SARTORIUS

Application Note

November, 2025

Keywords or phrases:
SIMCA®, SIMCA®-online, BioPAT® Viamass, Biobrain® Supervise, rAAV vector production, continuous monitoring, process control, process consistency, process analytical technology, PAT

Enhanced rAAV Production Monitoring With the BioPAT® Platform Via the Umetrics® Digital Twin AI Ecosystem and SIMCA®-online

Izabella Surowiec[1], Eva Wagner[2], Martin Döring[3], Mathias Nilsson[1], Erik Johansson[1], Chloe Lang[1], Suneetha Nunna[2], Magda Tomala[2]
[1]Sartorius Stedim Data Analytics AB, Östra Strandgatan 24, 90333, Umeå, Sweden
[2]Sartorius Stedim Biotech GmbH, August-Spindler-Str. 11, 37079 Göttingen, Germany
[3]Sartorius Stedim Systems GmbH, Robert-Bosch-Str. 5-7, 34302 Guxhagen, Germany

Contact: izabella.surowiec@sartorius.com

Abstract

Recombinant adeno-associated virus (rAAV) vector production is a complex process in which robust cultivation of human embryonic kidney cells (HEK293) is critical to generate high-quality viral vectors. Process analytical technology (PAT) sensors provide real-time, in situ measurements of critical process parameters that can be used to monitor and automate the process via control software. Enhanced utilization of the information from these sensors can be achieved using tools with multivariate data analysis capabilities, such as SIMCA®-online, part of the Umetrics® Digital Twin AI Ecosystem. These tools condense complex data into easy-to-understand plots for on-time decision making. Integrated application of these advanced technologies improves the consistency of production and ensures higher viral vector quality and yield.

This study provides guidelines for successful integration of the BioPAT® Viamass sensor, Biobrain® Supervise software, and SIMCA®-online to provide enhanced understanding and monitoring of rAAV upstream production. We highlight the critical importance of such technologies to meet the growing demand for process optimization, reduce process variability, and ensure compliance with regulatory standards for gene therapy processes. Ultimately, this example demonstrates the potential of integrated solutions to achieve more effective, reliable, and scalable gene therapy applications that contribute to improved patient outcomes and broader therapeutic possibilities.

Introduction

In the emerging field of gene therapy, recombinant adeno-associated virus (rAAV) vectors serve as key tools for the precise and safe delivery of genetic material to cells. The production of rAAV vectors is a complex process involving human embryonic kidney (HEK293) cells[1,2] expressing the adenoviral helper genes E1A | E1B to improve rAAV titer yields.[3] However, the production of rAAV vectors involves significant challenges, particularly maintaining consistent quality and efficiency.

To address these challenges, modern laboratories have implemented process analytical technology (PAT) as a concept for designing, monitoring, and controlling manufacturing processes through timely measurements of critical process values that affect process efficiency and critical quality attributes of the product. The Food and Drug Administration (FDA) PAT Initiative stresses the importance of introducing efficient inline or online monitoring and control strategies as a crucial prerequisite to achieve improved and robust production processes.[4]

Multivariate models that support process optimization and facilitate real-time process monitoring and control play a crucial role in the execution of these FDA guidelines. Integrated PAT combined with advanced analytics has been successfully assimilated in CHO-based production; however, the application of multivariate models utilizing data from bioprocess control systems and integrated PAT sensors to ensure reliable monitoring of rAAV production is still scarce.

In this example, we demonstrate how multivariate models can be developed and evaluated for rAAV production and demonstrate the benefits of implementing these models in an integrated approach for real-time monitoring of key stages of the production process.

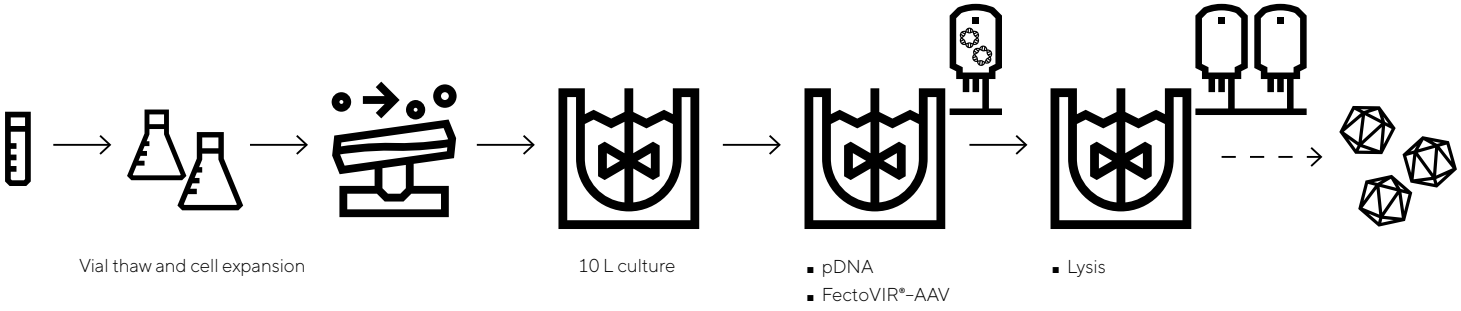
Methods

Production of rAAV-8 and reference analytics
After standard thawing procedures, HEK293 cells underwent sequential seed train expansion and were inoculated into the Univessel® Glass 10 L system. To initiate rAAV-8 vector expression, cells were seeded at a target of 0.3×10^6 viable cells/mL and cultivated in batch mode. The main bioreactor process values were set to a temperature of 37.0 °C, 40% dissolved oxygen, stirrer speed of 201 rpm, total gas flow rate of 0.01 vvm, and pH ≤ 7.25, and were controlled by a Biostat® control unit.

Transient transfection was performed approximately 48 hours post-inoculation using FectoVIR®-AAV transfection reagent and previously optimized plasmid combinations. Cultures were maintained for 72 hours post-transfection to enable viral particle assembly and maturation, followed by standardized cell lysis and harvest procedures for AAV8 recovery. The most crucial steps in the rAAV production process applied in this study are summarized in Figure 1.

A total of twelve independent rAAV-8 production batches were run. Capacitance and conductivity signals were recorded using a BioPAT® Viamass system equipped with a BioPAT® Viamass 12 mm annular probe for 10 L bioreactors. Cell attributes (viable cell concentration, viability, and average cell diameter) were analyzed offline once or twice per day using a Cedex® HiRes Analyzer (Roche).

Figure 1: Overview of the 10 L rAAV-8 production process used in this study



Data acquisition and analysis

Biobrain® Supervise (formerly known as BioPAT® MFCS) is a software platform designed to enhance process efficiency and consistency by offering precise monitoring, control, and automation capabilities. The software provides comprehensive solutions for managing various process values in real-time and allows seamless integration of PAT sensors and analyzers (for example, BioPAT® Viamass) into the bioprocess.

In the current study, rAAV production experiments were controlled on a Biostat® B unit or Biostat® B-DCU unit and recorded in Biobrain® Supervise 4 or its predecessor BioPAT® MFCS/win 3. The Biobrain® Supervise 4 software was used to gather all process data and enable data export via the MFCS 4 SimApi for subsequent modeling in SIMCA® 18.1, part of the Umetrics® Digital Twin AI Ecosystem.

Building the multivariate model

Batch evolution models (BEM) are advanced analytical tools used to summarize the time-evolution data for a process and predict the progression of a batch. BEM focus on capturing the dynamic changes and interactions throughout the duration of a batch. In contrast, batch level models (BLM) are designed to summarize the overall performance of the batch to enable a comparison with other batches or predict process outcomes.

Relevant score plots are used to identify patterns, trends, and outliers within the data, and provide insights into the relationships between samples. The corresponding loading plots are used to interpret the influence of individual variables | process parameters on the model, which helps to understand the variables or parameters that are most significant for explaining the variation in the data.

In the current study, process data were obtained from twelve rAAV-8 upstream production batches. All process variables were exported from BioPAT® MFCS 4 software using MFCS 4 SimApi directly to SIMCA® 18.1 and used to build OPLS® BEM and PCA BLM models.

Separate models were generated for three main process phases: cell growth (twelve batches, nine variables), virus production (nine batches, nine variables), and cell lysis (eight batches, seven variables). The lower number of batches and variables in the subsequent phases was related to process failures and | or limited data recorded in the later stages of the process (especially during cell lysis). UV scaling was applied to all variables, and models were generated using the AutoFit function in SIMCA®.

Finally, the OPLS® BEM models were deployed to SIMCA®-online 18 to illustrate how the models can be used for real-time monitoring of rAAV production.

The following model parameters were used to evaluate the validity of the models: R2X(cum), which is the cumulative sum of squares of the entire X matrix explained by all of the model's components (max = 1); R2Y(cum), which is the cumulative sum of squares of all y-variables explained by the model's components (max = 1); and Q2(cum), which is the cumulative fraction of the total variation of X and Y matrices that can be predicted by the model's components (max = 1).

SIMCA®-online

SIMCA®-online is a software solution designed for real-time monitoring and control of complex industrial processes. By continuously feeding data from the process to the deployed multivariate models, SIMCA®-online condenses process information into easy-to-understand graphics that enable the detection of process deviations in real time and allow process-critical actions to be taken. This enhances process efficiency, ensures high product quality, and supports compliance with industry regulations.

Results

Evaluation of process data
Access to historical data from 'good' processes is a prerequisite for building successful multivariate models. For this project, data were gathered from twelve rAAV production batches executed over a period of six months in the Application and Testing, Separation Technologies Laboratory at Sartorius. Data evaluation involved selecting process values relevant to process monitoring for inclusion in the multivariate model. The selected variables were:

- O_2 — the flow rate of added (sparged) oxygen (ccm)
- pO_2 — the partial oxygen pressure in the bioreactor (% of air saturation)
- CO_2 — the flow rate of added carbon dioxide (ccm)
- N_2 — the flow rate of added nitrogen (ccm)
- AIR — the flow rate of added air (ccm)
- T — temperature (°C)
- pH
- capacitance at 580 kHz (pF/cm)
- the conductivity of the medium (S/m)

Cell growth phase modeling
Process data from twelve batches were used to build OPLS® BEM models. In the cell growth phase, one batch showed slightly deviating time trajectories on both the predictive and orthogonal score plots (Figure 2, A-B). Hotelling's T2 Range and DModX approaches were applied to further investigate this deviation. The same batch was identified as deviating using both of these approaches (Figure 2, C-D). The 'drill-down' functionality of SIMCA® allowed rapid visualization of the deviating process values within the limits described for 'good' batches (Figure 3).

The deviations in the affected batch were attributed to the pO_2, O_2, N_2, and air flow rates. The atypical behavior of these process values could be directly related to technical constraints. An oxygen sensor was not employed during this cultivation batch; therefore, no data for the parameter pO_2 was available in the control software (zero values were recorded). The lack of a dissolved oxygen sensor required manual adjustments to the rates of nitrogen, oxygen, and air flow sparged into the culture for this batch, which resulted in the deviation compared to the other batches.

Overall, the plots produced by SIMCA® can be easily interpreted to provide a better understanding of the production process and enable rapid interpretation of the differences between process batches. SIMCA® plots also streamline the selection of the batches relevant for the generation of the "golden batch" model, which consists of data from representative "good" processes.

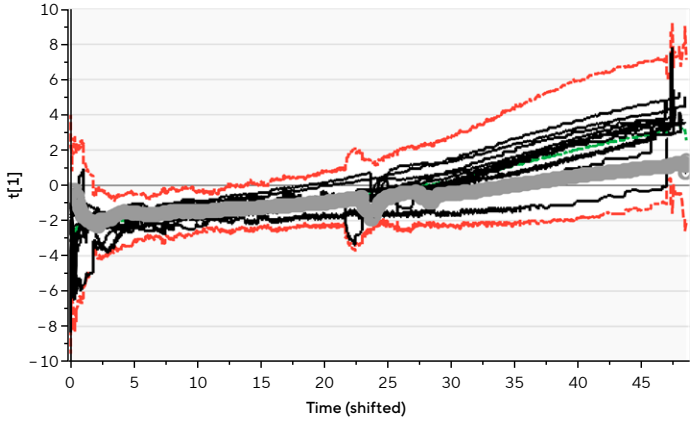
The deviating batch that was not representative was removed from the final model. The remaining batches were incorporated into a "golden batch" BEM model for cell growth. The model loadings indicate how the process values change over time (positive or negative correlations) and which variables are most important for explaining the behavior of the process with time (as indicated by high absolute loading values and small confidence intervals).

The BEM loading plot (Figure 4) shows that oxygen pressure and temperature as well as medium conductivity remained stable during the process. The oxygen sparger flow rate increased with time (more cells, more oxygen needed), air and nitrogen sparger flows decreased with time, as they were negatively correlated with the sparged oxygen flow (as the total flow rate of all three gases was programmed to be constant as part of the dissolved oxygen regulation strategy). The pH value decreased over time as the cells grew and produced CO_2, which triggered a reduction in the CO_2 sparger flow as part of the pH control strategy ($pH \leq 7.25$). The capacitance signal increased over time due to the increase in the volume of viable cells.

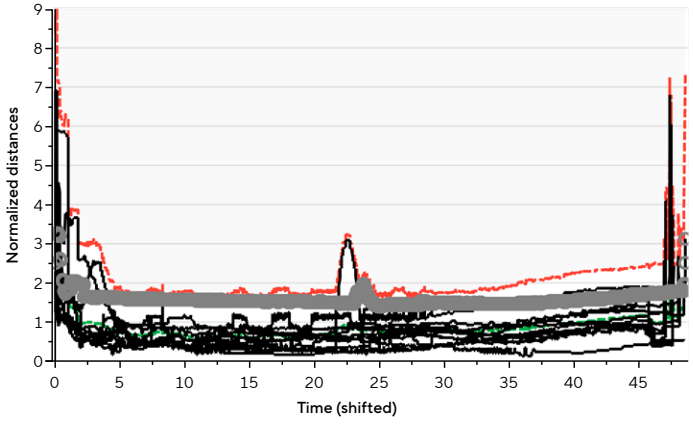
Figure 2: Cell growth model (1+1 OPLS® model: R2X(cum) = 0.58; R2Y(cum) = 0.75; Q2(cum) = 0.76) for twelve batches

Each batch is represented by a black line; the lines show the evolution of each batch over time (x-axis) in crucial model plots compared to other batches; the deviating batch is marked in grey. **(A)** BEM predictive score plot; the green dotted line represents the average value; **(B)** BEM orthogonal score plot; the green dotted line represents the average value; in both **(A)** and **(B)**, the red dotted lines display confidence intervals corresponding to ± 3 standard deviations; **(C)** DModX plot (used to detect outliers outside the correlation structure captured by the model); the green dotted line represents the average value; the red dotted line represents the confidence interval (± 3 standard deviations); **(D)** Hotelling's T2 plot (used to detect outliers within the correlation structure captured by the model); the green dotted line represents the average; the orange and red dotted lines represent the T2 critical values for the 95% and 99% confidence levels, respectively

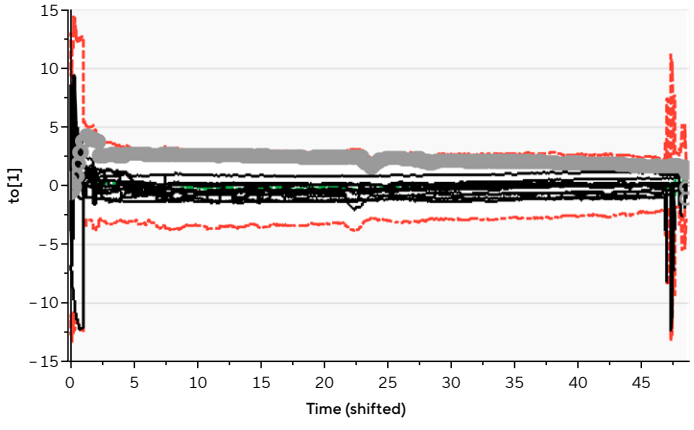
A BEM predictive score plot



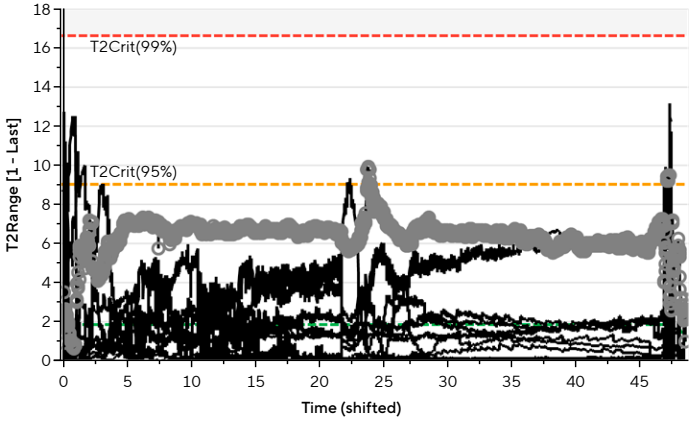
C DModX plot



B BEM orthogonal score plot



D Hotelling's T2 plot



Production phase modeling

The OPLS® BEM model for the rAAV production phase did not indicate any deviating batches (Figure 5A). The model also demonstrated less time-related variation in the data for this phase compared to the previous cell growth phase, as indicated by lower values for all model parameters (1+0 OPLS® model: R2X(cum) = 0.38; R2Y(cum) = 0.42; Q2(cum) = 0.22).

The loading plot of the model (Figure 5B) showed that only five process values changed significantly over time. The reduction in the total number of live cells due to cell death led to a total lower oxygen demand in the culture. As a result, the oxygen pressure increased over time, which triggered a decrease in the oxygen flow and hence an increase in the air flow rate.

However, in contradiction to these results, the capacitance loading had a significant positive value, indicating the volume of viable cells increased over time. In-depth analysis of the capacitance plot showed that although the cells were still growing during the first hours after transfection, cell growth stopped, and a decline in the viable cell volume was visible towards the end of the phase.

Figure 3: Contribution plot showing the variables most responsible for the behavior of the deviating batch during the cell growth phase (variables outside the range of their ± 3 standard deviation are colored orange): pO_2, O_2, N_2, and AIR

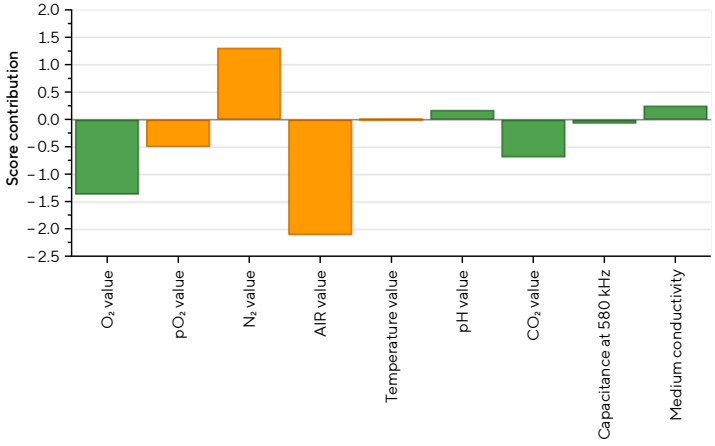


Figure 4: BEM loading plot for the cell growth model (1+1 OPLS® model, eleven batches, nine process variables: R2X(cum) = 0.44; R2Y(cum) = 0.75; Q2(cum) = 0.74)

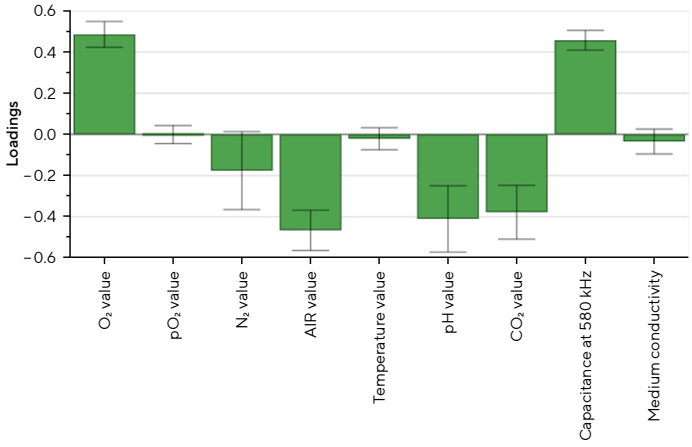
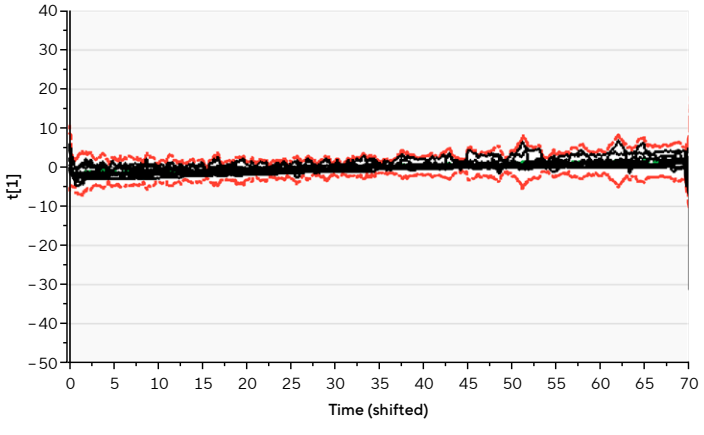


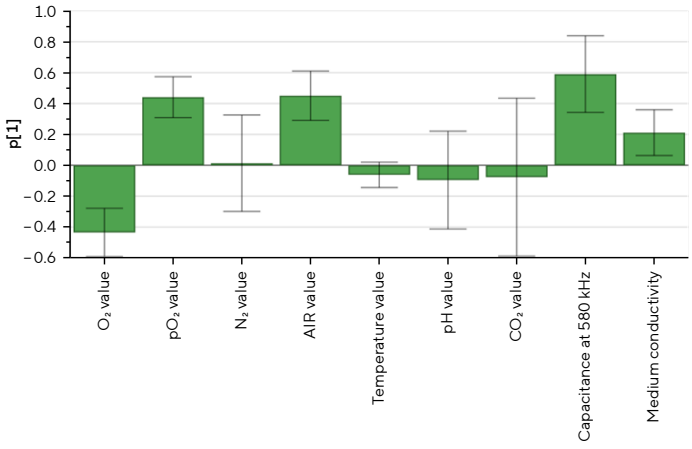
Figure 5: (A) Score plot for the production phase

Each batch is represented by a black line. The green dotted line represents the average value; the red dotted lines represent the confidence interval (± 3 standard deviations). **(B)** BEM loading plot for the production phase model (1+1 OPLS® model; nine batches, nine process variables: R2X(cum) = 0.38; R2Y(cum) = 0.42; Q2(cum) = 0.22)

A



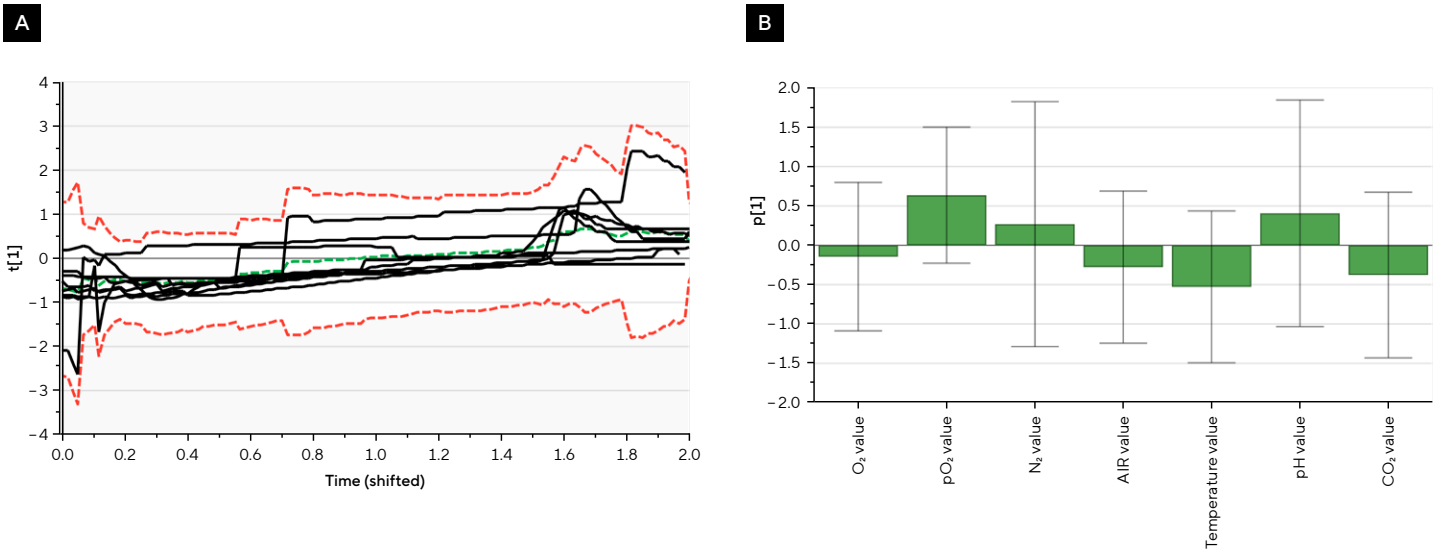
B



Cell lysis phase modeling

The BEM model for cell lysis did not reveal any deviating batches (Figure 6A) and demonstrated less time-related variation in the process values for the cell lysis phase compared to previous phases (1+0 OPLS® model: R2X(cum) = 0.27; R2Y(cum) = 0.44; Q2(cum) = 0.13). The loading plot of the model (Figure 6B) showed that no process values changed significantly over time during the cell lysis phase. In line with routine procedures, a BioPAT® Viamass signal was not recorded during the cell lysis phase, and the O_2 flow, pH control, and CO_2 flow were also stopped.

Figure 6: (A) Score plot for the cell lysis model. Each batch is represented by a black line. The green dotted line represents the average value; the red dotted lines represent the confidence interval (± 3 standard deviations). **(B)** BEM loading plot for the cell lysis model (1+1 OPLS® model; eight batches, seven process variables: R2X(cum) = 0.27; R2Y(cum) = 0.44; Q2(cum) = 0.13)



Principal component analysis BLM models for all phases
Principal component analysis (PCA) BLM models allow users to obtain a holistic | high-level overview of each batch by reducing each batch to one point on a PCA score plot, which simplifies comparison of the behavior of a given batch with other batches. The cell growth model PCA BLM score plot (Figure 7A) indicated that the same batch identified as an outlier in the BEM cell growth model was a strong outlier.

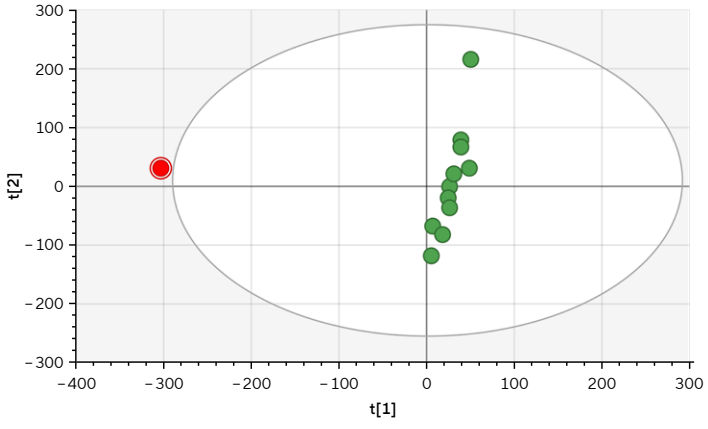
In addition, the PCA contribution plot (Figure 7B) identified the same variables were responsible for the deviating behavior as the variables identified by analysis of the BEM model (Figure 3).

The final PCA BLM models (data not shown) confirmed that all batches exhibited the expected behavior during the cell growth (after outlier removal), production and cell lysis phases, with no outliers visible on the PCA score plots.

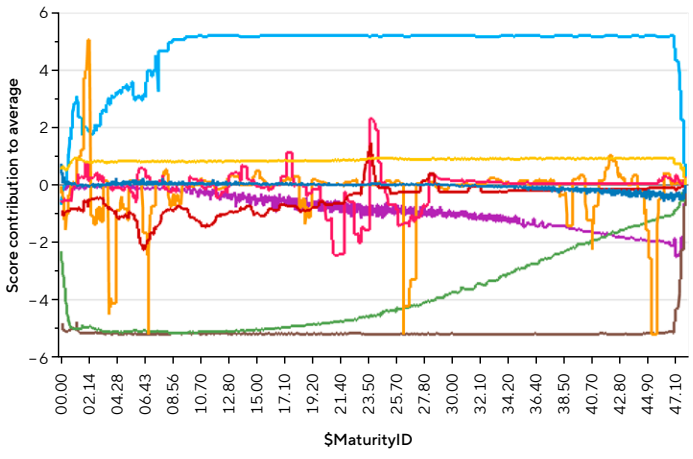
Over the long term, BLM analysis can be used in real time to identify differences between new cell cultures and the "good" batches used for model building, and can also detect systematic process deviations over time.

Figure 7: (A) PCA BLM score plot for the twelve batches in the cell growth phase. The deviating batch is marked in red. **(B)** PCA contribution plot for the cell growth phase of the deviating batch, showing the process values that contributed most to the deviating behavior were the air flow rate (green), nitrogen flow rate (light blue), oxygen flow rate (violet), and oxygen pressure (brown); the other parameters are capacitance at 580 kHz (dark blue), medium conductivity (yellow), temperature (orange), and pH (pink)

A



B



Deployment of the golden batch model to SIMCA®-online

The golden batch model developed in this study has been deployed in SIMCA®-online 18 for real-time monitoring of rAAV production. Therefore, SIMCA®-online provides a continuous visual snapshot of rAAV production and helps operators identify when a process trajectory changes and rapidly react to process deviations. Example BEM score plots for incoming AAV batches, produced in real-time using SIMCA®-online, are presented in Figure 8.

Figure 8: Score plot for the cell growth (left) and virus production (right) BEM models deployed in SIMCA®-online. The plot visualizes both previous batches and ongoing (light blue) batch in relation to the behavior of the "golden batch" (green line) and the model limits (red lines)

Conclusion

The development of efficient and robust processes is critical to guarantee the quality and increase the yield of viral vectors, which are essential components in the advancement of gene therapy applications. This study demonstrates how the integration of PAT sensors, Biobrain® Supervise software, and advanced analytics models offers a robust framework to advance the understanding and monitoring of rAAV production processes. In addition, the capability to deploy multivariate analytical models in real-time with SIMCA®-online enables real-time process insights that facilitate process management and on-time decision-making and ultimately achieve enhanced process control. This integrated approach has great potential to improve the consistency, efficiency, and reliability of rAAV production.

References

1. Wang, D., Tai, P. W. L., & Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nature Reviews Drug Discovery, 18(5), 358–378. https://doi.org/10.1038/s41573-019-0012-9

2. Atchison, R. W., Casto, B. C., & Hammon, W. M. D. (1965). Adenovirus-associated defective virus particles. Science, 149(3685), 754–756. https://doi.org/10.1126/science.149.3685.754

3. Tan, E., Chin, C. S. H., Lim, S. F. S., & Ng, S. K. (2021). HEK293 cell line as a platform to produce recombinant proteins and viral vectors. Frontiers in Bioengineering and Biotechnology, 9, 796991. https://doi.org/10.3389/fbioe.2021.796991

4. U.S. Department of Health and Human Services. (2004). Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance. U.S. Food and Drug Administration. https://www.fda.gov/media/71012/download

Germany
Sartorius Stedim Biotech GmbH
August-Spindler-Strasse 11
37079 Göttingen
Phone +49 551 308 0

USA
Sartorius Stedim North America Inc.
565 Johnson Avenue
Bohemia, NY 11716
Toll-Free +1 800 368 7178

For more information, visit
https://www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/real-time-process-monitoring-software/simca-online

Last modified: 01 | 2026