Review

Pharmaceutical BIOPROCESSING

Pharm. Bioprocess. (2013) 1(2), 159–166

Considerations in upstream bioprocess monitoring and statistical data analysis in the context of process analytical technology and quality by design

Economic and regulatory drivers tend to move the manufacturing of biopharmaceuticals from an empiricism-driven art to a science and engineering-based technology. Therefore, the US FDA started the process analytical technology and quality by design initiative to modernize risk-based regulation of quality. Industry implementation is encouraged by increased flexibility and knowledge-based approval procedures. In brief, the concept is based on advanced process monitoring and control, quality built in by design and the definition of the design space. Major obstacles in the translation of the process analytical technology concept are the complexity of biological systems and their limited online observability. To circumvent the lack of physiology-relevant online sensors and to improve process understanding either new online sensor devices are implemented, such as real-time analysis of volatile compounds, or multivariate statistical methods are employed to extract hidden information from acquired data sets, thereby enabling the prediction of non-measurable variables in real time.

Unlike the majority of the manufacturing industries, validation procedures in biopharmaceutical manufacturing evolved quite differently and in such a way that processes strictly adhere to predefined conditions without gaining detailed insight into the entire metabolic reactions. This means that even minor changes or even optimization of the process are hardly possible. Hence, validation in upstream processing is typically derived from the control of a few state variables such as pH, pO₂, temperature and stirrer speed (rpm), none of which provide significant information on the physiological state of the cells and quality assurance is shifted to in-depth characterization of the final product. As a whole, such an approach cannot provide the required process knowledge and does not allow for a reduction in variability or an improvement in reproducibility of biotechnological production processes.

Therefore, in 2002 the US FDA started a pivotal initiative to modernize the regulation of pharmaceutical quality and to encourage industry to implement this new concept: 'guidance for industry, process validation: general principles and practices' [101]. The keys are the implementation of risk- and science-based approaches for regulatory decision making. In this development process, pharmaceutical engineering should evolve from an empiricism driven art - 'quality by quality control' - to a science and engineering-based technology. Thus, the advanced science and engineering knowledge will improve efficiency of manufacturing and regulatory decision procedures. Hence, the initiative can be seen as a paradigm change for biopharmaceutical manufacturing from 'quality after design' to 'quality by design' (QbD). To achieve these goals a set of tools comprising process analytical technology (PAT) [1,2], QbD and design space are

Gerald Striedner & Karl Bayer*

University of Natural Resources & Life Sciences, Vienna, Department of Biotechnology, Muthgasse 18, A-1190 Vienna, Austria *Author for correspondence: Tel.: +43 1 47654 6220 Fax: +43 1 47654 6675 E-mail: karl.bayer@boku.ac.at



Key Terms

Quality by design: A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management.

Process analytical technology: Defined by the US FDA as a mechanism to design, analyze and control pharmaceutical manufacturing processes through the measurement of critical process parameters that affect critical quality attributes.

Bioprocess monitoring: The supervision of bioprocesses

performed by measuring critical parameters such as temperature, pH, O_2 and quality variables.

Statistical modeling: Comprises a framework of formalization of relationships between variables in the form of mathematical equations to identify causal dependencies and to enable prediction of complex variables.

Proton transfer reaction MS: A very sensitive technique for online monitoring of volatile organic compounds in ambient air.

suggested and need to be implemented. Put simply, PAT aims to improve analytical and monitoring capabilities and the ObD concept is based on understanding of the product and process, along with knowledge of the risks involved in manufacturing, and the design space defines valid operational conditions to meet the product quality attributes [3]. Hence, the validation of product quality is provided by mutual variation of a multitude of interdependent variables contained in the particular data set. As a whole, the defined design space provides more flexibility in manufacturing and enables process optimization within the solution space, without violating the validation status of the process.

Meanwhile, the new quality concept has been further developed and corresponding guidelines have been issued [4] in which the authorities strongly emphasize to provide a comprehensive understanding of the product and manufacturing process for reviewers and inspectors, which in turn can create a basis for more flexible regulatory approaches predicated on the level of scientific

knowledge. Risk-based regulatory decisions, primarily reviews and inspections, facilitate improvements of the manufacturing process within the approved design space as prescribed in the dossier without further regulatory review. Moreover, reduction of post-approval submissions and real-time quality control will lead to a reduction of end-product release testing, in addition to shortening the time to market.

Irrespective of the permanent needs to improve the monitoring capabilities of bioprocessing, the recent change of regulatory issues also implies a strong commitment to improve process knowledge and understanding in up- and down-stream processing. Due to the specific challenge of working with living cells, this article is focused on upstream processing. However, the implementation of process control and automation is in contrast to other manufacturing industries. The major reasons for this are:

- » The complexity and fragmentary understanding of the biological production systems;
- Limited online access to physiologically relevant molecules due to their inclusion in the cellular matrix;

- » The heterogeneity of cellular populations;
- » Lack of meaningful sensors fulfilling specific requirements in bioprocessing, such as maintenance of sterility, fouling and leaching of compounds [5].

However, implementation of QbD concepts and definition of the design space requires the availability of physiologically relevant information in real time. The solution to this problem can be accomplished by combining highly specific offline data [6-12] and online data acquired with a broad spectrum of established and new sensors/analyzers [5,13-15] using advanced computational modeling techniques [16-19]. In terms of this conceptual framework process understanding can be improved, which, in turn, supports the identification of the most meaningful process variables.

Concepts for acquisition of meaningful process variables

Bioprocess-specific online sensor requirements are described in detail by Sonnleitner [5]. Conduction of measurements in a solid-liquid-gas suspension under sterile process conditions represents one of the major challenges. Another problem is that physiologically relevant metabolites and compounds are not directly accessible since they are enclosed in the cellular matrix. Consequently, upstream bioprocess monitoring is always a combination of offline monitoring methods with sampling and subsequent time-delayed laboratory analysis and online in situ or ex situ monitoring strategies. In situ measurement signifies insertion of a sensor into the bioreactor with either direct contact to the fermentation broth (invasive) or separated by, for instance, a glass membrane (noninvasive). The ex situ configuration describes techniques with sampling and direct in-line analysis. Lastly, there are also intermediate approaches with sampling and proximate, timely analysis known as at-line techniques [14]. In this context reliable sampling systems and appropriate equipment must comply with bioprocess-specific features such as maintenance of sterility, low sampling volume, and rapid sampling procedure to assure authentic sample composition. Sampling techniques [20-22], commonly used ex situ analytical methods [5,23] and in situ monitoring strategies have been comprehensively reviewed [5,14,24,25]. In the pharmaceutical industry IR, mid-IR, and in particular near-IR (NIR), are the most frequently employed monitoring instruments in combination with chemometric-based signal processing for data interpretation. PAT compliance technology, as well as applications, have been reviewed in detail [14,26,27]. However, wide application of NIR in upstream processing is impaired by strong absorption

of water in the NIR region or by interaction with gas bubbles and matrix effects of nutrient media [14].

The challenge for process engineers is to select an accurate combination of sensor and analyzer systems that comply with challenges specific to the process they are intended for, and that enable acquisition of meaningful process variables, and combinations thereof, in real-time.

The potential of a versatile monitoring platform in combination with statistical modeling applied to advanced control of *Escherichia coli*-based recombinant protein production processes is shown and discussed in the 'Case study' section.

In addition to classical process variables, such as agitator speed, flow rates, temperature, pH, pO₂, CO₂ and O₂ in the off-gas, base consumption and medium feed, the established monitoring platform consists of advanced monitoring devices to gain physiologically relevant information. The quantification of biomass and its physiological activity in real-time are among the most important parameters in upstream bioprocesses, however, ex situ cell dry-mass determination is still frequently used, even in industrial applications. Optical in situ probes for optical density measurements in transmittance mode are an alternative, but in general they show a limited linearity range; they are sensitive to bubbles, suspended solids and changes in cell morphology and there is no differentiation between viable cells, non-viable cells and cell debris [13,28]. This backlog is also valid for NIR technology employed for cell density measurement [29]. Hence, dielectric spectroscopy offers the advantage of acquisition of living, physiologically active biomass; however, signal quality is affected by cell size [30,31]. Concentrations of medium compounds (glucose, amino acids) and by-products of metabolism (acetate, ethanol) are low in carbon-limited feed-batch cultivations with defined media and, therefore, NIR was not considered as additional process analyzer. Alternatively, emission-excitation multiwavelength fluorescence spectroscopy was applied to gain signals regarding intracellular components.

However, direct assignment of the fluorescence signal to specific analytes is only feasible with reporter proteins. Hence, the multiwave excitation of intrinsic fluorescence requires elaborate statistical data analysis and modeling to assign fluorescence signals to specific variables [32–36].

Considering the gas phase as source for process, information quantification of oxygen consumption and carbon dioxide evolution were, for a long time, the exclusively acquired data.

Recently, a highly promising method for direct measurement of volatile metabolites has been adapted for bioprocess monitoring, the proton transfer reaction MS (PTR MS) [37]. The major benefit of PTR MS lies in the acquisition of previously unavailable data in real time. Specific compounds are tightly associated with the cell metabolism and can be directly assigned to individual pathways to uncover bottlenecks and/or overflow of metabolites. Moreover, PTR MS does not require complex sample treatment. In brief, PTR MS consists of an ion source to produce hydronium ions (H_2O^+) connected to a drift tube to react with volatile organic compounds, which are analyzed in a quadrupole MS [38,39]. The major advantages of PTR MS instruments in upstream monitoring are the short response time of approximately 100 ms, enabling realtime measurement, the very low detection limit in the single digit parts per trillion by volume (pptv) region, direct, noninvasive sampling downstream, the sterile head space filter, quantification by calibration and easy interpretation of obtained mass spectra.

Thus, the increased process understanding provides valuable information for targeted host-vector system improvement and modification. Above all, depending on the physiological role of particular compounds, meaningful control loops can be established.

The role of statistics in data processing

Although in recent times a couple of new sophisticated sensor systems, qualified for use in the upstream processing mentioned above, have appeared on the market, the controllability of upstream processing is still impaired by a rather low observability of the entire biochemical process. Due to the complex nature of biological systems and the particular features of bioprocesses a significant extension of the online monitoring capabilities is not to be expected in the near future. Hence, the implementation of PAT, QbD and even the definition of the design space is impaired by these circumstances.

Statistics has a long tradition in the life sciences by quantitatively assessing relationships and unraveling hidden interdependencies. Overall, the role and evolution of statistics in bioprocess engineering reaches from assessment of data quality, to scrutinizing relations within sets of variables and discovery of interdependencies between variables of a data set. A comprehensive compilation of methods used in multivariate data analysis and prediction is provided by Backhaus et al. [40] and by Johnson and Wichern [41]. Concerning data quality the primary goal is to assure usability of the acquired data. Individual steps include checks for missing data, identification of ranges of individual variables, use of control charts and device-specific checks. These tasks are also crucial in view of minimizing error propagation. Further steps of data pre-processing encompass filtering and time alignment.

The potential of multivariate data analysis using support vector regression and partial least square regression for prediction of animal cell cultivation processes was reported by Li *et al.* [42]. Therein, data sets comprising 134 temporal online and offline parameters derived from 243 production runs were analyzed. As an important result, the association of growth, lactate metabolism and product titer could be identified.

» Development of statistics in process control

In the manufacturing industry, statistical concepts were further developed to statistical quality control. In the course of this approach, quality criteria were specified and even monitored over time by control charts. Industry then switched over to statistical process control and, thereby, process abnormities could be detected at an early stage and tracked back to the source of the problem. However, the complexity and high interactivity of biological processes made process supervision cumbersome since several of these charts trigger alarms in a short period and the operator cannot isolate the source of the problem. The solution was the transition from univariate to multivariate methods to meet the demands of bioprocess data analysis and modeling [43]. By application of multivariate statistics and chemometrics correlations between the fragmentary information derived from the low number of online measurable variables and the broad spectrum of offline data sets, such as DNA microarrays, 2D electrophoresis and mass spectrometry can be established. According to Herman Wold, chemometrics is defined by: "how to get chemically relevant information out of measured chemical data, how to represent and display this information and how to get such information into data" [44]. For further reading refer to [45].

The enormous progress in computer science supported the rapid implementation of multivariate statistical analytical methods and chemometrics, which are nowadays indispensable tools in research and development of life science. Multivariate statistics largely contribute to improved process understanding of biological systems and, thereby, support bioprocess monitoring and control and follow the requirements outlined in the PAT initiative exactly.

Variable selection & data compression

As mentioned before, the overwhelming complexity and high interactivity of cellular production systems largely impedes monitoring and control in real time, which, in turn, prevents substantial improvement of process understanding. In order to extract the most relevant information in situations where several hundreds of variables are available, variable selection methods are applied. For a detailed overview of typically used methods, such as stepwise selection methods, best-subset regression or genetic algorithms, see [17,46].

Multivariate projection methods are applied to reduce the dimensionality of the problem, taking into account the correlated nature of the data sets. Thereby new variables, the so-called latent variables, are calculated. The most frequently applied method is principal component analysis (PCA). Latent variables from PCA optimally represent the distances between the objects in the high-dimensional variable space. PCA is very useful for transformation of highly correlated x-variables into a smaller set of uncorrelated latent variables. The results of a PCA are usually discussed in terms of component scores - the transformed variable values corresponding to a particular data point - and loadings - the weight by which each standardized original variable should be multiplied to get the component score [47].

Another method is projection to latent structures (PLS), first introduced by Herman Wold [48] and further developed by his son Svante Wold [49]. PLS is used to find the fundamental relations between two matrices (X and Y; i.e., a latent variable approach to modeling the covariance structures in these two spaces).

Both methods, PCA and PLS, are frequently and successfully applied in data mining in the life sciences to identify the variables with utmost relevance. Multiway PLS was applied for real-time monitoring of fedbatch penicillin fermentation [50]. Due to the complexity of biological systems, the set-up of control regimes often fails because the most relevant variables are not known and, therefore, setup of corresponding control loops is impaired.

Prediction of key variables

As mentioned above, most of the currently implemented control variables in bioprocess control are state variables with rather low relevance to the physiology of the cell factory. However, prediction methods are able to extract relationships between easy to acquire online signals and physiologically relevant directly nonmeasurable variables. For an extended overview of prediction methods in chemometrics see [17,19,51]. The most frequently applied methods include PLS regression, linear weighted regression [52], artificial neural networks (ANNs) [53] and support vector machines regression [42,54].

In the authors' laboratory, PLS regression and ANNs are currently in use. ANNs are systems composed of highly interconnected processing elements – artificial neurons – organized in layers that operate in parallel and that are tied together with weighted connections analogous to synapses. The correlations between input and output data are achieved by adjustment of weight functions implemented in the artificial neurons. Typically easy to acquire online signals, such as oxygen uptake, rate of carbon dioxide production rate, alkaline base consumption, fluorescence emission units and/or mass units of PTR MS are assigned to offline determined cellular key variables. Among these are biomass and product concentration, plasmid copy number and so forth.

Nonetheless, the complexity of the ANN modeling process led to some skepticism and impaired the broad implementation in industry thus far. Meanwhile, the regulatory authorities explicitly recommend the application of multivariate statistics and simulation techniques to obtain non-measurable, but highly relevant, process variables. Another reason for the increased perception of these technologies in process monitoring and control is that in recent time, alternative machine learning techniques are assessed by which the obtained results can be cross-checked [55].

Case study

The proof-of-concept to predict complex process variables in real time was shown with E. coli laboratory-scale fed batch cultivation for the production of a recombinant protein. The intention of the 'case study' is to show that the described monitoring setup is useful and that application of multivariate data analysis allows for prediction of variables not directly accessible in real time. Although the results were obtained in an academic environment, one of the goals of this article is to encourage the enhanced application of statistics to make use of hidden information in existing data sets.

Online data sets comprising O₂ off-gas, CO₂ offgas, base consumption, dielectric spectroscopy (capacity, conductivity) and 2D multiwave fluorescence signals gained from three replicate processes were used to predict the ex situ analyzed offline variables: cell dry mass, the amount of recombinant protein and the plasmid copy number [56,57]. PLS and radial basis function ANN (RBF-ANN) were used to predict the above-mentioned parameters. In order to avoid over-fitting of the models, the input data of the fluorescence data set were pre-selected based on Pearson correlation coefficients. The quality of prediction was compared by the root mean squared error of prediction (RMSEP). The RBF-ANN model showed the best performance (i.e., the lowest RMSEP) by using the pre-selected data sets, whereas the PLS model could not take advantage of the data pre-processing step. When using the whole data sets, the RMSEP of the PLS model is lower than that of the RBF-ANN model, which is in accordance with Svante Wold's statement that the precision of the PLS model improves with an increasing number of relevant input signals [49].

The availability of such highly relevant process variables enables the calculation of quality parameters, such as actual growth rate (μ), specific product formation rate (qP), specific respiration rate qO_2 or oxygen uptake rate and carbon dioxide production rate. Furthermore, design of advanced process control strategies based on predicted variables is facilitated. For instance, calculation of substrate and/or inducer feed streams based on predicted cell dry mass will enable further improvement of the transcription tuning concept [58]. Moreover, due to the timely prediction of biomass and product concentration, process perturbations can be immediately recorded and measures to compensate can be initiated without delay.

The results obtained from laboratory-scale experiments prove the potential of statistical modeling and advanced monitoring techniques to predict complex variables, such as biomass and product titer in real time. Moreover, the study also revealed important information on contribution and performance of individual sensor systems to the overall modeling performance. The detailed analysis of dielectric spectroscopy-derived signals showed that differentiation between living and dead E. coli cells is not possible. Small cell size and compactness of E. coli cell membranes are identified as the main sources of this problem. Prediction accuracy of recombinant proteins is strongly improved by 2D multiwave fluorescence signals but general conclusions on other proteins cannot be drawn as GFP was used as recombinant model protein. In summary, additional experiments are required to show that prediction of process variables also works under varying process conditions (for instance, temperature or growth rate), prediction of proteins without autofluorescence is possible and sensor setup and strategy is also valid on a larger scale.

Future perspective

There is a great deal of evidence that chemometricbased approaches will substantially enhance the capabilities of bioprocess monitoring and that implementation of advanced physiology-based bioprocess control regimes in upstream processing will occur in the near future. Hence, the implementation of PAT and QbD and definition of the design space is nearing realization. For the time being, a major proportion of biopharmaceutical companies are still hesitant with the implementation of the PAT and QbD concept. However, with the increasing number of marketing authorization applications, companies will be more imposed to demonstrate comprehensive process and product understanding, and consequently the benefits of PAT and QbD approaches will become more important. The overall optimistic view of forced

implementation of knowledge-based approaches is grounded on the following supporting pillars:

- » Further exploration of physicochemical phenomena to gain new information of biological systems, for instance the measurement of inelastic, scattering of photons by Raman spectroscopy;
- » Rapid improvement of offline bioanalytical technologies, such as polyomics technologies (combining genomics, transcriptomics, proteomics and metabolomics), sequencing methods and automated micro-structured devices, which, in turn, provide enhanced insight into the cell factories;
- » Further development and improvement of statistical modeling techniques due to further progress in computer science and statistical methods.

Therefore, the improved knowledge and understanding is highly supportive for definition of the design space. Previously unavailable, highly relevant variables in real-time enable mathematical modeling of individual subprocesses that, in turn, can strongly support the quantitative determination of the key variables of the design space. The finally determined operational limiting values are the result of a mutual iterative process of modeling and interpretation of experiments. The flexibility in process operation gained by definition of the design space allows continuous process optimization (clearly) beyond previous or even current operation and validation concepts.

As a whole, it can be said that in the near future bioprocess monitoring and control in upstream processing will advance to the current standards in the manufacturing industry, to improve the productivity of biopharmaceuticals for the benefit of human society.

Financial & competing interests disclosure

Financial support was provided by the Austrian Center of Biopharmaceutical Technology. The authors have no other 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 apart from those disclosed.

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

References

- 1 US FDA. Guidance for Industry: PAT a framework for innovative pharmaceutical development, manufacturing, and quality assurance. US FDA, MD, USA.
- 2 Robert G, Gert T. Implementation of process analytical technologies. In: Process Analytical Technology Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries. Bakeev KA (Ed.). Wiley, Chichester, UK, 13–38 (2010).
- 3 Food and Drug Administration HHS. International Conference on Harmonisation; guidance on Q8(R1) pharmaceutical development; addition of annex; availability. Notice. *Fed. Regist.* 74(109), 27325–27326 (2009).
- 4 Food and Drug Administration HHS. International Conference on Harmonisation; guidance on Q10 Pharmaceutical Quality System; availability. Notice. *Fed. Regist.* 74(66), 15990–15991 (2009).
- 5 Sonnleitner B. Instrumentation of biotechnological processes. In: *Bioanalysis and Biosensors for Bioprocess Monitoring*. Springer, NY, USA, 1–64 (2000).
- 6 Dragosits M, Stadlmann J, Graf A *et al.* The response to unfolded protein is involved in osmotolerance of Pichia pastoris. *BMC Genomics* 11, 207 (2010).
- 7 Yoon SH, Han M-J, Lee SY, Jeong KJ, Yoo J-S. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. *Biotechnol. Bioeng.* 81(7), 753–767 (2003).

- 8 Fritzsch FS, Dusny C, Frick O, Schmid A. Single-cell analysis in biotechnology, systems biology, and biocatalysis. Annu. Rev. Chem. Biomol. Eng. 3, 129–155 (2012).
- 9 Kim I-K, Roldão A, Siewers V, Nielsen J. A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res.*12(2), 228–248 (2012).
- 10 Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* 93(3), 917–930 (2012).
- 11 Markely LR, Ong BT, Hoi KM, Teo G, Lu MY, Wang DI. A high-throughput method for quantification of glycoprotein sialylation. *Anal. Biochem.* 407(1), 128–133 (2010).
- 12 Hossler P. Protein glycosylation control in Mammalian cell culture: past precedents and contemporary prospects. *Adv. Biochem. Eng. Biotechnol.* 127, 187–219 (2012).
- 13 Vojinovic V, Cabral JMS, Fonseca LP. Real-time bioprocess monitoring: part I: *In situ* sensors. *Sens. Actuators B Chem.* 114(2), 1083–1091 (2006).
- 14 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).
- 15 Beutel S, Henkel S. *In situ* sensor techniques in modern bioprocess monitoring. *Appl. Microbiol. Biotechnol.* 91(6), 1493–1505 (2011).

PAT & QbD: upstream bioprocess monitoring & statistical data analysis

- 16 Pomerantsev AL, Rodionova OY. Process analytical technology: a critical view of the chemometricians. *J. Chemom.* 26(6), 299–310 (2012).
- 17 Varmuza K, Filzmoser P. *Introduction to Multivariate statistical analysis in chemometrics*. Taylor and Francis, FL, USA (2008).
- 18 Gabrielsson J, Lindberg N-O, Lundstedt T. Multivariate methods in pharmaceutical applications. J. Chemom. 16(3), 141–160 (2002).
- 19 Miller CE. Chemometrics in process analytical chemistry. In: Process Analytical Technology Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries. Bakeev KA (Ed.). Wiley, Chichester, UK, 226–328 (2010).
- 20 Schädel F, Franco-Lara E. Rapid sampling devices for metabolic engineering applications. *Appl. Microbiol. Biotechnol.* 83(2), 199–208 (2009).
- 21 van Gulik WM. Fast sampling for quantitative microbial metabolomics. *Curr. Opin. Biotechnol.* 21(1), 27–34 (2010).
- 22 Schügerl K, Hitzmann B, Jurgens H, Kullick T, Ulber R, Weigal B. Challenges in integrating biosensors and FIA for online monitoring and control. *Trends Biotechnol.* 14(1), 21–31 (1996).
- 23 Vojinovic V, Cabral JMS, Fonseca LP. *Ex situ* bioprocess monitoring techniques. *Chem. Ind. Chem. Engin. Quarterly* 13(2), 103–116 (2007).
- 24 Esbensen K, Kirsanov D, Legin A *et al.* Fermentation monitoring using multisensor systems: feasibility study of the electronic tongue. *Anal. Bioanal. Chem.* 378(2), 391–395 (2004).
- 25 Wolfbeis OS. Fiber-optic chemical sensors and biosensors. Anal. Chem. 74(12), 2663–2677 (2002).
- 26 Coates JP, Shelley PH. Infrared spectroscopy in process analysis. In: *Encyclopedia of Analytical Chemistry*. John Wiley and Sons, Ltd, Hoboken, NJ, USA (2006).
- 27 Bakeev KA, Wiley InterScience (Online service). Process Analytical Technology Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries. Wiley, Chichester, UK (2010).
- 28 Hatch RT, Veilleux BG. Monitoring of Saccharomyces cerevisiae in commercial bakers' yeast fermentation. Biotechnol. Bioengineer. 46(4), 371–374 (1995).
- 29 Arnold SA, Gaensakoo R, Harvey LM, McNeil B. Use of at-line and *in situ* near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. *Biotechnol. Bioeng*, 80, 405–413 (2002).
- 30 Arnoux AS, Preziosi-Belloy L, Esteban G, Teissier P, Ghommidh C. Lactic acid bacteria biomass monitoring in highly conductive media by permittivity measurements. *Biotechnol. Lett.* 27, 1551–1557 (2005).
- 31 Carvell JP, Dowd JE. Online measurement and control of viable cell density in cell culture manufacturing processes using RFI. *Cytotechnology* 50, 35–48 (2006).
- 32 Stärk E, Hitzmann B, Schügerl K et al. In situ-fluorescenceprobes: a useful tool for noninvasive bioprocess monitoring. Adv. Biochem. Eng. Biotechnol. 74, 21–38 (2002).

- 33 Bogomolov A, Grasser T, Hessling M. In-line monitoring of *Saccharomyces cerevisiae* fermentation with a fluorescence probe: new approaches to data collection and analysis. *J. Chemom.* 25(7), 389–399 (2011).
- 34 Haack MB, Lantz AE, Mortensen PP, Olsson L. Chemometric analysis of in-line multiwavelength fluorescence measurements obtained during cultivations with a lipase producing *Aspergillus oryzae* strain. *Biotechnol. Bioeng.* 96(5), 904–913 (2007).
- 35 Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. Advances in online monitoring and control of mammalian cell cultures: supporting the PAT initiative. *Biotechnol. Adv.* 27(6), 726–732 (2009).
- 36 Teixeira AP, Portugal CAM, Carinhas N et al. In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures. *Biotechnol. Bioengin.* 102(4), 1098–1106 (2009).
- 37 Luchner M, Gutmann R, Bayer K *et al.* Implementation of proton transfer reaction-mass spectrometry (PTR-MS) for advanced bioprocess monitoring. *Biotech. Bioeng.* 109(12), 3059–3069 (2012).
- 38 Hansel A, Jordan A, Holzinger R, Prazeller P, Vogel W, Lindinger W. Proton transfer reaction mass spectrometry: online trace gas analysis at the ppb level. *Int. J. Mass Spectrom. Ion Proc.* 149–150, 609–619 (1995).
- 39 Lindinger W, Hansel A, Jordan A. Proton-transfer-reaction mass spectrometry (PTR-MS): online monitoring of volatile organic compounds at pptv levels. *Chem. Soc. Rev.* 27(5), 347–354 (1998).
- 40 Backhaus K, Erichson B, Plinke W, Weiber R. Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Springer-Verlag Berlin, Heidelberg, Germany (1996).
- 41 Johnson RA, Wichern DW. *Applied multivariate statistical analysis.* Pearson Prentice Hall, NJ, USA (2007).
- 42 Li Y, Yuan J. Prediction of key state variables using support vector machines in bioprocesses. *Chem. Eng. Technol.* 29(3), 313–319 (2006).
- 43 Kourti T. The process analytical technology initiative and multivariate process analysis, monitoring and control. *Anal. Bioanal. Chem.* 384(5), 1043–1048 (2006).
- 44 Wold S. Chemometrics; what do we mean with it, and what do we want from it? *Chemom. Intell. Lab. Syst.* 30, 109–115 (1995).
- 45 Rathore AS, Bushban N, Hadpe S. Chemometrics applications in biotech processes: a review. *Biotechnol. Prog.* 27(2), 305–315 (2011).
- 46 Andersen CM, Bro R. Variable selection in regression a tutorial. J. Chemom. 24(11–12), 728–737 (2010).
- 47 Shaw PJA. *Multivariate Statistics for the Environmental Sciences.* Hodder Arnold, London, UK (2003).
- 48 Wold H. Partial least squares. Johnson SK (Ed.). Wiley, NY, USA (1985).
- 49 Wold S, Sjöström M, Eriksson L. Partial Least Squares Projections to Latent Structures (PLS) in Chemistry. Schleyer PVR, Allinger NL, Clark T et al. (Eds). Wiley, Chichester, UK (1999).

- 50 Ündey C, Tatara E, ÇInar A. Intelligent real-time performance monitoring and quality prediction for batch/ fed-batch cultivations. *J. Biotechnol.* 108(1), 61–77 (2004).
- 51 Rajalahti T, Kvalheim OM. Multivariate data analysis in pharmaceutics: a tutorial review. *Int. J. Pharm.* 417(1–2), 280–290 (2011).
- 52 Odman 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).
- 53 Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 22(5), 717–727 (2000).
- 54 Le H, Kabbur S, Pollastrini L *et al.* Multivariate analysis of cell culture bioprocess data – lactate consumption as process indicator. *J. Biotechnol.* 162(2–3), 210–223 (2012).
- 55 Glassey J. Multivariate data analysis for advancing the interpretation of bioprocess measurement and monitoring data. *Adv. Biochem. Eng. Biotechnol.* 132, 167–191 (2013).

- 56 Clementschitsch F, Kern J, Poetschacher F, Karl B. Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. *J. Biotechnol.* 120, 183–196 (2005).
- 57 Clementschitsch F, Bayer K. Improvement of bioprocess monitoring: development of novel concepts. *Microb. Cell Fact.* 5, 19 (2006).
- 58 Striedner G, Cserjan-Puschmann M, Potschacher F, Bayer K. Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration. *Biotechnol. Prog.* 19(5), 1427–1432 (2003).

» Website

 101 Guidance for Industry. Process Validation: General Principles and Practices.
www.fda.gov/downloads/Drugs/ GuidanceComplianceRegulatoryInformation/Guidances/ UCM070336.pdf