Review

Pharmaceutical BIOPROCESSING

Application of spectroscopic methods for monitoring of bioprocesses and the implications for the manufacture of biologics

The ability to monitor and control bioreactor processes is an integral component to the implementation of Process Analytical Technology and Quality by Design principles. Desirable attributes of monitoring methods include the ability to monitor multiple analytes in real time with little to no sample processing. Spectroscopic methods fit these criteria and significant advancements in their application have been made. However, implementation of these systems has been hampered by their complexity. Here, we present an overview of near IR, mid-IR, Raman and fluorescence spectroscopy technologies, and the steps taken to enable their implementation as effective bioprocess monitoring tools. Specific applications for monitoring of microbial and mammalian cell bioreactors, and screening and classification of raw materials are discussed.

Nicholas R Abu-Absi*[,1], Richard P Martel[1], Amanda M Lanza[1], Stacey J Clements[2], Michael CBorys[1] & Zheng Jian Li[1]

[1]Cell Culture Process Development, Biologics Process Development, Bristol-Myers Squibb, Devens, MA, USA

[2]Development Pilot Plant, Biologics Process Development, Bristol-Myers Squibb, East Syracuse, NY, USA

*Author for correspondence: nicholas.abuabsi@bms.com

Background

It has been almost a decade since the the US FDA outlined Process Analytical Technology (PAT) principles, and later Quality by Design (QbD), meant to encourage drug manufacturers to introduce and embrace cutting edge technologies in manufacturing processes. However, the inherent complexity of industrial bioprocesses has slowed the widespread application of these principles in the manufacture of biologics. To reap the benefits of PAT and QbD, a detailed understanding of the interaction between the cells and the bioreactor environment and methods to monitor and control important process variables is required. A variety of advanced analytical methods, including HPLC, NMR, LC–MS, flow cytometry and the various 'omics approaches [1–6], have been successfully applied to advance the development of robust, high-yielding bioprocesses. It is preferable to monitor parameters inline, thus enabling their control in real time. These types of measurements are routine for physical parameters such as temperature, pH and dissolved oxygen; however, the monitoring of nutrients, metabolites and cellular components are considerably more difficult. Growth media may consist of undefined additives, such as protein hydrolysates, and chemically defined media for mammalian cell culture can consist of upwards of 50 components. Thus, sensors that are employed for bioreactor monitoring must be capable of precisely measuring low concentrations of various nutrients or metabolites without interference from the complex, multiphasic matrix inherent in bioreactor processes.

Despite the complexities of monitoring bioreactor systems outlined above, optical sensors appear to be well suited for the task [7,8]. They offer the advantages of being non-invasive, nondestructive and capable of taking measurements both offline and inline enabling quantitative analysis of multiple components in real time. A further advantage of these methods is that the measured spectra also capture global fingerprints of culture dynamics that can be used to assess current and future status of bioprocesses. The use of spectroscopy for monitoring bioreactors has been under development for over 20 years [9]. The vast majority of studies published in the literature involve applications of near IR (NIR) spectroscopy primarily due to the simplicity of the instrumentation involved.

FUTURE SCIENCE part of fsg

Mid-IR (MIR) and Raman spectroscopy applications began to gain prominence with the advent of improved optics and smaller, more powerful lasers and detectors. Other applications are emerging that apply 2D fluorescence spectroscopy for bioprocess monitoring. Each of these spectroscopic methods has advantages and disadvantages and is best suited for analysis of different components and bioreactor matrices.

Implementation of these technologies is not trivial, and several aspects inherent to the use of spectroscopic technologies must be considered. First, measurement conditions must be optimized, especially in the case of highly agitated and aerated cultures that attain high cell densities, to suit the type of technology and instrument being used, the types of components being analyzed and the nature of the sample matrix. Second, calibrations must be developed between the analytes of interest and the spectra collected from the experimental system. This involves the generation and selection of an appropriate dataset that will accurately capture the variations and complexity of the system being studied. Multivariate chemometric methods are then used to develop models transforming the measured spectra into useful information [10].

While bioprocessing applications utilizing spectroscopic methods have increased, these analytical techniques have yet to live up to their potential for innovation and control, and still lag behind applications in the manufacture of small-molecule therapeutics. The aim of this review is to summarize and evaluate the development of spectroscopic methods as they apply to bioprocesses, and highlight areas that still require particular focus to enable more widespread adoption of QbD and PAT. The principles behind the different spectroscopic technologies will be introduced and the general methods used to construct calibration models are described. Next, we will provide a summary on the development of quantitative applications for microbial and mammalian systems that are relevant to the biotech industry, then address specific applications on the production of recombinant proteins, including analysis of product titer and quality attributes. Finally, qualitative analyses for the evaluation of raw materials and process trajectories are presented.

Key Terms

Rayleigh scattering: Scattering of light by particles much smaller than the wavelength of the light. In this type of interaction, there is no change in the energy state of scattered photons but there is a change in their direction.

Raman scattering: Refers to the small fraction of photons that are scattered at different energy states than the incident light. When energy is transferred from the light to the material, it is scattered at longer wavelengths, and this process is referred to as Stokes Raman scattering. When energy is transferred from the material to the light, it is scattered at shorter wavelengths and this process is referred to as anti-Stokes Raman scattering.

Overview of spectroscopic methods

Spectroscopy is the study of the interaction between matter and electromagnetic radiation. The electromagnetic spectrum ranges from low frequency radio waves through high frequency γ-rays. Chemical bonds have characteristic vibrational frequencies that absorb energy from specific wavelengths of radiation. The study of this phenomenon is known as vibrational spectroscopy, which consists of NIR, MIR and Raman. Another technique that has been demonstrated to be of value for bioprocess monitoring, 2D fluorescence spectroscopy, is also considered in this review. The theory behind each method and the data analysis techniques that are required to interpret the spectra will be briefly introduced here. Instrumentation that is used for each of these techniques has been discussed in detail elsewhere [9,11–12], and will not be considered here.

NIR and MIR spectroscopy measure absorption, in either transmission or reflectance modes, from dispersed polychromatic radiation. The NIR region of the spectrum corresponds with wavelengths of 780–2500 nm and the MIR region is between wavelengths of 2500 and 40,000 nm. Absorbances are also commonly expressed as a function of the wavenumber, which is directly proportional to the frequency, in units of inverse centimeters. NIR spectra consist of absorbances due to the combinations and overtones of fundamental vibrations. Most of the information contained in these spectra result from a change in the dipole moment of a molecule due to the stretching and bending of covalent hydrogen bonds (i.e., C-H, N-H and O-H bonds). Therefore, aqueous solutions are dominated by water peaks, which can interfere with the spectra of the analytes of interest, in both the NIR and MIR regions [13]. Absorbances occurring in the NIR region are typically 10–100-times less intense than those occurring in the MIR region. While NIR spectra consist of broad overlapping peaks, MIR spectra consist of well-defined peaks due to absorbances from characteristic bonds, which makes quantitative information on specific analytes more readily available [14]. There are two types of IR instruments used – dispersive and Fourier transform (FT)IR. FTIR instruments have several distinct advantages over the dispersive type including higher throughput and accuracy [9,11].

Raman spectroscopy is complementary to MIR, yielding different intensities and selectivity. Raman measures inelastic scattering from a monochromatic radiation source (i.e., lasers). It may take as many as 10^8 excitation photons to give rise to one Raman pho-

ton, therefore Rayleigh scattering, which dominates Raman scattering, must be filtered out. However, Raman spectra are not subject to large interference from polar molecules such as water, and the fiber optic cables used are not as expensive and fragile as MIR fibers [12]. Selection of appropriate excitation wavelength involves consideration of the Raman signal intensities (and potential impacts to exposure times to compensate) and the potential for interference from background fluorescence. For example, fluorescence can be minimized by utilizing excitation wavelengths in the NIR region; however, exposure times may need to increase to compensate for reduced Raman signals. Thus, in the case of some high cell density microbial systems with high nutrient consumption rates, the use of long wavelength lasers and higher exposure times may not be compatible with the dynamics of the system.

A fourth spectroscopic method applied to bioreactor monitoring is fluorescence spectroscopy. When a fluorophore absorbs a photon, it is excited to a higher energy state. As the energy of the molecule drops to a lower energy state, it emits a photon at a different energy, and thus frequency. The range of frequencies of emitted photons from a particular excitation frequency is known as the emission spectrum. Constructing a matrix consisting of the emission spectrum for a range of excitation frequencies is known as 2D fluorescence. This method can be used to collect information on all the fluorescent compounds in a sample.

Analysis of spectroscopy data

The application of spectroscopic methods for the analysis of bioreactors results in complex spectra with contributions from an aqueous phase, biomass and gas bubbles. Although chemically defined media formulations are becoming more common for the production of recombinant proteins, in the case of mammalian cell cultures, these formulations may consist of more than 50 components. The methods under discussion here produce datasets with overlapping spectra from all of these components in addition to cellular metabolites, biomass (proteins, nucleic acids, lipids and so on) and other matrix components (e.g., antifoam, microcarriers and bubbles). In general, multivariate data analysis methods, such as partial least squares (PLS) regression or principal component regression, are required to extract meaningful information from the acquired spectra. This requires the use of a calibration dataset to train the models to best fit the system.

In depth discussions of the chemometric approaches used for data analysis have be presented elsewhere [10–11,15–16] and are beyond the scope of this review. The general calibration procedure that is followed can be broken up into six categories for the purpose of discussion as shown in Figure 1. The work that has been presented in the literature will be discussed in the context of these individual procedures. The calibration scheme refers to the strategy used to generate samples for the calibration dataset. Three different strategies have been summarized [16]: the use of synthetic mixtures of the components of interest, the standard approach using real process samples, and the so-called adaptive calibration that relies on the addition of various components to process samples to break correlations. Appropriate samples are generated according to the calibration scheme being used and measurements are taken using both spectroscopy and reference methods for the analytes of interest. The data are then generally preprocessed, which refers to any mathematical manipulations performed on the data geared towards the reduction or removal of irrelevant sources of variation that may mask the variation of interest [10]. Some preprocessing methods commonly employed are normalization, weighting, smoothing and baseline corrections. Multiplicative scatter correction and derivitives can be used for baseline correction of spectroscopy data. Mean centering subtracts the mean value for a given variable from all the elements resulting in easier discernment in small differences between samples [10]. Next, in order to improve the prediction accuracy of particular analytes, certain wavelength regions corresponding to a particular component, as determined by analysis of pure component spectra or various statistical approaches, may be specified for modeling. References that contain either MIR or Raman spectra for

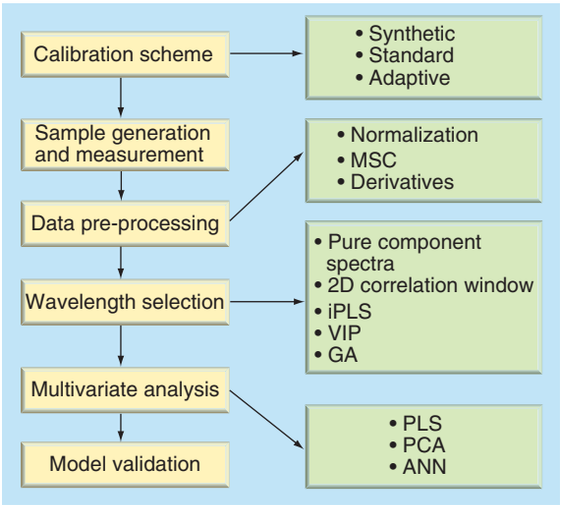

Figure 1. Procedures used for construction of models.
ANN: Artificial neural network; iPLS: Interval partial least squares; GA: Generic algorithm; MSC: Multiplicative scatter correction; PCA: Principal component analysis; PLS: Partial least squares; VIP: Variable influence on projection.

components of interest for bioprocess monitoring are shown in Table 1. Various statistical approaches, such as 2D correlation windows, interval PLS (iPLS), variable influence on projection and genetic algorithms, which look for model improvements by using optimal subsets of variables, can also be used for wavelength selection in models.

Multivariate analysis methods, such as PLS, principal component analysis (PCA), principal component regression or artificial neural networks (ANNs) are

Table 1. References containing pure component mid-IR or Raman spectra for analytes of interest for bioprocessing applications.

Chemical group	Components	MIR ref.	Raman ref.
Alcohol	• Ethanol	[17,18]	
Amino acids	• Alanine	[19,20]	[21,22]
	• Arginine; glutamic acid (glutamate); glycine; histidine; proline; serine; tryptophan; tyrosine; valine	[20]	[21,22]
	• Asparagine; glutamine	[19,20]	[21]
	• Aspartic acid; cysteine; isoleucine; leucine; lysine; methionine; threonine	[20]	[21]
	• Phenylalanine	[20]	[21–23]
DNA and RNA bases	• Adenine; cytosine; guanine; thymine; uracil		[22]
Fatty acids and fats	• 12-methly-tetradecanoic acid; 13-methylmyrisitic acid; 14-methylmyrisitic acid; 14-methylhexadecanoic acid; 15-methylpalmitic acid; vaccenic acid; glycerol; lauric acid; myrisitic acid; oleic acid; palmitic acid; stearic acid; triolein; trilinolein; trilinolenin		[22]
Hormones	• Human growth hormone		[24]
	• Insulin		[25]
Organic acids	• Acetate	[17]	[23]
	• Formate		[23]
	• Gluconic acid; keto-gluconic acid	[17]	
	• Lactate	[19,26]	[23]
Primary metabolites	• Acetoacetate; acetylcoenzyme a; coenzyme a; fumarate; malic acid; phosphenolpyruvate; pyruvate		[22]
	• Citric acid	[27]	
	• Succinic acid (succinate)		[22,23]
Salts	• Magnesium sulfate; nitrate; potassium phosphate monobasic		[28]
	• Sulfate		[23]
Sugars	• *N*-acetyl-D-glucosamine; amylopectin; amylose; arabinose; cellulose; chitin; dextrose; D-fructose-6-phosphate; fucose; galactosamine; lactose; mannose; trehalose; xylose		[22]
	• Fructose	[17]	[22,28]
	• Glucose	[17–19,26–27]	[22–23,28]
	• Sucrose	[17,27]	[28]
Other compounds	• β-carotene; ascorbic acid; glutathione; riboflavin		[22]
	• Ammonia	[19]	
	• Gluconatacean	[17]	
	• Imidazole	[20]	
	• Yeast	[18]	

MIR: Mid-IR.

then employed to correlate the measured concentrations of components of interest with the acquired spectra. The final step involves validation of the models by comparing model predictions with reference method measurements on independent and representative samples. Rigorous validation is a critical step in ensuring that predictions are accurate under the range of expected process variations This point is emphasized by Shaw and coworkers who stated, "It is not difficult to carry out statistical analysis on multivariate data in such a way that the results appear much more impressive than they really are" [15]. As such, spectroscopic monitoring of bioprocesses is discussed here in the context of the various approaches taken to ensure that high-quality spectroscopic measurements and model predictions can be attained for a particular system.

Monitoring of bioreactors using spectroscopic methods

Spectroscopic methods are of particular importance to the biotechnology industry since they are simple to operate and implement inline for real-time measurements. Furthermore, a single system generates global fingerprints of bioreactor cultures that can be analyzed to produce quantitative predictions of specific components. These applications have been reviewed by multiple authors [11,14,29–31]. The reviews by Scarff *et al.* in 2006 and Cervera *et al.* in 2009 [9,16] provide thorough discussions on the practical aspects of implementing NIR spectroscopy for bioreactor monitoring. Although these reviews focused solely on NIR applications, the practical sampling and calibration considerations and chemometric methods apply to other spectroscopic methods. The review by Scarff *et al.* was a comprehensive analysis of NIR applications organized by complexity of the bioreactor matrix from simple (anaerobic biotransformations) to complex (highly agitated, aerated cultures of filamentous fungi). Cervera *et al.*'s review presented various strategies used to address practical issues encountered when using NIR for bioprocess monitoring. The intent of this review was not to repeat what was presented in previous reviews, but to complement these publications by focusing on MIR and Raman applications and including only the most recent NIR applications.

Microbial fermentation

Microbial spectroscopic applications focus on analysis of carbon sources (glucose, glycerol, fructose and starches), nitrogen sources (ammonium), secondary metabolites (carotenoids) and metabolic byproducts (ethanol and acetate). A summary of the literature references describing implementation of spectroscopic methods for monitoring of microbial processes is listed in Table 2. This includes information on the organism cultured, spectroscopy type, sampling method and sample matrix for each application. Results, such as standard errors of prediction, from each of the models are not included in this table, but have been summarized elsewhere [16,32]. While microbial media tends to contain fewer components, a unique challenge associated with microbial systems is the diversity of sample matrices encountered from different organisms, and the large changes in biomass that can occur over relatively short times. These issues are reflected in the large variety of approaches taken for constructing calibration matrices. A few of the unique approaches employed to improve model robustness and accuracy are described below.

Development of accurate and robust calibration models requires appropriate selection of the spectroscopic method, calibration strategy and chemometric tools to suit the system. Aspects to consider include instrument cost and availability, where measurements will be performed (offline, online or inline), the sample matrix, spectra of the analytes of interest relative to the sample matrix, and concentrations of the analytes of interest. Cannizzaro and coworkers reported on a system producing carotenoids [39] where the choice of Raman spectroscopy was straight forward due to a resonance effect that enhances the Raman signals of carotenoids. Unfortunately, the choice of spectroscopic methods is rarely that simple and is difficult in part because there is a lack of peer-reviewed studies that directly compare the relative merits of different measurement methods on the same system. Sivakesava and coworkers published some relatively early studies evaluating FT-MIR and FT-Raman for monitoring of ethanol production using yeast [38] and also FT-MIR, NIR and FT-Raman for analysis of lactic acid fermentation [37]. In both cases, the authors concluded that MIR was the more promising technique due to its specificity and stronger signals, although it is unclear if recent advances in Raman detectors would affect this conclusion. Another study also compared FT-MIR and Raman measurements for monitoring of a gibberellic acid process [36]. Both methods yielded models with similar performance, with the authors preferring the Raman measurement as it required no sample processing. Work by Crowley and coworkers developing FT-MIR calibrations for yeast cultures suggest that quantitative MIR models are less complex, and thus, more robust than NIR models [35].

Microbial cultures generally use media containing a small number of components and have well-characterized metabolism, which can allow the use of simple calibration schemes for some applications. Examination of pure component spectra often yields informa-

Table 2. Literature employing vibrational spectroscopy for monitoring of microbial bioreactors.

Organism	Method	Measurement mode	Sample matrix	Analytes	Ref.
Escherichia coli	MIR and pyrolysis MS	Offline	Cell paste and supernatant	IFN-α2 production	[33]
Saccharomyces cerivisiae	Raman	Online	Supernatants	Ethanol and glucose	[34]
Pichia pastoris	MIR	Offline	Frozen filtrates	Methanol, glycerol and product (heterologous protein)	[35]
Gibberella fujikuroi	Raman, MIR and pyrolysis MS	Offline	Fermentation broth	Gibberellic acid	[36]
Lactobacillus casei	Raman, NIR and MIR	Offline	Fermentation broth	Glucose, lactic acid and biomass	[37]
S. cerevisiae	Raman and MIR	Offline	Fermentation broth	Glucose, ethanol and biomass	[38]
Phaffia rhodozyma	Raman	Online	Fermentation broth	Intracellular carotenoid	[39]
Gluconacetobacter xylinus	MIR	Online *in situ* probe	Fermentation broth	Fructose, acetate, ethanol and gluconacetan	[40]
G. xylinus	MIR	Online	Fermentation broth	Fructose, acetate and gluconacetan	[41]
G. xylinus	MIR	Online	Fermentation broth	Fructose, acetic acid, ethanol, gluconacetan, phosphate and ammonium	[42]
E. coli	Raman	Online	Fermentation broth	Glucose, lactate, formate, acetate and phenylalanine	[43]
Streptomyces clavuligerus	MIR	At-line	Fermentation broth and cell filtrate	Glycerol and clavulanic acid	[44]
P. pastoris	MIR	Online	Fermentation broth	Methanol	[45]
S. cerevisiae	MIR	Inline ATR probe	Fermentation broth	Glucose, glycerol, ethanol, acetate and ammonium	[46]
E. coli	MIR	Online and offline	Fermentation broth and filtered supernatant	Glucose, glycerol, ammonium, and acetate	[13]
E. coli	MIR	Online and inline	Fermentation broth	pH	[47]
S. cerevisiae	MIR	Online	Fermentation broth	Glucose, ethanol, ammonium, phosphates, glycerol and acetic acid	[48]
Streptomyces coeliicolor	Multiwavelength fluorescence	Online	Fermentation broth	Biomass and casamino acid	[49]

ATR: Attenuated total reflection; MIR: Mid-IR; NIR: Near IR.

Table 2. Literature employing vibrational spectroscopy for monitoring of microbial bioreactors (cont.).

Organism	Method	Measurement mode	Sample matrix	Analytes	Ref.
S. cerevisiae	Raman	Online and offline	Fermentation broth	Glucose, ethanol and glycerol	[50]
E. coli, P. pastoris, Streptomyces toxitricini and Aspergillus niger	NIR	Inline	Fermentation broth	Linoleic acid, oleic acid and ammonia	[32]
P. pastoris	NIR	Online	Fermentation broth	Glycerol and biomass	[51]
S. cerevisiae	Raman	Online probe	Fermentation broth	Starch, dextrins, maltotriose, maltose, glucose and ethanol	[52]
Microalgae	Raman	Offline	Fermentation broth	Glucose	[53]

ATR: Attenuated total reflection; MIR: Mid-IR; NIR: Near IR.

tion to help determine best practices for generating calibration datasets and modeling. For example, the strong peak located at 883 cm^{-1} in the Raman spectrum of ethanol allowed Gray and coworkers to apply a univariate model for this analyte, while predictions of total starch, dextrins, maltotriose, maltose, and glucose required multivariate methods due to spectral overlap [52].

The schemes employing multivariate modeling generally require generation of large datasets for model calibrations. A method that updates calibrations in real time was implemented to reduce the number of samples required in the original dataset [41]. An initial model was constructed using just 14 samples in the calibration, and additional spectra collected during process operation were added. The consistency of the added data is then checked using elemental balances of carbon and oxygen from off-gas analyses. The authors demonstrated improvements in model accuracy using real-time model adaptation for both normal processing conditions and cases where new substrates were added. This method was further developed through the use of carbon, nitrogen, degree of reduction and charge balances. These balances were performed using online estimation of nutrients, metabolites, biomass, base consumption and off-gas concentrations [48].

The accuracy of multivariate models is vulnerable when samples with characteristics outside those included in the calibration dataset are encountered. However, in process development settings conditions are constantly varied and different cell lines are used. In these cases, a reduction in the number of calibration samples and the design of calibration strategies capable of accurately predicting analytes under a range of process conditions could make the implementation of these technologies more widespread. Lee and coworkers presented a method to address this issue using a library of pure component spectra to predict analyte concentrations [43]. *In situ* Raman spectroscopy was used to predict concentrations by assuming that measured spectra were the result of a linear combination of the pure component spectra. Schenk and coworkers also adopted this approach for FT-MIR measurements [46] and extended it by adding drift spectra obtained by factor analysis to the library to compensate for signal intensity drifts and noise problems [13]. The authors demonstrated that this method can successfully be applied to scenarios commonly encountered in a process development setting. The robustness of this approach was shown using nine different process conditions consisting of different medium formulations, and three different *Escherichia coli* strains.

An example of using a unique calibration dataset consisting of both synthetic mixtures and bioreactor samples to improve model predictions was presented by Kornmann and coworkers [40]. This was demonstrated by comparing the performance of models calibrated using three different datasets. These datasets were made up of either 12 fermentation samples; 12 fermentation samples and 42 synthetic mixtures consisting of modeled analytes; or 12 fermentation samples and 76 synthetic mixtures consisting of model analytes and two additional byproducts. Models were improved by addition of synthetic standards and were most accurate when the synthetic standards included two byproducts that were not modeled. It is not clear how much of the model improvement is due to the different types versus the number of standards employed.

The choice of chemometric methods is also difficult to determine *a priori* and often requires significant trial

and error to find models that best fit the system under study. Selection of suitable preprocessing is best when it is based on knowledge of the chemistry of the system, but optimization will require iteration. A variety of data preprocessing methods have been described [10] and illustrations of the impact of various pretreatments on spectroscopy data have been shown [34]. Pretreatments are often applied to reduce the effects of the sample matrix on the analytes of interest. In some cases, cells are removed prior to analysis to eliminate these effects [44]; however, it is unclear whether or not appropriate preprocessing can achieve the same result to enable inline monitoring applications. Similarly, in the case of Raman instruments, the wavelength of the excitation laser is often chosen in the NIR region to minimize interference from fluorescence signals, although Oh and coworkers report the use of rolling circle filters to remove these background effects [53]. McGovern and coworkers generated models for gibberellic acid production using both PLS and ANN, but pointed out that when all of the collected data are used during modeling, it is difficult to determine which regions of the spectra are relied upon for the model predictions [36]. Thus, it is difficult to confirm that the models are using regions of the spectra that correspond with the analytes of interest. The accuracy of multivariate models can be improved through examination of pure component spectra to specify wavelength regions to include for particular components [44,54] or to identify regions to exclude from analysis due to colinearity of multiple components [52]. Different variable selection methods applied to spectroscopy data have also been directly compared with genetic algorithms and PLS bootstrap methods, both improving model accuracy [36,55]. This also allowed for the determination of the input variables that contributed the most to the models enabling verification that models are based on the analytes in question.

While there are a large number of publications describing procedures for the development of spectroscopy as a process monitoring tool, there are few publications describing the practical use of these data, although these are beginning to emerge. Fazenda and coworkers describe the application of NIR measurements to close mass balances for metabolic flux analysis [51]. While this successfully demonstrated that the information provided by measured spectra is useful for better process understanding, the ultimate goal of developing these technologies is to use the information to manipulate bioreactor inputs for process improvement. A simple application was presented using univariate calibration of methanol concentrations with FT-MIR spectra to control a methanol feed for a *Pichia pastoris* process [45]. Methanol is used as both a carbon source and an inducer of protein expression, but concentrations must be maintained below inhibitory levels. An attenuated total reflection FT-MIR probe was used to generate a two-point calibration, which was performed at the beginning of the run. This calibration scheme was possible since methanol was shown to be the component that significantly influenced MIR spectra in this system. Although no productivity gains were realized in this case, predicted methanol concentrations were used to control the feed rate such that methanol was maintained at a desired concentration.

Similarly, Kornmann and coworkers described a *Gluconacetobacter xylinus* culture using ethanol as a substrate for conversion to acetate, which is then metabolized to form biomass [42]. Fructose was also supplied as a substrate, which was converted into the product of interest, gluconacetan. The authors used a set of synthetic calibration standards containing fructose, ethanol, acetate, gluconacetan, ammonium and phosphate to construct PLS models from *in situ* MIR spectra. These predictions were used to adjust ethanol feeding to maintain constant acetate concentrations so as to prevent buildup of ethanol to toxic levels. At the end of the growth phase, feeding was switched from ethanol to fructose with the aim of maintaining constant fructose levels. Precise control of a fed batch process using MIR resulted in a 60% increase in productivity over a batch process. Successful demonstration of this feeding strategy shows the robustness and accuracy of the MIR predictions.

In one of the better illustrations of spectroscopy for bioreactor monitoring and control, Tiwari and coworkers describe a *Streptomyces toxitricini* fermentation that requires a specific ratio of linoleic acid to ammonium to maintain productivity of lipstatin [32]. Using the conventional control strategy, it was difficult to mantain optimal nutrient ratios resulting in periodic, 20% decreases in productivity. Calibration models were constructed for prediction of linoleic acid, oleic acid and ammonium concentrations from NIR spectra measured inline using PLS. The authors used a calibration matrix consisting of fermentation supernatants, biomass and analytes of interest mixed in different proportions and under different measurement conditions. This attention to detail resulted in extremely accurate predictions of linoleic acid and oleic acid at very low levels (below 0.1 g/l). The data collected inline were used to manipulate nutrient feeds to maintain the linoleic acid to ammonium ratio at the optimal level resulting in more consistent process performance.

Cell culture applications

The development of spectroscopic methods for mammalian cell culture applications is not as advanced compared with microbial systems. This is probably

due to cell culture processes being used exclusively for the production of recombinant proteins, whereas microbial systems are used in multiple industries. Nonetheless, the demand for lower cost biologics for the treatment of a variety of diseases drives the need for higher yielding cell culture processes that consistently produce proteins with desired quality profiles. It is expected that spectroscopic methods will play a critical role in enabling more efficient monitoring and control of these processes. A summary of the literature describing implementation of spectroscopic methods for monitoring of cell culture bioreactors is listed in Table 3. Initial studies geared towards developing techniques for spectroscopic monitoring of cell culture bioreactors have been reviewed [9,16]. A major advancement in the technology was made by Arnold and coworkers in 2003 with the demonstration of *in situ* monitoring of CHO cell cultures using NIR [56]. Models were successfully developed for glucose, lactate, glutamine and ammonium using calibration samples derived from inline monitoring of the actual process. The authors argue that adopting this methodology gives greater confidence that the models generated are valid, and this seems to have become the accepted approach. Synthetic calibration schemes are less likely to capture the complexity of cell culture systems and adaptive calibration strategies are difficult to implement inline since this procedure would involve manipulating the cell culture environment and risk creating unfavorable and unrepresentative culture phenotypes. Furthermore, while adaptive calibration strategies have been used to decrease the numbers of factors used in models, this can skew the calibration dataset towards higher concentrations of components and is counter to the general cell culture strategy of maintaining lower nutrient and waste product levels in the cultures [57].

Subsequent publications evaluating spectroscopy as a monitoring tool for cell culture processes follow the practice of using bioreactor samples for calibration and validation datasets. Specifically, several publications demonstrating the utility of NIR monitoring have emerged [58–59,66]. The number and types of components being monitored expanded with Card and coworkers who developed calibration models for cell density and pH in addition to the more commonly modeled glucose, lactate, ammonia and glutamine [59]. The pH predictions were probably the result of deprotonation equilibria affecting the spectra [47]. While first and second derivative pretreatments are routinely employed to normalize baseline drifts and enhance spectral features, the cell density models were developed with no derivative applied to the spectra. This suggests that the information contained in the baseline could have been used to model cell density. Limited information on these models was presented making evaluation of the performance of this model difficult.

A subsequent publication addressed measurement of VCD using NIR, suggesting that spectroscopy may be able to detect cellular changes associated with changes in intracellular pH as a result of loss of cell membrane activity [60]. An alternate mechanism presented by the authors is that NIR may be able to detect the loss of complex lipid raft structures that follows cell death. Regardless of the mechanism, the authors generated NIR models capable of predicting VCD from microcarrier cultures, which is particularly laborious to measure using traditional techniques. Of note, second derivative preprocessing of the spectra was determined to be capable of minimizing light scattering effects of the microcarriers, while first derivative pretreatment ensured that light scattering from cells would still be detectable.

In the last few years, alternative means to generate calibration datasets for NIR spectra have been presented. While developing calibrations for glucose and lactate for Vero cultures grown on microcarriers, researchers showed that due to changes in the composition and matrix of samples measured inline versus offline, it was necessary to construct models using only inline samples that accurately reflect the bioreactor matrix [62]. The dynamics of fed-batch systems and technology transfer to large-scale systems can result in variations that may not have been accounted for in calibration procedures applied at small scale. Thus, the authors implemented a novel approach for introducing variability into the calibration dataset by intentionally manipulating bioreactor conditions by altering feed protocols and using different microcarrier concentrations. These changes resulted in different glucose, lactate and VCD profiles presumably increasing the robustness of the resulting models.

A sophisticated calibration approach was taken by Hakemeyer and colleagues to develop a general model for offline analysis of CHO culture supernatants using FT-NIR [66]. Here, the authors describe different analyses of over 1000 samples collected from 100 bioreactor runs performed at different scales. With these data, the authors were able to extract both qualitative and quantitative information from supernatant samples. First, the authors used multiway PCA to demonstrate that the score plots obtained from sample analysis using eight different reference methods were similar to those obtained from the NIR measurements. This suggests that process trajectories can be mapped from NIR samples and allow speedy evaluation of run status to determine if further action is needed to correct potential performance deviations. The authors then went on to describe development of quantitative models to pre-

Table 3. Literature employing vibrational spectroscopy for monitoring of cell culture bioreactors.

Sampling	Sampling matrix	Spectroscopy type	Analytes	Ref.
Inline	Bioreactor samples	NIR	Glucose, lactate, glutamine and ammonium	[56]
Inline	Bioreactor samples	NIR	Glucose, ammonium, product titer, methionine, lactate, glutamate and glutamine	[58]
Online	Bioreactor samples	NIR	Glucose, lactate, glutamine, ammonium, pH and VCD	[59]
Inline	Bioreactor samples	NIR	VCD with microcarriers	[60]
Offline (96 well)	Dried supernatants	MIR	Glucose, antibody titer and lactate	[61]
Inline	Bioreactor (microcarriers)	NIR	Glucose and lactate	[62]
Inline	Bioreactor samples	Raman	Glucose, lactate, glutamine, glutamate, ammonium, VCD, TCD and viability	[63]
Inline	Synthetic mix with media	MIR	Glucose, glutamine, glutamate, lactate, sodium carbonate, potassium phosphate and ammonium sulfate	[64]
Inline	Bioreactor samples	Raman	Glucose, glutamine, glutamate, lactate, ammonium, VCD and osmolality	[65]
At-line	Supernatants	NIR	Product titer, glucose, glutamate, glutamine, lactate and osmolality	[66]
Inline	Bioreactor samples	Raman	Glucose, glutamine, glutamate, lactate, ammonium, TCD and VCD	[67]
Inline	Bioreactor samples	MIRS NIRS	VCD, glucose, product titer and LDH	[68]
Offline	supernatant	Raman	Glucose, lactate, antibody titer	[69]

LDH: Lactate dehydrogenase; MIR: Mid-IR; NIR: Near IR; TCD: Total cell density; VCD: Viable cell density.

dict analyte concentrations from the NIR spectra. A subset partitioning of x and y spaces algorithm [70] was used to divide the data into calibration and validation sets such that the calibration set covers all sources of variation in the data. Quantitative models of samples from a single cell line and media were obtained using PLS for eight parameters and validated using samples from three separate runs. Wavelength selection for each parameter was performed using both iPLS and variable influence on projection with iPLS being the preferred method. General models using data from three cell lines using similar processes were developed for product, glucose, glutamate, glutamine, lactate and osmolality. It was also noted that accurate models for lactate dehydrogenase, ammonium and viability were not achieved from this dataset. Given the large amount of data included, this suggests that there could be a limit in terms of quantitative modeling of parameters using NIR, especially if one assumes that the number of accurate models generated will decrease in more complex matrix-containing cells.

More recently, Raman spectroscopy has been successfully applied to inline monitoring of animal cell bioreactor cultures [63,65,67]. The first such demonstration was by Abu-Absi and coworkers [63]. In this case, the calibration dataset consisted of normal runs and a run where inoculum conditions were manipulated to yield different parameter profiles. Furthermore, component concentrations were calculated after feed additions to increase the number of samples and their variability in the dataset. Models were developed using PLS and yielded accurate predictions of glucose, lactate, glutamine, glutamate, ammonium, VCD and viability. Moretto and coworkers used an almost identical approach with the introduction of an autosampler to quickly generate large datasets that capture variations through the entire run [65]. Whelan and coworkers described the application of Raman spectroscopy to improve a process [67]. Models were first generated using bioreactor cultures receiving daily bolus feed additions and then used in a feedback control loop with continuous feeds to maintain glucose at 11 mM. Data collected from the continuous feed runs were then used to update the models to improve their predictive capacity.

A thorough and systematic evaluation of the potential for MIR for PAT was conducted by Foley and coworkers in 2012 [64]. This work included determina-

tion of the LOD for glucose, glutamine, ammonium, phosphate, glutamate, lactate, 4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid and bicarbonate. In addition, a comparison between a flexible fiber optic and a fixed conduit for light collection was made, and a study was conducted of how various matrix factors influence the predictions. In this work, synthetic mixtures were used in the calibration dataset while the validation dataset consisted of synthetic mixtures and cell culture medium spiked with known concentrations of various model analytes. The authors showed that determination of LOD of pure component spectra using PLS is significantly lower than visual evaluation of spectra by eye, and that LOD increased when using cell culture media rather than pure component spectra. Evaluation of the optics revealed that the fiber optic performed similarly or worse than the fixed conduit for every analyte. The external influence analysis indicated that agitation speed and antifoam have no impact on predictions, while biomass, pH, temperature and pluronic all can impact predictions. The authors conclude that MIR has limited potential as a process monitoring tool for animal cell cultures.

Given the complex development process of the tools described above and the lack of clear methods to validate the accuracy and robustness of models, it is important to be aware of the limitations inherent to these systems. Primarily, measured spectra are correlated to measurements taken using reference methods and the models that are generated are predictions and not direct measurements of components of interest. As such, model accuracy will at best approach the accuracy of the reference methods employed. Furthermore, there are numerous other important medium components, such as trace elements, that have profound effects on culture systems and are not measured by the technologies described here.

Monitoring protein titers & quality attributes using spectroscopy

The primary products of cell culture processes are recombinant proteins. The ability to monitor protein titers and quality attributes inline would significantly enhance the ability to produce high protein yields with desired properties. Current methods for quantifying protein attributes include ELISA, circular dichroism, MS and chromatography; however, none of these techniques are capable of yielding real-time bioreactor data. The use of spectroscopic methods has recently been demonstrated for these types of measurements to analyze protein formulations, thus representing an important area of potential innovation for bioreactor monitoring and control [71,72]. Several cases have recently demonstrated the application of spectroscopy for accurate measurement of protein concentration in bioreactors. Sellick and coworkers applied FT-IR spectroscopy to mammalian cell culture samples to measure the concentration of a secreted monoclonal antibodies while simultaneously measuring glucose and lactate in the media [61]. Supernatant was collected from four monoclonal antibody-producing cell lines and two untransfected control cell lines over a 7-day period. FT-IR spectra were collected for each sample and significant spectral changes were found in the amide I and polysaccharide regions using PCA and discriminant functional analysis. Using PLS regression, models were developed to calculate protein, glucose and lactate concentrations from FT-IR spectra, and those models accurately predicted concentrations for all three attributes with an error less than 10%.

Another approach for online monitoring of protein concentration involves 2D fluorescence. While this technique is well established for cofactor quantification and has been used to monitor microbial growth, it has only recently been explored with mammalian bioprocesses [73,74]. Researchers evaluated the applicability of 2D fluorescence to measure recombinant protein concentration and viable cell density using an IgG-producing baby hamster kidney cell line. Fluorometric maps were collected from batch and fed-batch cultures, resulting in a VCD range of 1.8–4.5 million cells and titer from 5 to 20 mg/l. A PLS model was generated linking the fluorescence data with VCD and protein concentration, which were definitively determined by a trypan blue exclusion method and ELISA. This model resulted in 91 and 99% accuracy for VCD and protein concentration, respectively. Next, they evaluated the use of synchronous fluorescence spectroscopy, in conjunction with chemometrics, to accurately measure VCD and protein titer [75]. Data collection with synchronous fluorescence spectroscopy is faster because excitation and emission wavelengths are simultaneously scanned. Finally, the same researchers augmented this work by introducing a 96-well format for predicting VCD and protein concentration from 2D fluorometric mapping of cell culture samples [74]. Their method was evaluated using three Chinese hamster ovary (CHO) cell clones, and gave good predictions even for those concentrations outside the calibration range of the assay. An earlier study also showed the incorporation of a multiwell plate with Raman spectroscopy for qualitative analysis of aqueous, but cell-free, solutions [76]. This work dem-

Key Term

Chemometrics: Use of statistical and mathematical techniques to analyze chemical data.

onstrates how a spectroscopic method can be adapted to a cell culture process, and used to both rapidly and accurately measure critical culture attributes. Furthermore, the incorporation of a 96-well format represents a step towards high-throughput, real-time monitoring of cell culture conditions, which is applicable to early-stage bioprocess development.

Protein structure and post-translational modifications (PTMs) [77] are other critical factors ideal to monitor in real time in a recombinant protein bioprocess. Although a protein may be formed correctly when it is expressed, misfolding and aggregate formation can occur during the bioreactor, purification or formulation phases of a manufacturing process [78]. Traditional methods for measuring aggregates and PTMs are generally time consuming, require extensive sample preparation and can only be implemented offline. Spectroscopy is emerging as a potential means for monitoring protein structures and PTMs [72]. Raman spectroscopy, in particular, can be used to probe secondary and tertiary protein structures including amides, aromatic side chains and sulfur bridges [79]. A study was conducted using vibrational spectroscopy to examine the structures of monoclonal antibody aggregates and unfolded species generated through several stresses that occur during protein formulation processes [80]. Raman spectroscopy has also been used to predict concentrations of phosphorylated protein [81] and to determine glycosylation structure [77]. For the determination of glycosylation, the authors selected bovine pancreatic ribonuclease proteins RNase A and RNase B as model proteins since they have identical sequence and structure except for a single N-linked glycan at the asparagine 34 residue of RNase B. They prepared 21 mixtures of RNase A and B at varying concentrations and used PLS regression to build a model to predict glycosylation levels. Although this work is limited to one protein and a single glycan structure, it demonstrates the application of Raman spectroscopy to monitor and predict glycosylation levels of a protein. While these approaches have not yet been applied to bioreactor systems, their potential to provide real-time predictions of aggregate formation and PTMs has been demonstrated. However, more work must be carried out to adapt these techniques to the complexity of cell culture broth.

Key Term

Post-translational modifications: Modifications that are made to proteins including phosphorylation, acetylation, trimethylation, ubiquitination and, most commonly, glycosylation, which are critical considerations for therapeutic proteins as these modifications can affect protein stability, function, pharmacokinetics and immunogenicity.

Utilizing spectroscopy for raw material screening

Spectroscopic methods have significant utility for online monitoring of microbial and mammalian cell culture, as demonstrated above. However, the applications of these methods extend beyond live cultures into the equally critical areas of raw material and media analysis. Despite best efforts to control media components, it is well known that differences can arise between vendors, and even different lots from the same supplier, that can impact process performance [82]. Although chemically defined media is relatively consistent, it is highly complex and composed of many components including vitamins, minerals, amino acids and sugars. Furthermore, some cell culture media components are not chemically defined, such as hydrolysates, and high levels of variation are normal [83–85]. Identifying and controlling this variability by characterizing raw materials and media would be advantageous. While current methods are available to do this, they typically utilize NMR, MS or HPLC technology, which is prohibitively time consuming, expensive and labor intensive [84,86–87]. For this reason, there is significant interest in developing rapid methods for accurately assessing the quality of cell culture media components.

A generic approach to raw material screening was recently demonstrated with a handheld Raman spectroscopy device [88]. This instrument is small, easy to use and the researchers outline a generic approach that can be applied to any material of interest. The authors present approaches for identification of raw materials from a library and for analysis of quality or purity. One such raw material used in industrial bioprocess is yeastolate, a complex biological material prepared from lysed yeast strains. A recent study compared two methods, surface-enhanced Raman spectroscopy (SERS) and excitation–emission matrix (EEM) spectroscopy, for characterization of yeastolate samples [83]. SERS relies on the absorption of molecules onto rough metal surfaces or nanoparticles to enhance the signal and reduce background compared with Raman spectroscopy, which is of particular use when low signals are encountered. When applied to yeastolate samples, researchers found that SERS provided more information compared with EEM, can be used to characterize and distinguish yeastolate sources and lots, and monitor changes resulting from storage. SERS proved to be a highly reproducible approach that is both rapid and inexpensive compared with current techniques.

Recently, researchers developed and implemented a Raman spectroscopy method for identifying and monitoring quality of aqueous, chemically defined cell culture media [86]. This method utilized spectra collected from five proprietary, industrial, chemically defined media

with unique formulations to build a PCA model. This model was then used to accurately identify and assess the quality of 336 sample measurements. While this work sought to assess the holistic attributes of a media sample, it is also possible to use spectroscopic methods to quantify specific components present in the mixture. The authors also developed a method utilizing EEM spectroscopy and chemometrics to quantify medium concentrations of tryptophan and tyrosine in approximately 5 min [87]. It was determined that the models were able to predict tryptophan concentration to within 5.5% and tyrosine concentration to within 4.5%.

Another important factor in cell culture media is the degradation of important medium components, which can result from long-term storage, changes in temperature and exposure to light. The use of transparent storage containers, glass bioreactors and the frequent presence of light-sensitive riboflavin in cell culture media indicate that monitoring photodegradation of cell culture media is especially valuable. This was recently addressed using EEM spectroscopy, which was selected over Raman spectroscopy owing to its sensitivity and ability to detect small changes in photosensitive analytes [89]. Media samples containing riboflavin, folic acid and pyridoxine were stored in both warm and cold, and light and dark conditions with aliquots taken over a 32-day period. A model linking EEM spectra of media samples and concentrations measured using HPLC was developed using parallel factor analysis and PLS. This model was able to predict pyridoxine and riboflavin concentration with less than 5% error, and folic acid concentration with less than 10% error. This work represents the first application of a spectroscopic technique to determine light-induced damage to cell culture media, and is a rapid alternative to HPLC or MS [89]. Collectively, these studies are promising examples of how spectroscopic methods can be used to rapidly and accurately characterize raw materials.

Predicting cell culture performance using spectroscopy & chemometrics

In addition to measuring raw material quality and composition, spectral methods have applications in long-term monitoring of culture attributes, and utilizing that information for predictive purposes [90–92]. Typically, industrial bioprocesses have many steps spanning several weeks starting with media preparation and vial thaw, a seed train, a production bioreactor run and downstream processing. The ability to monitor these steps in real time and accurately assess process attributes such as chemical composition of media, inoculum quality and cell metabolism, and then to use this information to predict product yield and quality would be advantageous. Using chemometrics, spectral features associated with both beneficial (e.g., high titer) and detrimental (e.g., by-product formation) outcomes can be identified.

Recently, researchers sought to monitor a complete fed-batch CHO cell process from inoculum expansion through a production bioreactor [90]. They collected clarified supernatant at 12 different time points from 37 production runs, from which Raman spectra and protein yield were measured. Using PLS regression, a model was developed to predict protein yield from the Raman spectrum of clarified culture broth. This model was further enhanced using a technique known as competitive adaptive reweighted sampling to remove excess spectral information resulting in an improved predictive ability and a final relative error below 4%.

Prior to this work, the same group demonstrated the use of EEM and chemometrics to monitor the quality of blended cell culture media, and predict protein yield from changes in media composition [91]. Spectra were collected for 33 different media lots prepared in two batches and sampled at four time intervals over a 20-day period. Using multiway robust principal component analysis, different blends and batches were distinguished with less than 3% error. Furthermore, using *n*-way PLS discriminant analysis, media sampled at 0, 5, 10 and 20 days could be differentiated. This is an important finding and indicates that such technology could be used to quickly screen batches for potentially detrimental degradation products or other storage-induced effects. Finally, using measured protein yields for the media batches and an *n*-way PLS calibration model, a predictive protein yield model was developed with more than 94% agreement.

The impact of storage-induced media degradation was also examined using NIR and EEM [92]. Researchers prepared both basal and feed media powder suitable for monoclonal antibody production in CHO cells. Spectra were collected over a period of 12 weeks (84 days), and in conjunction with PCA, PLS models were developed for each spectroscopic method. While this particular feed media proved to be less sensitive to aging over the period evaluated, both spectroscopic methods were able to accurately predict storage time from measured spectra with error no greater than 11%. Aqueous media was then prepared from powder at various ages and used in a shake flask fed-batch experiment. A multivariate linear regression was used to quantify the impact of media age on both the integral of viable cells (IVC) and product titer. Media age did not seem to impact titer; however, a strong, detrimental correlation was evident for IVC, indicating that this could impact process performance.

NIR spectroscopy was also used to characterize and predict performance of soy hydrolysates used in CHO

cell culture media [84]. A total of 15 soy hydrolysate samples were used in media prepared for an IgG-producing CHO cell line. Culture performance was assessed at 7 days by measuring IVC and titer. NIR spectra for all 15 hydrolysate lots at varying concentrations were collected. Using PCA, physicochemical characteristics distinguishing different hydrolysate lots and vendors were identified. Finally, using PLS regression, a predictive model linking hydrolysate spectra, IVC and IgG was developed. Prediction accuracy did vary with hydrolysate dosage, but at high dosages the error rate was consistently below 10%. This group then extended their work utilizing NIR, Raman, EEM and x-ray fluorescence spectra [85]. Spectra for all hydrolysate lots were collected using all four techniques. Predictive models linking spectra, IVC and IgG were established for each technique individually, and then using an ensemble PLS algorithm, a data fusion model was developed. The fusion model's predictions outperformed all of the single technique models. This is not surprising, given the different advantages and disadvantages intrinsic to each technique.

Finally, another group sought to characterize and predict cell culture performance using spectra of cell culture media [82]. Specifically, they employed both NIR and EEM to analyze eight lots of a carbon-nitrogen peptone-based complex medium and five lots of a defined chemical medium. Protein titers of CHO cells grown in each media lot were measured using HPLC. Through PLS modeling, NIR spectroscopy was able to accurately predict performance of the complex media, but not defined media. Alternatively, the models developed using EEM accurately predicted performance of chemically defined media but not the complex media. These differences in performance are not unexpected, as NIR is a vibrational technique sensitive to physical changes and EEM is an optical technique sensitive to fluorophores [82]. Finally, the researchers used PCA and PLS to develop a combined model that utilized both NIR and EEM for performance prediction. The combined model was more accurate and robust than models using a single technique. This work again demonstrates the potential synergistic effects that can be achieved by utilizing multiple spectroscopic techniques in tandem.

Future perspective

The ultimate goal of PAT and QbD principles is to build sufficient knowledge of a culture system such that process conditions can be manipulated based on real-time measurements to consistently produce high yields of products with desired quality attributes. A key aspect to achieve this is the availability of tools able to yield detailed culture information in real time. Spectroscopic methods have the potential to provide this information, but implementation in industrial settings for the production of recombinant proteins has been slow. A review of the literature suggests that a primary reason for this is the complexity and diversity of cell culture systems employed. Different cell lines, media, feeding strategies and bioreactor setups are each unique and introduce their own complexities to process monitoring. These all have the ability to impact how to best generate calibration models capable of accurately predicting analytes of interest, the designation of which may also vary from process to process. Furthermore, the chemometric methods employed to extract relevant information out of the measured spectra require specialized expertise, as well as knowledge of the experimental system, to perform well. Most of the applications presented in the literature focus on proof-of-concept studies that lack rigorous model validation. Others have made incremental advancements in one or more of the aspects of the measurement methods, calibration strategy or chemometrics. Robust application of spectroscopy requires attention to detail of each aspect of the experimental system, analytical method, sampling and modeling.

As instrumentation and understanding of experimental systems advance, significant applications of spectroscopy for bioreactor monitoring and control are beginning to emerge. Such applications are using large sets of industrially relevant data for calibration and validation of models [66,86,90–91], and are generally applicable to enable control of bioreactor conditions based on real-time feedback of analytes [32]. Trends are also emerging in terms of which types of instruments are best for particular analytes in a given system and how best to combine them and maximize the collection of meaningful information [68,85]. These successful demonstrations of the potential of these measurements, as well as use of spectroscopic data to enable sophisticated analysis of cell metabolism [51], will lead to earlier implementation of these technologies by process development groups and further advancement of these methods. Recent advancements in the use of spectroscopy for measurements of cell productivity and product quality [71,77,80] and an increasing understanding of how medium compositions impact cell metabolism will further enable the routine application of PAT and QbD principles in manufacturing settings.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background
- Implementation of Process Analytical Technology and Quality by Design principles have been hampered by the lack of real-time process monitoring technologies.
- Near IR, mid-IR, Raman and fluorescence spectroscopy are able to provide global measurements of bioreactor parameters, although advanced chemometric methods are required to interpret data.

Development of model calibrations
- Multivariate models are used to predict analyte concentrations from measured spectra.
- The process for model generation involves selection of an appropriate instrument, development of a calibration strategy, generation of calibration samples, data preprocessing, multivariate analysis and model validation.

Bioreactor monitoring applications
- Applications for bioreactor monitoring and control strategies utilizing spectroscopy have been demonstrated for both microbial and mammalian cell systems.
- Other applications involving screening and classification of complex and chemically defined media are emerging.

Future perspectives
- Demonstration of successful application on large industrially relevant datasets will enable more wide spread adoption of these methods earlier in the development cycle.
- The different spectroscopic methods yield highly complementary information and the fusion of these technologies will probably yield the most accurate and useful information.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Casci T. Bioinformatics: next-generation omics. *Nat. Rev. Genet.* 13(6), 378 (2012).

2 Datta P, Linhardt RJ, Sharfstein ST. An 'omics approach towards CHO cell engineering. *Biotechnol. Bioeng.* 110(5), 1255–1271 (2013).

3 Dietmair S, Nielsen LK, Timmins NE. Mammalian cells as biopharmaceutical production hosts in the age of omics. *Biotechnol. J.* 7(1), 75–89 (2012).

4 Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* 24(6), 1102–1107 (2013).

5 Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell. Biol.* 13(4), 263–269 (2012).

6 Wuest DM, Harcum SW, Lee KH. Genomics in mammalian cell culture bioprocessing. *Biotechnol. Adv.* 30(3), 629–638 (2012).

7 Becker T, Hitzmann B, Muffler K *et al.* Future aspects of bioprocess monitoring. *Adv. Biochem. Eng. Biotechnol.* 105, 249–293 (2006).

8 Ulber R, Frerichs JG, Beutel S. Optical sensor systems for bioprocess monitoring. *Anal. Bioanal. Chem.* 376(3), 342–348 (2003).

9 Scarff M, Arnold SA, Harvey LM, McNeil B. Near infrared spectroscopy for bioprocess monitoring and control: current status and future trends. *Crit. Rev. Biotechnol.* 26(1), 17–39 (2006).

•• **Provides a thorough review of near IR applications and is organized in terms of the challenges presented by different sample matrices and how to overcome them.**

10 Beebe KR, Pell RJ, Seasholtz MB. *Chemometrics: a Practical Guide.* John Wiley & Sons, Inc., NY, USA (1998).

• **Practical guide for implementation of chemometric tools.**

11 Lourenço ND, Lopes JA, Almeida CF, Sarraguça MC, Pinheiro HM. Bioreactor monitoring with spectroscopy and chemometrics: a review. *Anal. Bioanal. Chem.* 404(4), 1211–1237 (2012).

12 Roychoudhury P, Harvey LM, McNeil B. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Anal. Chim. Acta* 571(2), 159–166 (2006).

13 Schenk J, Viscasillas C, Marison IW, Von Stockar U. On-line monitoring of nine different batch cultures of *E. coli* by mid-infrared spectroscopy, using a single spectra library for calibration. *J. Biotechnol.* 134(1–2), 93–102 (2008).

14 Landgrebe D, Haake C. Höpfner T *et al.* On-line infrared spectroscopy for bioprocess monitoring. *Appl. Microbiol. Biotechnol.* 88(1), 11–22 (2010).

15 Shaw AD, Winson MK, Woodward AM *et al.* Rapid analysis of high-dimensional bioprocesses using multivariate spectroscopies and advanced chemometrics. *Adv. Biochem. Eng. Biotechnol.* 66, 83–113 (2000).

16 Cervera AE, Petersen N, Lantz AE, Larsen A, Gernaey KV. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. *Biotechnol. Prog.* 25(6), 1561–1581 (2009).

•• **Thorough review of near IR applications providing in-depth discussions of practical issues encountered when implementing the technology.**

17 Kornmann H, Rhiel M, Cannizzaro C, Marison I, Von Stockar U. Methodology for real-time, multianalyte monitoring of fermentations using an *in-situ* mid-infrared sensor. *Biotechnol. Bioeng.* 82(6), 702–709 (2003).

18 Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B. On-line fermentation monitoring by mid-infrared spectroscopy. *Appl. Spectrosc.* 58(7), 804–810 (2004).

19 Rhiel MH, Amrhein MI, Marison IW, Von Stockar U. The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-infrared spectra of animal cell cultures. *Anal. Chem.* 74(20), 5227–5236 (2002).

20 Wolpert M, Hellwig P. Infrared spectra and molar absorption coefficients of the 20 alpha amino acids in aqueous solutions in the spectral range from 1800 to 500 $cm^{(-1)}$. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 64(4), 987–1001 (2006).

21 Candeloro P, Grande E, Raimondo R *et al.* Raman database of amino acids solutions: a critical study of extended multiplicative signal correction. *Analyst* 138(24), 7331–7340 (2013).

22 De Gelder J, De Gussem K, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* 38, 1133–1147 (2007).

23 Lee HLT, Boccazzi P, Gorret N, Ram RJ, Sinskey AJ. *In situ* bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy. *Vib. Spectrosc.* 35(1–2), 131–137 (2004).

24 Tensmeyer LG, Heathman MA. Analytical applications of Raman spectroscopy in the pharmaceutical field. *Trends Anal. Chem.* 8(1), 19–24 (1989).

25 Tensmeyer LG, Shields JE. Raman spectra of crystalline 4Zn, 2Zn, and Na insulin. *Proc. SPIE* 1336, 222–234 (1990).

26 Rhiel M, Ducommun P, Bolzonella I, Marison I, Von Stockar U. Real-time *in situ* monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. *Biotechnol. Bioeng.* 77(2), 174–185 (2002).

27 De Oliveira GA, De Castilhos F, Renard CM, Bureau S. Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Res. Int.* (In Press) (2013).

28 Willis L, Gil GA, Lee HLT *et al.* Application of spectroscopic methods for the automation of oil palm culture. *J. Oil Palm Res.* 20, 1–13 (2008).

29 Henriques JG, Buziol S, Stocker E, Voogd A, Menezes JC. Monitoring mammalian cell cultivations for monoclonal antibody production using near-infrared spectroscopy. 116, 73–97 (2010).

30 Marison I, Hennessy S, Foley R, Schuler M, Sivaprakasam S, Freeland B. The choice of suitable online analytical techniques and data processing for monitoring of bioprocesses. *Adv. Biochem. Eng. Biotechnol.* 132, 249–280 (2013).

31 Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative. *Biotechnol. Adv.* 27(6), 726–732 (2009).

32 Tiwari S, Suraishkumar GK, Chandavarkar A. Robust near-infra-red spectroscopic probe for dynamic monitoring of critical nutrient ratio in microbial fermentation processes. *Biochem. Eng. J.* 71, 47–56 (2013).

•• **Effective demonstration of how to implement near IR for microbial bioreactor monitoring and control.**

33 McGovern AC, Ernill R, Kara BV, Kell DB, Goodacre R. 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. *J. Biotechnol.* 72(3), 157–168 (1999).

34 Shaw AD, Kaderbhai N, Jones A *et al.* Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive Raman spectroscopy and chemometrics. *Appl. Spectrosc.* 53(11), 1419–1428 (1999).

35 Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). *Biotechnol. Lett.* 22(24), 1907–1912 (2000).

36 McGovern AC, Broadhurst D, Taylor J *et al.* Monitoring of complex industrial bioprocesses for metabolite concentrations using modern spectroscopies and machine learning: application to gibberellic acid production. *Biotechnol. Bioeng.* 78(5), 527–538 (2002).

37 Sivakesava S, Irudayaraj J, Ali D. Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques. *Process Biochem.* 37(4), 371–378 (2001).

38 Sivakesava S, Irudayaraj J, Demirci A. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. *J. Ind. Microbiol. Biotechnol.* 26(4), 185–190 (2001).

39 Cannizzaro C, Rhiel M, Marison I, Von Stockar U. On-line monitoring of *Phaffia rhodozyma* fed-batch process with *in situ* dispersive Raman spectroscopy. *Biotechnol. Bioeng.* 83(6), 668–680 (2003).

40 Kornmann H, Rhiel M, Cannizzaro C, Marison I, Von Stockar U. Methodology for real-time, multianalyte monitoring of fermentations using an *in-situ* mid-infrared sensor. *Biotechnol. Bioeng.* 82(6), 702–709 (2003).

41 Kornmann H, Valentinotti S, Marison I, Von Stockar U. Real-time update of calibration model for better monitoring of batch processes using spectroscopy. *Biotechnol. Bioeng.* 87(5), 593–601 (2004).

42 Kornmann H, Valentinotti S, Duboc P, Marison I, Von Stockar U. Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using *in situ* mid-IR spectroscopy. *J. Biotechnol.* 113(1–3), 231–245 (2004).

43 Lee HLT, Boccazzi P, Gorret N, Ram RJ, Sinskey AJ. *In situ* bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy. *Vib. Spectrosc.* 35(1–2), 131–137 (2004).

44 Roychoudhury P, Mcneil B, Harvey LM. Simultaneous determination of glycerol and clavulanic acid in an antibiotic bioprocess using attenuated total reflectance mid infrared spectroscopy. *Anal. Chim. Acta* 585(2), 246–252 (2007).

45 Schenk J, Marison IW, Von Stockar U. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2), 344–353 (2007).

46 Schenk J, Marison IW, Von Stockar U. Simplified Fourier-transform mid-infrared spectroscopy calibration based on a spectra library for the on-line monitoring of bioprocesses. *Anal. Chim. Acta* 591(1), 132–140 (2007).

47 Schenk J, Marison IW, Von Stockar U. pH prediction and control in bioprocesses using mid-infrared spectroscopy. *Biotechnol. Bioeng.* 100(1), 82–93 (2008).

48 Dabros M, Amrhein M, Bonvin D, Marison IW, Von Stockar U. Data reconciliation of concentration estimates from mid-infrared and dielectric spectral measurements for improved on-line monitoring of bioprocesses. *Biotechnol. Prog.* 25(2), 578–588 (2009).

49 Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE. Sensor combination and chemometric variable selection for online monitoring of *Streptomyces coelicolor* fed-batch cultivations. *Appl. Microbiol. Biotechnol.* 86(6), 1745–1759 (2010).

50 Ávila TC, Poppi RJ, Lunardi I, Tizei PA, Pereira GA. Raman spectroscopy and chemometrics for on-line control of glucose fermentation by *Saccharomyces cerevisiae*. *Biotechnol. Prog.* 28(6), 1598–1604 (2012).

51 Fazenda ML, Dias JML, Harvey LM *et al.* Towards better understanding of an industrial cell factory: investigating the feasibility of real-time metabolic flux analysis in *Pichia pastoris*. *Microb. Cell Fact.* 12(1), 51 (2013).

52 Gray SR, Peretti SW, Lamb HH. Real-time monitoring of high-gravity corn mash fermentation using *in situ* raman spectroscopy. *Biotechnol. Bioeng.* 110(6), 1654–1662 (2013).

53 Oh SK, Yoo SJ, Jeong DH, Lee JM. Real-time estimation of glucose concentration in algae cultivation system using Raman spectroscopy. *Bioresour. Technol.* 142, 131–137 (2013).

54 Roychoudhury P, Harvey LM, McNeil B. At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance-mid-infrared (ATR-MIR) spectroscopy. *Anal. Chim. Acta* 561(1–2), 218–224 (2006).

55 Ferreira AP, Alves TP, Menezes JC. Monitoring complex media fermentations with near-infrared spectroscopy: comparison of different variable selection methods. *Biotechnol. Bioeng.* 91(4), 474–481 (2005).

56 Arnold SA, Crowley J, Woods N, Harvey LM, Mcneil B. *In-situ* near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. *Biotechnol. Bioeng.* 84(1), 13–19 (2003).

57 Rhiel M, Cohen MB, Murhammer DW, Arnold MA. Nondestructive near-infrared spectroscopic measurement of multiple analytes in undiluted samples of serum-based cell culture media. *Biotechnol. Bioeng.* 77(1), 73–82 (2002).

58 Mattes RA, Root D, Chang D, Molony M, Ong M. *In situ* monitoring of CHO cell culture medium using near-infrared spectroscopy. *BioProcess Int.* 5(1 Suppl.), 46–50 (2007).

59 Card C, Hunsaker B, Smith T, Hirsch J. Near infrared spectroscopy for rapid, simultaneous monitoring. *BioProcess Int.* 6(3), 58–66 (2008).

60 Mattes RA, Root D. Monitoring viable cell density in bioreactors using near-infrared spectroscopy. *BioProcess. J.* 8(4), 48–51 (2009).

61 Sellick CA, Hansen R, Jarvis RM *et al.* Rapid monitoring of recombinant antibody production by mammalian cell cultures using Fourier transform infrared spectroscopy and chemometrics. *Biotechnol. Bioeng.* 106(3), 432–442 (2010).

62 Petiot E, Bernard-Moulin P, Magadoux T, Gény C, Pinton H, Marc A. *In situ* quantification of microcarrier animal cell cultures using near-infrared spectroscopy. *Process Biochem.* 45(11), 1832–1836 (2010).

63 Abu-Absi NR, Kenty BM, Cuellar ME *et al.* Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. *Biotechnol. Bioeng.* 108(5), 1215–1221 (2011).

64 Foley R, Hennessy S, Marison IW. Potential of mid-infrared spectroscopy for on-line monitoring of mammalian cell culture medium components. *Appl. Spectrosc.* 66(1), 33–39 (2012).

• **Describes rigorous evaluation of mid-IR spectroscopy for monitoring of mammalian cell cultures.**

65 Moretto J, Smelko JP, Cuellar M *et al.* Process Raman spectroscopy for in-line CHO cell culture monitoring. *Am. Pharm. Rev.* 14(3), 18–25 (2011).

66 Hakemeyer C, Strauss U, Werz S, Jose GE, Folque F, Menezes JC. At-line NIR spectroscopy as effective PAT monitoring technique in Mab cultivations during process development and manufacturing. *Talanta* 90, 12–21 (2012).

•• **Impressive description of the utility of near IR spectroscopy for monitoring of an industrial Chinese hamster ovary cell culture process.**

67 Whelan J, Craven S, Glennon B. *In situ* Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. *Biotechnol. Prog.* 28(5), 1355–1362 (2012).

68 Buziol S, Roth M, Jung C, Felizardo P, Mendes J, Menezes J. Combination of spectroscopic methods for inline monitoring and control of mammalian cell cultivations. In: *2012 Annual Meeting AIChE*, PA, USA, 11 January 2012.

• **Describes the implementation of several technologies for monitoring diffrent process parameters.**

69 Ashton L, Xu Y, Brewster VL *et al.* The challenge of applying Raman spectroscopy to monitor recombinant antibody production. *Analyst* 138(22), 6977–6985 (2013).

70 Galvao R, Araujo M, Jose GE, Pontes M, Silva E, Saldanha T. A method for calibration and validation subset partitioning. *Talanta* 67(4), 736–740 (2005).

71 Wen ZQ. Raman spectroscopy of protein pharmaceuticals. *J. Pharm. Sci.* 96(11), 2861–2878 (2007).

72 Wen ZQ, Cao X, Phillips J. Application of Raman spectroscopy in biopharmaceutical manufacturing. *American Pharmaceutical Review* 13(4), 46–53 (2010).

73 Teixeira AP, Portugal CaM, Carinhas N *et al. In situ* 2D fluorometry and chemometric monitoring of mammalian cell cultures. *Biotechnol. Bioeng.* 102(4), 1098–1106 (2009).

74 Teixeira AP, Duarte TM, Oliveira R, Carrondo MJT, Alves PM. High-throughput analysis of animal cell cultures using two-dimensional fluorometry. *J. Biotechnol.* 151(3), 255–260 (2011).

75 Teixeira AP, Duarte TM, Carrondo MJT, Alves PM. Synchronous fluorescence spectroscopy as a novel tool to enable PAT applications in bioprocesses. *Biotechnol. Bioeng.* 108(8), 1852–1861 (2011).

76 Ryder AG, Vincentis JD, Li B, Ryan PW, Sirimuthu NMS, Leister KJ. A stainless steel multi-well plate (SS-MWP) for high-throughput Raman analysis of dilute solutions. *J. Raman Spectrosc.* 41(10), 1266–1275 (2010).

77 Brewster VL, Ashton L, Goodacre R. Monitoring the glycosylation status of proteins using raman spectroscopy. *Anal. Chem.* 83(15), 6074–6081 (2011).

78 Gagnon P, Arakawa T. Editorial aggregation detection and removal biopharmaceutical proteins. *Curr. Pharm. Biotechnol.* 10(4), 347 (2009).

79 Tantipolphan R, Rades T, Medlicott NJ. Insights into the structure of protein by vibrational spectroscopy. *Curr. Pharm. Anal.* 4(2), 53–68 (2008).

80 Li CH, Li T. Application of vibrational spectroscopy to the structural characterization of monoclonal antibody and its aggregate. *Curr. Pharm. Biotechnol.* 10(4), 391–399 (2009).

81 Jarvis RM, Blanch EW, Golovanov AP, Screen J, Goodacre R. Quantification of casein phosphorylation with conformational interpretation using Raman spectroscopy. *Analyst* 132(10), 1053–1060 (2007).

82 Jose GE, Folque F, Menezes JC, Werz S, Strauss U, Hakemeyer C. Predicting Mab product yields from cultivation media components, using near-infrared and 2D-fluorescence spectroscopies. *Biotechnol. Prog.* 27(5), 1339–1346 (2011).

83 Li B, Sirimuthu NMS, Ray BH, Ryder AG. Using surface-enhanced Raman scattering (SERS) and fluorescence spectroscopy for screening yeast extracts, a complex component of cell culture media. *J. Raman Spectrosc.* 43(8), 1074–1082 (2012).

84 Lee HW, Christie A, Liu JJ, Yoon S. Estimation of raw material performance in mammalian cell culture using near infrared spectra combined with chemometrics approaches. *Biotechnol. Prog.* 28(3), 824–832 (2012).

85 Lee HW, Christie A, Xu J, Yoon S. Data fusion-based assessment of raw materials in mammalian cell culture. *Biotechnol. Bioeng.* 109(11), 2819–2828 (2012).

•• **Demonstrates the effectiveness of combining several, complementary spectroscopic methods.**

86 Li B, Ryan PW, Ray BH, Leister KJ, Sirimuthu NMS, Ryder AG. Rapid characterization and quality control of complex cell culture media solutions using Raman spectroscopy and chemometrics. *Biotechnol. Bioeng.* 107(2), 290–301 (2010).

87 Calvet A, Li B, Ryder AG. Rapid quantification of tryptophan and tyrosine in chemically defined cell culture media using fluorescence spectroscopy. *J. Pharm. Biomed. Anal.* 71, 89–98 (2012).

88 Bakeev K, Chimenti R. Pros and cons of using correlation versus multivariate algorithms for material idnetification via handheld spectroscopy. *American Pharmaceutical Review*, 1 October 2013.

89 Calvet A, Li B, Ryder AG. A rapid fluorescence based method for the quantitative analysis of cell culture media photo-degradation. *Anal. Chim. Acta* 807, 111–119 (2013).

90 Li B, Ray BH, Leister KJ, Ryder AG. Performance monitoring of a mammalian cell based bioprocess using Raman spectroscopy. *Anal. Chim. Acta* 796, 84–91 (2013).

91 Ryan PW, Li B, Shanahan M, Leister KJ, Ryder AG. Prediction of cell culture media performance using fluorescence spectroscopy. *Anal. Chem.* 82(4), 1311–1317 (2010).

92 Hakemeyer C, Strauss U, Werz S, Folque F, Menezes JC. Near-infrared and two-dimensional fluorescence spectroscopy monitoring of monoclonal antibody fermentation media quality: Aged media decreases cell growth. *Biotechnol. J.* 8(7), 835–846 (2013).