Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2014, **139**, 1661

Comprehensive, quantitative bioprocess productivity monitoring using fluorescence EEM spectroscopy and chemometrics†

Boyan Li,[a] Michael Shanahan,[a] Amandine Calvet,[a] Kirk J. Leister[b] and Alan G. Ryder*[a]

This study demonstrates the application of fluorescence excitation-emission matrix (EEM) spectroscopy to the quantitative predictive analysis of recombinant glycoprotein production cultured in a Chinese hamster ovary (CHO) cell fed-batch process. The method relies on the fact that EEM spectra of complex solutions are very sensitive to compositional change. As the cultivation progressed, changes in the emission properties of various key fluorophores (*e.g.*, tyrosine, tryptophan, and the glycoprotein product) showed significant differences, and this was used to follow culture progress *via* multiple curve resolution alternating least squares (MCR-ALS). MCR-ALS clearly showed the increase in the unique dityrosine emission from the product glycoprotein as the process progressed, thus provided a qualitative tool for process monitoring. For the quantitative predictive modelling of process performance, the EEM data was first subjected to variable selection and then using the most informative variables, partial least-squares (PLS) regression was implemented for glycoprotein yield prediction. Accurate predictions with relative errors of between 2.3 and 4.6% were obtained for samples extracted from the 100 to 5000 L scale bioreactors. This study shows that the combination of EEM spectroscopy and chemometric methods of evaluation provides a convenient method for monitoring at-line or off-line the productivity of industrial fed-batch mammalian cell culture processes from the small to large scale. This method has applicability to the advancement of process consistency, early problem detection, and quality-by-design (QbD) practices.

Received 3rd January 2014
Accepted 27th January 2014

DOI: 10.1039/c4an00007b

www.rsc.org/analyst

Introduction

Some of the most common methods for producing therapeutics *via* mammalian cell culture are the complex fed-batch fermentation processes. Their operations are governed by multiple process parameters (*e.g.* feed quality, feeding strategy, inoculum age, harvest point *etc.*) that determine product yield and quality.[1–3] Accurately monitoring the progress and productivity of these processes requires informative analytical methods to provide detailed information for the accurate control and management of production.[1–9] Once a process seed reactor has been transferred to the large-scale manufacturing bioreactor stage, most of the process operational parameters will have been fixed, except for feed quality, which can vary substantially. The bioreactor broth is a chemically complex environment, which comprises the cell culture media, metabolites, product and host cell protein, whole cells and cell debris.[1,10,11]

Detailed chemical analysis of these complex bioreactor broths is challenging and high performance liquid chromatography (often coupled with mass spectrometry),[12–15] or high-field NMR[16,17] can be used. However, both approaches typically require time-consuming sample preparation, high capital cost, frequent maintenance and often highly skilled, labour-intensive/time-consuming data analysis. For bioprocess monitoring, there are two options to consider: to use an in-reactor probe for online monitoring, or aseptically sample the bioreactor for at-line or off-line analysis.[5,8,13,18–21] For online spectroscopic monitoring of bioprocess, one can implement vibrational or fluorescence spectroscopy based methods.[8,22–24] Each method has its own unique advantages; however, both near-infrared (NIR) and mid-infrared (MIR) are not always suited to aqueous sample analysis because of the strong water signals. Raman spectroscopy on the other hand is good for aqueous sample analysis, but analyte signals can be too weak to be clearly discriminated, and fluorescence interference can be an issue.[25–27] In the context of bioreactor broth analysis, many significant chemical species (*e.g.*, tryptophan, tyrosine, riboflavin, pyridoxine *etc.*) fluoresce and can be easily detected at micro-molar concentrations. The complex interplay between energy transfer and quenching effects in complex multi-fluorophore mixtures has been exploited using fluorescence

[a]*Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Ireland. E-mail: alan.ryder@nuigalway.ie; Fax: +353 91 495576; Tel: +353 91 492943*

[b]*Bristol-Myers Squibb, Process Analytical Sciences, Syracuse, New York, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4an00007b

EEM spectroscopy for raw materials[28] and cell culture media analysis.[29] The unique EEM spectra produced as the sample composition changes have also been employed for in-reactor bioprocess monitoring.[30–34] More recently, synchronous fluorescence spectroscopy has been successfully demonstrated for the quantitative monitoring of cell density and antibody titer.[35] A 2D fluorescence technique like EEM offers some other advantages in terms of minimal sample handling, relatively fast analysis time, ease of implementation/use, inexpensive, low-maintenance instrumentation, and the ability to be used for a wide variety of analytical problems. Finally, one can use EEM for the simultaneous analysis of multiple fluorophores in complex samples such as cell culture media.[36]

One of the key factors driving the adoption of multivariate spectroscopic based methods has been the increased use of chemometrics to extract useful quantitative and qualitative information from the data.[37–40] In the context of quantitative bioreactor broth analysis, chemometrics has generally been used for metabolite or nutrient quantification,[24,31,34] or more holistically predict the final yield.[32] Partial least-squares regression (PLS)[39,41,42] is one of the most important chemometric tools used for quantitative analysis, however, with some spectral data there are a lot of uninformative variables present which can degrade model performance. To improve PLS model performance, there are numerous variable selection methods, which can be used to remove the uninformative data.[43–46]

There are a wide variety of variable selection methods available.[46–58] Competitive adaptive reweighted sampling (CoAdReS)[59] and ant colony optimization (ACO)[60,61] have previously been used on Raman data collected from this sample set.[27] Both methods use Monte Carlo (MC) strategies to select the key variables from the multivariate spectral data, and thus generate more accurate chemometric models. While the variable selection method can remove uninformative spectral data, it can, if the analyst is not careful contribute to model over-fitting, and thus there is a need to combine the variable selection with a robust assessment of model complexity. If there are sufficient samples available, then an independent test set validation is the best option. However, in the initial steps of determining if a new analytical method is feasible there may not be sufficient samples available. Thus chemometric methods have to be implemented to prevent over-fitting. One of the best methods for use with complex spectra such as EEM is the randomization method,[62] which can help offset the issue of low sample numbers.

The goal of this study was to develop a rapid fluorescence based methodology for both monitoring and predicting glycoprotein production in a fed-batch, mammalian cell culture process from the initial small, litre-scale right up to the final large-scale fermenter. This was achieved by undertaking three separate studies. First determine the optimal sample preparation conditions for generating informative EEM data from each set of bioprocess broth samples. Second use multiple curve resolution-alternating least squares (MCR-ALS) to deconvolute the gross compositional information contained in these EEM data, and then use this to track the production of the glycoprotein product over the course of the bioprocess.[63–65] Finally to determine the feasibility of accurately predicting the final glycoprotein yield from the EEM data using variable selection and other chemometric methods at each stage in the bioprocess.

Materials and methods

Materials

L-Tyrosine (≥98%), L-tryptophan (≥98%) L-phenylalanine, pyridoxine, and folic acid dehydrate (97%) were purchased from Sigma-Aldrich and used without further purification to prepare solution standards (ESI†). The bioprocess broth samples were extracted from an industrial CHO cell based bioprocess producing a recombinant glycoprotein.[27] Samples were acquired when available from a continuous batch campaign of 40+ batches. The bioprocess was operated in fed-batch mode using proprietary basal and feed media formulations. The bioprocess was sampled at twelve different stages during the fermentation process. This gave dataset sizes of between 17 and 37 (Table 1) samples. The samples were first centrifuged and sterile filtered to remove any whole cells, before being carefully handled and stored at −70 °C (for the sake of clarity these samples are referred to as bioprocess broths).[27] Samples were randomly removed from cold storage, defrosted at room temperature, and all EEM data collected within 6 hours. It was also important to note that at each transfer point between bioreactors various filtration steps were implemented, thus for example, the DS7 and DS8 samples will have a different composition.

The DS9–12 samples follow the final stages of fermentation up to the harvest point and during this phase, feed media was also added. Protein yield was measured using a spectrophotometric measurement (A_{280}) method which was validated to ICH standards and has been described elsewhere.[27]

Instrumentation and data collection

Samples for chemometric analysis were first diluted by mixing 50 μL of sample solution (DS1–12) with ultrapure water (18 MΩ resistivity) to a final volume of 1 mL. This mixture was then pipetted directly into a semi-micro quartz cuvette (Lightpath Optical Ltd., UK) sealed under aseptic conditions.[28] Fluorescence measurements were carried out at 25 °C using a Cary Eclipse (Varian, now Agilent) fluorescence spectrometer fitted

Table 1 Details of the bioprocess samples used in this study[27]

Dataset ID	Bioreactor content description	Bioreactor vol.	Sample size
DS1	Media start	2 L	21
DS2	Media end	2 L	17
DS3	Cells + spent basal media	2 L	17
DS4	Cells + spent & fresh basal media	100–200 L	31
DS5	Cells + spent basal media	100–200 L	31
DS6	Cells + spent & fresh basal media	1000 L	31
DS7	Cells + spent basal media	1000 L	34
DS8	Cells + spent basal media	5000 L	37
DS9	Cells + spent & fresh basal media	5000 L	29
DS10	Day 5 post inoculation	5000 L	35
DS11	Day 10 post inoculation	5000 L	34
DS12	Prior to transfer for harvest	5000 L	33

with a thermostatted multi-cell holder. EEMs were obtained with a λ_{ex} range of 230–520 nm, 5 nm interval, and a λ_{em} range of 270–600 nm (excitation and emission bandwidths = 5 nm). Samples were allowed to thermally equilibrate for several minutes prior to measurement. For each sample, the EEM data were collected in triplicate using fresh aliquots over a period of 10 weeks. A smaller portion of the EEM data (λ_{ex} = 230–315 nm and λ_{em} of 270–465 nm) was used for chemometric analysis because the major fluorescence bands appeared in this region. Therefore, each fluorescence data matrix comprised of 18 excitation and 40 emission wavelengths.

Data analysis and chemometric methodology

To minimize baseline, scatter, and instrumental effects/errors the data was subjected to pre-processing for scattering and baseline removal, normalization and auto-scaling for PLS modelling.[29] Rayleigh scattering bands (including the second order peaks) were replaced with a curve fit, connecting points either side of the peak using imputation.[66–68] It was not possible to remove the Raman bands as they overlapped the much stronger fluorescence signal. To better understand the causes of the observed changes in the fluorescence EEM, the MCR-ALS method was used to deconvolve the spectra.[63,64,69] This provided information about the evolution of EEM measurements, inner filter effects (IFE), and compositional changes through pure component contributions. MCR-ALS was used because it can more robustly account for variations in the pure component spectra induced by compositional heterogeneity and energy transfer effects than other factor analysis based methods.[63] The criterion used for selecting the correct number of significant components was based on the noise perturbation in functional principal component analysis (NPFPCA) method.[70] When implemented here for all MCR-ALS models, a perturbation level of 1% (of EEM maximum intensity) noise with 500 repetitions was used (data not shown).

Variable selection on the EEM data was performed using CoAdReS and ACO methods, the selected variables were then used to build PLS models to predict glycoprotein yield at each stage of the bioprocess. For each dataset 200 CoAdReS sampling runs were performed and for each sampling run, a PLS model was constructed using 83% of the samples, which were randomly selected. CoAdReS then generated sequentially 200 subsets of variables (720 in run 1, only 2 in run 200) and regression coefficients for each variable were obtained from the PLS models. The variable selection process was based first on the magnitude of the regression coefficients, and second on the reduction rate, for example in the i^{th} sampling run, the ratio of variables to be kept (r_i) was given by: $r_i = ae^{-ki}$ (a = 1.0234, k = 0.0232, i = 1, 2, …, 200). The variables with low regression coefficients were weighted to zero, and the significant variables to be retained were weighted with a value related to their absolute regression coefficient value. These retained variables were then used for PLS modelling in the next sampling run, and so on.[59] Once the 200 subsets were generated, the remaining samples (17%) were employed for cross validation on each CoAdReS sampling run, and the root mean square error of cross-validation (RMSECV) was calculated for this cross-validation. The optimal subset of variables (from the 200) was the subset with the lowest RMSECV value.

The basis of ACO refinement is completely different to CoAdReS and ACO was implemented with the following conditions: ρ (rate of pheromone evaporation) = 0.65, N (number of ants) = 400, w (sensor width) = 1, a maximum number of time steps of 50, and 100 repeated MC calculation cycles were run to build a histogram of variable selection probability for each sample set. The selected variables were then used to build the PLS models. All calculations were performed using MATLAB,[71] PLS_Toolbox,[72] and in-house-written MATLAB routines. ACO MATLAB code was generously provided by Prof. A. C. Olivieri (Universidad Nacional de Rosario, Argentina).

Calibration and validation samples

Twelve spectral datasets were obtained after averaging the replicate spectra, from the various sample sets (Table 1) but only sample sets (DS4–12) were used for chemometric modelling. For PLS modelling, each sample set was randomly split into a calibration (between 20 and 26 samples) and a test set (always five samples) in a ~80 : 20 split using a Monte Carlo based sampling protocol. To ensure robustness, the calibration/test set selection was repeated 500 times and a PLS model run on each unique selection. PLS model quality was assessed using a combination of parameters including: root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV) for validation/test set, relative error of cross-validation (RECV% = 100 × RMSECV/$\bar{y}_{cal}$, where $\bar{y}_{cal}$ is the mean calibration value of the product titre), and the square of the correlation coefficient (R^2) between predicted and nominal titres for validation set. Finally, a randomization test method[62] was deployed to determine the model complexity and ensure that the optimal number of latent variables were used in the PLS models (see ESI for details†). This step was critical as the overall sample numbers were low, the spectral data was complicated and thus any potential over-fitting had to be minimised.

Results and discussion

Spectral analysis

Fig. 1 shows typical EEM spectra of diluted bioprocess broths from the start (DS1) and end (DS12) stages of the cell culture. All samples have similar features, indicating the presence of multiple fluorophores from the media, metabolites, and glycoprotein product. It was not possible to identify specific fluorophores from

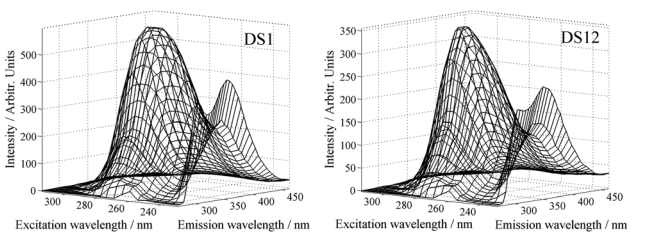

Fig. 1 EEM landscapes of two typical, 1 : 20 diluted bioprocess samples (DS1 and DS12) selected from the same production lot.

the spectra visually due to spectral overlap, nevertheless when compared to the spectra of standard fluorophores, it was clear that the strongest fluorescence in all cases originated from tryptophan and tyrosine, with weaker contributions from a variety of other fluorophores (Fig. S-1, ESI†). The major differences between the DS1 and DS12 spectra were a ~40% drop in overall fluorescence intensity, large changes in band shape, and a small band shift in the maximum of the main band ($\lambda_{ex}/\lambda_{em} = 275/355$ nm for DS1 and $\lambda_{ex}/\lambda_{em} = 280/355$ nm for DS12). These changes were caused by the combination of inner filter (IFE), energy transfer (ET), and quenching effects induced by the changing chemical composition in the bioreactor as the bioprocess evolves. Ascribing the changes simply to consumption of the media fluorophores and formation of the glycoprotein product, was probably not valid as the production of photophysically active metabolites cannot be distinguished from the EEM data. Without chemometric analysis (*vide infra*) was impossible to assign clear spectral changes due to the product glycoprotein which was present at a relatively low concentration (always between 0.67 and 0.92 g L^{-1}) compared to all the other fluorophores present *e.g.* media, metabolites, host cell proteins, which when considered all together were present at much higher concentration particularly during the early stages of the bioprocess.

Dilution effect

Analysis of the undiluted samples through the bioprocess showed relatively small spectral changes. This was due to the relatively high chromophore and fluorophore concentrations causing high rates of IFE, ET, and quenching, which in combination reduce the magnitude of observable spectral change. One way to reduce IFE, ET, and quenching effects was to dilute the solution, which was expected to resolve the spectral changes more clearly. To determine the optimum dilution factor one typical DS9 sample was taken and mixed aliquots of 2, 5, 10, 20, 40, 50, 60, 80, and 100 μL of the original solution with high purity water to give a final volume of 1 mL (*i.e.* dilution factors of between 1 : 500 and 1 : 10). Fluorescence EEM data were acquired from these nine samples which were then analysed empirically by MCR-ALS (*vide infra*). The overall EEM spectral intensity increased as the dilution factor decreased from 1 : 500 to 1 : 10. Taking a slice through the EEM at $\lambda_{ex} = 275$ nm (Fig. 2) shows how the main Trp emission band changed as the dilution factor decreased to 1 : 10. The effect appears less dramatic for Tyr band (Fig. 2b) because the Tyr signal was much weaker compared to Trp, however the rate was the same for both amino acids when one looks at the normalised intensity ratio ($I_X/I_{1:10}$). With more concentrated samples, dilution factors below 1 : 10, the increased rate of quenching and ET, IFE *etc.* caused band intensity to decrease once more. Balancing the competing demands for high signal intensity and spectral detail, it was decided that a dilution factor of 1 : 20 was appropriate for generating EEM data suitable for chemometric analysis.

To better understand the biomass concentration induced spectral changes MCR-ALS was used to extract the changing fluorophore contributions from the EEM as the broth composition changed during the bioprocess. For the MCR-ALS, a

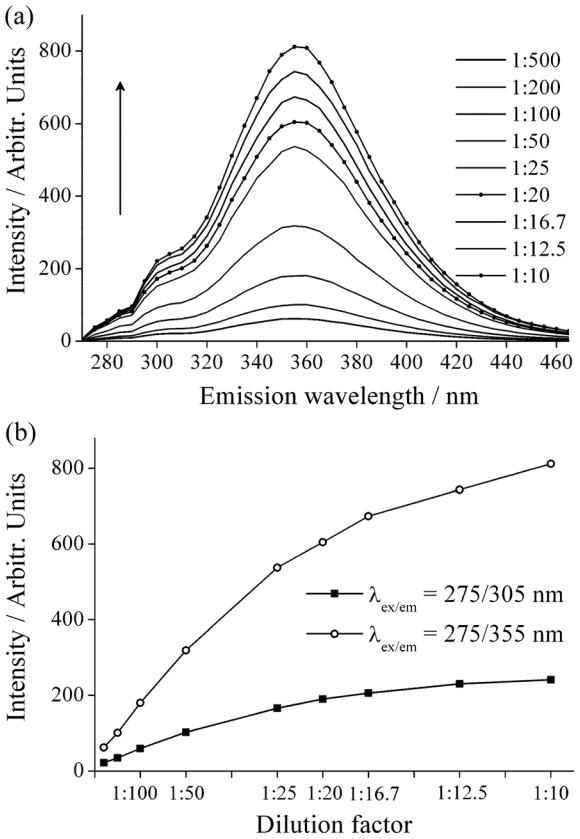

Fig. 2 (a) Emission spectra of the nine diluted solutions of a DS9 sample, and (b) the intensity changes at the Trp band at $\lambda_{ex}/\lambda_{em} = 275/355$ nm and Tyr band at 275/305 nm.

single sample-EEM augmented data matrix $D_g(18 \times 9, 40)$ was generated using these 9 dilution samples, with 18 excitation and 40 emission wavelengths. In the MCR model, $D_g = S_g^{ex}S^{emT} + E$, the $S_g^{ex}(18 \times 9, n)$ matrix contained the excitation spectra (18 wavelengths) obtained at the various dilutions for each n component, E was the residual matrix describing the variance not explained by $S_g^{ex}S^{emT}$ and T meant the transpose of matrix.

In this MCR-ALS model the emission spectra S^{em} of the n components, common to all measurements, were constrained to a fixed profile (*i.e.* spectral shape) for each component but variable intensity. In reality, this was not true for heterogeneous fluorophore mixtures because IFE and ET will cause distortions in the emission profile. However, this was a necessary assumption required to implement MCR-ALS on this type of EEM data.[69] These distortions in the data structure were commonly manifested by the appearance of additional components in MCR-ALS models of complex samples.[65]

IFE usually affects shorter wavelength regions most (due to higher absorbance in these regions); consequently the long wavelength emission was less likely to be affected by IFE than the excitation profile. By inspecting the excitation spectra one obtained a better understanding of the IFE induced by varying the dilution factor.

Here MCR-ALS decomposed the dilution dataset into five components (Table 2) using non-negativity constraints. The

Table 2 Significance of each component decomposed from the EEM spectra of nine dilute solutions (from 1 : 500 up to 1 : 10) by the MCR-ALS model

		Percent variance captured by MCR-ALS model		
# Component	λ_{em} max. (nm)	Fit (% model)	Fit (% spectra)	Cumulative fit (% spectra)
1	355 nm	96.06	96.05	96.05
2	305 nm	3.69	3.69	99.74
3	385 nm	0.19	0.19	99.93
4	410 nm	0.04	0.04	99.97
5	290 nm	0.02	0.02	99.99

NPFPCA method[70] was applied to select five MCR-ALS components (Fig. S-3, ESI†). Three of the components (Comps 1, 2, 3) were significant (Fig. 3), and by comparison with the spectra of known fluorophores (ESI†) it was established that Comp1 was Trp and Comp2 was Tyr.[36] The identity of the weak Comp3 was less certain, but it was likely to comprise of emission from several media components present in low (<5 μm) concentration (*e.g.* pyridoxine and riboflavin[36]) and/or the glycoprotein product (*vide infra*).

Comp1 (Trp) and Comp2 (Tyr) dominated the EEM (Table 2) and these components comprised of emission from a variety of sources including: the free amino acids, a small proportion from peptides, the product glycoprotein, and host cell proteins.

The excitation spectra of Comp1 (Trp) resolved by MCR-ALS had an excitation band maximum at 275 nm when highly diluted (1 : 500/1 : 50) and this band shifted to ~285 nm as the biomass concentration increased (Fig. 3b). This excitation red-shift was indicative of significant IFE for Trp which was not as evident for Comp2 (Tyr) and Comp3 where the excitation spectra band maxima were constant at 275 and 290 nm, respectively. The fact that the Trp undergoes much larger IFE was unsurprising since its absorption band overlaps the emission band of Tyr significantly and thus undergoes significant emission enhancement *via* energy transfer. The change in the shape of the excitation profile for Comp1 was due to the overlap of the absorption spectra of Trp and Tyr. This MCR-ALS study of a diluted DS9 sample indicates that one can easily extract information as to compositional changes.

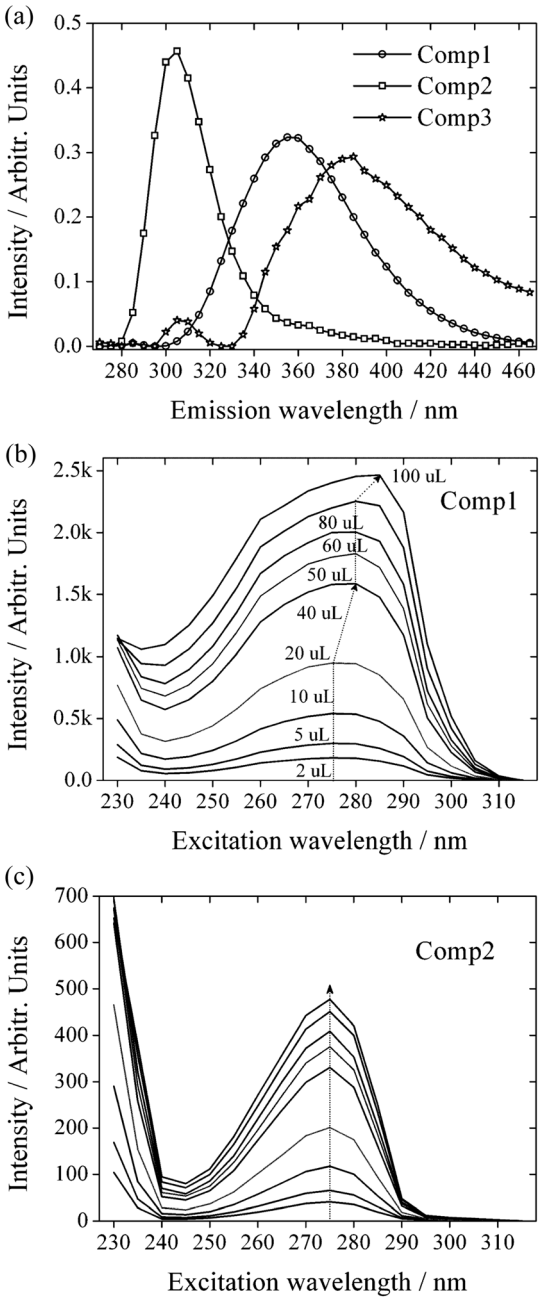

Fig. 3 (a) Emission spectra (area normalised) of the three most significant components deconvoluted from the dilution dataset by MCR-ALS, and (b and c) their individual excitation spectra (from 1 : 500 up to 1 : 10 dilutions), recovered by MCR-ALS using normalised emission spectra. The arrows show the direction of change with dilution.

Bioprocess evolution

The next step was to see if MCR-ALS could be used to evaluate the compositional changes across a series of production lots, and more importantly observe any correlation with product formation. For these bioprocess evolution studies, a 1 : 20 sample dilution was used as this gives a strong EEM signal in which the individual components are reasonably well-resolved. The use of the diluted bioprocess broths and/or media was a departure from our previous methods where media were analysed in its prepared state.[28,29,73] The approach was justified here because one needed to clearly observe the spectral contribution of the glycoprotein product, and this was easier to do once IFE has been reduced.

To demonstrate the method, EEM spectra collected from the 12 samples (DS1–12) for a single production lot were selected. The triplicate EEM spectra were submitted to MCR-ALS with non-negativity constraints and five significant constituents were extracted (Table 3). Fig. 4 shows the individual emission and excitation spectral profiles generated by the MCR-ALS model.

As before, the first two components extracted, Comp1 and Comp2, were Trp and Tyr respectively. Comp3 showed very

Table 3 Significance of each constituent decomposed by the MCR-ALS model of the EEM spectra from samples pulled from a single complete production lot sampled at 12 stages (DS1 to DS12)

# Component	$\lambda_{ex}/\lambda_{em}$ max. (nm)	% Variance captured by MCR-ALS model		
		Fit (% model)	Fit (% spectra)	Cumulative fit (% spectra)
1	275/355 nm	58.38	58.36	58.36
2	275/305 nm	8.33	8.33	66.69
3	275/365 nm	32.28	32.27	98.97
4	240/385 nm	0.37	0.37	99.34
5	265/350 nm	0.63	0.63	99.97

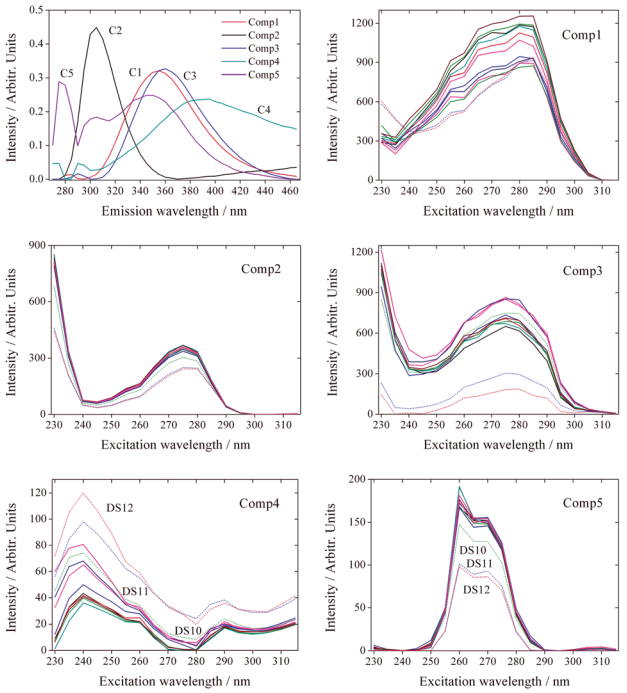

Fig. 4 Emission spectra (top left) of the five significant constituents deconvoluted from the 2D fluorescence spectra of the bioprocess broths pulled from a single complete production lot sampled at 12 stages (from DS1 up to DS12) by the MCR-ALS model. The other plots show the changes in the individual excitation spectra for each component.

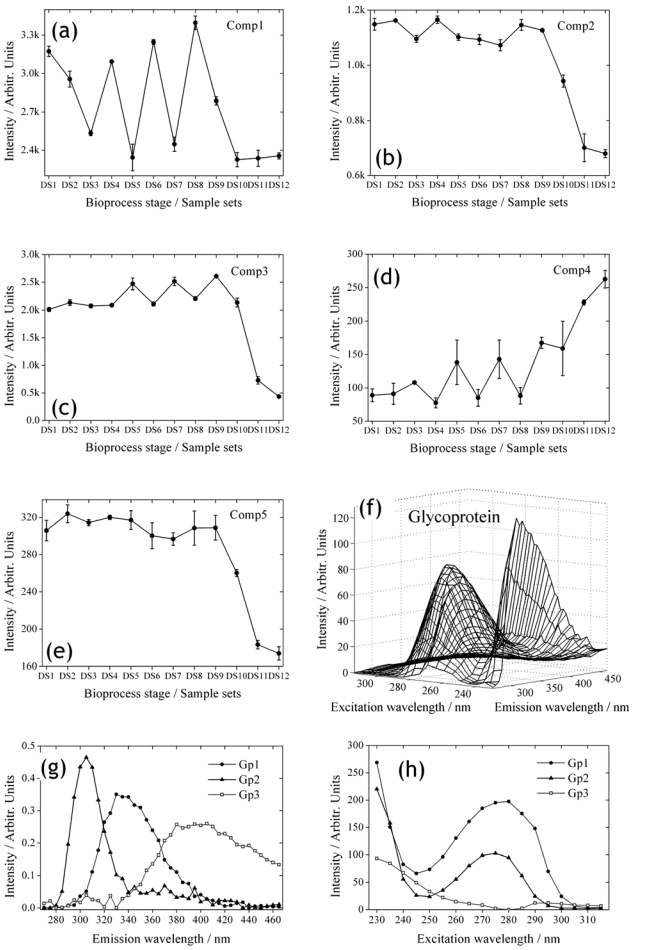

Fig. 5 (a–e) Temporal concentrations of the significant constituents in bioprocess broths pulled from a single complete production lot sampled at 12 target stages over four bioreactors, which were resulted from the 2D fluorescence spectra by MCR-ALS model. The ordinate was represented after both the emission and excitation spectra were normalized. Error bars are standard error for $n = 3$. (f) EEM landscape of a 10^{-2} g L^{-1} solution of the pure product glycoprotein solution. (g and h) The excitation and emission spectra extracted from the EEM data of the product glycoprotein solution by MCR-ALS: ○—tryptophan, Δ—tyrosine, □—Gp3, and the MCR-ALS model details.

similar emission and excitation spectral profiles to Trp (but red-shifted emission) and this constituent had the second highest apparent spectral concentration in the samples. Comp4 was tentatively assigned to the glycoprotein product which contains 16 tyrosine and 4 tryptophan residues. MCR-ALS modelling of a pure solution of the glycoprotein product (Fig. 5g and h) shows 3 components (Gp1–3); the first two being Trp and Tyr, and the third component matches Comp4. The first contributor (Gp1) comprised a broad emission band at the 335 nm and an excitation band at 280 nm. This was very similar to the pure Trp EEM (Fig. S-1†) but with a blue-shifted emission and red-shifted excitation indicative of the Trp being located in a more hydrophobic environment. The second factor, Gp2 was definitely tyrosine because of its characteristic bands appearing at the $\lambda_{ex}/\lambda_{em}$ = 230/305 nm and 275/305 nm. Gp1 and Gp2 were two main contributors explained 68.73% and 25.27% respectively of the EEM data variance. The final factor, Gp3, weak ~5%, gave a broad emission band from 320 to 465 nm with a maximum at ~395 nm, and its excitation spectrum showed two local maxima at 230 and 290 nm respectively. It was suspected that this emission was probably from dityrosine because there was one pair of adjacent Tyr residues in the product sequence. This was confirmed *via* fluorescence lifetime measurements using 295 nm excitation made on a solution of the pure glycoprotein (ESI†). In these measurements, a long lifetime component of ~4.3 ns was observed (along with two 0.7 and 1.49 ns components) and the fractional component increased to ~20% at 400 nm. This long lifetime component was indicative of

dityrosine.[74] Thus Comp4 tracks the evolution of the dityrosine emission which originates from the glycoprotein product and was thus a unique marker of product formation.

The final component, Comp5 was likely to be a composite of the other fluorophores which were present in low concentration, *e.g.* pyridoxine, folic acid, riboflavin and others which emit at these longer wavelengths. However, with these very complex samples, low sample numbers, and the domination of the Trp/Tyr signals it was not possible (or necessary) to resolve any further.

When the individual emission and excitation spectra of each constituent were normalized, a relative concentration (namely scores) profile was generated for each constituent at each stage of the bioprocess (Fig. 5). These curves show the evolution of the bioprocess in terms of comparative change. At the present time it's not feasible to convert these changes into accurate concentration values, because there were no in-process Trp/Tyr or glycoprotein product concentration measurements with which to standardise the scores. It was also important to note that these scores do not represent the precise relative concentration changes either, as they do not take into account for the differences in quantum yield, molar absorptivity, and absolute concentration of each fluorophore type. However, they do provide a facile method for evaluating the changes in bioreactor composition. In many cases, this may be all that is required to monitor the health and progress of the process. If exact concentrations of Trp/Tyr were required then orthogonal methods like off-line HPLC or a modified standard addition EEM based method could be implemented.[36]

Analysis of the changes in the component intensity uncovered some interesting trends. Comp1 shows the effect of the addition of fresh media very clearly, whereas Comp2 and Comp5 show very little change. The feed addition effect in terms of component signal increase was less strong for Comp3 and Comp4, indicating that these were not media components but rather a metabolite and the glycoprotein product respectively. Looking carefully at the change in signal, a decrease in intensity was observed when the fresh basal media was added, and thus the saw tooth pattern was ascribed as being a simple dilution effect in these cases (Fig. 5c and d).

DS3/5/7 samples were solutions of cells and spent basal media just prior to transfer to next, larger-sized bioreactor, and this explains why Comp1 (Trp) followed the indicated trend, *e.g.*, it was being consumed by cell growth.[35] It was suspected that this component was free Trp in solution. The spikes in the Comp1 scores for the DS4/6/8 datasets were therefore clearly due to the addition of extra Trp in the fresh basal media added to advance process scale-up. The overall downward trend in Comp1 was indicative of consumption due to cell growth. Comp3 which looks very like Trp, also followed a similar trend to Comp1 with an increase with the addition of fresh media and a large decrease in the final growth phase of the bioreactor. But critically it was time-shifted and the maxima correspond to the DS5/7 samples which were cells and spent media, and also the DS9 samples which were cells with spent & fresh basal media.

It was possible that this component was a fluorescent metabolite of tryptophan, *e.g.* 5-hydroxytryptophan (5-HTP).[15] The change in Comp2 (Tyr) was less striking during the early bioreactor stages, but once the growth phase was induced in DS9 a dramatic decrease occurred due to cell consumption of the amino acid. In general for all components the biggest changes can be seen over the DS9–12 samples, which follow the final stages of fermentation up to harvest. During this phase, feed media was also added at specific times, and it was interesting to note that there were no major changes that can be ascribed to these additions.

Overall MCR-ALS provides a good insight to the evolutionary changes as the bioprocess progresses, and although the method does not provide accurate quantitative estimates of specific component change, it does provide a useful tool for process monitoring.

Correlation with product yield

The next stage in the method was to see if it was possible to predict the final glycoprotein yield from the EEM data collected. PLS regression was applied to each individual sample dataset using the pre-processed spectra in order to associate fluorescence spectra with the glycoprotein yield (Table 4). The calibration

Table 4 Summary of PLS models generated using 9 different sample sets. RMSEC/RMSECV errors are given in g L^{-1} of the final protein product titre. Values are the mean values obtained from 500 different individual models (see main body text for details). In each case five samples were used for the test set. Dataset sample size (in parentheses) varied according to sample availability

Data set	# Vars.	LVs	RMSEC	RMSECV	RECV%	R^2
Full spectrum based models						
DS4(28)	720	8	0.006 ± 0.001	0.086 ± 0.018	10.73	0.26
DS5(28)	720	7	0.009 ± 0.002	0.078 ± 0.019	9.61	0.24
DS6(28)	720	6	0.014 ± 0.002	0.078 ± 0.017	9.61	0.26
DS7(29)	720	7	0.008 ± 0.001	0.062 ± 0.010	7.74	0.39
DS8(31)	720	8	0.007 ± 0.001	0.071 ± 0.019	8.59	0.27
DS9(25)	720	7	0.010 ± 0.002	0.080 ± 0.015	9.99	0.21
DS10(31)	720	7	0.012 ± 0.001	0.076 ± 0.016	9.42	0.23
DS11(30)	720	7	0.013 ± 0.001	0.071 ± 0.014	8.76	0.33
DS12(28)	720	9	0.004 ± 0.001	0.049 ± 0.010	6.04	0.65
CoAdReS based models						
DS4(28)	57	8–7	0.011 ± 0.001	0.030 ± 0.006	3.79	0.96
DS5(28)	79	7	0.010 ± 0.001	0.037 ± 0.008	4.63	0.94
DS6(28)	70	6	0.013 ± 0.001	0.031 ± 0.010	3.68	0.94
DS7(29)	80	7	0.008 ± 0.001	0.022 ± 0.005	2.80	0.98
DS8(31)	86	8–5	0.010 ± 0.001	0.032 ± 0.009	3.97	0.92
DS9(25)	68	7	0.010 ± 0.001	0.030 ± 0.007	3.64	0.97
DS10(31)	71	7	0.012 ± 0.001	0.029 ± 0.007	3.58	0.95
DS11(30)	64	7–6	0.017 ± 0.001	0.035 ± 0.009	4.22	0.92
DS12(28)	90	9	0.006 ± 0.001	0.018 ± 0.004	2.30	0.98
ACO based models						
DS4(28)	82	8–4	0.011 ± 0.001	0.032 ± 0.009	3.89	0.96
DS5(28)	111	7	0.013 ± 0.001	0.032 ± 0.007	3.98	0.96
DS6(28)	129	7	0.010 ± 0.001	0.030 ± 0.006	3.71	0.96
DS7(29)	148	8	0.007 ± 0.001	0.024 ± 0.005	3.02	0.98
DS8(31)	101	8–5	0.010 ± 0.001	0.028 ± 0.007	3.49	0.96
DS9(25)	140	7	0.010 ± 0.001	0.034 ± 0.008	4.25	0.95
DS10(31)	138	8	0.010 ± 0.001	0.030 ± 0.007	3.66	0.96
DS11(30)	71	7	0.017 ± 0.001	0.033 ± 0.008	4.06	0.93
DS12(28)	129	9–3	0.006 ± 0.001	0.020 ± 0.004	2.54	0.98

models were then validated using the test sets, and the appropriate PLS component number was determined using the randomization test (Table S-2, ESI†). When the full EEM spectrum (720 variables) was used for each dataset, the resulting PLS models were poor with $R^2 < 0.7$. In addition, the very low RMSEC values indicated an over optimistic calibration model, which may have arisen from the high variable-to-sample ratio pertaining to the full spectrum data. It was also suspected that useful information in the EEM data was being swamped by the large number of uninformative signal/variables present. The overlap between glycoprotein product emission and other bioprocess broth signals (media, metabolites, host cell protein) was very significant since the glycoprotein yield range for these samples was between 0.67 and 0.92 g L^{-1}, while the dissolved solid concentration of the original media was significantly greater.

Therefore variable selection was implemented to try and first extract the informative variables from the EEM data and second reduce the variable/sample ratio to generate more reliable, less optimistic calibration models. To ensure a robust selection of key variables, a histogram of variable selection probability was built for each sample set by repeating both the CoAdReS calculation 500 times and the ACO 100 times in the MC manner.

Variable selected PLS models

CoAdReS was implemented on each dataset and the selected variables were then used to generate quantitative PLS models (Table 4). To ensure robustness CoAdReS was rerun 500 times using random calibration/test sample combinations (selected using MC), generating 500 sets of different key variables. Statistical analysis of these 500 variable sets generated a normalized histogram, and Fig. 6a shows the variables with histogram values greater than a 0.24 limit (*vide infra*). Most of these variables are associated with low intensity bands in the EEM. To determine the number of the selected variables to be used for the optimal chemometric model, leave-one-out cross validation[75] PLS modelling was performed with trial numbers of selected variables from 10 up to 300. In practice, all the variables selected for the 500 sets were ranked according to the magnitude of the histogram values from largest to lowest. Then, a number of the selected variables (from 10 to 300) were picked for PLS modelling and then the RMSECV values calculated. The result (Fig. 6b) showed that the RMSECV decreased to a minimum of 0.022 g L^{-1} with 57 variables before increasing to ~0.065 g L^{-1} when more than 282 variables selected.

This suggested that 57 variables should be used for an optimal PLS model which corresponded to a threshold histogram of 0.24. Similarly, the procedure was completed on all the other datasets to select the optimal number of variables. This rather computationally intensive approach was necessary because of the complexity of the sample matrix, the relatively low number of samples, and the high variable-to-sample ratio in the full range EEM data.

With the CoAdReS selected variables, the improvement in PLS model quality was dramatic: R^2 was >0.9, the RMSECV : RMSEC ratios were low ~2–3, and the RECV% values were low at between 2.3 and 4.6% (Table 4). This significant improvement benefited

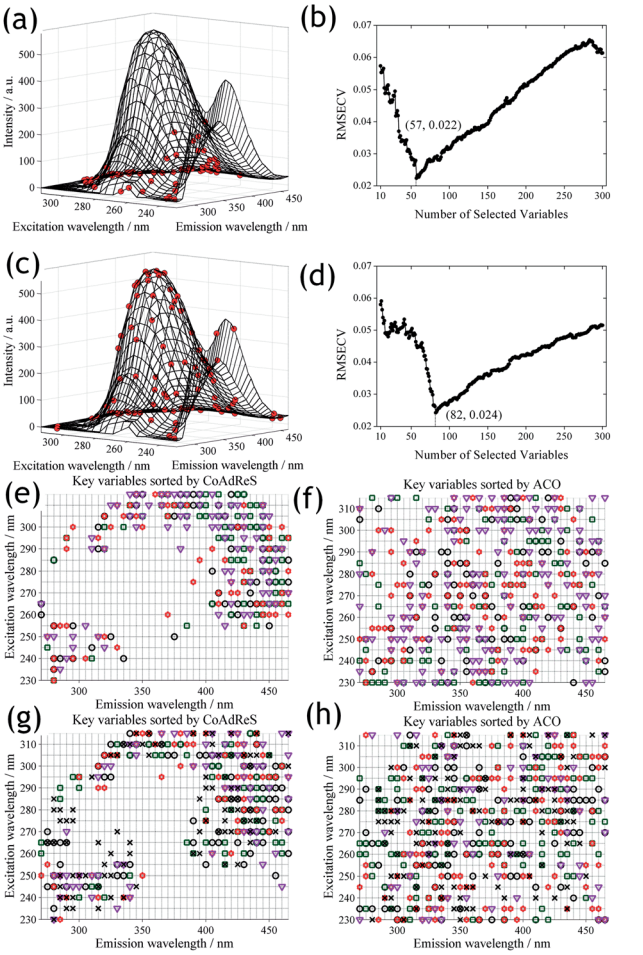

Fig. 6 (a) CoAdReS variable selection result for the sample set DS4 (red markers = variables with histogram values ≥0.24). Superimposed mesh is the mean scattering-corrected EEM landscape in arbitrary vertical scales; (b) determination of number of the selected variables by means of LOOCV with CoAdReS-selected variables; (c) ACO variable selection result for the DS4 (red markers = variables with histogram values ≥0.35); (d) determination of number of the selected variables by means of LOOCV with ACO-selected variables. Relevant variables selected by CoAdReS (e) and ACO (f) for the sample sets: DS4 to DS7 (○—DS4, □—DS5, ☆—DS6, and ▽—DS7). Relevant variables selected by CoAdReS (g) and ACO (h) for the sample sets: DS8 to DS12 (○—DS8, □—DS9, ☆—DS10, ▽—DS11, and ×—DS12).

from the removal of redundant variables (or more correctly those with low information content relating to product yield). For example a large proportion of the measured fluorescence signal originates from tryptophan, tyrosine and other nutrient sources which were present in the highest concentration (typical concentration ranges of 100–600+ μM for Trp and Tyr),[36,76] and were unlikely to show signal variances associated with the glycoprotein product. The high variable reduction factor of ~1 in 10 indicated that the vast majority of the fluorescence signal was as expected not related directly or indirectly to the product yield.

It was interesting to note that for both sets of PLS models (Table 4) the RMSEC values were very close when the same number of PLS components was used for each sample set. This implied that the variables which had the greatest contribution

were present in the entire and CoAdReS selected variable datasets, but that their contribution to the PLS models when the entire spectra were used, was swamped by the mass of irrelevant variables, leading to very poor RMSECV values.

For DS4 when a histogram threshold of 0.35 was set, ACO selected 82 variables and a minimum RMSECV of 0.024 g L^{-1} was achieved (Fig. 6c and d). Overall, the ACO–PLS and CoAdReS–PLS models had very similar performance in terms of RMSEC/RMSECV/RECV values. Both approaches (Table 4) significantly improved the predictive ability compared to those resulted from the use of entire spectral data, with the ACO derived models having the R^2 values >0.93, the RMSECV : RMSEC ratios around ~2–3, and the RECV% values between 2.5 and 4.3%.

In general, ACO selected approximately 1.5–2 times as many variables as CoAdReS except for DS11 where the variable numbers were very similar (Table 4). Furthermore, ACO selected variables were located within the major high intensity fluorescence bands. For the CoAdReS selected variables, these seem to correlate well with the emission properties of Comp4 which was ascribed to the glycoprotein product. However, since fluorescence emission bands are not very specific, we thus cannot discount the possibility that the variables selected for the PLS models may be attributed to a metabolite whose emission signal was correlated with glycoprotein concentration…*i.e.* a secondary correlation.

Conclusions

The use of EEM spectroscopy coupled with both factor and regression based chemometric techniques provides a convenient method for monitoring (at-line or off-line) the progress of mammalian cell culture based processes. The method uses small quantities of clarified bioprocess broth (<100 μL) which was diluted and then measured using standard off-the-shelf fluorescence spectrometers. MCR-ALS provides an interpretive tool for visualising the changes in the bioprocess in terms of several fluorescent components *e.g.* tyrosine, tryptophan, and the glycoprotein product *via* its dityrosine emission component. The relatively high resolution EEM spectra collected enabled the application of factor-based methods (MCR-ALS and among others) for more detailed analysis of bioprocess induced changes, something which was not as feasible using lower resolution filter-based spectrometers.[24] Compared to the use of Raman spectroscopy for bioprocess monitoring,[27] this method provided a clearer correlation with a product signal because there are fewer species contributing to the fluorescence emission and the dityrosine signal was thus easily discriminated.

The use of PLS regression on variables selected by either CoAdReS or ACO then provided a predictive tool for evaluating bioprocess performance in terms of glycoprotein yield. This was significant because these predictive models can be generated at the early, small scale stages of the industrial production process. Here there were ~30 samples which represent only a year's production. In practice one would recalculate the models regularly as more samples/data becomes available, and thus within a relatively short time (in the context of a manufacturing lifespan of decades) one would have a tool which can be used for early-stage process intervention. A key advantage in being able to predict the end yield at all stages from the 100 to 5000 L scale bioreactors is that one could use this prediction to intervene and stop a non-productive process at an early stage, thus saving process scale up costs. Conversely, if a process was predicted to have a potential high yield one could optimise downstream purification processes in advance, which would provide significant benefits.

Some of the other advantages with using this offline method are that it avoids issues with probe fouling in bioreactors, it generates better quality spectra in terms of signal-to-noise by virtue of using a dedicated bench-top spectrometer, and the method is more easily validated in a laboratory than on-site. These advantages outweigh the problems associated with sterile sampling and low throughput; however, these are not that significant in practice when one is monitoring a process over a long time period as is the case with many mammalian cell culture processes. Thus the time-lag associated with this method may not be very important in the context of a 10–30 day bioprocess. Finally, this method is virtually identical to the EEM methods for the analysis of the raw materials,[28,73] and cell culture media,[29] and so providing a single integrated analytical platform.

Acknowledgements

This work was undertaken as part of the Centre for BioAnalytical Sciences, funded by the Irish Industrial Development Authority (IDA) and Bristol-Myers Squibb (BMS). We thank Ms Lindy Smith (BMS) for her work in sample organization.

Notes and references

1 T. Chattaway, G. A. Montague and A. J. Morris, in *Bioprocessing, Biotechnology*, ed. H. J. Rehm and G. Reed, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2nd edn, 2008, vol. 3, pp. 319–354.
2 A. S. Rathore, R. Bhambure and V. Ghare, *Anal. Bioanal. Chem.*, 2010, **398**, 137–154.
3 J. Glassey, K. V. Gernaey, C. Clemens, T. W. Schulz, R. Oliveira, G. Striedner and C. F. Mandenius, *Biotechnol. J.*, 2011, **6**, 369–377.
4 P. L. Pham, A. Kamen and Y. Durocher, *Mol. Biotechnol.*, 2006, **34**, 225–237.
5 T. Becker, B. Hitzmann, K. Muffler, R. Portner, K. F. Reardon, F. Stahl and R. Ulber, in *White Biotechnology*, Springer-Verlag Berlin, Berlin, 2007, vol. 105, pp. 249–293.
6 L. Baldi, D. L. Hacker, M. Adam and F. M. Wurm, *Biotechnol. Lett.*, 2007, **29**, 677–684.
7 D. L. Hacker, M. De Jesus and F. M. Wurm, *Biotechnol. Adv.*, 2009, **27**, 1023–1027.
8 N. D. Lourenco, J. A. Lopes, C. F. Almeida, M. C. Sarraguca and H. M. Pinheiro, *Anal. Bioanal. Chem.*, 2012, **404**, 1211–1237.
9 J. Y. Kim, Y. G. Kim and G. M. Lee, *Appl. Microbiol. Biotechnol.*, 2012, **93**, 917–930.
10 T. Cartwright and G. P. Shah, in *Basic Cell Culture*, ed. J. M. Davis, Oxford University Press Inc., New York, 2002, pp. 69–106.

11 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.
12 L. Olsson, U. Schulze and J. Nielsen, *TrAC, Trends Anal. Chem.*, 1998, **17**, 88–95.
13 K. N. Baker, M. H. Rendall, A. Patel, P. Boyd, M. Hoare, R. B. Freedman and D. C. James, *Trends Biotechnol.*, 2002, **20**, 149–156.
14 M. Milburn, *BioPharm Int.*, 2009, 28–34.
15 L. Zang, R. Frenkel, J. Simeone, M. Lanan, M. Byers and Y. Lyubarskaya, *Anal. Chem.*, 2011, **83**, 5422–5430.
16 S. A. Bradley, A. Ouyang, J. Purdie, T. A. Smitka, T. Wang and A. Kaerner, *J. Am. Chem. Soc.*, 2010, **132**, 9531–9533.
17 E. K. Read, S. A. Bradley, T. A. Smitka, C. D. Agarabi, S. C. Lute and K. A. Brorson, *Biotechnol. Prog.*, 2013, **29**, 745–753.
18 K. Schugerl, *J. Biotechnol.*, 2001, **85**, 149–173.
19 K. A. Bakeev, *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*, Blackwell Publishing Ltd., Oxford, 2005.
20 J. S. Alford, *Comput. Chem. Eng.*, 2006, **30**, 1464–1475.
21 A. P. Teixeira, R. Oliveira, P. M. Alves and M. J. Carrondo, *Biotechnol. Adv.*, 2009, **27**, 726–732.
22 H. Kornmann, M. Rhiel, C. Cannizzaro, I. Marison and U. von Stockar, *Biotechnol. Bioeng.*, 2003, **82**, 702–709.
23 G. Jain, G. Jayaraman, O. Kokpinar, U. Rinas and B. Hitzmann, *Biochem. Eng. J.*, 2011, **58–59**, 133–139.
24 K. Hantelmann, A. Kollecker, D. Hull, B. Hitzmann and T. Scheper, *J. Biotechnol.*, 2006, **121**, 410–417.
25 B. Li, P. W. Ryan, B. H. Ray, K. J. Leister, N. M. S. Sirimuthu and A. G. Ryder, *Biotechnol. Bioeng.*, 2010, **107**, 290–301.
26 A. G. Ryder, J. De Vincentis, B. Y. Li, P. W. Ryan, N. M. S. Sirimuthu and K. J. Leister, *J. Raman Spectrosc.*, 2010, **41**, 1266–1275.
27 B. Li, B. H. Ray, K. J. Leister and A. G. Ryder, *Anal. Chim. Acta*, 2013, **796**, 84–91.
28 B. Li, P. W. Ryan, M. Shanahan, K. J. Leister and A. G. Ryder, *Appl. Spectrosc.*, 2011, **65**, 1240–1249.
29 P. W. Ryan, B. Li, M. Shanahan, K. J. Leister and A. G. Ryder, *Anal. Chem.*, 2010, **82**, 1311–1317.
30 A. Hagedorn, R. L. Legge and H. Budman, *Biotechnol. Bioeng.*, 2003, **83**, 104–111.
31 G. Wolf, J. S. Almeida, J. G. Crespo and M. A. M. Reis, *J. Biotechnol.*, 2007, **128**, 801–812.
32 J. I. Rhee and T. H. Kang, *Process Biochem.*, 2007, **42**, 1124–1134.
33 M. B. Haack, A. E. Lantz, P. P. Mortensen and L. Olsson, *Biotechnol. Bioeng.*, 2007, **96**, 904–913.
34 A. P. Teixeira, C. A. M. Portugal, N. Carinhas, J. M. L. Dias, J. P. Crespo, P. M. Alves, M. J. T. Carrondo and R. Oliveira, *Biotechnol. Bioeng.*, 2009, **102**, 1098–1106.
35 A. P. Teixeira, T. M. Duarte, M. J. T. Carrondo and P. M. Alves, *Biotechnol. Bioeng.*, 2011, **108**, 1852–1861.
36 A. Calvet, B. Li and A. G. Ryder, *J. Pharm. Biomed. Anal.*, 2012, **71**, 89–98.
37 H. Martens and T. Naes, *Multivariate Calibration*, Wiley, New York, 2nd edn, 1991.
38 J. A. Lopes, P. F. Costa, T. P. Alves and J. C. Menezes, *Chemom. Intell. Lab. Syst.*, 2004, **74**, 269–275.
39 M. J. Adams, *Chemometrics in analytical spectroscopy*, Royal Society of Chemistry, Cambridge, 2nd edn, 2004.
40 Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond and N. Jent, *J. Pharm. Biomed. Anal.*, 2007, **44**, 683–700.
41 S. Wold, M. Sjöström and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.
42 M. Andersson, *J. Chemom.*, 2009, **23**, 518–529.
43 J. P. Gauchi and P. Chagnon, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 171–193.
44 R. K. H. Galvao and M. C. U. Araujo, in *Comprehensive Chemometrics*, ed. S. D. Brown, R. Tauler and B. Walczak, Elsevier, Amsterdam, 2009, vol. 3, pp. 233–283.
45 X. B. Zou, J. W. Zhao, M. J. W. Povey, M. Holmes and H. P. Mao, *Anal. Chim. Acta*, 2010, **667**, 14–32.
46 N. Sorol, E. Arancibia, S. A. Bortolato and A. C. Olivieri, *Chemom. Intell. Lab. Syst.*, 2010, **102**, 100–109.
47 A. J. Burnham, J. F. MacGregor and R. Viveros, *J. Chemom.*, 2001, **15**, 265–284.
48 R. F. Teofilo, J. P. A. Martins and M. M. C. Ferreira, *J. Chemom.*, 2009, **23**, 32–48.
49 C. D. Brown and R. L. Green, *TrAC, Trends Anal. Chem.*, 2009, **28**, 506–514.
50 V. Centner, D. L. Massart, O. E. deNoord, S. deJong, B. M. Vandeginste and C. Sterna, *Anal. Chem.*, 1996, **68**, 3851–3858.
51 Q. J. Han, H. L. Wu, C. B. Cai, L. Xu and R. Q. Yu, *Anal. Chim. Acta*, 2008, **612**, 121–125.
52 I. G. Chong and C. H. Jun, *Chemom. Intell. Lab. Syst.*, 2005, **78**, 103–112.
53 L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen, *Appl. Spectrosc.*, 2000, **54**, 413–419.
54 R. Leardi and A. L. Gonzalez, *Chemom. Intell. Lab. Syst.*, 1998, **41**, 195–207.
55 H. C. Goicoechea and A. C. Olivieri, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1146–1153.
56 J. H. Jiang, R. J. Berry, H. W. Siesler and Y. Ozaki, *Anal. Chem.*, 2002, **74**, 3555–3565.
57 H. C. Goicoechea and A. C. Olivieri, *J. Chemom.*, 2003, **17**, 338–345.
58 L. Xu, J. H. Jiang, H. L. Wu, G. L. Shen and R. Q. Yu, *Chemom. Intell. Lab. Syst.*, 2007, **85**, 140–143.
59 H. D. Li, Y. Z. Liang, Q. S. Xu and D. S. Cao, *Anal. Chim. Acta*, 2009, **648**, 77–84.
60 M. Shamsipur, V. Zare-Shahabadi, B. Hemmateenejad and M. Akhond, *J. Chemom.*, 2006, **20**, 146–157.
61 F. Allegrini and A. C. Olivieri, *Anal. Chim. Acta*, 2011, **699**, 18–25.
62 S. Wiklund, D. Nilsson, L. Eriksson, M. Sjostrom, S. Wold and K. Faber, *J. Chemom.*, 2007, **21**, 427–439.
63 A. de Juan and R. Tauler, *Crit. Rev. Anal. Chem.*, 2006, **36**, 163–176.
64 J. da Silva, M. Tavares and R. Tauler, *Chemosphere*, 2006, **64**, 1939–1948.
65 J. C. G. E. da Silva and R. Tauler, *Appl. Spectrosc.*, 2006, **60**, 1315–1321.

66 M. Bahram, R. Bro, C. Stedmon and A. Afkhami, *J. Chemom.*, 2006, **20**, 99–105.
67 R. D. JiJi, G. G. Andersson and K. S. Booksh, *J. Chemom.*, 2000, **14**, 171–185.
68 A. Rinnan and C. M. Andersen, *Chemom. Intell. Lab. Syst.*, 2005, **76**, 91–99.
69 M. C. G. Antunes and J. da Silva, *Anal. Chim. Acta*, 2005, **546**, 52–59.
70 Y. Hu, B. Y. Li, H. Sato, I. Noda and Y. Ozaki, *J. Phys. Chem. A*, 2006, **110**, 11279–11290.
71 *Matlab*, Mathworks Inc., Cambridge, MA, 1994–2008.
72 *PLS_Toolbox*, Eigenvector Research Inc., 3905 West EaglerockDrive, Wenatchee, WA.
73 B. Li, N. M. S. Sirimuthu, B. H. Ray and A. G. Ryder, *J. Raman Spectrosc.*, 2012, **43**, 1074–1082.
74 G. S. Harms, S. W. Pauls, J. F. Hedstrom and C. K. Johnson, *J. Fluoresc.*, 1997, **7**, 283–292.
75 D. M. Haaland and E. V. Thomas, *Anal. Chem.*, 1988, **60**, 1193–1202.
76 Z. Z. Xing, B. Kenty, I. Koyrakh, M. Borys, S. H. Pan and Z. J. Li, *Process Biochem.*, 2011, **46**, 1423–1429.