduction to Multivariate Analysis. Chapman and Hall, London, 1980, pp. 57–81.
[109] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.
[110] D.R. Causton, A Biologist's Advanced Mathematics, Allen and Unwin, London, 1987, pp. 48–72.
[111] B. Flury and H. Riedwyl, Multivariate Statistics: A Practical Approach. Chapman and Hall, London, 1988, pp. 181–233.
[112] B.S. Everitt, Cluster Analysis, Edward Arnold, London, 1993.
[113] H.J.H. MacFie, C.S. Gutteridge and J.R. Norris, J. Gen. Microbiol., 104 (1978) 67–74
[114] W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 387–391.
[115] J.C. Gower, Biometr., 27 (1971) 857–874
[116] P.J. Werbos, The Roots of Back-Propagation: From Ordered Derivatives to Neural Networks and Political Forecasting, Wiley, Chichester, 1993.
[117] P.J. Brown, J. Chemom., 6 (1992) 151–161.
[118] M.B. Seasholtz and B. Kowalski, Anal. Chim. Acta, 277 (1993) 165–177.
[119] J.E. Moody, in J.E. Moody, S.J. Hanson and R.P. Lippmann (Eds.), NIPs 4, Morgan Kaufmann, San Mateo, CA, 1992.
[120] K. Hornik, M. Stinchcombe and H. White, Neural Networks, 3 (1990) 551–560.
[121] H. White, Artificial Neural Networks: Approximation and Learning Theory, Blackwell, Oxford, 1992.