Rapid Screening for Metabolite Overproduction in Fermentor Broths, Using Pyrolysis Mass Spectrometry with Multivariate Calibration and Artificial Neural Networks

Royston Goodacre,[1*] Sally Trew,[2] Carys Wrigley-Jones,[2†] Mark J. Neal,[1] John Maddock,[3] Thomas W. Ottley,[3] Neil Porter,[2] and Douglas B. Kell[1]

[1]Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, United Kingdom; [2]Xenova, Ltd., 545 Ipswich Road, Slough, Berkshire SL1 4EQ, United Kingdom; and [3]Horizon Instruments, Ltd., Ghyll Industrial Estate, Heathfield, East Sussex TN 21 8AW, United Kingdom

Binary mixtures of model systems consisting of the antibiotic ampicillin with either *Escherichia coli* or *Staphylococcus aureus* were subjected to pyrolysis mass spectrometry (PyMS). To deconvolute the pyrolysis mass spectra, so as to obtain quantitative information on the concentration of ampicillin in the mixtures, partial least squares regression (PLS), principal components regression (PCR), and fully interconnected feedforward artificial neural networks (ANNs) were studied. In the latter case, the weights were modified using the standard backpropagation algorithm, and the nodes used a sigmoidal squashing function. It was found that each of the methods could be used to provide calibration models which gave excellent predictions for the concentrations of ampicillin in samples on which they had not been trained. Furthermore, 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 *S. aureus* background, illustrating the robustness of ANNs to rather substantial variations in the biological background. The PyMS of the complex mixture of ampicillin in bacteria could not be expressed simply in terms of additive combinations of the spectra describing the pure components of the mixtures and their relative concentrations. Intermolecular reactions took place in the pyrolysate, leading to a lack of superposition of the spectral components and to a dependence of the normalized mass spectrum on sample size. Samples from fermentations of a single organism in a complex production medium were also analyzed quantitatively for a drug of commercial interest. The drug could also be quantified in a variety of mutant-producing strains cultivated in the same medium. The combination of PyMS and ANNs constitutes a novel, rapid, and convenient method for exploitation in strain improvement screening programs. © 1994 John Wiley & Sons, Inc.

Key words: pyrolysis mass spectrometry • artificial neural networks • fermentor broths • regression analysis • chemometrics

---

* To whom all correspondence should be addressed.

† Present address: Department of Biological and Health Sciences, University of Westminster, 115 New Cavendish Street, London W1M 8JS, U.K.

INTRODUCTION

There is a large and growing interest in the screening of microbial cultures for the production of biologically active metabolites (e.g., refs. 9, 10, 35, and 36), which can provide structural templates for synthetic programs using rational methods of drug design. Modern screens for such metabolites are targeted on the modulation of particular biochemical steps in the disease process and show a high degree of both specificity and sensitivity. This sensitivity means that metabolites showing activity during screening need only be produced in very small amounts by the organism. In such cases, increasing the titer of the metabolite is vital to provide enough material for further biological evaluation and chemical characterization.

The process of titer improvement will frequently involve the search for overproducing mutants derived from the original producing organism (e.g., refs. 1, 3, 10, 27, and 48). Titer-improving mutants are rare, typically at frequencies around $10^{-4,37}$ and therefore, many thousands of mutants need to be screened in search of an overproducing strain. Previous methods of high-throughput mutant screening have relied on assessing antibiotic activity of the metabolites (e.g., ref. 4) or use rapid thin layer chromatography (e.g., ref. 43). Such methods were designed to accommodate 15,000 and 40,000 isolates, respectively, per month.

The ideal method for culture screening would have minimum sample preparation; would analyze samples directly (i.e., be reagentless); would give information about recognizable chemical characters; and would be rapid, automated, quantitative, and (at least relatively) cheap. Pyrolysis mass spectrometry (PyMS) is an automated, instrument-based technique for which, given an initial outlay of some £50,000 on the machinery, running costs are inexpensive, typically about £1 per sample. It is rapid (the typical sample time is less than 2 min) and is automated such that the throughput of samples in a working day may be 300 or more. We therefore considered that PyMS might be a very suitable method for exploitation in the rapid screening of microbial and other cultures.

Pyrolysis is the thermal degradation of complex molecules in a vacuum causing their cleavage to smaller, volatile fragments 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 analyzed. PyMS has been applied to the characterization and identification of a variety of microbial systems over a number of years[12,20,34,40] and, because of its high discriminatory ability,[13] presents a powerful fingerprinting technique, which is applicable to any organic material. Only rarely, however, has the method been used in quantitative (bio)chemical analysis.

The pyrolysis mass spectra of complex organic mixtures may be expressed in the simplest terms as subpatterns of spectra describing the pure components of the mixtures and their relative concentrations.[47] This may not always be true because during pyrolysis, intermolecular reactions can take place in the pyrolysate,[41,44] leading to a lack of superposition of the spectral components and to a possible dependence of the mass spectrum on sample size. However, suitable (nonlinear) numerical methods could still be employed to measure the concentrations of biochemical components from pyrolysis mass spectra of complex mixtures.

Our own aims have therefore been to exploit PyMS for the *quantitative* analysis of the chemical constituents of microbial and other samples. To this end, we have sought to apply fully interconnected feedforward artificial neural networks (ANNs) (see refs. 11, 21, 26, 38, and 45 for excellent introductory surveys), and the multivariate linear regression techniques of partial least squares regression (PLS) and principal components regression (PCR) (see refs. 6 and 30–33 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 *E. coli* grown on media incorporating various amounts of tryptophan,[15] to estimate the amount of casamino acids in mixtures with glycogen,[14] and to deconvolute the pyrolysis mass spectra of complex biochemical and microbiological mixtures.[18] With regard to classifications and discriminations, we have also exploited the combination of PyMS and ANNs for the rapid and accurate assessment of the presence of lower-grade seed oils as adulterants in extra virgin oils,[16,17] and for the identification of strains *Propionibacterium* spp.[19] Chun et al.[8] have also used the combination of PyMS and ANNs for the discrimination of strains of *Streptomyces*.

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 optimize control of the fermentation process. Whereas on-line mass spectrometry (MS) has been used to analyze fermentation broths for flavones,[29] the majority of MS applications during fermentations have been for the analysis of gases and volatiles produced over the reactor,[25] or by employing a membrane inlet probe for volatile compounds dissolved in the broths (e.g., refs. 5, 23, and 28). It is obvious that more worthwhile information would be gained by measuring the nonvolatile components of fermentation broths. Indeed, Heinzle et al.[24] were able to characterize the states of fermentations using off-line by PyMS, and this technique was extended to on-line analysis.[39] The authors, however, were not very satisfied with their system and, although they have continued to use mass spectrometry for the analysis of volatiles produced using fermentation[22,25] the analysis of nonvolatiles by PyMS does not seem to have been investigated further.

The fermentor may be treated as a complex mixture of bacteria, growth medium, and product. Using this model as a starting point, we used PyMS to analyze a mixture of ampicillin added to suspensions of the bacteria *E. coli* or *S. aureus*, as representative of pharmaceutical product and microbial background, and exploited ANNs, PCR, and PLS to estimate the amount of ampicillin in "unknown" (i.e., unseen) spectra. We also evaluated the ability of the ANNs further to generalize by creating prediction models using pyrolysis mass spectra solely of mixtures of ampicillin in *E. coli* and interrogated with pyrolysis mass spectra from mixtures of ampicillin and *S. aureus*. We observed that the normalized spectra of ampicillin mixed in *E. coli* was not the same as the appropriately weighted sum of the individual normalized spectra of ampicillin plus *E. coli*. We therefore studied the dependence of the mass spectra on sample size.

Finally, in addition to the above fermentation model we also studied samples from real fermentations producing a molecule of commercial interest. The aim was to demonstrate that PyMS, with multivariate calibration and ANNs, can be used to predict rapidly the amount of determined in fermentor broths.

MATERIALS AND METHODS

Preparation of the Ampicillin Mixture in Bacteria

The bacteria used were *Staphylococcus aureus* NCTC6571 and *Escherichia coli* W3110.[15] These strains are ampicillin-sensitive, indicating that any spectral features observed are not due, for instance, to β-lactamase activity. Both strains were grown in 4 L liquid media (glucose [BDH], 10.0 g; peptone [LabM], 5.0 g; beef extract [LabM], 3.0 g; H_2O, 1 L) for 16 h at 37°C in a shaker. After growth, the cultures were harvested by centrifugation and washed in phosphate-buffered saline (PBS). The dry weight 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.

Pyrolysis Mass Spectrometry

Five-microliter aliquots of the bacterial suspensions were evenly applied onto iron–nickel foils. Prior to pyrolysis the samples were oven dried at 50°C for 30 min. Samples were run in triplicate.

The pyrolysis mass spectrometer used for the main part of this study was the Horizon Instruments PYMS-200X (Aberystwyth), as initially described by Aries et al.[2] The Horizon Instruments RAPyD-400 (Heathfield) was also used for one experiment. For full operational procedures see Goodacre and Kell[15] and Goodacre et al.[14,18] The sample tube carrying the foil was heated, prior to pyrolysis, at 100°C for 5 s (This was found to be extremely helpful in ensuring the reproducibility of the spectra obtained, and this was particularly true for high-molecular-weight fragments).

Curie-point pyrolysis was at 530°C for 3 s, with a temperature rise time of 0.5 s. Data were collected over the m/z range 51 to 200 and normalized to the total ion count.

Data Analysis

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, while the ordinate contains information on the ion count for any particular m/z value ranging from 51 to 200. Data were normalized as a percentage of total ion count to remove the influence of sample size per se.

The data (normalized as above but not weighted by their standard deviations) were then analyzed by principal components analysis (PCA) using the Unscrambler package
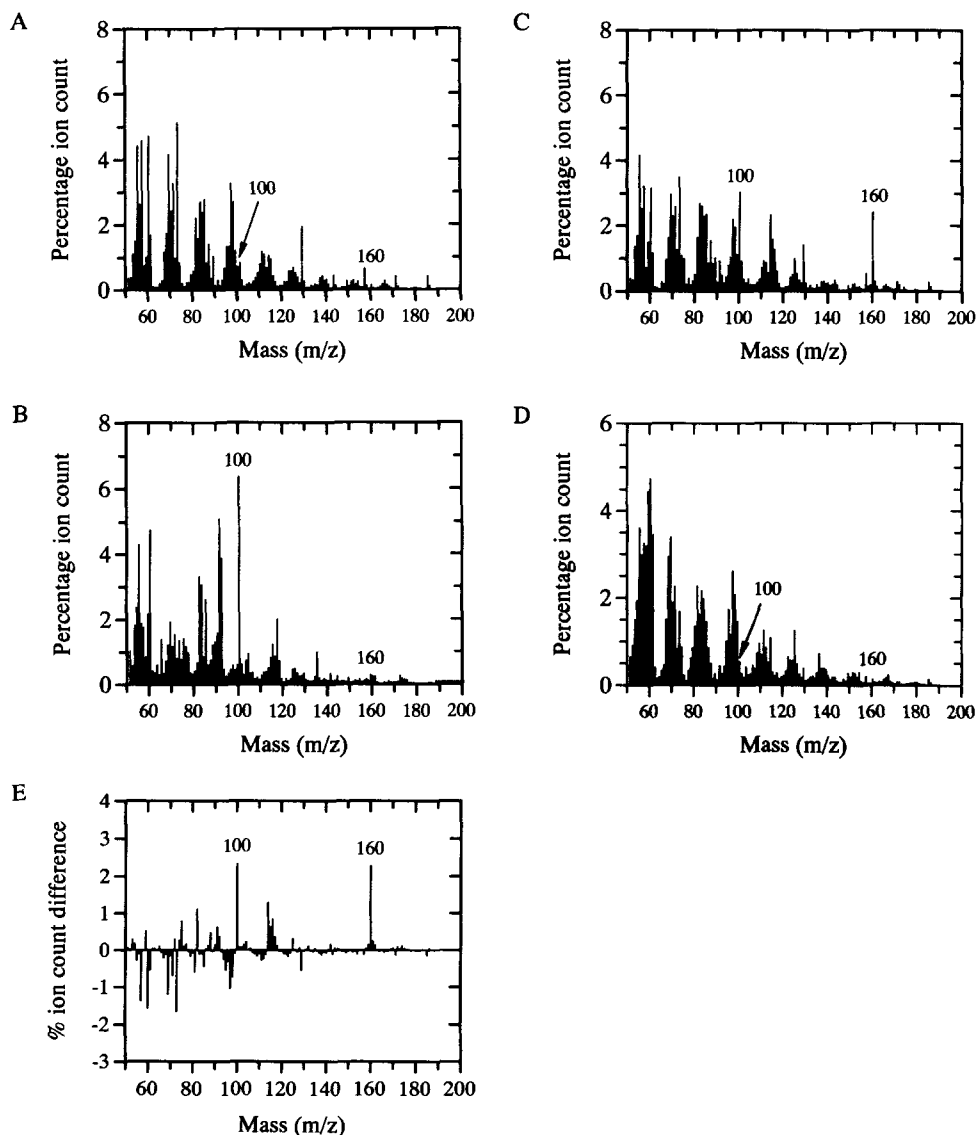

Figure 1. Normalized pyrolysis mass spectra of 200 μg *E. coli* (A), 50 μg ampicillin (B), of 5 μL of a sample comprising 2500 μg/mL ampicillin mixed in 40 mg/ml *E. coli* (C), and of 200 μg *S. aureus* (D). E is the subtraction spectrum of the normalized average of three pyrolysis mass spectra of *E. coli* (A) from the equivalent normalized average spectra from (C).

(vide infra and see ref. 31). PCA is a well-known technique for reducing the dimensionality of multivariate data while preserving most of the variances[7,20,31] and is an excellent technique for observing the *natural* relationships between samples.

For deconvolution of the pyrolysis mass spectra by ANNs, PLS, or PCR the training data (ANNs) or the *X*-variables (PLS and PCR) were the three normalized replicate pyrolysis mass spectra derived from the mixtures containing 0, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, and 5000 μg/mL ampicillin, with the output (ANN) or *Y*-variables (PLS and PCR) being the actual (true) concentration of ampicillin in the mixtures.

Artificial Neural Networks

All ANN analyses were carried out using a user-friendly, neural network simulation program, NeuralDesk (Version 1.2; Neural Computer Sciences, Lulworth Business Centre, Nutwood Way, Totton, Southampton, Hants SO1 0JR, U.K.), which runs under Microsoft Windows 3.1 on an IBM-compatible PC. For in-depth descriptions of the *modus operandi* the reader is referred to Goodacre and Kell[15] and Goodacre et al.[14,18]

The algorithm used was standard back-propagation (BP).[38,46] This algorithm employs processing nodes (neurons or units) connected using abstract interconnections (connections or synapses). Connections each have an associated real value, termed the weight, that scale signals passing through them. Nodes 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.

For the training of the ANN each input (i.e., normalized pyrolysis mass spectrum) is paired with a desired output (i.e., the amount of 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.

The structure of the ANN used in this study to analyze pyrolysis mass spectra therefore consisted of 3 layers containing 159 nodes made up of the 150 input nodes (normalized pyrolysis mass spectra), 1 output node (amount of determinand), and 1 "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, which in turn were connected to the output node. In addition, the hidden layer and output node were connected to the bias, making a total of 1217 connections, whose weights will be altered during training. Before training commenced, the values applied to the input nodes were normalized between 0.4 and 0.6 for each node,[18] and the connection weights were set to small random values.[45] The output layer was scaled to exploit less than the full range of the normalized scale between 0 and 1[14] and was scaled from −2500 to 7500. Each epoch represented 1217 connection weight updatings and a recalculation of the root mean squared (RMS) error between the true and desired outputs over the entire training set. A plot of the RMS error versus the number of epochs represents the "learning curve," and was used to estimate the extent of training. Finally, after training, all pyrolysis mass spectra of ampicillin in *E. coli* or *S. aureus* were used as the "unknown" inputs (test data); the network then output its estimate in terms of the concentration of ampicillin in the mixtures.

Principal Component Regression and Partial Least Squares

All PCR and PLS analyses were carried out using the program Unscrambler II (Version 4.0; (CAMO A/S, Olav Tryggvasonsgt. 24, N-7011 Trondheim, Norway) (and see ref. 31) which runs under Microsoft MS-DOS 6.2 on the 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 normalized triplicate pyrolysis mass spectra; and *Y*-, which represents the concentration of ampicillin in the bacterial slurry. The *Y*-matrix contains one *Y*-variable of 0 to 5000 μg/mL (in steps of 500 μg/mL) ampicillin mixed in 40 mg/mL *E. coli* or *S. aureus* (i.e., 33 objects representing 11 triplicate concentrations). 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 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.[31] Because 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 redetermined 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., to neither under- nor overfit predictions of unseen data).[31,42] 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, both for the known and unknown mass spectra, and plotted these RMS errors vs. the number of PCs or PLS factors used in predictions. We also generated plots of the RMS error of the test set versus the error in the training set to assist in determining the calibration model that best generalized. Using this approach, after calibration, to choose the optimal number of PCs or PLS factors to use in the prediction, all pyrolysis mass spectra of the mixtures (0 to 5000 µg/mL) were used as the "unknown" inputs (test data); the model then gave its prediction in terms of the concentration of ampicillin in the bacterial slurry.

RESULTS AND DISCUSSION

Pyrolysis mass spectral fingerprints of *E. coli*, ampicillin, ampicillin mixed with *E. coli*, and *S. aureus* are shown in Figure 1. These pyrolysis mass spectra are fairly complex, and there is quite a lot of difference between the spectra of *E. coli* (Fig. 1a) and of ampicillin (Fig. 1b), notably the very large peak at m/z 100 in the mass spectra of ampicillin. The spectrum of the mixture of ampicillin in *E. coli* (Fig. 1c) is rather similar to the spectrum of *E. coli* (Fig. 1a) alone, the exceptions being most noticeable at m/z 100 and m/z 160. The occurrence of a strong peak at m/z 160 is perhaps surprising, because this peak is not very intense in either the spectra of pure ampicillin (Fig. 1b) nor of *E. coli* (Fig. 1a).

Figure 1e shows a simple subtraction of the normalized averages of three spectra of *E. coli* from those of the above *E. coli*/ampicillin mixture. The positive half of the graph indicates the peaks that are more intense in the ampicillin spectra. The main similarity to the pyrolysis mass spectrum of pure ampicillin (Fig. 1b) is again the strength of the peak at m/z 100; although (in the difference spectrum) m/z 160 also seems to feature as a major component arising from ampicillin (alone), Figure 1b shows that this is not the case. Subtractions of the spectra of *S. aureus* (Fig. 1d), which did not contain strong peaks at m/z 100 or m/z 160, from those of *S. aureus*/ampicillin mixtures also suggested that m/z 100 and m/z 160 were the major components from ampicillin (data not shown).

It seems likely, therefore, that m/z 100 has arisen in the spectra of the mixture mainly because of its genesis in the pyrolysate of ampicillin. However, in contrast to this, m/z 160 is present in the spectra of the mixture but is not a major component in the spectra either of *E. coli* or of pure ampicillin; therefore, this peak must have resulted from intermolecular reactions which took place in the pyrolysate. One might predict, therefore, that this lack of superposition of the spectral features of the components in the mass spec-

trum of the mixture should depend on the size of sample analyzed, because larger sample sizes would be expected to promote greater intermolecular reactivities in the pyrolysate. (One might also comment that a more detailed analysis of these peaks might have been effected using tandem MS–MS; however, this facility was not available to us.)

We therefore prepared a mixture of 40 mg/mL *E. coli* containing 2500 µg/mL ampicillin and analyzed various sample sizes of this suspension using PyMS. The sample size varied from 42.5 to 850 µg, and Figure 2 shows the intensity of m/z 100 (open circles) and m/z 160 (closed circles) plotted against the sample size (Fig. 2). In addition, the total ion count (open squares) was also plotted against the same abscissa. It can be seen that relationship between the total ion count and sample size is linear from 42.5 to 400 µg after which the graph tends to level off. This is due to saturation effects on the intense peaks in the spectrum, which can be seen when the smaller peaks increase in size (data not shown), and is because the electron multiplier in the mass spectrometer is being overloaded with ions. The graph (Fig. 2) thus shows that provided the working range for the sample size is kept between 42.5 and 400 µg then saturation should not occur. Not surprisingly (given the arguments above) the percentage intensity of m/z 100 stays constant within this range; this is what one would predict for a peak that is derived from a single component in a mixture (here mainly ampicillin) and which does not undergo any pyrolysate–pyrolysate reactions. However, the percentage intensity of m/z 160 starts at about 1% and then increases approximately linearly to about 2.5% (at 212.5 µg sample size). When samples of greater than 200 µg were analyzed, the percentage intensity of m/z 160 was constant at around 2.5%. This result strongly suggests that not only is m/z 160 a result of intermolecular reactions (between fragments derived from both ampicillin and *E. coli*) within the pyrolysate
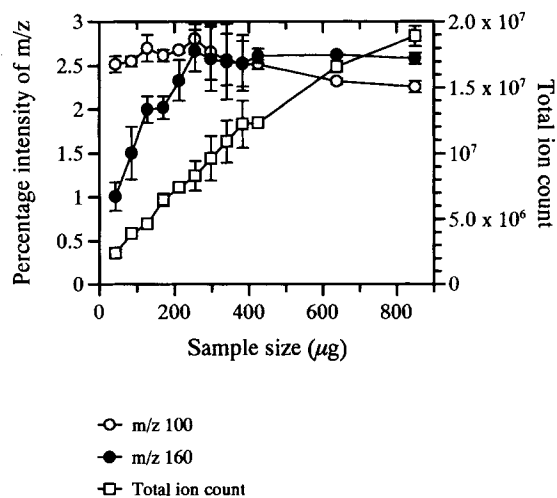

-O- m/z 100

-●- m/z 160

-□- Total ion count

Figure 2. Effect of sample size (micrograms) on the percentage intensity of m/z 100 (open circles), m/z 160 (closed circles), and the total ion count (open squares) in pyrolysis mass spectra of suspensions of 40 mg/mL *E. coli* containing 2500 µg/mL ampicillin. Error bars show the standard deviation.

but that the extent of this reaction *does* depend on the amount of the mixture analyzed (at least until the sample size is >200 μg, when the reaction has presumably gone to completion). Samples of ⩾200 μg were therefore used for the rest of this work.

If these masses can be considered characteristic for ampicillin, the intensities should alter linearly depending on the relative proportion of ampicillin mixed in *E. coli* (because the concentration of ampicillin is much lower than that of *E. coli* and any interreaction might be expected to be pseudo-first order with respect to the ampicillin concentration). A plot of the average intensities of the two masses, m/z 100 and 160, against the amount of ampicillin in the mixtures, with standard error bars and the best linear fits, is shown in Figure 3. It can be seen that m/z 100 and m/z 160 do indeed alter in a fashion that is approximately linear with the amount of ampicillin. Similar results were observed when ampicillin in *S. aureus* was examined in a similar way (data not shown).

One might presume that it could be possible simply to use the normalized intensities of these two peaks to estimate the relative amount of ampicillin in these mixtures. However, m/z 160 could not be used for this because the variation in intensity (i.e., noise) is very large (Fig. 3), while there are two problems associated with using m/z 100. The first is that there is some variation in the intensity of m/z 100; in the region above 1500 μg ampicillin the standard deviation error bars in fact overlap from one data point to the next. The other problem is that although the relationship between the percentage ion count with m/z 100 and the amount of ampicillin is (approximately) *linear*, it is not *proportional* (i.e., the line does not pass through the origin), which means that the source of m/z 100 is not purely from ampicillin, and there is some contribution from *E. coli* (as is also clear from the data in Fig. 1a).
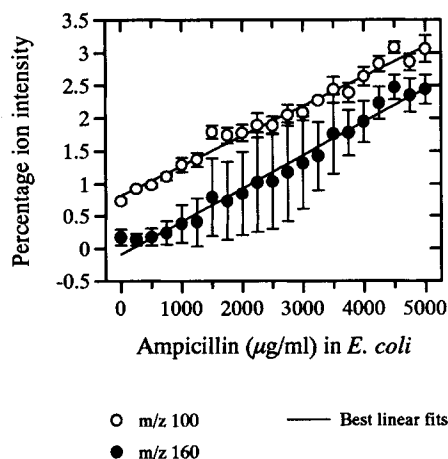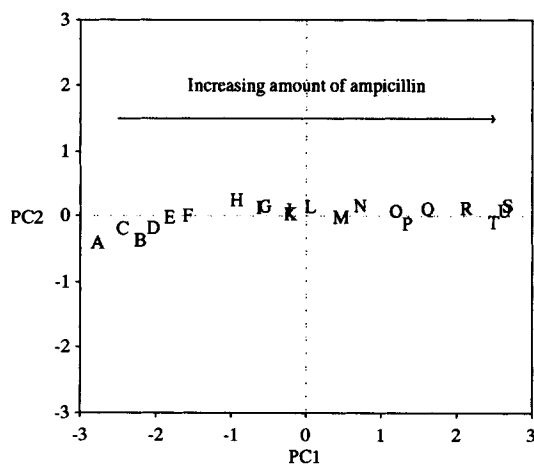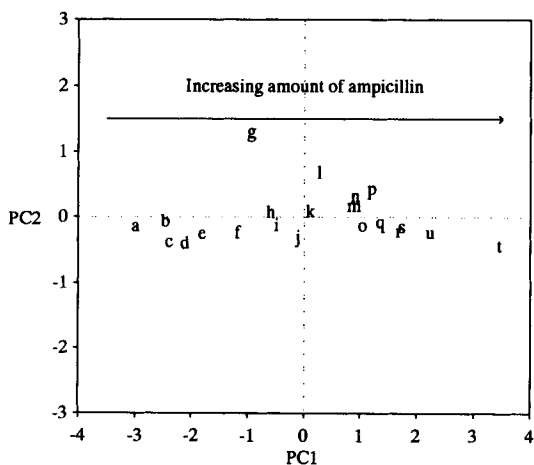
The next stage was to look at the relationship between the pyrolysis mass spectra of the various amounts of ampicillin mixed in the two bacteria using principal components analysis (PCA). PCA is the best method for reducing the dimensionality of multivariate data while preserving most of the variance; in our pyrolysis mass spectral data this reduction will be from the 150 m/z values to two or three principal components (PCs). Plots of the first two PCs of the variance in the PyMS for ampicillin in *E. coli* (Fig. 4a) and for ampicillin in *S. aureus* (Fig. 4b) show that the great bulk of the variation was indeed preserved in the first PC; this in fact accounted for 91.9% and 86.1% of the variance, respectively. It was also evident that the first PC largely served to account for or describe the difference in the amount of ampicillin in the bacterial backgrounds. Both sets of mixtures were then analyzed by PCA and the resulting three-dimensional ordination plot is shown in Figure 4c. Not surprisingly, the first PC now described the difference between *E. coli* and *S. aureus* and accounted for 69.7% of the total variation, whereas the second PC, which accounted for 25.9% of the total variance, was related to the influence of ampicillin on the pyrolysis mass spectra. Finally, the same samples were reanalyzed with the pyrolysis mass spectra of pure ampicillin, and the PCA plot shown in Figure 4d. This plot demonstrates further that the spectrum of pure ampicillin is only weakly related to the spectra of increasing amounts of ampicillin in *E. coli* and *S. aureus*, and that the addition of ampicillin to a cellular background changes the mass spectra in a *nonlinear* fashion; this is due to the intermolecular reactions that had occurred in the pyrolysate of ampicillin mixed with *E. coli* and *S. aureus*.

We therefore trained ANNs, using the standard backpropagation algorithm, with normalized ion intensities from the pyrolysis mass spectra from 0, 500, 1000, up to 4500 and 5000 μg/mL ampicillin in 40 mg *E. coli* as inputs and the stated concentrations of ampicillin as outputs, the latter being scaled between −2500 and 7500. The effectiveness of training was expressed in terms of the RMS error between the actual and desired network outputs; this "learning curve" is shown in Figure 5a (open circles). The learning curve of the test data (closed circles) is also shown in Figure 5a; it can be seen that, whereas the learning curve of the training set continues to decrease during training, the test set's learning curve initially decreases for approximately 10^3 epochs and then increases. This indicates that the ANN was being overtrained, and it is important not to overtrain ANNs, because (by definition) the network will not generalize well.[15,21] This overtraining appears even more marked when the RMS error of the test set is plotted against the RMS error of the training set (Fig. 5b); the minimum RMS error in the test set was reached (1.14%) when the RMS error of the training set was 1.00% and optimal training had occurred. The ANN was then interrogated with the training and test sets and a plot of the network's estimate versus the true concentration of ampicillin mixed in *E. coli* (Fig. 6) gave a linear fit which was indistinguishable from the expected proportional fit (i.e., $y = x$). It was therefore evident
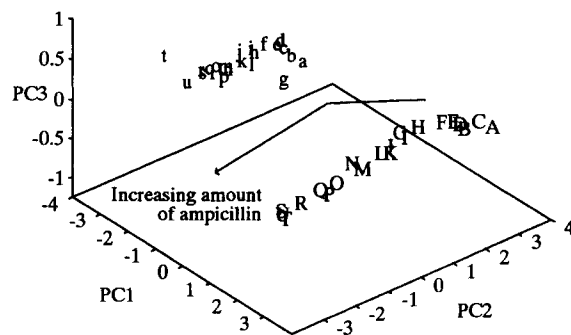

O m/z 100   ——— Best linear fits
● m/z 160

Figure 3. Effect of the concentration of ampicillin on the percentage intensity of m/z 100 and 160 in the pyrolysis mass spectra of mixtures of ampicillin in *E. coli*. Ampicillin solutions were added to give a final concentration of 0 to 5000 μg/mL (in steps of 250 μg/mL), in a constant cellular background of 40 mg/ml *E. coli*. Error bars show standard deviation. The best linear fits are shown.
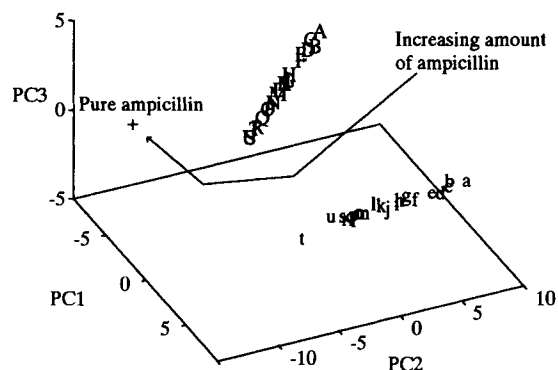
A

B

C

D

Figure 4. Principal components plots based on PyMS data showing the relationship between the ampicillin mixed in *E. coli* (A) or *S. aureus* (B). The first two principal components are displayed which for (A) account for 91.9% and 3.8% of the total variation, and for (B) 86.1% and 9.1%. (C) Ordination plot of the first three principal components (containing 69.7%, 25.9%, and 1.5% of the total variation) showing the relationships between ampicillin mixed in *E. coli* and *S. aureus*. Pure ampicillin was also analyzed by PCA with both of the mixtures (D); the first three latent variables are shown, which account for 57.8%, 29.6%, and 9.1% of the total variation. The upper case letters (A, B, C, . . . U) represent suspensions of 40 mg/ml *E. coli* containing ampicillin from 0 to 5000 μg/mL in steps of 250 μg/mL. In the same manner, the lower case letters symbolize the same concentrations of ampicillin mixed in 40 mg/mL *S. aureus*.

that the network's estimate of the quantity of ampicillin in the mixtures was very similar to the true quantity, both for spectra that were used as the training set (which is not surprising) and, most importantly, for the "unknown" pyrolysis mass spectra.

In other studies, ANNs were set up using the same architecture as the one used above except that they were trained on mixtures of ampicillin *S. aureus*, with the same concentrations of ampicillin as above. Networks were able to converge and optimal training occurred when the RMS error of the training set was 1.3%, the RMS error in the test set then being 3.35% (data not shown).

As outlined above, PCR and PLS were also used to create calibration models, using the same data that were used to train ANNs, to predict the amount of ampicillin (0 to 5000

μg/mL) mixed in 40 mg *E. coli* or *S. aureus*. Table I gives the percentage RMS error on the predictions produced by PCR and PLS on both the training and test sets for ampicillin mixed in *E. coli* and *S. aureus* and is compared with results from ANNs. The number of latent variables used to obtain optimal calibration models are also given, the values in brackets being the points at which Unscrambler stated that the optimal model should be formed, and one would have presumed that using more than three factors would cause overfitting[42] (i.e., inaccurate predictions on the test data). It is therefore perhaps surprising that optimal calibration always occurred using more latent variables, a phenomenon that has, however, been seen previously[18] and which usually implies that there are nonlinear relationships within the pyrolysis mass spectral data.[31] It can be seen

A

Number of epochs
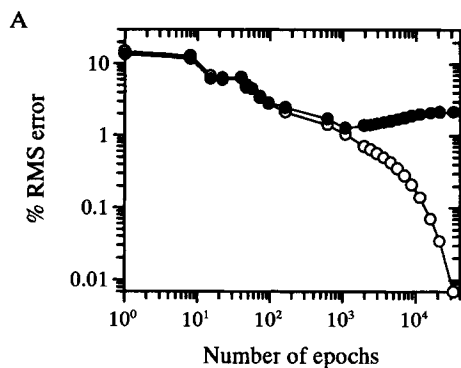
-O- Training set (data used to train ANNs)
-●- Test set (data not shown to the ANNs)

B
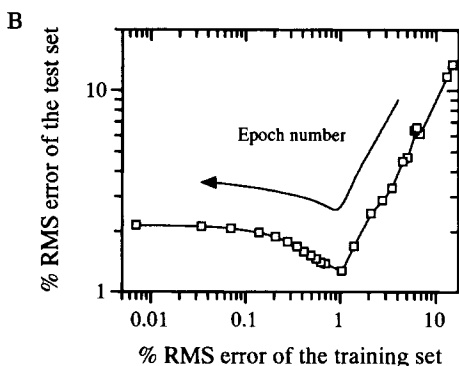

% RMS error of the training set

Figure 5. 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 ampicillin (μg) in *E. coli* (A). The open circles represent the percentage RMS error of the data used to train the neural network (the training 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 RMS error of the training set (B) shows that optimal training (to produce a network which generalized well) occurred at 1.0% RMS error; the number of epochs (and, hence, extent of training) increases from right to left.

(Table I) in all instances that the number of PCs used to give optimal calibration models in PCR was higher than the number of PLS factors needed, but that the percentage RMS error on prediction was approximately the same, typically between 1% and 1.9% for the training set and 1.3% to 3.5% for the "unknown" mass spectra in the test set. These values were very similar to the estimates obtained using ANNs and illustrates that all three methods could be used to provide an accurate deconvolution of the pyrolysis mass spectra of ampicillin mixed in *E. coli* or *S. aureus*.

It has previously been shown that to obtain good network generalization, the training set should normally consist of a reasonable number of samples equally spaced over the desired concentration range.[14] Sufficient generalization is needed because it is undesirable to have to collect new pyrolysis mass spectra of the product against a variable microbial background, e.g., different metabolic state or changes in strain populations. Therefore, the question arises as to whether ANNs trained to predict the amount of ampi-
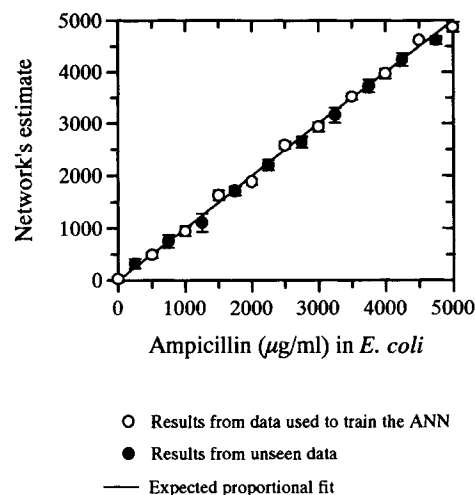

O Results from data used to train the ANN
● Results from unseen data
— Expected proportional fit

Figure 6. The estimates of trained 150-8-1 neural networks versus the true amount of ampicillin (0 to 5000 μg/mL in steps of 250 μg/mL) mixed in 40 mg/mL *E. coli*. Networks were trained using the standard back-propagation algorithm, to 1.0% RMS error (the point at which Fig. 5 indicated that optimal training took place). Data points are the averages of the triplicate pyrolysis mass spectra. Open circles represent spectra that were used to train the network and closed circles indicate "unknown" spectra which were not in the training set. Error bars show standard deviation. The expected proportional fit is shown.

cillin in a background of *E. coli* would be able to generalize sufficiently to estimate the amount of ampicillin against a *S. aureus* background.

Thus, ANNs were set up using the same architecture as the one used above, and trained on mixtures of ampicillin in *E. coli*. After training to 1.00% RMS error of the training set, the point at which optimal generalization occurred previously (Fig. 5 and 6), the ANNs were interrogated with the pyrolysis mass spectra of ampicillin mixed in 40 mg *S. aureus*. Table II gives the percentage RMS error on the predictions produced by ANNs on both the training and test sets for ampicillin mixed in *E. coli* and *S. aureus*. It can be seen that the RMS error in predicting ampicillin in *S. aureus* was only 5.16% and that the network had generalized to a rather remarkable degree.

ANNs were then trained on the 11 triplicate spectra of ampicillin in *E. coli* and the spectra of *S. aureus*. It was interesting to observe that using only this small piece of extra information the RMS error in estimating ampicillin in *S. aureus* had now reduced to 4.05%. When 5000 μg/mL ampicillin in *S. aureus* was also used in network training, generalization improved again and the error in the test set was now 3.27%, which was equivalent to ANNs trained on only examples of ampicillin in *S. aureus* (Table I). When all 11 triplicates of ampicillin in *E. coli* and *S. aureus* were used to train ANNs the RMS error of prediction of ampicillin in *S. aureus* was 2.26% and was much improved over using only ampicillin in *S. aureus*; however, this was at the expense of the network's ability to estimate ampicillin in *E. coli*, which had 2.17% RMS error in the test set compared with 1.14% observed previously. Nevertheless, it was evident that, for the deconvolution of the pyrolysis mass spec-

Table I. Comparison of artificial neural network calibration with principal components regression and partial least squares in the deconvolution of pyrolysis mass spectra from ampicillin mixed in either *E. coli* or *S. aureus*.

Calibration method	No. of epochs	No. latent variables	%RMS error Training set	%RMS error Test set
For the prediction of the amount of ampicillin mixed in *E. coli*				
ANNs	1140	—	1	1.14
PLS[a]	—	7 (2)	1.00 (2.34)	1.30 (3.90)
PCR[a]	—	10 (1)	1.35 (3.93)	1.49 (4.57)
For the prediction of the amount of ampicillin mixed in *S. aureus*				
ANNs	670	—	1.3	3.35
PLS	—	3 (2)	1.87 (2.16)	3.44 (3.46)
PCR	—	3[b]	2.03	3.18

[a]The values in brackets are the optimal number of latent variables for PCR or PLS calibration models as predicted by Unscrambler II, and the %RMS errors when predicted with models containing these numbers of latent variables.

[b]The optimal number of latent variables calculated was the same as those predicted by Unscrambler II.

tra of the determinand ampicillin, that ANNs are robust enough to generalize well to new biological backgrounds.

The use of the combined technique of PyMS and ANNs was then extended to the off-line analysis of fermentation broths. The identity of the producing organism and the structure of the metabolite of interest are confidential; the microorganism and the product are therefore coded M and P respectively.

Samples were taken aseptically from fermentations and frozen until they were analyzed by PyMS. The medium used to grow organism M was a complex medium containing mixed sugars and hydrolyzed protein, and samples were taken at different times. The amount of P was determined using high-performance liquid chromatography (HPLC). The error in these values was typically 2% to 5%.

Figure 7a shows a PCA plot of the pyrolysis mass spectra of the samples analyzed; it was again evident that the first PC, which accounted for 84.9% of the total variation, was largely related to the amount of P.

ANNs were then trained, using the standard back-propagation algorithm, with normalized ion intensities from the triplicate pyrolysis mass spectra from 0, 20.46, and 48.77 μg/mL P in M as inputs and the stated concentrations of P (as assessed by HPLC) as outputs, the latter being scaled between − 25 and 75. In these networks only three nodes were used in the hidden layer as an attempt at limiting the ANNs from learning the three examples perfectly (i.e., overfitting) and not generalizing. The effectiveness of training was expressed in terms of the RMS error between the actual and desired network outputs; to assess the extent of training, the RMS error of the test set was plotted against the RMS error of the training set (Fig. 7b). The minimum RMS error in the test set was reached (6.91%) when the RMS error of the training set was 3.00% and optimal training had occurred, which was after approximately 2.5×10^3 epochs. These percent RMS (%RMS) errors are higher than those achieved using ampicillin mixed in either *E. coli* or *S. aureus* and could be because of the lower numbers of ex-

Table II. Artificial neural network generalization. %RMS errors of test sets of ampicillin in *S. aureus* using networks trained on pyrolysis mass spectra of ampicillin in *E. coli* with increasing numbers of ampicillin in *S. aureus* examples in the training set.

Set	Composition of training set	No. of epochs[a]	Predictions of ampicillin in	%RMS error Training set	%RMS error Test set
A	Zero to 5000 μg/mL ampicillin (amp) (in 250-μg/mL increments) mixed in *E. coli*	1120	*E. coli*	1.00	1.14
			S. aureus	—	5.16
B	Set "A" and 0 μg/mL amp in *S. aureus*	1590	*E. coli*	1.03	1.23
			S. aureus	0.47	4.05
C	Set "A" and 0 and 5000 μg/mL amp in *S. aureus*	1620	*E. coli*	1.06	1.32
			S. aureus	0.52	3.27
D	Set "A" and 0, 2500, and 5000 μg/mL amp in *S. aureus*	2660	*E. coli*	1.11	1.39
			S. aureus	0.38	3.70
E	Zero to 5000 μg/mL amp (in 250-μg/mL increments) mixed in *E. coli* and *S. aureus*	3500	*E. coli*	1.18	2.17
			S. aureus	0.78	2.26

[a]ANNs were trained to 1% RMS error in the training set.
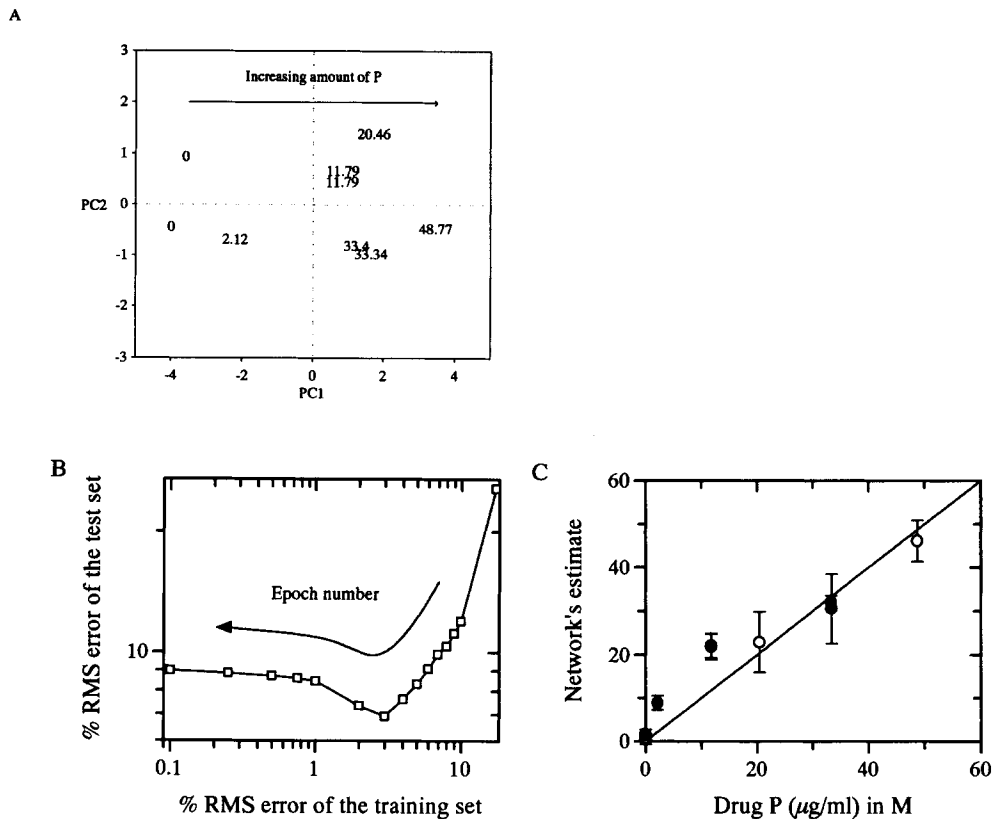
Figure 7. Relationship between the pyrolysis mass spectra of substance P mixed in organism M. (A) Principal components plots based on PyMS data analyzed by Unscrambler; the first two principal components account for 84.9% and 10.8% of the total variation. The amount of P in micrograms per milliliter is shown on the graph; this was established using HPLC. ANNs were trained using the standard back-propagation algorithm, employing one hidden layer consisting of eight nodes, to estimate the amount of P (μg/mL) in M. The generalization of the ANNs and the extent of training was elucidated by plotting the percentage RMS error of the test set versus the percentage RMS error of the training set (B); this shows that optimal training occurred at 3.00% RMS error. The number of epochs (and, hence, extent of training) increases from right to left. At this point the network was interrogated and the estimates of trained 150-8-1 neural networks versus the true amount of P (μg/mL) in M (as estimated by HPLC) were plotted (C). Data points are the averages of the triplicate pyrolysis mass spectra. Open circles represent spectra that were used to train the network and closed circles indicate "unknown" spectra which were not in the training set. Error bars show standard deviation. The expected proportional fit is shown.

emplars in the training set or, alternatively, due to addictive effect of inaccuracies in the HPLC analyses. The ANN was then interrogated with the training and test sets and a plot of the network's estimate versus the true concentration of P in M (Fig. 7c) gave an approximately linear fit. The network's estimate of the quantity of drug P in the fermentation broth was very similar to the true quantity, as judged by HPLC analysis, both for spectra that were used as the training set and, most importantly, for the "unknown" pyrolysis mass spectra.

The next stage was to assess the ability of PyMS and ANNs to indicate new strains producing drug P. If successful, then this approach could be used as a screening technique in addition to fermentation monitoring.

Mutants derived from organism M were grown under the same fermentation conditions as used previously. Samples were taken aseptically from the fermentations, treated to release the product from the biomass, and the supernatants frozen prior to analysis by PyMS. The amount of P was previously determined using HPLC. Because this experiment was to screen for overproducing strains, in addition to the above cultures, samples were also prepared by taking M

(or its mutants) and spiking with P. This would enable the neural network training to include these levels. The maximum amount of P produced in the last experiment was 48.77 μg/mL; the maximum amount of P spiked in M was 86.59 μg/mL.

ANNs were then trained, using the standard back-propagation algorithm, with normalized ion intensities from 17 triplicate pyrolysis mass spectra as inputs and the concentrations of P (as assessed by HPLC) as outputs, the latter being scaled between −40 and 120. These networks contained eight nodes in the hidden layer. The effectiveness of training was again expressed in terms of the RMS error between the actual and desired network outputs; to assess the extent of training, the RMS error of the test set (containing 22 samples) was plotted against the RMS error of the training set (data not shown). The minimum RMS error in the test set was reached (9.25%) when the RMS error of the training set was 5.00% and optimal training had occurred, which was after approximately 7 × 10³ epochs. The ANN was then interrogated with the training and test sets and a plot of the network's estimate versus the true concentration of P in mutants of M (Fig. 8) gave an approximately
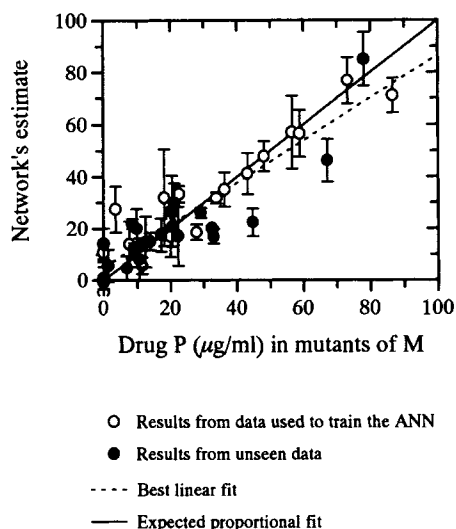
Figure 8. The estimates of trained 150-8-1 neural networks versus the amount of drug P (μg/mL) in a variety of mutants of M; the quantity of P was established using HPLC. Networks were trained using the standard back-propagation algorithm, to 5% RMS error; this was after approximately 7×10^3 epochs. Data points are the averages of the triplicate pyrolysis mass spectra. Open circles represent spectra that were used to train the network and closed circles indicate "unknown" spectra which were not in the training set. Error bars show standard deviation. The best linear fit is shown and the slope of this line is 0.82. The expected proportional fit is also shown.

linear fit, the slope of the linear regression line on the network's estimates was 0.82 and this intersected the ordinate at 5.1 μg/mL. It was therefore evident that the network's estimate of the quantity of drug P was very similar to the true quantity, as judged by HPLC, both for spectra that were used as the training set (open circles) and, most importantly, for the "unknown" pyrolysis mass spectra (closed circles). This demonstrates great promise for the application of PyMS and ANNs to the rapid screening of many cultures for the overproduction of specific metabolites.

CONCLUSIONS

We have shown that the combinations of PyMS with multivariate calibration (PLS and PCR) and ANNs were able quantitatively to analyze the PyMS of mixtures of ampicillin in *E. coli* or *S. aureus*. Further, ANNs trained to predict the amount of ampicillin in *E. coli* were able to generalize sufficiently to predict ampicillin in the completely different *S. aureus* background. This illustrates the robustness of ANNs to variations in the biological background.

It is also evident that the PyMS of complex organic mixtures can not necessarily be expressed solely in terms of linear superpositions of subpatterns of spectra describing the pure components of the mixtures and their relative concentrations as suggested by Windig and Meuzelaar.[47] During pyrolysis, intermolecular reactions do take place in the pyrolysate, as is particularly evident from the occurrence of

m/z 160 in the mass spectra of mixtures of ampicillin in *E. coli* and *S. aureus*. Such reactions lead to a lack of superposition of the spectral components and also to a dependence of the mass spectrum on sample size.

Fermentation broths could also be analyzed quantitatively for the determinand P using PyMS and ANNs, and this combination of techniques was used to screen many mutants (or new isolates) for the (over-)production of the metabolite of interest. One might also remark that the pyrolysis mass spectra contains information on the entire molecular constitution of the samples. Thus, if subsequent analyses showed the presence of a different "target" molecule in the isolates, one could train a small subset of the samples to learn the concentration of the "new" molecule, and use the rest of the entire data as the test set, without having to repeat the pyrolysis mass spectrometry.

PyMS is rapid (the typical sample time is less than 2 min) and automated; the present system allows 300 samples to be analyzed in 8 h 45 min. Thus, in a working day of two shifts (and allowing for 2 days down-time per month), one might expect to be able to analyze some 12,000 isolates per month. Furthermore, after the initial outlay of some £50,000 on the instrumentation and software, running costs are relatively cheap, typically about £1 per sample. We conclude that the combination of PyMS and ANNs constitutes a novel, rapid, and convenient method for exploitation in microbial fermentation development programs generally.

References

1. An, G. -H., Bielich, J., Auerbach, R., Johnson, E. A. 1991. Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting. Bio/Technology **9**: 70–73.
2. Aries, R. E., Gutteridge, C. S., Ottley, T. W. 1986. Evaluation of a low-cost, automated pyrolysis-mass spectrometer. J. Anal. Appl. Pyrol. **9**: 81–98.
3. Bailey, J. E. 1991. Toward a science of metabolic engineering. Science **252**: 1668–1675.
4. Ball, C., McGonagle, M. P. 1987. Development and evaluation of a potency index screen for detecting mutants of *Penicillium chrysogenum* having increased penicillin yield. J. Appl. Bacteriol. **45**: 67–74.
5. Bohatka, S., Langer, G., Szilagyi, J., Berecz, I. 1983. Gas concentration determination in fermentors with quadrupole mass spectrometer. Int. J. Mass Spect. **48**: 277–280.
6. Brereton, R. G. 1992. Multivariate pattern recognition in chemometrics. Elsevier, Amsterdam.
7. Chatfield, C., Collins, A. J. 1980. pp. 57–81. Introduction to multivariate analysis. Chapman & Hall, London.
8. Chun, J., Atalan, E., Ward, A. C., Goodfellow, M. 1993 Artificial neural network analysis of pyrolysis mass-spectrometric data in the identification of *Streptomyces* strains. FEMS Microbiol. Lett. **107**: 321–325.
9. Crittenden, P. D., Porter, N. 1991. Lichen-forming fungi—potential sources of novel metabolites. Trends Biotechnol. **9**: 409–414.
10. Crueger, W., Crueger, A. 1989. Biotechnology: a textbook of industrial microbiology. Sinauer Associates Inc., Sunderland, MA.

11. Gallant, S. I. 1993. Neural network learning. MIT Press, Cambridge, MA.

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

13. Goodacre, R. Berkeley, R. C. W. 1990. Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry. FEMS Microbiol. Lett. 71: 133–138.

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

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

16. Goodacre, R., Kell, D. B., Bianchi, G. 1992. Neural networks and olive oil. Nature 359: 594.

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

18. Goodacre, R., Neal, M. J. & Kell, D. B. 1994. 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.

19. Goodacre, R., Neal, M. J., Kell, D. B., Greenham, L. W., Noble, W. C., Harvey, R. G. 1994. Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. J. Appl. Bacteriol. 76: 124–134.

20. Gutteridge, C. S. 1987. Characterization of microorganisms by pyrolysis mass spectrometry. Meth. Microbiol. 19: 227–272.

21. Hecht-Nielsen, R. 1990. Neurocomputing. Addison-Wesley, Reading, MA.

22. Heinzle, E. 1992. Present and potential application of mass spectrometry for bioprocess research and control. J. Biotechnol. 25: 81–114.

23. Heinzle, E., Kramer, H., Dunn, I. J. 1985. State analysis of fermentation using a mass spectrometer with membrane probe. Biotechnol. Bioeng. 27: 238–246.

24. Heinzle, E., Kramer, H., Dunn, I. J. 1985. Analysis of biomass and metabolites using pyrolysis mass spectrometry. In: A. Johnson (ed.), Modelling and control of biotechnological processes. First IFAC Symposium. Pergamon, Oxford.

25. Heinzle, E., Oeggerli, A., Dettwiler, B. 1990. On-line fermentation gas-analysis—error analysis and application of mass spectrometry. Anal. Chim. Acta 238: 101–115.

26. Hertz, J., Krogh, A., Palmer, R. G. 1991. Introduction to the theory of neural computation. Addison-Wesley, Redwood City, CA.

27. Hutter, R. 1986. Overproduction of microbial metabolites, pp. 3–17. In: H. J. Rehm and G. Reed (eds.), Biotechnology, vol 4. VCH Publishers, New York.

28. Lauritsen, F. R., Nielsen, L. T., Degn, H., Lloyd, D., Bohatka, S. 1991. Identification of dissolved volatile metabolites in microbial cultures by membrane inlet tandem mass-spectrometry. Biol. Mass Spect. 20: 253–258.

29. Lee, M. S., Hook, D. J., Kerns, E. H., Volk, K. J., Rosenberg, I. E. 1993. Rapid screening of fermentation broths for flavones using tandem mass spectrometry. Biol. Mass Spect. 22: 84–88.

30. Malinowski, R. R. 1991. Factor analysis in chemistry. Wiley, New York.

31. Martens H., Næs, T. 1989. Multivariate calibration. Wiley, New York.

32. Martin, K. A. 1992. Recent advances in near-infrared reflectance spectroscopy. Appl. Spect. Rev. 27: 325–383.

33. McAvoy, T. J., Su, H. T., Wang, N. S., He, M. 1992. A comparison of neural networks and partial least squares for deconvoluting fluorescence spectra. Biotechnol. Bioeng. 40: 53–62.

34. Meuzelaar, H. L. C., Haverkamp, J., Hileman, F. D., 1982. Pyrolysis mass spectrometry of recent and fossil biomaterials. Elsevier, Amsterdam.

35. Omura, S. 1986. Philosophy of new drug discovery. Microb. Rev. 50: 259–279.

36. Porter, N., Fox, F. 1993. Diversity of microbial products and applications. Pestic. Sci. 39: 161–168.

37. Rowlands, R. T. 1984. Industrial strain improvement: mutagenesis and random screening procedures. Enz. Microb. Technol. 6: 3–10.

38. Rumelhart, D. E., McClelland, J. L., and the PDP Research Group. 1986. Parallel distributed processing. Experiments in the microstructure of cognition, vols. I and II. MIT Press, Cambridge, MA.

39. Sandmeier, E. P., Keller, J., Heinzle, Dunn, I. J., Bourne, J. R. 1988. Development of an on-line pyrolysis mass spectrometry system for the on-line analysis of fermentations, pp. 209–215. In: E. Hienzle and M. Reuss (eds.), Mass spectrometry in biotechnological process analysis and control. Plenum Press, New York.

40. Sanglier, J. J., Whitehead, D., Saddler, G. S., Ferguson, E. V., Goodfellow, M. 1992. Pyrolysis mass-spectrometry as a method for the classification, identification and selection of actinomycetes. Gene 115: 235–242.

41. Schulten, H. -R., Lattimer, R. P. 1984. Applications of mass-spectrometry to polymers. Mass Spec. Rev. 3: 231–315.

42. Seasholtz, M. B., Kowalski, B. 1993. The parsimony principle applied to multivariate calibration. Anal. Chim. Acta 277: 165–177.

43. Spagnoli, R., Cappelletti, L. 1981. New screening method for rapid identification of blocked mutants: application to erythromycin and oleandomycin producing strains. Agric. Biol. Chem. 45: 761–763.

44. Van de Meent, D., de Leeuw, J. W., Schenck, P. A., Windig, W., Haverkamp, J. 1982. Quantitative analysis of polymer mixtures by pyrolysis mass spectrometry/discriminant analysis. J. Anal. Appl. Pyrol. 4: 133–142.

45. Wasserman, P. D. 1989. Neural computing: theory and practice. Van Nostrand Reinhold, New York.

46. Werbos, P. J. 1974. Beyond regression: new tools for prediction and analysis in the behavioral sciences, Masters thesis, Harvard University, Cambridge, MA, USA.

47. Windig W., Meuzelaar, H. L. C. 1984. Non-supervised numerical components extraction from pyrolysis mass spectra of complex mixtures. Anal. Chem. 56: 2297–2303.

48. Wittrup, K. D., Bailey, J. E. 1988. A single-cell assay of β-galactosidase activity in *Saccharomyces cerevisiae*. Cytometry 9: 394–404.