

Direct Monitoring of Upstream Bioprocesses Using Raman Ready-To-Use Embedded Models and Transfer Learning Algorithm

Introduction

The bioprocessing industry is constantly looking for innovative, reliable, and easy-to-implement solutions to improve process monitoring and control, particularly in upstream operations. Raman spectroscopy is an increasingly adopted Process Analytical Technology (PAT) tool for bioprocess monitoring, enabling in-line, real-time monitoring of Critical Process Parameters (CPPs) such as glucose and lactate concentrations and Critical Quality Attributes (CQAs) such as protein titer.

To effectively apply this technology to process monitoring, a calibration phase is required. This involves the development of chemometric models to correlate the Raman spectra with reference measurements of the parameters of interest, typically across multiple batches for high reliability and accuracy. However, this model building process can be resource-intensive and time-consuming. Additionally, generating sufficient data to calibrate robust multivariate models is challenging due to the complexity of bioprocesses and biological variabilities, which can lead to data scarcity. These factors may hinder the implementation of Raman technology for some users.

Ready-to-use models embedded into the Raman analyzer simplify the standard calibration workflow by eliminating lengthy calibration phases. Users can bypass extensive model development and proceed directly to monitoring for immediate insights. This application note outlines the performance of embedded models applied during a fed-batch cell culture run for real-time measurement of glucose, lactate, and monoclonal antibody (mAb) titer. Following the initial run, the models were updated using a proprietary Transfer Learning Algorithm (TLA) that incorporated data from that run to enhance accuracy and extend

monitoring capabilities to additional parameters. The performance of all three parameters was validated, yielding relative errors of 12–15% for the initial ready-to-use models, and 6–12% after refinement with the TLA.

What Are ProCellics™ Raman Analyzer's Embedded Models?

The plug-and-play models of ProCellics™ Raman Analyzer embedded into its Bio4C® PAT Raman Software are designed to enhance your bioprocess monitoring experience. Developed through collaboration across our global R&D and application laboratories, these models are built on extensive data generated using qualified instruments, resulting in a proprietary library of embedded, pre-trained models.

Our team of chemometric experts focuses on optimizing calibrations to achieve high accuracy and robustness. The entire process, from data generation to model development, undergoes a thorough qualification and validation, involving specialists from bioprocessing, data science, manufacturing, and quality engineering. This comprehensive approach ensures that the ready-to-use solution meets specific and documented performance specifications, providing you with reliable insights in real-world scenarios.

By utilizing these embedded models, you can streamline your monitoring process, avoiding the lengthy calibration phase while still achieving accurate results and in-depth process understanding. For detailed information about the model library and how it can benefit your process application, please refer to the dedicated application sheet.

Materials and Methods

Cell culture experiment

The cell culture batches were conducted using a CHO DG44 cell line and EX-CELL® Advanced CHO Fed-batch Medium with Cellvento® ModiFeed Prime COMP. The cells were cultivated in a 1 L glass bioreactor over 14-day period as a fed-batch process.

A total of 26 samples were manually collected from the first batch as off-line references, and 23 samples from the second batch, to measure the concentrations of the targeted parameters, including glucose, lactate (BioProfile® Flex2, Nova Biomedical, Waltham, MA, USA), and mAb titer (Octet® K2 for Protein A, Pall Corp., Port Washington, NY, USA). Viable Cell Density (VCD) was also monitored (Vi-CELL XR, Beckman Coulter, Inc., Brea, CA, USA).

Raman spectroscopy

The ProCellics™ Raman Analyzer was used in a single channel configuration. The instrument has a 785 nm laser with laser power around 350 mW at the

probe-tube output. The probe tube of 225 mm immersion length obtained measurements directly within the bioreactor using PG13.5 port adaptors. For this study, the bioreactor was protected from all light sources to avoid any potential interference. Spectral data acquisition, reference association, and data processing were controlled using Bio4C® PAT Raman Software and its Bio4C® PAT Chemometric Expert module. The Bio4C® PAT Raman Software's advanced features of straylight noise detection and removal were activated during the batches to mitigate any potential impact of external light in case of incomplete cover of the bioreactor.

A total acquisition time of 8.5 minutes was determined per measurement, consisting of an integration time of 17 seconds and an average of 30 spectra. As shown in **Figure 1**, a standard IQ/OQ was performed before the first run to ensure proper qualification of the instrument. The system was delivered with pre-installed embedded models, making it immediately ready-to-use for real-time measurements.

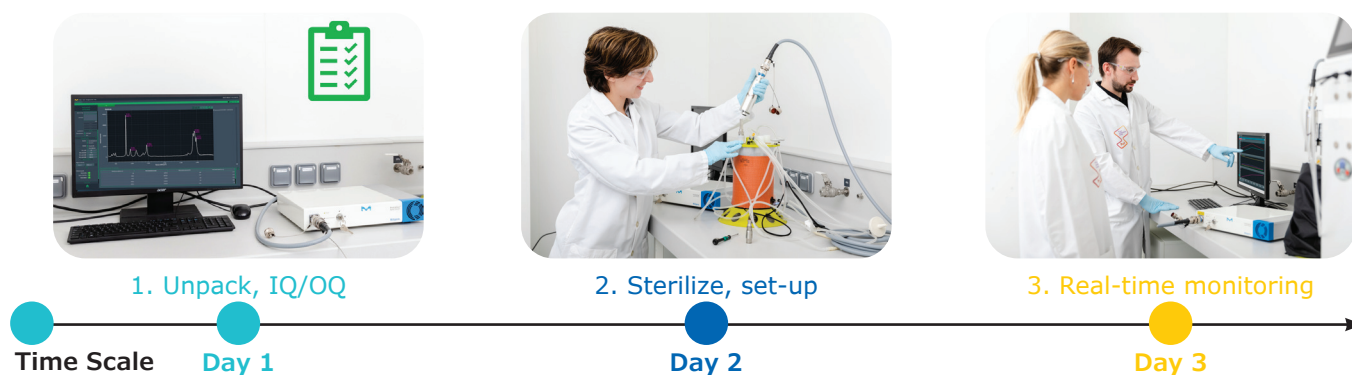


Figure 1: Plug-and-Play ProCellics™ Raman Analyzer implementation steps.

Results and Discussions

Plug-and-play monitoring of the first batch

Figure 2 illustrates the use of the ready-to-use models for monitoring the three target parameters (glucose, lactate, and mAb titer).

- **Glucose:** The in-line Raman measurements demonstrated good performance in capturing the feeding strategy and overall glucose kinetics with 14% relative error. A one-point calibration based on the first off-line reference effectively mitigated slight biases from the process. As a result, this approach achieved an impressive accuracy of 94% (6% residual error), demonstrating the strong capability of the ready-to-use model for the glucose parameter.
- **Lactate:** The in-line Raman readings were closely aligned with the off-line references during the cell growth phase up to day 11, achieving a relative error of 15%. After day 11, the model had difficulty accurately tracking the continued rise in lactate level. This late stage of the process has shown a decrease in VCD, which is indicative of cell death. This decline in cell viability likely contributed to the discrepancies observed in the measurements during this period.
- **mAb Titer:** The in-line Raman readings for mAb titer were consistent with the off-line reference measurements. Although there were minor variations in the kinetics, the overall relative error was 12%, indicating satisfactory accuracy for real-time monitoring of this quality attribute throughout this first process.

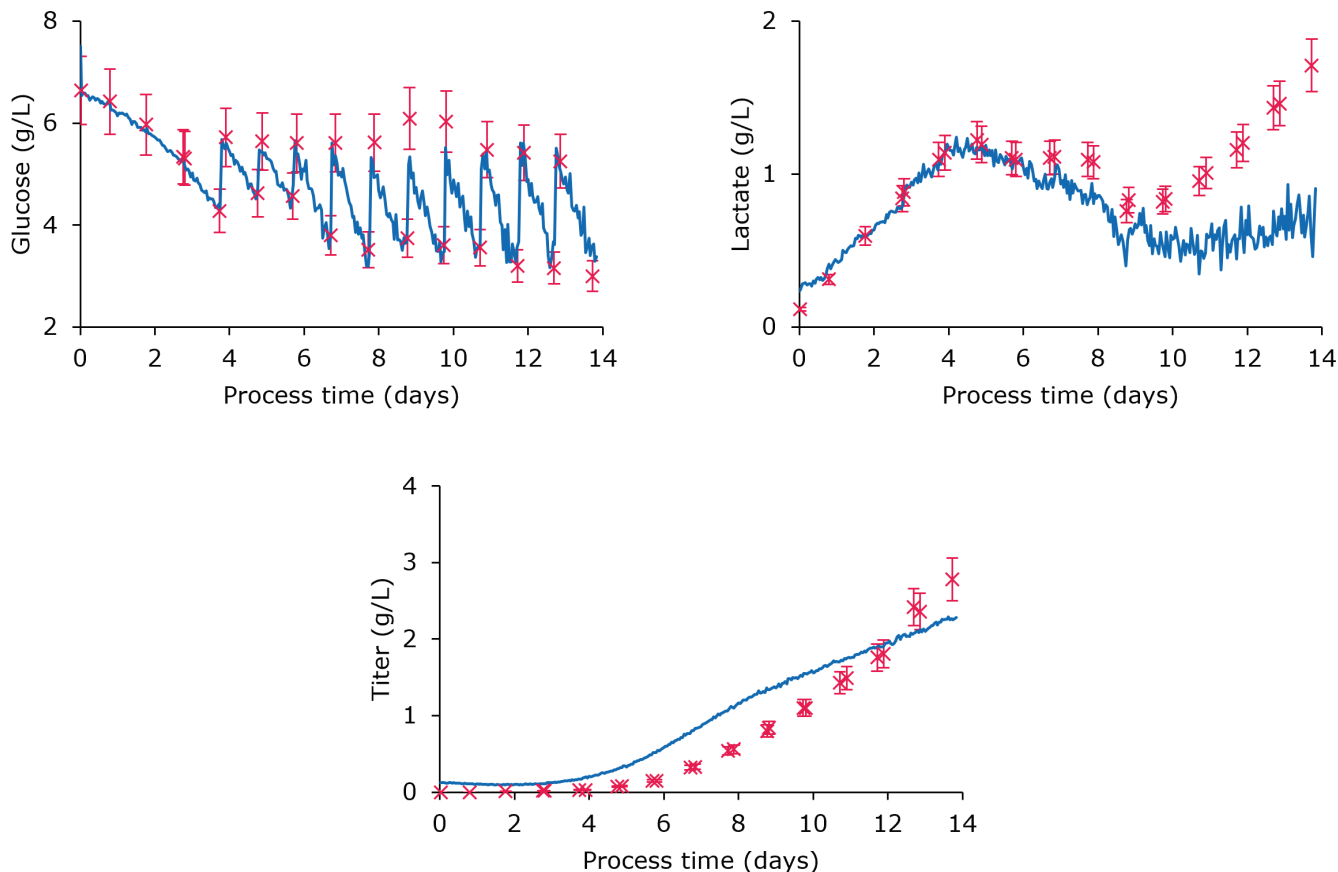


Figure 2: Embedded models monitoring plots based on the first batch for glucose, lactate, and mAb titer. The blue continuous lines (-) represent the embedded real-time Raman monitoring values (glucose monitoring after one-point adjustment on the first off-line), the red crosses (x) represent the off-line references with their respective error bars.

From the first batch, the plug-and-play solution enables capture of monitoring trends comparable to those achieved through offline analysis, giving scientists the ability to gather valuable process information right from the start.

Refined models for the subsequent batches

For the second batch, the initial embedded models were refined using the Transfer Learning Algorithm (TLA), leveraging data from the first monitoring run to further improve monitoring accuracy. Additionally, the TLA solution allows monitoring of VCD by leveraging spectral data and off-line reference values from the first batch to extend the scope of monitoring.

The second batch monitoring plots are displayed in **Figure 3** below.

- Glucose:** The refined glucose model effectively captured the feeding strategy and showed significant improvement, achieving a relative error of 6% in real-time monitoring, without the application of any one-point calibration.
- Lactate:** The updated model provided acceptable performance during the first days of cell growth and improved monitoring during the plateau phase. It achieved a relative error of 12% over the whole process duration, with an even better relative error of 9% when considering the process up to day 9.
- mAb Titer:** The refined model showed enhanced accuracy with a relative error of 9%, down from the already good 12% error obtained during the initial monitoring run with the plug-and-play strategy.
- VCD:** The model demonstrated an enhanced ability to monitor VCD up to 30 Mcells/mL with a relative error of 8%, benefiting from the one-point calibration approach based on the inoculation reference value.

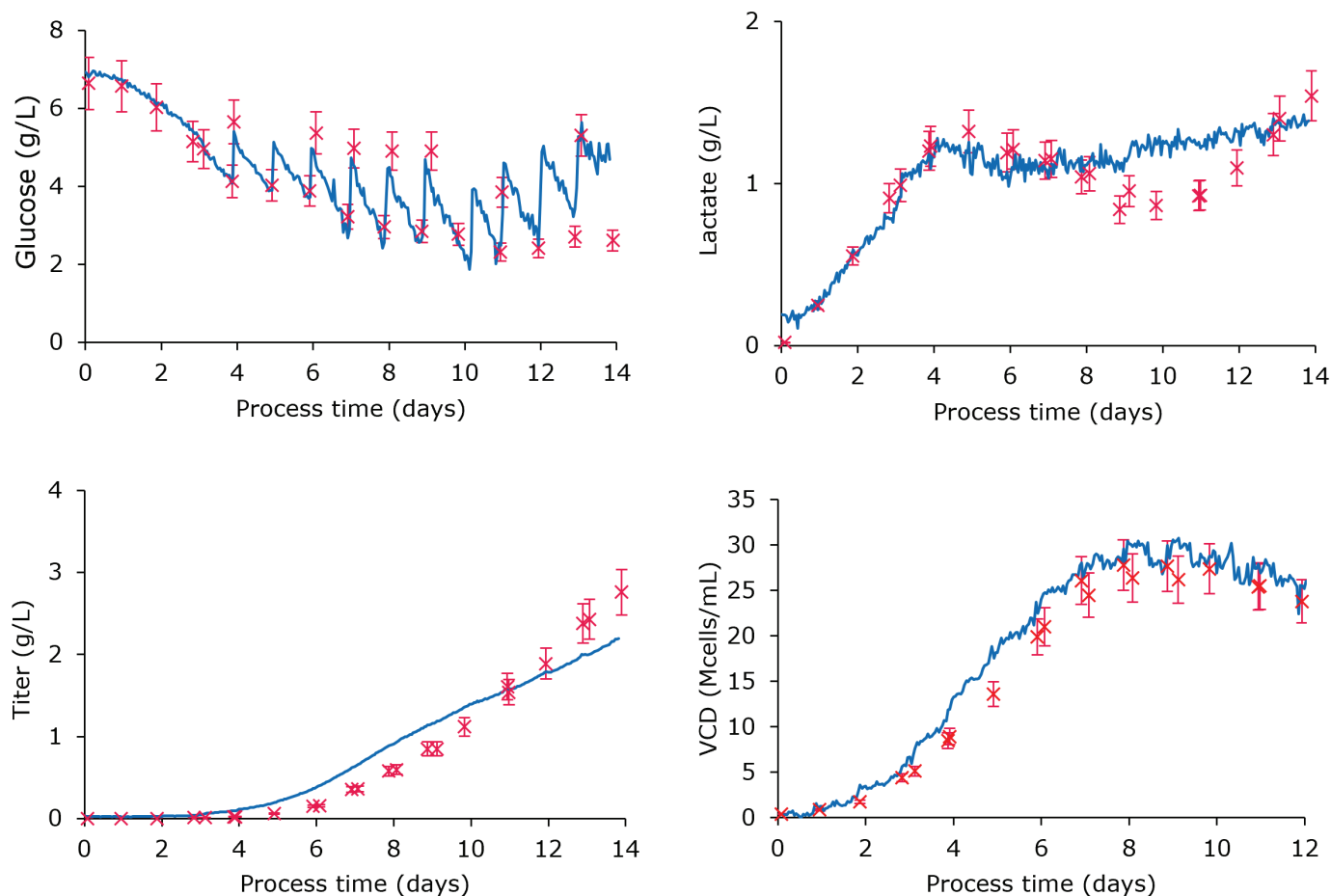


Figure 3: Refined models monitoring plots based on the second batch for glucose, lactate, mAb titer, and VCD. The blue continuous line (-) represents the real-time Raman monitoring values, the red crosses (x) represent the off-line references with their respective error bars.

The refinements made to the embedded models, using data from the first monitoring batch and the TLA algorithm, enabled more accurate monitoring right from the start of the second batch.

This application note demonstrates that it is possible to achieve above 85% accuracy overall on key parameters out of the box, with the potential for quick refinement of the models to further improve accuracy in

subsequent batches. This is especially beneficial during the latter stages of the process, leading to more reliable data for decision-making in bioprocess management.

Results summary

The resulting summary table of performances and bar graph summary plots are respectively shown in **Table 1** and **Figure 4** below.

Table 1: Summary table of Raman monitoring results for first and second batches, respectively using ready-to-use models (that could benefit from one-point calibration* based on the first off-line reference) and TLA-refined models (considering cell growth phase** before cell death) - RMSEP: Root Mean Square Error of Prediction

		Glucose (g/L)	Lactate (g/L)	Titer (g/L)
1 st monitoring on 1 st batch Plug-and-Play Raman models	Range	3.0–6.6	0.1–1.7	0.0–2.8
	RMSEP	0.38*	0.26**	0.34
	Relative Error	6%*	15%**	12%
2 nd monitoring on 2 nd batch Refined Raman models	Range	2.3–6.6	0.0–1.5	0.0–2.8
	RMSEP	0.36	0.17	0.22
	Relative Error	6%	12%	9%

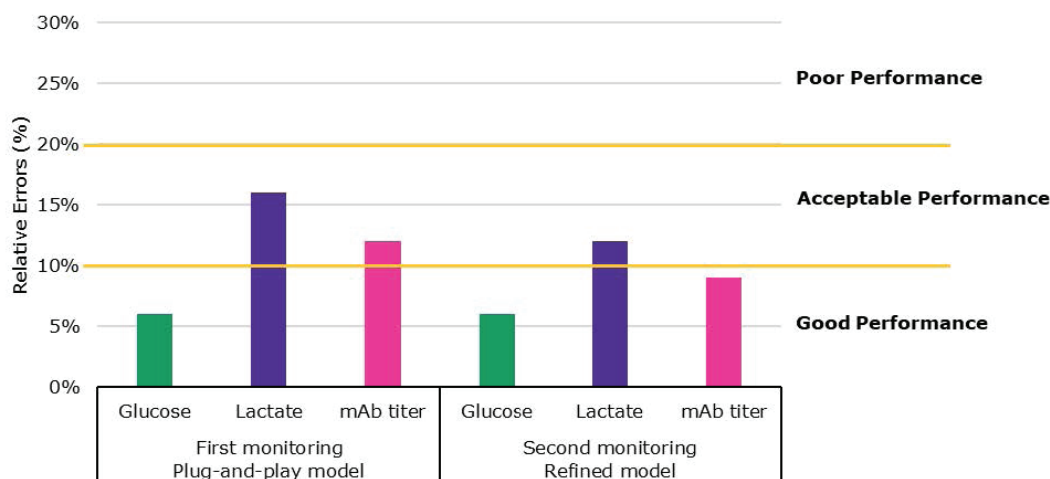


Figure 4: Summary plot of relative errors for both batches.

Conclusion

This application note highlights the effectiveness of the ready-to-use embedded models of ProCellics™ Raman Analyzer in monitoring CPPs and CQAs during upstream bioprocessing, achieving over 85% accuracy across glucose, lactate and mAb titer parameters. These models are ready-to-use on the first batch, requiring no additional efforts from the end user. The insights gained from the initial batch informed refinements for the second batch using the TLA, resulting in improved accuracy and reliability during the second monitoring run.

Continuous evaluation and adjustment of these models will be crucial for optimizing their monitoring capabilities in future bioprocess applications and extending their use to other important compounds in upstream processing, such as cell viability, amino

acids, and glycoforms. Additionally, there is significant potential for applying a ready-to-use model strategy in downstream processing like Tangential Flow Filtration (TFF) and novel modalities like mRNA and Viral Gene Therapy (VGT) products.

By eliminating the complexities of model building, ProCellics™ Raman Analyzer's ready-to-use models streamline implementation, improve process understanding, and pave the way for faster automation strategies. This application note serves as a foundation for further exploration and refinement of monitoring techniques in bioprocessing, emphasizing the importance of real-time data in enhancing process efficiency and product quality.

ProBioGen
Intelligent Biopharmaceutical Solutions

This study was supported by ProBioGen AG (Berlin, Germany), a recognized leader in Contract Development And Manufacturing Organization (CDMO) services and pioneering technology provision. Learn more at www.probiogen.de



To place an order or receive technical assistance:
SigmaAldrich.com/support



For local contact information:
SigmaAldrich.com/offices

Merck KGaA
Frankfurter Strasse 250
64293 Darmstadt, Germany
SigmaAldrich.com

The Life Science business of Merck operates as MilliporeSigma in the U.S. and Canada.

© 2025 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved. Merck, the vibrant M, BioReliance, Millipore, Milli-Q, SAFC, Sigma-Aldrich, Supelco, ProCellics, Bio4C, EX-CELL, and Cellvento are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources.

MK_AN14623EN Ver. 1.0 63751 07/2025

