Analytica Chimica Acta, 279 (1993) 17–26
Elsevier Science Publishers B.V., Amsterdam

Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production

Royston Goodacre and Douglas B. Kell

Department of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA (UK)

(Received 27th September 1992; revised manuscript received 1st December 1992)

Abstract

In pure form indole, when subjected to pyrolysis mass spectrometry (PyMS), gave a pattern of peaks at m/z 117, 90, 89 and a murmur at 63. Significant differences in the magnitudes of these peaks were observed between strains of *Escherichia coli* which were grown on nutrient agar and which differed solely in whether a transposon had been inserted into the tryptophanase gene or elsewhere within the genome. We applied artificial neural networks (ANNs) to the deconvolution of pyrolysis mass spectra. The combination of ANNs and PyMS was able quantitatively to detect the component indole when a single strain of *E. coli*, containing the tryptophanase gene, was grown on a minimal supplemented salts medium incorporating various amount of tryptophan, in the range 0–253 mg/l. This approach constitutes a novel, powerful and interesting technology for the analysis of the concentrations of appropriate substrates, metabolites and products in chemical and bioprocesses generally.

Keywords: Mass spectrometry; Artificial neural networks; Indoles; Neural networks; Pyrolysis MS

Pyrolysis is the thermal degradation of complex material in an inert atmosphere or a vacuum. It causes molecules to cleave at their weakest points to produce smaller, volatile fragments called pyrolysate [1,2]. 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) so as to produce a pyrolysis mass spectrum, which can then be used as a "chemical profile" or fingerprint of the complex material analysed.

Within microbiology, this technique, called pyrolysis mass spectrometry (PyMS), has largely been applied to the characterisation of *bacterial* systems (for reviews see Refs. 3–5). In particular, PyMS, because of its high discriminatory ability, has been successfully applied to the inter-strain comparison of a wide range of bacterial species and groups, including: *Bacillus* [6], *Corynebacterium* [3], *Escherichia coli* [7,8], *Legionella* [9], mycobacteria [10–12], salmonellae [13] and streptococci [14], highlighting the usefulness of this technique in the detection of small differences between microbial samples. Furthermore, one of the major advantages that PyMS has over other diagnostic methods, such as ELISA [15] and nucleic acid probing [16], is that it is rapid, both for a single sample and with respect to the (automated) throughput of samples. Typical sample time is less than 2 min.

PyMS of complex organic mixtures can be expressed in subpatterns of spectra describing the pure components of the mixtures and their relative concentrations [17]; here the authors success-

Correspondence to: R. Goodacre, Department of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA (UK).

fully used factor and discriminant analyses [18,19] to uncover the concentration of components (expressed in the form of "variance diagrams") from various sets of simulated mixtures (biopolymers, lignites and grass leaves). It is plausible that such an approach would be successful in estimating the concentrations of biochemical components from pyrolysis mass spectra of microorganisms (simply another form of complex mixture).

Chemometrics is the discipline concerned with the application of statistical and mathematical methods to chemical data, typically via the transformation of multivariate spectral inputs into the concentrations of target determinands [20,21]. A related approach is the use of (artificial) neural networks (ANNs), which are, by now, a well-known means of uncovering complex, nonlinear relationships in multivariate data. ANNs can be considered as collections of very simple "computational units" which can take a numerical input and transform it (usually via summation) into an output (see Refs. 22–27 for excellent introductions). The relevant principle of supervised learning in ANNs is that the ANNs take numerical inputs (the training data) and transform them into desired predetermined outputs. The input and output nodes may be connected to the "external world" and to other nodes within the network. The way in which each node transforms its input depends on the so-called "connection weights" (or "connection strength") and "bias" of the node, which are modifiable. The output of each node to another node or the external world then depends on both its weight strength and bias and on the weighted sum of all its inputs, which are then transformed by a (normally) nonlinear weighting function referred to as its activation function. For present purposes, the great power of neural networks stems from the fact that it is possible to "train" them. Training is effected by continually presenting the networks with the "known" inputs and outputs and modifying the connection weights between the individual nodes and the biases, typically according to some kind of back-propagation algorithm [28], until the output nodes of the network match the desired outputs to a stated degree of accuracy. The network, the effectiveness of whose training is usually determined in terms of the root mean square (RMS) error between the actual and the desired outputs averaged over the training set, may then be exposed to "unknown" inputs and will then immediately" output the globally optimal best fit to the outputs.

The reason this method is so attractive for the quantitative analysis of PyMS data is that it has been shown mathematically [29–31] that a neural network consisting of only one hidden layer, with an arbitrarily large number of nodes, can learn any, arbitrary (and hence nonlinear) mapping to an arbitrary degree of accuracy. ANNs are also considered to be robust to noisy data, such as those which may be generated by PyMS.

In this study the combination of PyMS and ANNs was evaluated for the possible use of these techniques quantitatively to analyse biological samples for the presence of unknown concentrations of determinands. Although ANNs have been applied to analyses for the presence of functional groups in the mass spectra of purified compounds [32], we believe this to be the first demonstration of the ability of ANNs quantitatively to analysis pyrolysis mass spectra in terms of the concentrations of target determinands.

EXPERIMENTAL

Bacterial strains

The three strains used in this study were *E. coli* W3110, a laboratory strain [33], ECO861 and ECO883. ECO861 and ECO883 have the common *E. coli* parent strain ECO80, a nalidixic acid-resistant mutant of a natural isolate from chicken litter, ECO70 [34], into which the transposon Tn1732 [35] was inserted [36]. In the case of ECO883, Tn1732 was inserted into the tryptophanase gene (tnaA), inactivating the gene, thus making this strain indole-negative. In ECO861, Tn1732 was inserted into another chromosomal region and the strain remained an indole producer. Strains were maintained on nutrient agar slopes (LabM) at 4°C.

Growth conditions

For the preliminary analyses of strains by PyMS cultures were grown on nutrient agar (LabM) for

16 h at 37°C. In the experiment investigating the production of indole, strains were grown for 16 h at 37°C on a minimal salts supplemented media [MSSM: K_2PO_4 (BDH), 7.0 g; KH_2PO_4 (BDH), 3.0 g; $(NH_4)_2SO_4$ (BDH), 1.0 g; sodium citrate (BDH), 0.5 g; $MgSO_4 \cdot 7H_2O$ (BDH), 0.25 g; casamino acids (Difco), 5.0 g; glucose (BDH), 4.0 g; agar (Oxoid), 16.5 g; H_2O, 1 l] with increasing amounts of DL-tryptophan (BDH) incorporated, ranging from 0 to 253 mg/l (0, 19, 38, 56, 74, 92,

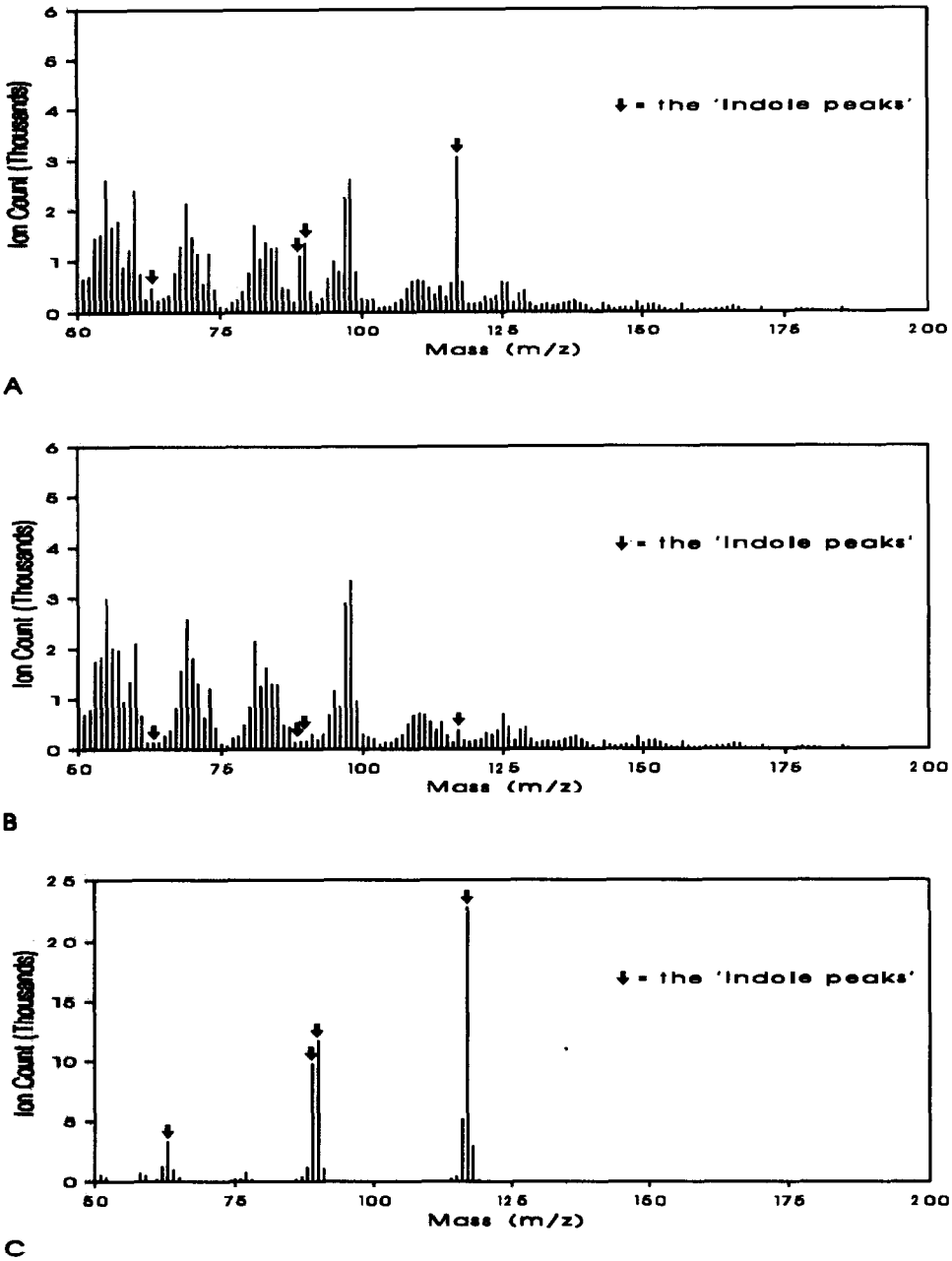


Fig. 1. Pyrolysis mass spectra of (A) *E. coli* ECO861, (B) *E. coli* ECO883, (C) and indole.

109, 127, 143, 160, 176, 192, 208, 223, 238 and 253 mg/l).

Sample preparation for pyrolysis mass spectrometry

Clean iron–nickel foils (Horizon Instruments, Heathfield) were inserted, using clean forceps, into clean pyrolysis tubes (Horizon Instruments), so that 6 mm was protruding from the mouth of the tube. After incubation, bacteria were picked up carefully from the top of one or more well-isolated colonies, avoiding the plate surface, by means of disposable plastic loops and smeared on 5 mm of a protruding foil to give a thin uniform surface coating. The samples were dried by vacuum desiccation for 20 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. Cultures to be analysed were grown in duplicate and two samples were prepared from each, giving four replicates for each culture. The samples were then analysed immediately. For the analysis of indole by PyMS: indole was dissolved in warm (50°C) distilled water, to an unknown concentration, and a 5 ml aliquot applied to a pyrolysis foil.

Pyrolysis mass spectrometry

The pyrolysis mass spectrometer used in this study was the Horizon Instruments PYMS-200X as described by Aries et al. [37]. The pyrolysate was generated in a vacuum by the heating of a ferro-magnetic foil carrying the sample. Heating was achieved by passing a radiofrequency current for 3 s through a pyrolysis coil which surrounds the sample-coated alloy foil. The foil and sample heated rapidly, 0.5 s, to the temperature corresponding to the Curie-point of the iron–nickel foil. At this temperature, 530°C, the alloy ceased to exhibit ferro-magnetic properties and heating finished; on cooling below the Curie-point, inductive heating resumed, so that the foil-pyrolyser system acted as a thermostatic switch maintaining the sample at the Curie-point, until current ceased to flow through the pyrolysis coil. 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.

The pyrolysate was bombarded with low energy electrons (25 eV) producing both molecular and fragment ions (because low energy was used the majority carried only a single positive charge). Non-ionised molecules were deposited on a cold trap, cooled by liquid nitrogen. The ionised fragments were focussed by the electrostatic lens of a set of source electrodes, accelerated and directed into a quadrupole mass filter. The ions were separated by the quadrupole, on the basis of their mass-to-charge ratio, and detected and amplified with an electron multiplier. The mass spectrometer scans the ionised pyrolysate 160 times at 0.2 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.

Data analysis

The data from PyMS may be displayed as quantitative pyrolysis mass spectra (Fig. 1). The x-axis represents the m/z ratio and the y-axis contains information on the ion count for any particular m/z value ranging from 51 to 200. Data were normalised to a total ion count of 2^{16}.

All ANN analyses were carried out using a user-friendly, neural network simulation program, NeuralDesk (Neural Computer Sciences, Southampton), which runs under Microsoft Windows/3.1 on an IBM-compatible PC. To ensure maximum speed, an accelerator board for the PC (NeuSprint) based on the AT&T DSP32C chip, which effects a speed enhancement of some 100-fold, permitting the analysis (and updating) of some 400 000 weights per second, was used. Data were also manipulated prior to analysis using the Microsoft Excel 4.0 spreadsheet.

For training the ANN, the inputs were the

averages of the four normalised replicate pyrolysis mass spectra derived from *E. coli* W3110 grown on MSSM containing tryptophan at concentrations of 0, 38, 74, 109, 143, 176, 208, 238 and 253 mg/l, with the output nodes being the actual (true) initial tryptophan concentration.

The primary algorithm used was standard back-propagation (BP) [28], running on the accelerator board. As indicated above this algorithm employs processing nodes (neurons or units), connected using abstract interconnections (connections or synapses). The format (topology) of the network is that of a directed acyclic graph. 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 with a sigmoidal shape.

The training of the network consists of the preparation of a set of pairs of patterns where one half of the pair is input to the network and the other represents the known or expected response. The stimulus pattern is applied to the network, which is allowed to run until an output is produced at each output node. The differences between the actual output and that expected, taken over the entire set of patterns 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.

For any given network, set of weight values, and set of training patterns there exists an overall RMS error value. If one dimension in a multidimensional space is put aside for each weight, and one more for the RMS error, one can construct an error surface. The BP algorithm performs gradient descent on this error surface by modifying each weight in proportion to the gradient of the surface at its location. Two parameters, *learning rate* and *momentum*, control this process. Learning rate scales the size of the step down the error surface taken each iteration, and momentum acts like a low pass filter, smoothing out progress over small bumps in the error surface.

It is known that gradient descent can cause the network to get "stuck" in a depression in the error surface should such a depression exist. These are termed "local minima" [24,26]. However, it has been found empirically that these are seldom problematic for larger networks, since the chances of encountering a multidimensional depression that is bounded in every dimension are small.

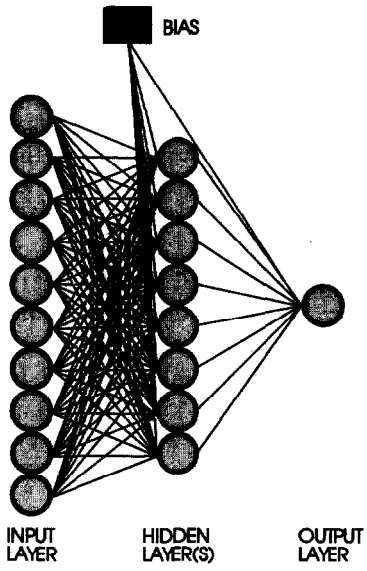


Fig. 2. A neural network consisting of 10 inputs (data herein actually consisted of 150 inputs/masses) and 1 output (tryptophan concentration) connected to each other by 1 hidden layer consisting of 8 nodes. In the architecture shown, adjacent layers of the network are fully interconnected, although other architectures are possible.

The structure of the ANN used in this study therefore consisted of 3 layers containing 159 nodes made up of the 150 input nodes (normalised pyrolysis mass spectra), 1 output node (initial tryptophan concentration), and one "hidden" layer containing 8 nodes. 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 (for a diagrammatic representation see Fig. 2). Before training commenced the input and output nodes were normalised between 0 and +1, and the connection weights were set to small random values, except the bias

which was always set to +1 [24]. Each epoch (one complete calculation in the network) 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 vs. the number of epochs represents the "learning curve", and was used to estimate the extent of training. Finally during training, all the 16 spectra (the averages of the four normalised replicate spectra) from *E. coli* W3110 grown on MSSM containing tryptophan (0–253 mg/l) (a mixture of seen and unseen data) were used as the "unknown" inputs (test data); the network then output its estimate (best fit) in terms of the initial tryptophan concentrations.

RESULTS AND DISCUSSION

Pyrolysis mass spectral fingerprints of *E. coli* ECO861, *E. coli* ECO883 and indole are shown in Fig. 1. When indole was analysed by PyMS, peaks at m/z 117, 90, 89 and a "murmur" at 63 can be seen (Fig. 1C). These we designate the "indole peaks". The analysis of an indole-producing strain of *E. coli*, ECO861, showed that these "indole peaks" were present in its pyrolysis mass spectrum (Fig. 1A), but were absent, or at least marginal, in spectra from ECO883, an indole-negative strain (Fig. 1B).

The significance of these changes is clear from Fig. 3 which shows a simple subtraction of the normalised averages of four spectra of ECO883 from ECO861. The positive half of the graph indicates the peaks that are more intense in ECO861 and shows many similarities to the pyrolysis mass spectrum of pure indole (Fig. 1C). ECO861 and ECO883 are genotypically very similar; they both arise from the same ECO80 parent strain and both have the Tn1732 transposon inserted, so they contain the same DNA. Genotypically they differ only in that ECO861 has an active tryptophanase gene (when this strain is cultivated on tryptophan-containing media it has the indole-positive phenotype), which in ECO883 has been inactivated by insertion of Tn1732, giving an indole-negative phenotype.

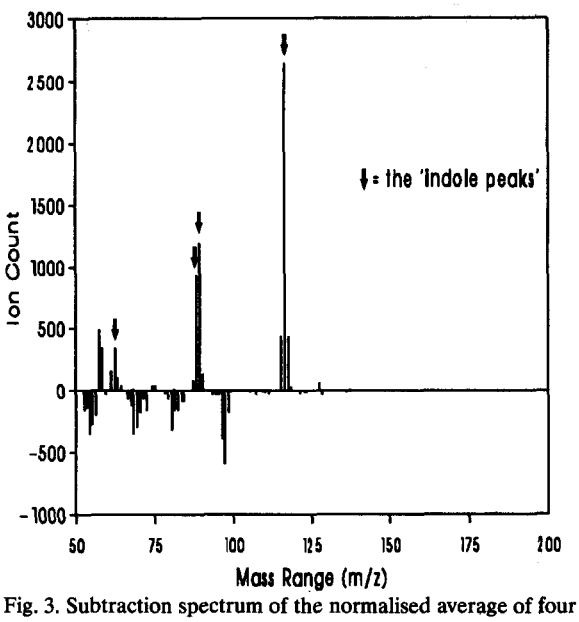


Fig. 3. Subtraction spectrum of the normalised average of four *E. coli* ECO883 pyrolysis mass spectra from the average of four *E. coli* ECO861 spectra, showing the masses that are more intense in *E. coli* ECO861 (positive half) and ECO883 (negative half).

This demonstrates that the presence of the single cellular component indole gives a large and clearly visible spectral change. It is plausible that indole gives such an easily detectable signal because it is preferentially vaporized on pyrolysis. Indole boils at 253°C *without* decomposition, a temperature which is lower than the Curie-point used (530°C). When indole was introduced by evaporation into the mass spectrometer without pyrolysis (a blank tube, with no foil, was loaded with 5 μl of indole solution, which will "sublime" when exposed to a vacuum) the same spectrum as that shown in Fig. 1C was observed (data not shown). Since the molecular weight of indole is 117.15, the peak at m/z 117 corresponds to the molecular ion. The peaks at m/z 90, 89 and 63 are likely to be produced by electron impact fragments, but the structures of these ions have not been elucidated.

The relationship between the relative amount of indole produced from *E. coli* ECO861, ECO883 and W3110, as an effect of altering the amount of tryptophan in MSSM from 0 to 253 mg/l is shown in Fig. 4. This is a plot of the

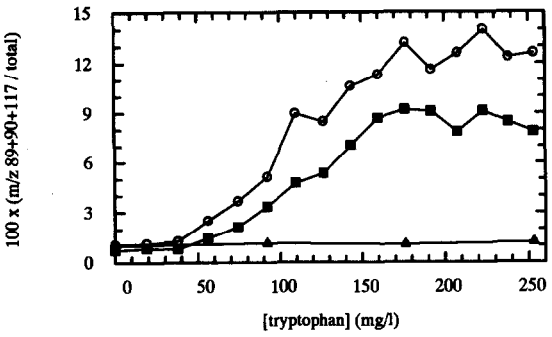


Fig. 4. Graph showing the percentage of the 3 major "indole peaks" (masses 117, 90 and 89) for *E. coli* ECO861, ECO883 and W3110 when grown on minimal media with increasing concentrations of tryptophan.

relative percentage of the indole peaks in the pyrolysis mass spectra, i.e., the ion counts at m/z 89 + 90 + 117 (63 was omitted because of its low intensity) over the total ion count against the true tryptophan concentration. It can be seen that increasing the amount of tryptophan in the medium has no effect on the indole-negative *E. coli* ECO883 strain, but in the stated representation (Fig. 4) gives an S-shaped curve with both *E. coli* ECO861 and W3110. In the region from 0 to 50 mg/l added tryptophan no indole was apparently produced, either because the tryptophanase gene had not yet been activated or (more likely) because the tryptophan was being used by the bacteria in biosynthetic processes. At greater values of added tryptophan the tryptophanase gene appears to be induced and indole was produced in an amount approximately linear with respect to the amount of tryptophan added, until at about 175 mg/l tryptophan the graph levelled off, showing that indole production had reached a maximum. This phenotypic change in *E. coli* ECO861 and W3110 can be attributed to the production of indole from tryptophan-containing medium, since these bacteria possess the tryptophanase gene. It is clear that alterations in the amount of tryptophan in the growth medium can cause measurable phenotypic changes in these bacteria.

As described above, the neural network was then trained with the various spectral inputs and the effectiveness of training determined in terms of the RMS error between the actual and the desired outputs; this "learning curve" is shown in Fig. 5. Training was effected five times; because the five curves were found to superimpose, despite the randomised starting connection weights, it is clear that training was executed in a rather reproducible manner. At various points during training, the network was interrogated both with spectra that were used to train the network (closed circles) and with "unknown" spectra (open circles) which were not in the training set; these are displayed in Fig. 6. Each plot consists of the five replicate trainings of Fig. 5 (although the initial random weightings on the connections in the network will have been different), shown as an average with standard error bars.

These experiments display some very interesting neurodynamics. It can be seen in the learning curve (Fig. 5) that the network very quickly reached a plateau after 100 epochs and between this time and some $2–3 \times 10^4$ epochs training appeared to have finished. When the network was interrogated in the middle of this plateau, a plot of the network's estimate vs. the true output (the initial concentrations of tryptophan) (Fig. 6A) gave a sigmoidal plot, and it was evident that although the network had made some sort of estimate of the tryptophan concentration, training was not yet finished, i.e. (Fig. 5), the net appeared to have found a very flat area in weight space. When the network was left to train a bit further the RMS error rather "suddenly" and reproducibly decreased (between 3×10^3 and $2 \times$

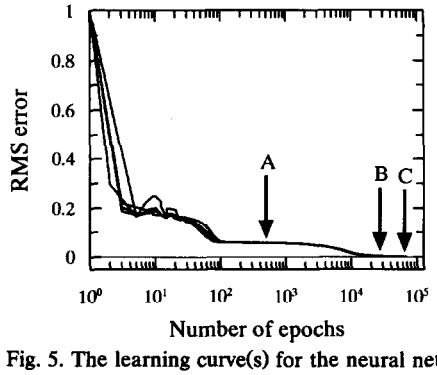
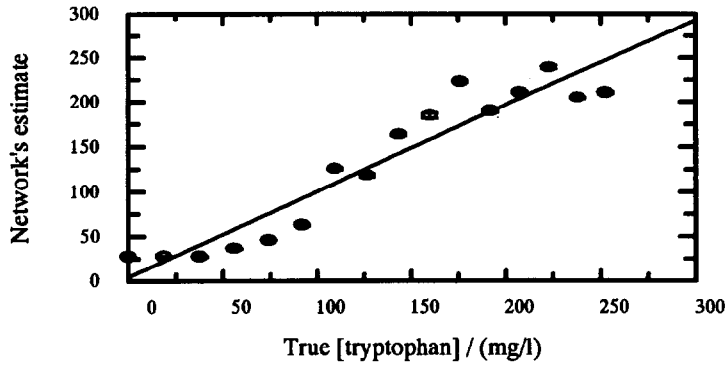
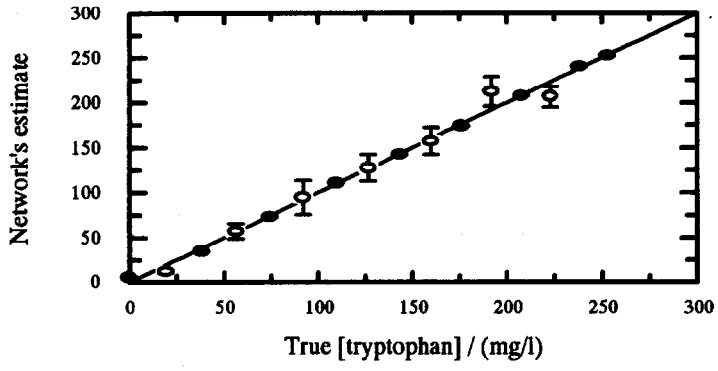


Fig. 5. The learning curve(s) for the neural network using the standard-back propagation algorithm with 1 hidden layer consisting of 8 nodes. For details see text.

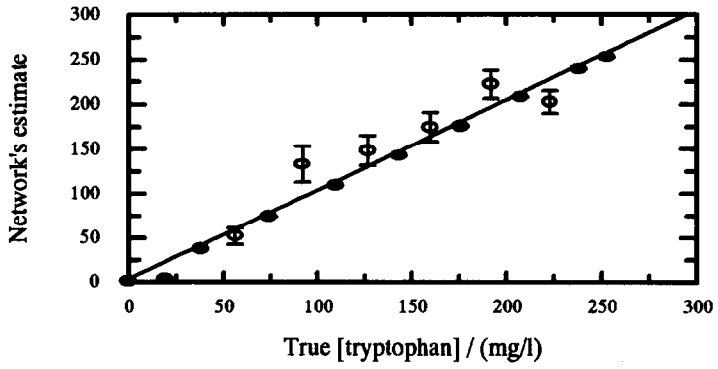
10^4 epochs) to approximately 0.005 and there was an approximately flat area until the error slowly diminished to 0.001 (Fig. 5). At an RMS error of 0.005 (Figs. 5 and 6B) the network's estimate of initial tryptophan concentration was very similar to the true concentrations, both for spectra that were used as the training set and the "unknown" spectra. If training was allowed to proceed further to an RMS error of 0.001, however, a different trend was seen: the network's estimate of the



A



B



C

Fig. 6. Results of the estimates of trained neural network against true initial tryptophan concentration, at interrogation points A, B and C from Fig. 5; points are the average of five trainings. Closed circles represent spectra that were used to train the network and open circles indicate "unknown" spectra which were not in the training set. Error bars show standard deviation. The best linear fits were calculated using all data.

training set is very good, whilst the "unknown" spectra (test set) are not nearly as well estimated. This indicated that the network had now been over-trained.

That optimal training was achieved when an RMS error of 0.005 was reached is most important for studies of the present type. For ANNs accurately to learn the concentrations of determinands in biological systems the network must be trained to the correct point (and must be trained with the appropriate number of standards [38]. Therefore it is imperative that ANNs should be trained several (perhaps many) times to ascertain whether they reproducibly converge, in concert with appropriate multivariate statistical techniques such as "leave-one-out" [39].

In other studies ANNs were set up with the same architecture as the ones used above except that they contained fewer, including no, hidden layers. It was interesting to observe that the networks were still able to converge, i.e. they were successfully trained (data not shown), indicating in the boundary case that the differences due to indole in these pyrolysis mass spectra were linearly separable in 150-dimensional space.

In summary, we have shown that the combination of PyMS and ANNs was able quantitatively to detect the component indole, which in pure form gave a pattern of peaks at m/z 117, 90, 89 and a murmur at 63, and was important in the detection of a single genotypic difference, attributed to tryptophanase, between *E. coli* ECO861 and ECO883. It was also demonstrated that varying the amount of tryptophan in a minimal supplemented salts medium gave a measurable phenotypic change in indole-positive strains. It should be obvious that this approach might be exploited, inter alia, for the analysis of any fermentation or biotransformation of interest, and that the combination of PyMS and ANNs constitutes a novel, powerful and interesting technology for the analysis of the concentrations of appropriate substrates, metabolites and products in chemical processes generally.

We thank Roger Berkeley for his help with PyMS, Mark Bale for providing the ECO strains, and Andy Edmonds for constructive criticism of the manuscript. This work is supported under the terms of the UK SERC LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, ICI Biological Products and Fine Chemicals, and Neural Computer Sciences.

Note added in proof

D. Lloyd et al. [40] have observed the tryptophan-enhanced production of indole by *Trichomonas vaginalis*, using membrane-inlet electron impact mass spectrometry and with a mass spectrum very similar to that shown in Fig. 1C.

REFERENCES

1 D.B. Drucker, Methods Microbiol., 9 (1976) 51.
2 W.J. Irwin, Analytical Pyrolysis: A Comprehensive Guide, Marcel Dekker, New York, 1982.
3 H.L.C. Meuzelaar, J. Haverkamp and F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials, Elsevier, Amsterdam, 1982.
4 C.S. Gutteridge, Methods Microbiol., 19 (1987) 227.
5 R.C.W. Berkeley, R. Goodacre, R.J. Helyer and T. Kelley, Lab. Pract., 39 (1990) 81.
6 L.A. Shute, C.S. Gutteridge, J.R. Norris and R.C.W. Berkeley, J. Gen. Microbiol., 130 (1984) 343.
7 G. Wieten, H.L.C. Meuzelaar and K. Haverkamp, in G. Odham, L. Larsson and P.A. Mårdh (Eds.), Gas Chromatography/Mass Spectrometry: Applications in Microbiology, Plenum Press, New York, 1984, pp. 335–380.
8 R. Goodacre and R.C.W. Berkeley, FEMS Microbiol. Lett., 71 (1990) 133.
9 R. Kajioka and P.W. Tang, J. Anal. Appl. Pyrol., 6 (1984) 59.
10 H.L.C. Meuzelaar, P.G. Kistemaker, W. Eshuis and H.W.B. Engel, Rapid Methods and Automation in Microbiology, Learned Information, Oxford, 1976, pp. 225–230.
11 G. Wieten, K. Haverkamp, H.B.W. Engel and L.G. Berwald, Rev. Infect. Diseases, 3 (1981) 871.
12 G. Wieten, K. Haverkamp, H.L.C. Meuzelaar, H.B.W. Engel and L.G. Berwald, J. Gen. Microbiol., 122 (1981) 109.
13 R. Freeman, M. Goodfellow, F.K. Gould, S.J. Hudson and N.F. Lightfoot, J. Med. Microbiol., 32 (1990) 283.
14 J.T. Magee, J.M. Hindmarch, I.A. Burnett and A. Pease, J. Med. Microbiol., 30 (1989) 273.
15 S.M. Chantler and M.B. McIllmurray, Methods Microbiol., 19 (1987) 273.
16 A. Saano and K. Lindström, Symbiosis, 8 (1990) 61.
17 W. Windig and H.L.C. Meuzelaar, Anal. Chem., 56 (1984) 2297.

18 N.H. Nie, C.H.G. Hull, J.G. Jenkins, K. Steinbrenner and W.H. Brent, Statistical Package for the Social Sciences, McGraw-Hill, New York, 1975.
19 W. Windig, P.G. Kistemaker and J. Haverkamp, J. Anal. Appl. Pyrol., 3 (1981) 199.
20 D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufmann, Chemometrics: A Textbook, Elsevier, Amsterdam, 1988.
21 S.D. Brown, Anal. Chem., 64 (1992) 22R.
22 P.D. Wasserman and R.M. Oetzel, NeuralSource: the Bibliographic Guide to Artificial Neural Networks, Van Nostrand Reinhold, New York, 1989.
23 J.L. McClelland and D.E. Rumelhart, Explorations in Parallel Distributed Processing; A Handbook of Models, Programs and Exercises, MIT Press, Cambridge, MA, 1988.
24 P.D. Wasserman, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.
25 R.C. Eberhart and R.W. Dobbins, Neural Network PC Tools, Academic Press, London, 1990.
26 P.K. Simpson, Artificial Neural Systems, Pergamon Press, Oxford, 1990.
27 J. Hertz, A. Krogh and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City, 1991.
28 D.E. Rumelhart, J.L. McClelland and the PDP Research Group, Parallel Distributed Processing, Experiments in the Microstructure of Cognition Vols. I & II. MIT Press, Cambridge, MA, 1986.
29 K. Hornik, M. Stinchcombe and H. White, Neural Networks, 2 (1989) 359.
30 K. Hornik, M. Stinchcombe and H. White, Neural Networks, 3 (1990) 551.
31 H. White, Neural Networks, 3 (1990) 535.
32 B. Curry and D.E. Rumelhart, MSnet: A Neural Network that Classifies Mass Spectra, Hewlett Packard Technical Report HPL-90–161, 1990.
33 B.J. Bachmann, Bacteriol. Rev., 36 (1972) 525.
34 M.J. Bale, P.M. Bennett, M. Hinton and J.E. Beringer, in J.C. Fry and M.J. Day (Eds.), Bacterial Genetics in Natural Environments, Chapman and Hall, London, 1990, pp. 231–239.
35 D. Ubben and R. Schmitt, Gene, 41 (1986) 145.
36 S.J. Bale, personal communication.
37 R.E. Aries, C.S. Gutteridge and T.W. Ottley, J. Anal. Appl. Pyrol., 9 (1986) 81.
38 R. Goodacre, A.N. Edmonds and D.B. Kell, J. Anal. Appl. Pyrol., (1993) in press.
39 E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991.
40 D. Lloyd, F.R. Lauritsen and H. Degn, J. Gen. Microbiol., 137 (1991) 1743.