Integrating Raman Spectroscopy and Genome-Scale Modeling to Enhance AAV Upstream Manufacturing

I Arangoa Guelbenzu1, M Iglesias González1, A Apezteguia García1, J Keune1
1: Viralgen Vector Core S.L, San Sebastián, Gipuzkoa, 20009, Spain.

Abstract

Viralgen is a Contract Development and Manufacturing Organization (CDMO) focused on the production of adeno-associated viruses (AAVs). Viralgen relies on an innovative manufacturing platform based on Pro10™ cells, allowing the production of diverse AAVs and scaling up to 2000L bioreactors. This platform is structured into two main phases: Upstream (USP) and Downstream (DSP). In the final stage of USP, the process includes cell expansion and subsequent virus production, initiated by the transfection of expanded cells. Although the process is standardized, variability in Critical Quality Attributes (e.g.: AAV titer) arises from factors like the product itself, measurement inaccuracies, and the degrees of freedom of Critical Process Parameters (CPPs) which are measured off-line every 24 hours.

During these USP stages, the cell's performance in generating biomass and AAVs depends on the levels of precursors and energy produced by the cell metabolism, and at the same time, the level of precursors and energy produced should be related to the consumed metabolites from the media:

Media metabolite consumption (CPPs) → **Precursor + energy generation** → **Biomass + AAV proliferation (CQA)**

In this context, Viralgen aims to optimize the production process to reduce variability and maximize viral production. However, to carry out any type of process optimization, knowledge is required to target bottlenecks or other factors that may cause such variabilities and inefficiencies. Thus, Viralgen is working on the generation of a metabolic model based on the Human[1] model, which aims to model the consumption of metabolites together with the generation of precursors and energy, and therefore, the production of biomass and AAVs. Additionally, the Raman spectroscopy has been implemented for monitoring CPPs, which has helped on the model calibration.

Data gathering (RAMAN) → **Process Understanding (Metabolic Modelling)** → **Process Optimization**

The Raman Spectroscopy has been successfully implemented, and Viralgen is working on the contextualization and calibration of the metabolic model that, currently, only models biomass proliferation. Viralgen has reduced the lactate prediction error to a 30%, metabolite used for model validation. Still, extra calibration needs to be done to reduce this error and AAV proliferation mechanistic formulas need to be integrated. In the future, the model will be dynamized and simulations will be done to target possible optimizations and control strategies.

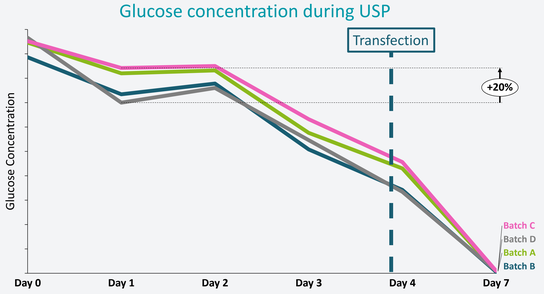
Introduction

Contract development and manufacturing organizations (CDMOs) face a significant challenge in the production of adeno-associated virus (AAV) vectors. As the demand for AAV-based therapies continues to rise, so does the pressure to optimize and maximize the production efficiency without compromising the quality and safety of the final product.

One of the primary obstacles in AAV production is the variability observed inside the permitted Normal Operating Ranges (NOR) in which CPPs move, which impact the Critical Quality Attributes (See Graph).

Currently, there are no control strategies for all the CPPs, as for example, for glucose consumption, since this derives from cellular metabolism and requires extensive knowledge to, first, understand the mechanistic relationship between these parameters and their influence on the CQA, and second, to define control strategies. Glucose feeding strategies has been tested in small bioreactors after evaluating significative correlations with AAV titer, but no optimization were seen in CQA (lack of underlaying knowledge).

To address this, on one hand, monitorization of the culture has been increased from once every 24 hours through the implementation of Raman spectroscopy, and on the other hand, a metabolic model has been contextualized and calibrated to describe intracellular activity that helps detecting bottlenecks and inefficiencies when biomass proliferation (not yet developed for AAV proliferation). This model will not only be fed by the data measured by Raman, but the reactions will also be regulated with a complementary transcriptomic study conducted with the Oxford Nanopore equipment.

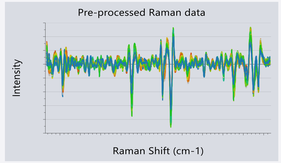


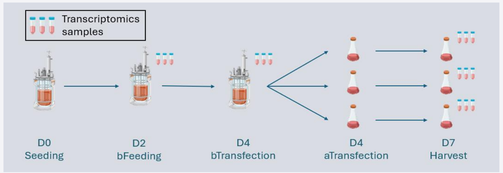
Methods

RAMAN implementation:

The RAMAN implementation involved 15 3L glass Thermo Fisher bioreactor runs to build a model for key metabolites, including glucose, lactate, glutamine, glutamate, ammonia, K+ and Viable Cell Density (VCD), along with offline measurements using NovaFlex and ABER (capacitance) probe, totaling 115 samples for the Partial Least Squares (PLS) model.

Data pre-processing and model building was performed using *SIMCA*[2], applying techniques such as spectrum cutting, smoothing, normalization, and second derivative. A PLS model was generated using cross-validation, with 80% of the data for training and 20% for testing





Transcriptomic analysis:

Samples were analyzed in triplicate, collected at three points in the process across three different transfections, resulting in a total of 15 samples.

Sequencing was performed using Oxford Nanopore, and the transcriptomic analysis was conducted using the EPI2ME pipeline by the AAV Development Department at Viralgen.

Metabolic Model:

Viralgen started with the Human[1] model and adjusted it to fit our specific case (only for biomass proliferation, AAV proliferation to be developed). Python[3] was used for the complete development.

1. Contextualization:

   The contextualization was made by eliminating reactions for metabolites not present in the medium based on the medium compounds. Moreover, reactions that require genetic components not present in the sequencing results were removed (Based on transcriptomics). Viralgen also utilized CORDA[4] to add essential reactions for biomass proliferation and to enable simulations (Used solver: Gurobi[5]).

2. Calibration:

   RAMAN measurements, which included all previously mentioned metabolites (glucose, lactate, glutamine, glutamate, ammonia, and VCD), were used to calibrate certain fluxes of specific reactions. On top of that, calibration of fluxes was done based on the expression of genes related to those reactions.

3. Simulation:

   Viralgen performed parsimonious Mass Flux Analysis (MFA) with L2 norm to simulate the remaining fluxes using CobraPy[6] (Used solver: Gurobi[5]).

4. Validation:

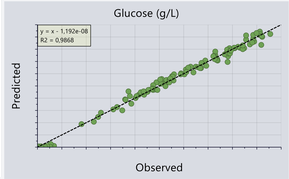
   Viralgen evaluated the model's accuracy by allowing the lactate flux to remain free and assessing the error in the predictions against an offline measurement

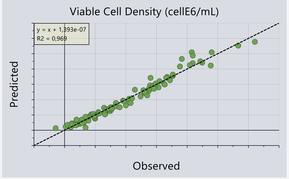
Results

RAMAN implementation:

The results based on the (Q2) from the cross-validation for the generation of PLS are as follows

Metabolite	Q2
Glucose (g/L)	0.98
Lactate (g/L)	0.96
Glutamine (mmol/L)	0.71
Glutamate (mmol/L)	0.91
NH4+ (mmol/L)	0.67
K+ (mmol/L)	0.74
VCD Viable (cellE6/mL)	0.94





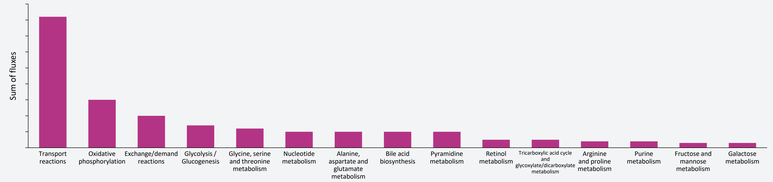
Transcriptomic analysis:

The total number of transcripts that were sequenced is 23,623 transcripts (≥1 transcript per million). These transcripts correspond to 8,140 genes associated with various cellular activities. Among these 8,140 genes, only 603 are linked to metabolic activity. Therefore, the results related to these genes have been used to calibrate the metabolic model.

Metabolic Model:

First, after contextualizing (with medium compounds and transcriptomics) and reconstructing the model with CORDA to achieve the minimum viable product, the generic model has been simplified to Viralgen production process environment, which decreases its degrees of freedom and facilitates obtaining more deterministic results when simulating. The number of reactions has been reduced from 12,972 to 2,340, which represents a reduction of around 80%.

Secondly, after calibrating the model with experimental values, simulations (MFA) were conducted in which, on one hand, the accuracy of the model was analyzed by estimating the error in the prediction of lactate, which was around 30% and on the other hand, the remaining fluxes of the network were estimated to analyze the most active subsystems for the time point with which the model was calibrated:



Discussion and Conclusion

Viralgen, as a Contract Development and Manufacturing Organization (CDMO) focused on the production of adeno-associated viruses (AAVs), faces significant challenges in optimizing production processes while maintaining high quality and safety standards. The variability in Critical Quality Attributes (CQAs) arises from multiple factors, including the dynamic nature of Critical Process Parameters (CPPs) during the manufacturing process.

Among the 15 most active subsystems, the four most active—transport reactions, oxidative phosphorylation, exchange/demand reactions, and glycolysis/gluconeogenesis—are all related to energy generation. This correlation suggests that cellular activity is primarily focused on producing energy, which is crucial for the targeted reaction (biomass proliferation). This indicates that, broadly speaking, the system is directing its activity appropriately; however, further analysis is needed for the subsystems that are less active. Additionally, the model currently has a prediction error, as demonstrated by the 30% error in lactate prediction. Therefore, further calibration is required to draw definitive conclusions.

Future Lines

Looking forward, several future directions for this work can be identified:

1. **Model Dynamization**: It is essential to enhance the model further, as the current analysis has been limited to a specific time point. This will involve expanding the model to incorporate additional time points and conditions.

2. **Accurate Calibration of Fluxes**: Efforts will focus on refining the calibration of fluxes to ensure that simulated values align closely with actual measurements. This will include rigorous testing of the model against real measured fluxes.

3. **Exploration of Control Strategies**: By leveraging the insights gained from the metabolic model, we will explore and implement control strategies that target specific bottlenecks and inefficiencies in the production process.

References

1) J. L. Robinson, P. Kocabas, H. Wang, P.-E. Cholley, et al. An atlas of human metabolism. Sci. Signal. 13, eaaz1482 (2020). doi:10.1126/scisignal.aaz1482
2) Umetrics. (n.d.). SIMCA. https://www.umetrics.com/simca
3) Python Software Foundation. (n.d.). Python. https://www.python.org
4) Gurobi Optimization, LLC. (n.d.). Gurobi optimizer. https://www.gurobi.com/
5) Corda. (n.d.). CORDA: A tool for metabolic modeling. PyPI. https://pypi.org/project/corda/
6) Ali Ebrahim, Moritz E. Beber, Synchon Mandal, Matthias König, Henning Redestig, Christian Diener, Dr. Scientist, Peter St. John, akaviaLab, Hemant_Yadav, Zak King, Nikolaus Sonnenschein, Maximilian Greil, JuBra, Maureen Carey, Ali Kaafarani, Benjamín Sánchez, Erik Cederstrand, Greg Medlock, … Achilles Rasquinha. (2024). opencobra/cobrapy: 0.29.1 (0.29.1). Zenodo. https://doi.org/10.5281/zenodo.13791512
7) Nguyen, T. N. T., Sha, S., Hong, M. S., Maloney, A. J., Barone, P. W., Neufeld, C., Wolfrum, J., Springs, S. L., Sinskey, A. J., & Braatz, R. D. (2021). Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. Molecular Therapy - Methods & Clinical Development, 21, 642-655. https://doi.org/10.1016/j.omtm.2021.04.006