ELSEVIER

Journal of Biotechnology 72 (1999) 157–167

JOURNAL OF Biotechnology

Rapid analysis of the expression of heterologous proteins in *Escherichia coli* using pyrolysis mass spectrometry and Fourier transform infrared spectroscopy with chemometrics: application to α2-interferon production

Aoife C. McGovern [a], Richard Ernill [b], Bo V. Kara [b], Douglas B. Kell [a], Royston Goodacre [a,*]

[a] *Institute of Biological Sciences, Cledwyn Building, University of Wales, Aberystwyth, Ceredigion, SY23 3DD, Wales, UK*
[b] *ZENECA Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK*

Received 20 August 1998; received in revised form 28 January 1999; accepted 12 April 1999

Abstract

Cell pastes and supernatant *Escherichia coli* samples, taken from an industrial bioprocess overproducing recombinant α2 IFN were analysed using pyrolysis mass spectrometry (PyMS) and diffuse reflectance-absorbance Fourier transform infrared spectroscopy (FT-IR). PyMS and FT-IR are physico-chemical methods which measure predominantly the bond strengths of molecules and the vibrations of bonds within functional groups, respectively. They therefore give quantitative information about the total biochemical composition of the bioprocess sample. The interpretation of these hyperspectral data, in terms of the quantity of α2 IFN in the cell pastes and supernatant samples was possible only after the application of the 'supervised learning' methods of artificial neural networks (ANNs) and partial least squares (PLS) regression. Both PyMS and FT-IR are novel, rapid and economical methods for the screening and the quantitative analysis of complex biological bioprocess over producing recombinant proteins. Models established using either spectral data set had a similarly satisfactory predictive ability. This shows that whole-reaction mixture spectral methods, which measure all molecules simultaneously, do contain enough information to allow their quantification when the entire spectra are used as the inputs to methods based on supervised learning. Moreover, this is the first study where FT-IR in the mid-IR range has been used to quantify the expression of a heterologous protein directly from fermentation broths and the first study to compare the abilities of PyMS and FT-IR for the quantitative analyses of an industrial bioprocess.

Keywords: PyMS; FT-IR; Bioprocess; Interferon; Heterologous proteins; Chemometrics

1. Introduction

Bioprocess control strategies are dependent on the observability of the culture; that is to say, the

* Corresponding author. Tel.: +44-1970-621-947; fax: +44-1970-622-354.
E-mail address: rrg@aber.ac.uk (R. Goodacre)

0168-1656/99/$ - see front matter
PII: S0168-1656(99)00128-5

acquisition and exploitation of signals (such as pH, O_2, CO_2, and temperature), even under complex environmental conditions such as those found in bioprocess reaction mixtures (Locher et al., 1992). The power of spectroscopic methods to analyse complex biological samples has been realised in the research laboratory (Marquardt et al., 1993; Hazen et al., 1994; Goodacre and Kell, 1996) and there is now an increasing interest in acquiring evidence regarding the applicability of these powerful analytical tools on an industrial scale. Although many physical parameters are measurable on-line, others such as the concentration of product(s) are difficult to determine and often require considerable pre-processing of the samples.

An ideal method for quantitative analysis of biologically important products in crude bioprocess samples would involve minimal sample preparation (i.e. be reagentless), would give information on recognisable chemical characters, would be rapid, quantitative, robust (Kell et al., 1990) and relatively inexpensive. Pyrolysis mass spectrometry (PyMS) and Fourier transform infrared spectroscopy (FT-IR) are valuable techniques that offer these advantages.

PyMS and FT-IR are physico-chemical methods which measure predominantly the bond strengths of molecules and the vibrations of bonds within functional groups, respectively (Meuzelaar et al., 1982; Griffiths and de Haseth, 1986; Schrader, 1995). Therefore they give quantitative information about the total biochemical composition of the bioprocess sample. However, the interpretation of these multidimensional spectra, or what are known as hyperspectral data (Goetz et al., 1985; Abousleman et al., 1994; Wilson et al., 1995), has conventionally been by the application of 'unsupervised' pattern recognition methods such as principal components (PCA), discriminant function (DFA) and hierarchical cluster (HCA) analyses. With 'unsupervised learning' methods of this sort the relevant multivariate algorithms seek 'clusters' in the data, thereby allowing the investigator to group objects together on the basis of their perceived closeness (Everitt, 1993); this process is often subjective because it relies upon the interpretation of complicated scatter plots and dendograms, which are not suitable for accurate quantitative interpretation.

More recently, various related but much more powerful methods, most often referred to within the framework of chemometrics (Massarrt and Buydens, 1987; Brown et al., 1996), have been applied to the 'supervised' analysis of these hyperspectral data (Goodacre et al., 1995); arguably the most significant of these is the application of 'intelligent' systems based on artificial neural networks (ANNs) (Collins, 1993; Zupan and Gasteiger, 1993; Widrow et al., 1994; Bishop, 1995; Goodacre et al., 1996a). PLS is another multivariate full spectrum method (Martens and Næs, 1989), which is based on inverse modelling, and is often used for estimation of the concentration of a determinand of interest in multicomponent mixtures (McAvoy et al., 1992; Bhandare et al., 1993; Song and Otto, 1995). Once calibrated these chemometric methods can effectively model the relationship between PyMS or FT-IR data and the concentrations of an unknown component(s), without the knowledge of the other constituents in the samples.

The combination of pyrolysis mass spectrometry (PyMS) and chemometrics has been shown to have the potential for the screening and quantitative analysis of microbial cultures (Goodacre et al., 1994a, 1995; Goodacre and Kell, 1996) including those that produce recombinant proteins such as cytochrome b_5 (Goodacre et al., 1994b). This technique is automated and rapid and typically permits the acquisition of 300 samples in a working day.

Diffuse reflectance-absorbance FT-IR is a more rapid, automated method which yields more detailed information about chemical structure than, for example, the rather slower UV absorbance spectroscopy typically used in HPLC analysis. The method can be employed simply after oven-drying bioprocess samples at 50°C (Goodacre et al., 1996b; Winson et al., 1997; Timmins et al., 1998). In particular because the sample presentation approach utilises the diffuse reflectance of a sample, held on an aluminium plate, it is very rapid (spectral acquisition is 10 s, and can be decreased by lowering the number of acquisitions co-added). Four hundred samples may be

analysed on a single plate, conveniently allowing in excess of 3000 samples per day to be collected. Indeed, we have shown previously (Winson et al., 1997) that this approach can be used accurately to determine the concentration of ampicillin added to *E. coli* and *Staphylococcus aureus*.

In the present study, we demonstrate the use of the PyMS and FT-IR techniques in combination with chemometrics for determining the accumulation of α2 IFN as the percentage of the total microbial protein (% TMP) in recombinant *E. coli* bioprocess samples. The bioprocess reaction mixtures were previously analysed by ZENECA Pharmaceuticals and α2 IFN levels determined by conventional methods.

2. Materials and methods

2.1. Model recombinant E. coli bioprocess

An *E. coli* K-12 derivative, transformed with a plasmid vector, expressing human interferon-α2 from the thermo-inducible λp_L promoter was cultured in 15 l of batch growth medium at 37°C, pH 6.7, 50% dOT (air saturation). The composition of the batch growth medium was (per litre): KH_2PO_4 3 g; $(NH_4)_2SO_4$ 10 g; NaCl 10 g; Na_2HPO_4 6 g; casein hydrolysate 2 g; glycerol 35 g; yeast extract 20 g; $MgSO_4 \cdot 7H_2O$ 0.5 g; $CaCl_2 \cdot H_2O$ 0.03 g; thiamine 0.008 g; $FeSO_4$/citric acid 40/20 mg; trace elements solution 0.5 ml; tetracycline 10 μg. A feed of yeast extract was supplied continuously (4.5 h post inoculation) to the bioreactor at 0.75 g l^{-1} per h. At the end of the batch phase (ca. 14 h, glycerol exhaustion), a fed-batch feed containing glycerol and ammonium sulphate (714 and 143 g l^{-1} respectively) was supplied to the bioreactor at a rate which maintained the oxygen uptake rate of the culture at maximum without exceeding the maximum oxygen transfer rate possible under the conditions described. At 15 h fermentation time, the bioreactor temperature was increased from 37 to 42°C to induce expression of the recombinant protein product.

UWA was provided with a supernatant sample (residual growth medium), and a cell-paste sample for each time point (h): 0, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 and 24. The corresponding values for accumulation of α2 IFN (as % TMP; total microbial protein) as calculated from Coomassie stained gels and protein estimations are shown in Table 1.

Table 1
Number of samples examined, indication of their age and quantity of α2 IFN present in cell pastes and in the bioprocess samples

Bioprocess time (h)	Accumulation of α2 IFN as % TMP in cell pastes	Accumulation of α2 IFN as % TMP in supernatant
0	Not detected	Not detected*
14	Not detected*	4.3 ± 1.0
15	Not detected	5.3 ± 0.4
16	3.8 ± 0.5*	5.5 ± 0.7
17	8.7 ± 0.6	6.1 ± 0.9*
18	12.1 ± 0.5*	6.0 ± 0.9
19	16.5 ± 0.7	5.7 ± 0.9*
20	16.9 ± 0.4*	5.5 ± 0.8
21	17.8 ± 0.2	4.8 ± 1.1*
22	17.3 ± 0.7*	5.5 ± 1.0
23	19.7 ± 0.1	4.6 ± 0.8*
24	20.7 ± 0.4*	5.2 ± 0.8

* Samples used in training sets for both PyMS and FT-IR.

2.2. 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. Five-microliter aliquots of cell pastes and supernatant were evenly applied onto the foils. The dry weights of the cells were used to adjust the weight of the final slurries with physiological saline to ≈ 25 mg ml^{-1}. Supernatant samples were applied unprocessed. Prior to pyrolysis, the samples were oven-dried at 50°C for 30 min, the foils were then 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 ≈ 1 mm from the mouth of each tube. Samples were run in triplicate.

Table 2
Comparison of PyMS and FT-IR, in combination with PLS regression and ANNs for the quantification of α2 IFN, in the cell pastes and the supernatant bioprocess samples

Sample type	Spectral method	RMS error in predictions	
		ANNs	PLS
Cell pastes	PyMS	2.27	1.78
	FT-IR	1.92	2.87
Supernatant	PyMS	0.53	0.56
	FT-IR	0.46	0.43

The pyrolysis mass spectrometer used for this study was a Horizon Instruments PYMS-200X. 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. Data were collected over the m/z range 51–200 and normalised as a percentage of total ion count.

2.3. Diffuse reflectance-absorbance FT-IR

Diffuse reflectance-absorbance FT-IR analysis was performed using a Bruker IFS28 infrared spectrometer equipped with a diffuse-reflectance TLC attachment (Bruker, Banner Lane, Coventry, UK) and a liquid N_2-cooled MCT (Mercury-Cadmium-Telluride) detector. Culture cell pastes and supernatant samples (5 μl; three replicates) were applied into the wells of a sand-blasted aluminium plate. After allowing the samples to oven dry at 60°C for 20 min, the plate was mounted on a motorised stage and infrared spectra were collected in the range of 4000–600 cm^{-1} with 256 co-adds, with a spectral resolution of 4 cm^{-1}. For chemometric processing, spectral data were converted into ASCII format, using the Opus software that controls the FT-IR instrument.

2.4. Discriminant function analysis

The initial stage involved the reduction of the dimensionality of the PyMS and FT-IR data by principal components analysis (PCA; (Jolliffe, 1986; Causton, 1987). PCA is a well known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance, and Matlab V.5 (Mathworks, Natick, MA) was employed to perform PCA according to the NIPALS algorithm (Wold, 1966). Discriminant function analysis (DFA) then discriminated between groups on the basis of the retained PCs and the a priori knowledge of which spectra were replicates (MacFie et al., 1978; Windig et al., 1983) and thus this process did not bias the analysis in any way. DFA was programmed according to Manly's principles (Manly, 1994). The objective of DFA is to maximise the ratio of the between-group to within-group variance (a group includes the spectral replicates), therefore a plot of the first two discriminant functions (DFs) displays the best 2-D representation of the *group* separation and can be used to observe any differences or similarities between the fermentation samples.

2.5. Chemometric processing using supervised learning

To gain quantitative information on α2 IFN production, ANNs and PLS models were established. Both were carried out using a user-friendly 'in-house' package developed by Jones et al. (1998).

ANNs and PLS are both well established chemometric techniques used for quantitative analyses of analytical chemical data. For a given analytical system there are some patterns (e.g. mass or IR spectra) which have desired responses which are known (i.e. the concentration of target determinands, α2 IFN). These two types of data form pairs which for the present purpose are called inputs and targets. The goal of *supervised* learning is to find a model or mapping that will correctly associate the inputs with the targets. Once the model has been formed it can then be challenged with new inputs (the mass or IR spectra) and will give its estimates of the concentration of α2 IFN (the target determinand).

The ANNs were trained by gradient descent using the standard back propagation (BP) algorithm (Rumelhart et al., 1986). Each input and

output variable was scaled between 0.2 and 0.8. The structure of the ANN used in this study to analyse pyrolysis mass spectra consisted of three layers containing 150 input nodes, one output node (amount of determinand) and one 'hidden' layer containing eight nodes (a 150-8-1 topology),

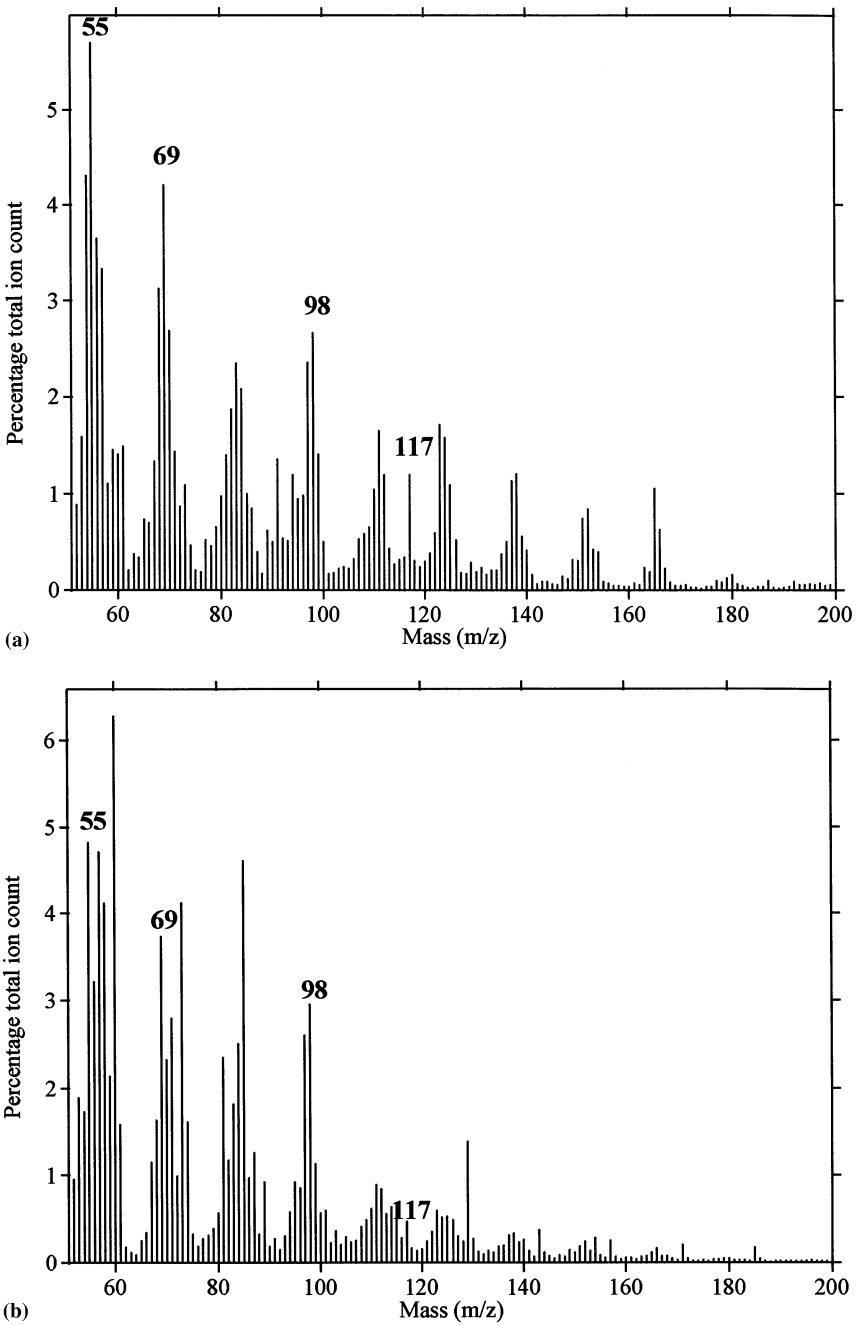

Fig. 1. Pyrolysis mass spectra of (a) α2 IFN, (b) cell paste alone, and (c) cell paste with 20.7 α2 IFN (as % TMP). These have been normalised so that the total ion count is 100%.

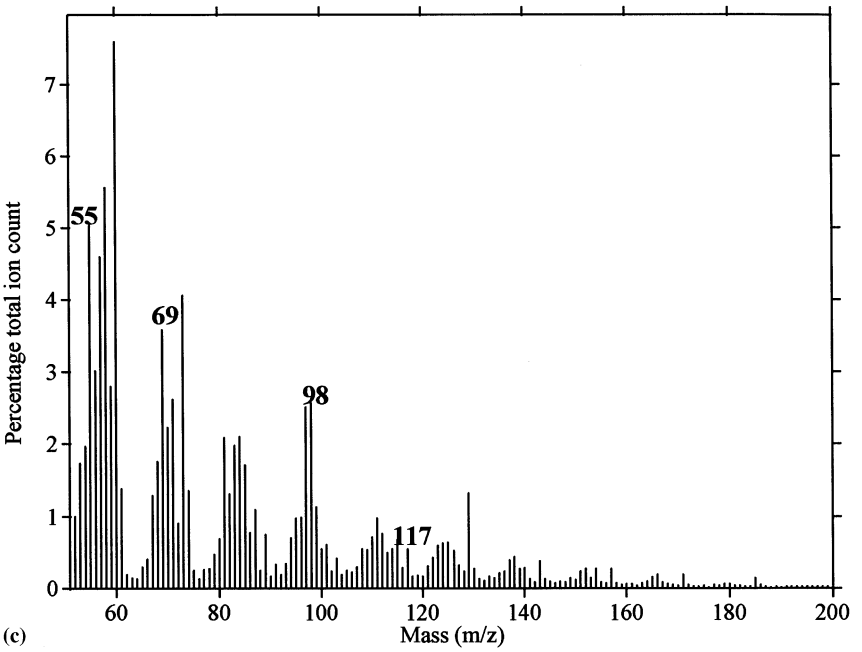

Fig. 1. (*Continued*)

whilst the architecture of the ANN for the FT-IR data was 883-10-1. The inputs and outputs of the PLS models were scaled to a standard deviation of one and mean centred (Martens and Næs, 1989).

During calibration of the ANNs and PLS, the root mean squared (RMS) error between the true and desired concentrations over the entire calibration model for the unknown spectra was calculated. The RMS errors vs. the number of epochs or factors used in predictions in the ANN and PLS models, respectively, were plotted. This allowed an indication of the optimal number of epochs or PLS factors to form the best general predictive model.

Training data for the construction of these models consisted of the PyMS and FT-IR data of every other sample from the bioprocess sample sets, whilst to avoid extrapolation, the extremes were always placed in the training set (cell pastes and supernatant were treated as independent sets; Table 2) together with the known quantity of α2 IFN present in those samples. After training, the remaining samples were used to test the calibrated models, and output their estimates in terms of the amount of α2 IFN present in the bioprocess samples.

3. Results and discussion

As can be seen from Table 1, a greater range of homologous protein production levels was observed in the cell pastes (% TMP range 3.8–20.7) compared to levels detected in the supernatant (% TMP range 4.3–6.1). Since the cell paste samples cover a wider range of total mass protein it is likely that their analysis will be most useful for assessing the use of the PyMS and FT-IR analytical techniques for the quantification of α2 IFN.

Pyrolysis mass spectra are commonly difficult to interpret and that of α2 IFN (Fig. 1a) was no exception; the supernatant (data not shown) and cell paste (Fig. 1b,c) spectra were equally complex. Mass 117, which is the molecular ion of indole, a breakdown product of tryptophan (Goodacre and Kell, 1993) and consequently a characteristic indicator of protein (Meuzelaar et

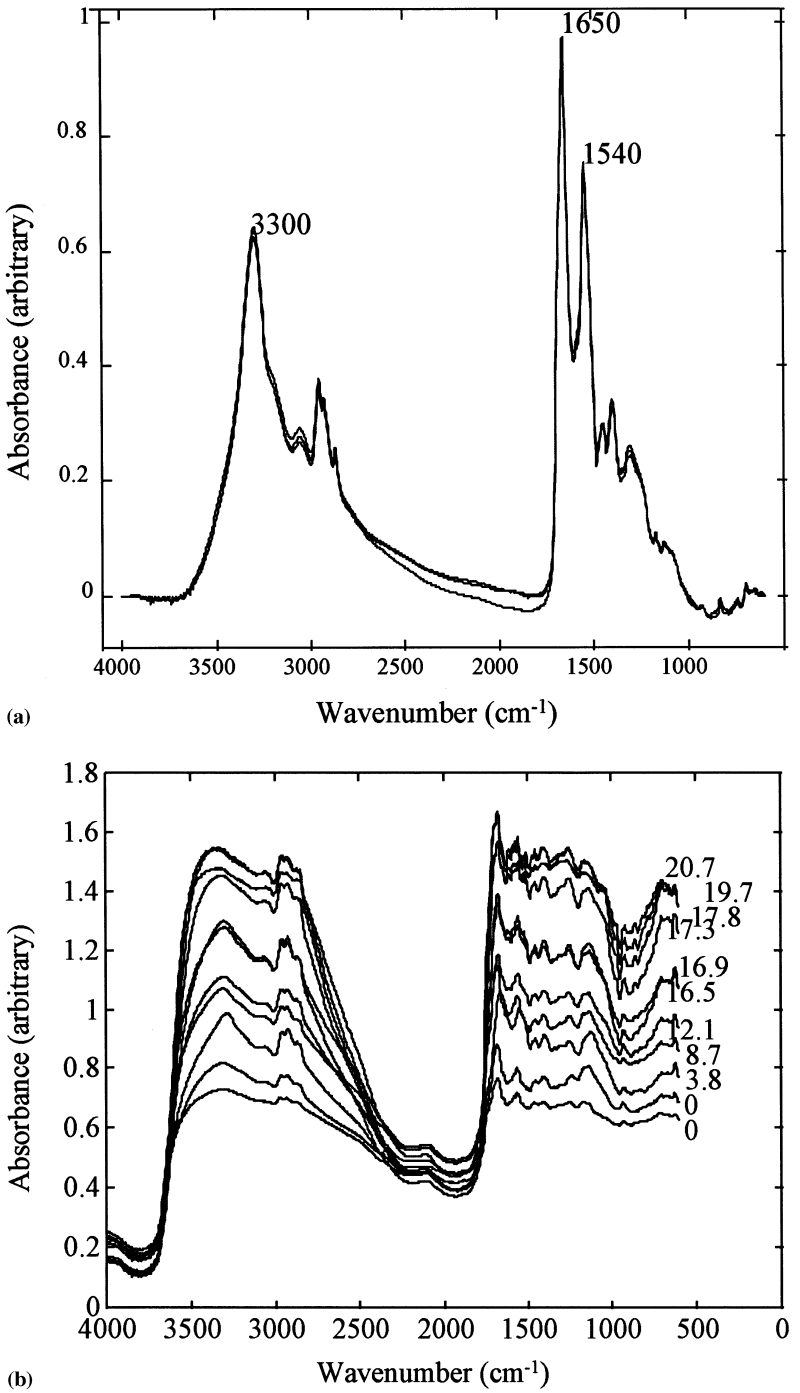

Fig. 2. Diffuse reflectance-absorbance FT-IR spectra (unmanipulated) of (a) α2 IFN (triplicates) and (b) cell pastes (average of triplicates) with different concentrations of α2 IFN.

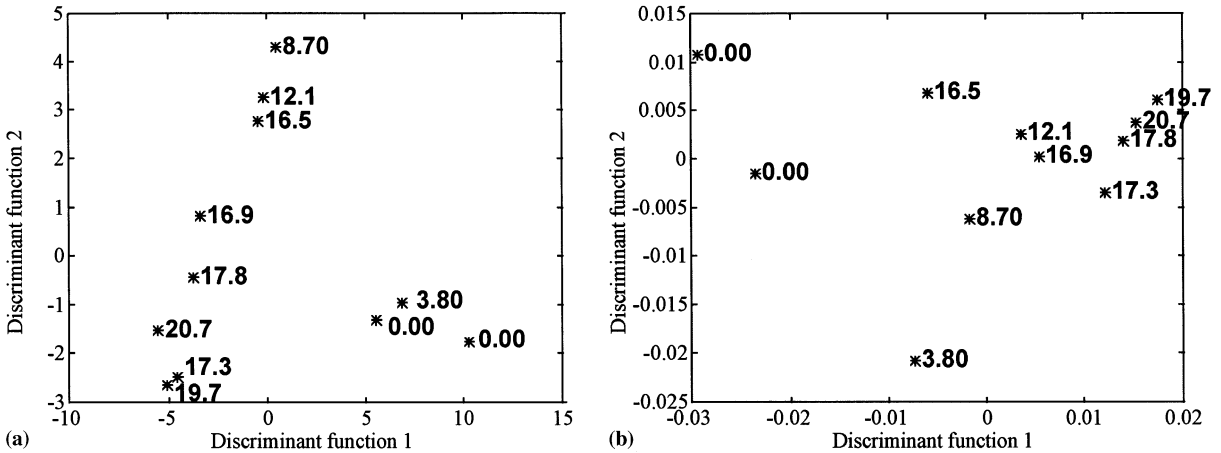

Fig. 3. Separation of cell pastes (a) PyMS data (b) FT-IR data by DFA. Samples are labelled quantitatively as detailed in Table 1.

al., 1982), was examined for increasing intensity over the period of increased expression of the recombinant protein of interest (data not shown). Although a roughly linear relationship was detected, there was a lot of scatter around the best fit line and this line did not pass through the origin. This is perhaps not surprising since α2 IFN is not the sole protein found in *E. coli*. No other single ion intensities were found to alter in a proportional or linear manner and thus these could not be used alone to estimate the quantities of the determinand.

The FT-IR spectra from these bioprocess samples were also complex (Fig. 2), and because of the multitude of cellular components, all with their own molecular vibrations capable of absorbing appropriate electromagnetic radiation, broad superimposed spectral bands were observed within the mid-infra range (4000–600 cm^{-1}). The infrared spectra of proteins exhibit strong amide I absorption bands at 1650 cm^{-1} associated with the characteristic stretching of C=O and C–N and the bending of the N–H bond (Stuart, 1996); however changes in the absorbance at 1650 cm^{-1} could not be used to quantify α2 IFN production.

Since both the PyMS and FT-IR spectra are complex and no single mass or wavenumber could be used to quantify the level of α2 IFN, it was therefore necessary to explore various chemometric approaches which use information from the *whole* spectrum.

The first stage in the chemometric process was to use the unsupervised analysis of discriminant function analysis (DFA) as detailed above. DFA on the pyrolysis mass spectra from the cell paste samples (Fig. 3a) separated the non-producers and the low producer (3.8 as % TMP) away from the other samples in the first discriminant function (DF); the second DF showed a trend which was correlated more with the age of the sample rather than the accumulation of α2 IFN. The same phenomenon was found on DFA of the supernatant samples (data not shown). DFA on the FT-IR spectra from the same cell pastes (Fig. 3b) showed a similar, but not so obvious, trend to that observed in Fig. 3a. It was evident that DFA alone would not be able to give accurate estimates of the accumulation of α2 IFN and additional chemometric methods which use supervised learning would need to be exploited.

To gain quantitative information on α2 IFN production, ANNs and PLS models were established using the PyMS and FT-IR data for both supernatant and cell pastes. After the ANN and PLS models were calibrated they were challenged with the training and test sets. Plots of the ANNs and PLS estimates versus the true amount of α2 IFN in the cell pastes (Fig. 4) gave linear fits (dashed lines) which were very close to the expected proportional fits (i.e. $y = x$, solid line). It was therefore evident that the neural network's and PLS estimates of the quantity of α2 IFN in

the cell pastes were very similar to the true quantity, both for spectra that were used as the training set and, *most importantly*, for the 'unknown' pyrolysis mass spectra (Fig. 4a,b) and FT-IR spectra (Fig. 4c,d). Table 2 gives details of the RMSEP for the test sets only for both the cell pastes and supernatant samples. It can be seen that these errors of prediction were satisfactorily low, highlighting that both PyMS and FT-IR could be used to give accurate estimates of α2 IFN levels in bioprocess reaction mixtures.

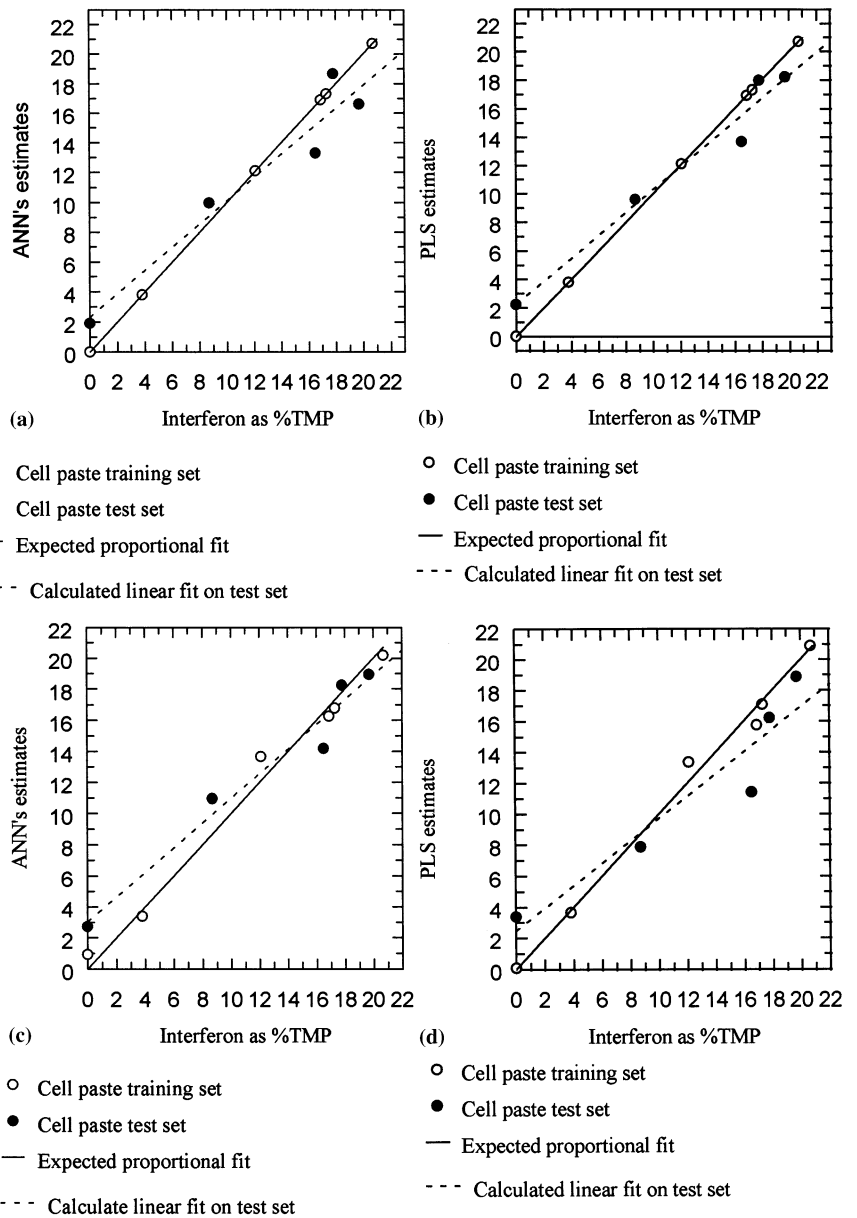

Fig. 4. Estimates of the levels of α2 IFN in cell pastes using PyMS data and ANNs (a), PyMS data and PLS regression (b), FT-IR data and ANNs (c) and FT-IR data and PLS regression (d).

4. Concluding remarks

In analysing complex fermentor reaction mixtures and supernatants, unlike simple mixtures (Goodacre et al., 1993) the multivariate analysis approach monitors not only the spectral contribution of the target molecule per se but may also take into account other features which are associated with the metabolite concentration. In this way the use of supervised methods such as PLS and ANNs can exploit differences in the organism or the supernatant which correlate with metabolite overproduction. This can provide a useful 'amplification' to the method when the target molecule concentrations are particularly low.

The object of these experiments has been to quantify production of the protein α2 IFN in crude bioprocess samples. The true test for spectroscopic techniques in combination with chemometrics for the quantitative analysis of fermentor reaction mixtures comes with real world problems, such as those tested in the experiment above. We consider that these approaches will prove to be valuable tools and the data accumulated will enable rapid and quantitative analysis of recombinant protein production in other hosts over-expressing a gene of interest.

The important conclusion to be drawn from the results so far, is that whole-reaction mixture spectral methods which measure *all molecules simultaneously*, do contain enough information to allow their quantification when the entire spectra are used as the inputs for modern chemometric methods, based on supervised learning.

In the future, we hope to form models for a much greater number of samples collected from a number of similar bioprocesses, varying only in the arbitrary experimental conditions. This should give an indication of the stability and robustness of these models to even more diverse data. Another possibility is to analyse the spectrum of pure determinands and to use this information (via appropriate variable selection methods) to 'weight' the spectra in favour of those masses most likely to contribute to a parsimonious model with which to obtain quantitative information.

In conclusion, these results demonstrate that modern analytical spectroscopies can provide rapid accurate quantitative estimates of the levels of heterologous proteins such as α2 IFN in *E. coli* bioprocesses, but only when combined with intelligent chemometric systems which perform supervised learning.

Acknowledgements

A.C. McGovern thanks ZENECA Pharmaceuticals for financial support, R. Goodacre is indebted to the Wellcome Trust for financial support (grant number 042615/Z/94/Z), and D.B. Kell thanks the Engineering and Biological Systems Directorate of the UK BBSRC for financial support.

References

Abousleman, G.P., Gifford, E., Hunt, B.R., 1994. Enhancement and compression techniques for hyperspectral data. Opt. Eng. 33, 2562–2571.

Bhandare, P., Mendelson, Y., Peura, R.A., Janatsch, G., Kruse-Jarres, J.D., Marbach, R., Heise, H.M., 1993. Multivariate determination of glucose in whole blood using partial least-squares and artificial neural networks based on mid-infrared spectroscopy. Appl. Spectrosc. 47, 1214–1221.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon, Oxford.

Brown, S.D., Sum, S.T., Despagne, F., Lavine, B.K., 1996. Chemometrics. Anal. Chem. 68, R21–R61.

Causton, D.R., 1987. A Biologist's Advanced Mathematics. Allen and Unwin, London.

Collins, M., 1993. Empiricism Strikes Back: Neural Networks in Biotechnology. Bio/Technology 11, 163–166.

Everitt, B.S., 1993. Cluster Analysis. Edward Arnold, London.

Goetz, A.F.H., Vane, G., Solomon, J., Rock, B.N., 1985. Imaging spectrometry for earth remote sensing. Science 228, 1147–1153.

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.

Goodacre, R., Kell, D.B., 1996. Pyrolysis mass spectrometry and its applications in biotechnology. Curr. Opin. Biotechnol. 7, 20–28.

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 amino acids in glycogen. J. Anal. Appl. Pyrol. 26, 93–114.

Goodacre, R., Karim, A., Kaderbhai, M., Kell, D.B., 1994a. Rapid and quantitative analysis of recombinant protein expression using pyrolysis mass spectrometry and artificial neural networks. J. Biotechnol. 34, 185–193.

Goodacre, R., Trew, S., Wrigley-Jones, C., Neal, M.J., Maddock, J., Ottley, T.W., Porter, N., Kell, D.B., 1994b. Rapid screening for metabolite overproduction in fermentor reaction mixtures, using pyrolysis mass spectrometry with multivariate calibration and artificial neural networks. Biotechnol. Bioeng. 44, 1205–1216.

Goodacre, R., Trew, S., Wrigley-Jones, C., Saunders, G., Neal, M.J., Porter, N., Kell, D.B., 1995. Rapid and quantitative analysis of metabolites in fermentor reaction mixtures using pyrolysis mass spectrometry with supervised learning: application to the screening of *Penicillium chrysogenum* bioprocesses for the overproduction of penicillins. Anal. Chim. Acta 313, 25–43.

Goodacre, R., Neal, M.J., Kell, D.B., 1996a. Quantitative analysis of multivariate data using artificial neural networks: A tutorial review and applications to the deconvolution of pyrolysis mass spectra. Zentralbl. Bakteriol. 284, 516–539.

Goodacre, R., Timmins, É.M., Rooney, P.J., Rowland, J.J., Kell, D.B., 1996b. Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks. FEMS Microbiol. Lett. 140, 233–239.

Griffiths, P.R., de Haseth, J.A., 1986. Fourier Transform Infrared Spectrometry. Wiley, New York.

Hazen, K.H., Arnold, M.A., Small, G.W., 1994. Temperature-insensitive near-infrared spectroscopic measurement of glucose in aqueous-solutions. Appl. Spectrosc. 48, 477–483.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer, New York.

Jones, A., Shaw, A.D., Salter, G.J., Bianchi, G., Kell, D.B., 1998. The exploitation of chemometric methods in the analysis of spectroscopic data: application to olive oils. In: Hamilton, R.J. (Ed.), Lipid Analysis of Oils and Fats. Chapman and Hall, London, pp. 317–337.

Kell, D.B., Markx, G.H., Davey, C.L., Todd, R.W., 1990. Real-time monitoring of cellular biomass: methods and applications. Trends Anal. Chem. 9, 190–194.

Locher, G., Sonnleitner, B., Fiechter, A., 1992. On-line measurement in biotechnology: Exploitation, objectives and benefits. J. Biotechnol. 25, 55–73.

MacFie, H.J.H., Gutteridge, C.S., Norris, J.R., 1978. Use of canonical variate analysis in differentiation of bacteria by pyrolysis gas–liquid chromatography. J. Gen. Microbiol. 104, 67–74.

Manly, B.F.J., 1994. Discriminant function analysis. In: Multivariate Statistical Methods. Chapman and Hall, London.

Marquardt, L.A., Arnold, M.A., Small, G.W., 1993. Near-infrared spectroscopic measurement of glucose in a protein matrix. Anal. Chem. 65, 3271–3278.

Martens, H., Næs, T., 1989. Multivariate Calibration. Wiley, New York.

Massarrt, D.L., Buydens, L.J., 1987. Chemometrics in pharmaceutical analysis. J. Pharm. Biomed. Anal. 6, 535–545.

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

Meuzelaar, H.L.C., Haverkamp, J., Hileman, F.D., 1982. Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials. Elsevier, Amsterdam.

Rumelhart, D.E., McClelland, J.L., Group, P., 1986. Parallel Distributed Processing, Experiments in the Microstructure of Cognition. MIT Press, Cambridge, MA.

Schrader, B., 1995. Infrared and Raman Spectroscopy: Methods and Applications. Verlag Chemie, Weinheim.

Song, C.Y., Otto, R., 1995. Rapid-determination of the value-determining constituents in sausage products by near-infrared transmission spectroscopy. Z. Lebensmittel-Untersuchung -Forsch. 201, 226–229.

Stuart, B., 1996. Modern Infrared Spectroscopy. Wiley, Chichester.

Timmins, É.M., Howell, S.A., Alsberg, B.K., Noble, W.C., Goodacre, R., 1998. Rapid differentiation of closely related *Candida* species and strains by pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. J. Clin. Microbiol. 36, 367–374.

Widrow, B., Rumelhart, D.E., Lehr, M.A., 1994. Neural networks: applications in industry, business and science. Commun. ACM 37, 93–105.

Wilson, T.A., Rogers, S.K., Myers, L.R., 1995. Perceptual-based hyperspectral image fusion using multiresolution analysis. Opt. Eng. 34, 3154–3164.

Windig, W., Haverkamp, J., Kistemaker, P.G., 1983. Interpretation of sets of pyrolysis mass spectra by discriminant analysis and graphical rotation. Anal. Chem. 55, 81–88.

Winson, M.K., Todd, M., Rudd, B.A.M., Jones, A., Alsberg, B.K., Woodward, A.M., Goodacre, R., Rowland, J.J., Kell, D.B., 1997. A DRASTIC (Diffuse Reflectance Absorbance Spectroscopy Taking in Chemometrics) approach for the rapid analysis of microbial bioprocess products: quantification of aristeromycin and neplanocin A in *Streptomyces citricolor* reaction mixtures. In: Kieslich, K., van der Beek, C.P., de Bont, J.A.M., van den Tweel, W.J.J. (Eds.), New Frontiers in Screening for Microbial Biocatalysts. Elsevier, Amsterdam, pp. 185–191.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. In: Krishnaiah, K.R. (Ed.), Multivariate Analysis. Academic Press, New York, pp. 391–420.

Zupan, J., Gasteiger, J., 1993. Neural Networks for Chemists: An Introduction. VCH, Weinheim.