Leveraging Historical Data and Percentile based Modeling for Improved AAV Production Scaling

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

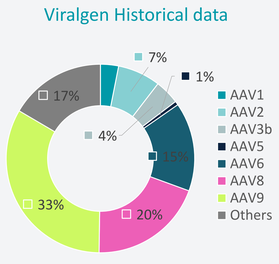
Abstract

Viralgen is a Contract Development and Manufacturing Organization (CDMO) dedicated to the production of adeno-associated viruses (AAVs). The organization employs an innovative and standardized manufacturing platform utilizing Pro10™ cells, which facilitates the production of diverse AAVs and allows scaling from 2L Erlenmeyer flasks to 2000L bioreactors. This universal platform for different AAV products has enabled the generation of data from over 1000 batches produced under the same process across various scales and serotypes, providing a wealth of valuable knowledge based on prior experience.

Scaling from smaller to larger volumes represents a critical factor impacting both CDMOs and their clients, as inaccurate estimations of required capacities to meet product demands can lead to significant and unwanted costs for both parties involved. Historical data from Viralgen over recent years indicates that productivity is scale-dependent, likely due to differences in bioreactor dynamics, geometry, and other operational parameters, which complicate the ideal scaling process. Contrary to the industry's conventional belief in linear scaling, this variability can be manageable if a suitable method is identified to describe the scaling pattern, enabling more accurate productivity predictions.

Currently, the approach for estimating or projecting productivity assumes a linear relationship between scales, applying a 1:1 factor for productivity predictions. However, the data science department at Viralgen has developed a percentile-based model that has successfully reduced prediction errors by 40-45% independently on the serotype, depending on the scales for the projection. In this model, productivity between scales is determined based on historical performance metrics, derived from past production data.

The model calculates the percentile of each product relative to already produced scales, providing insights into the product's performance on the platform. This information aids in decision-making regarding whether to optimize the product or process before initiating the scaling process. Once the performance is established, productivity estimates for the next scale are made using a linear regression model that incorporates historical data.

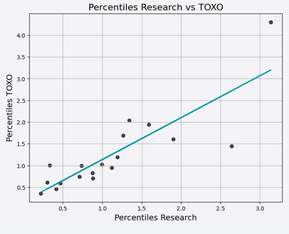


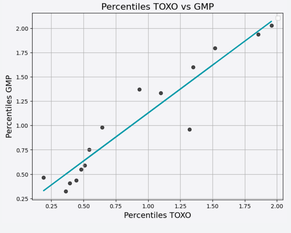
Introduction

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 scale production efficiently without compromising the quality and safety of the final product.

One of the primary obstacles in scaling AAV production is the lack of linearity observed in the scaling processes, particularly due to variations in the geometry and design of bioreactors. Changes in the scale of bioreactors can significantly affect cultivation conditions, such as mixing, oxygen transfer, and temperature control, which directly impact productivity and yield. Furthermore, the intrinsic variability of the production process, combined with the specific characteristics of each product, adds an additional layer of complexity to this challenge.

To address these difficulties, the motivation for employing regression mathematical models arises from the observed correlation between the percentiles of the same product at different scales. This correlation has proven to be a key factor in making accurate estimations regarding production outcomes. These percentiles have been calculated at the final Downstream sample points (Anion Exchange (AEX) for Toxicological (TOXO) to Good Manufacturing Practices (GMP) batch projections and Drug Product (DP) for Research to TOXO due to lack of data at AEX). The motivation to estimate final sample points is that from productivity at this point, vial counts are estimated using other client requirements. Additionally, predictions made at this stage tend to be more accurate than those made at other sample points.



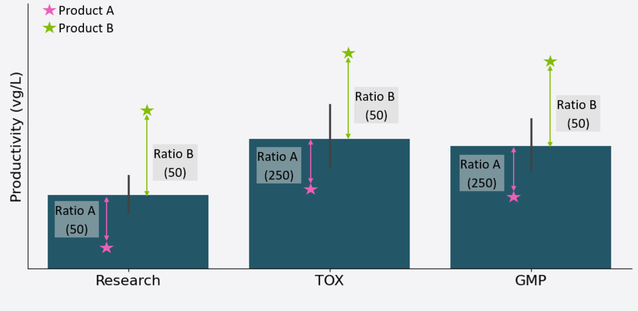


Methods

Dataset composition:

Research to GMP batches considered (plasmid DNA, nor ne-DNA). Moreover, all considered programs have been manufactured using the same manufacturing process generation and only products.

Percentile Representation



Computing methods:

Exploratory Data Analyses (EDA) and a Linear Regression Model (LRM) were conducted using Python[1] and JMP18[2] :

- Distribution analysis in EDA: Used to analyze percentile distribution. A state-of-the-art method was used as the criteria to determine weather to reject or not the null hypothesis about data being normally distributed.
- Correlation methods in EDA: Used to analyze the correlation between percentiles from contiguous scales considered in the analysis.
- Tests for mean difference significance: Conducted to analyze mean differences between contiguous scale ratios to evaluate if a linear ratio scale-up can be assumed or a regression model is required to estimate next scale ratio.
- Linear Regression Modelling: Conducted to predict percentiles of a given product at the contiguous scale based on historical data.

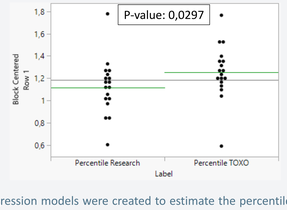
Results

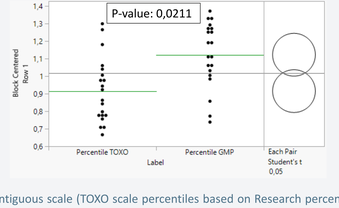
Based on the results of the distribution of the dataset used to generate the prediction model from Research to TOXO and TOXO to GMP, the correlation analysis methods were selected.

The results of the correlations between the percentiles Research VS TOXO and the percentiles TOXO VS GMP are statistically significant following their correlation analysis, demonstrating a good indicator for scaling estimation.

	Correlation method	P-value
Research VS TOXO	Method 1	<0,0001
TOXO VS GMP	Method 2	<0,0001

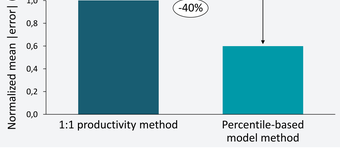
After analyzing the correlation between the percentiles, whether there was a significant difference between the means of the two groups was examined to decide if a linear ratio scale-up can be assumed or if a regression model is required to estimate the next scale ratio. The results showed that there was a significant difference, leading to the development of a linear regression model to estimate percentiles on the contiguous scale.



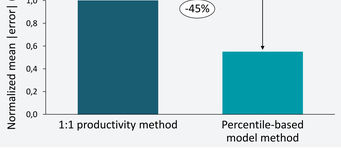


Linear regression models were created to estimate the percentiles on the contiguous scale (TOXO scale percentiles based on Research percentiles, and GMP percentiles based on TOXO percentiles). The error was reduced by 40 to 45%, as shown in the following graph:

Research to TOXO Prediction Error



TOXO to GMP Prediction Error



Discussion

The production of adeno-associated viruses (AAVs) presents a significant challenge for Contract Development and Manufacturing Organizations (CDMOs) like Viralgen, particularly as the demand for AAV-based therapies continues to grow. Traditional methods of scaling production have relied on the assumption of linear scalability, which often leads to inaccuracies in productivity predictions. The complexities introduced by variations in bioreactor design, operational parameters, and the intrinsic variability of the production process necessitate a more nuanced approach to scaling.

Viralgen's innovative percentile-based model represents a significant advancement in addressing these challenges. By leveraging historical performance metrics and calculating percentiles, the model provides a more accurate framework for estimating productivity across different scales. This method not only reduces prediction errors by 40-45% but also allows for better decision-making regarding product optimization and scaling strategies before production begins. The statistical significance of the correlations observed between percentiles at different scales further validates the model's effectiveness, demonstrating that a non-linear approach is essential for accurate scaling estimations.

Furthermore, the findings indicate that the nature of the dataset plays a crucial role in influencing the correlation analyses and subsequent predictive modeling. The ability to discern these differences highlights the importance of employing appropriate statistical methods tailored to the characteristics of the data. As CDMOs navigate the complexities of AAV production, adopting such data-driven approaches will be vital for enhancing productivity and ensuring the successful commercialization of AAV-based therapies.

Conclusion

In conclusion, the challenges associated with scaling AAV production in CDMOs are multifaceted and require a departure from traditional linear scaling assumptions. Viralgen's percentile-based productivity prediction model offers a robust solution, significantly improving the accuracy of productivity estimates and reducing associated costs. By integrating historical data into the scaling process, the model not only enhances the understanding of product performance across different bioreactor scales but also facilitates informed decision-making for process optimization.

As the demand for AAV therapies continues to rise, the implementation of advanced statistical methods, such as those developed by Viralgen, will be crucial in meeting production goals without compromising quality. The success of this model underscores the potential for data science to transform manufacturing processes in the biopharmaceutical industry, paving the way for more efficient and effective production strategies. Future research should focus on further refining these models and exploring their applicability across a broader range of products and scales, ultimately contributing to the advancement of AAV-based therapeutics.

References

1) Python Software Foundation. (n.d.). Python. https://www.python.org
2) JMP®, Version 18. SAS Institute Inc., Cary, NC, 1989–2025.