Molecular Therapy
Methods & Clinical Development
Original Article

# Unlocking DOE potential by selecting the most appropriate design for rAAV optimization

Konstantina Tzimou,[1,3] David Catalán-Tatjer,[1,3] Lars K. Nielsen,[1,2] and Jesús Lavado-García[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark; [2]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, QLD 4072, Australia

**Producing recombinant adeno-associated virus (rAAV) for gene therapy via triple transfection is an intricate process involving many cellular interactions. Each of the different elements encoded in the three required plasmids—pHelper, pRepCap, and pGOI—plays a distinct role, affecting different cellular pathways when producing rAAVs. The required expression balance emphasizes the critical need to fine-tune the concentration of all these different elements. The use of design of experiments (DOE) to find optimal ratios is a powerful method to streamline the process. However, the choice of the DOE method and design construction is crucial to avoid misleading results. In this work, we examined and compared four distinct DOE approaches: rotatable central composite design (RCCD), Box-Behnken design (BBD), face-centered central composite design (FCCD), and mixture design (MD). We compared the abilities of the different models to predict optimal ratios and interactions among the plasmids and the transfection reagent. Our findings revealed that blocking is essential to reduce the variability caused by uncontrolled random effects and that MD coupled with FCCD outperformed all other approaches, improving volumetric productivity 109-fold. These outcomes underscore the importance of selecting a model that can effectively account for the biological context, ultimately yielding superior results in optimizing rAAV production.**

## INTRODUCTION

Gene therapy is a revolutionary medical technology that involves the *in vivo* introduction, alteration, or deletion of genetic material to treat or prevent diseases. It aims to correct faulty genes, supplement missing or defective ones, or modulate gene expression to restore normal cellular function. Gene therapy has shown great potential, addressing a wide range of genetic disorders, such as cancer, blood disorders, muscular dystrophies, or cardiovascular diseases.[1] Its potential lies in providing a treatment for hitherto untreatable diseases or an alternative where existing treatments are accompanied by severe side effects that significantly compromise the patient's quality of life. Although only 10 gene therapy products have been approved by the FDA, there is a high number of gene therapies in clinical trials, paving the way for future personalized medicine.[2]

Several platforms are available for delivering therapeutic genes to the patient, including viral vectors, liposomes, inorganic nanoparticles, and cationic polymers. Recombinant adeno-associated viruses (rAAVs) have gained high attention in recent years as *in vivo* gene delivery vectors, due to their favorable safety profile, high efficiency in gene delivery, and broad tropism for specific tissues.[3] The predominant method to produce rAAVs to date involves triple transfection of mammalian cells. Meeting the high demand for rAAVs is impeded by the scalability of this production process and the challenges associated with large-scale transient gene expression (TGE).[4,5] Optimization is crucial but can be challenging due to the involvement of multiple parameters, including cell line, cell culture conditions such as culture medium and cell density, and transfection conditions such as total plasmid DNA amount, the plasmid ratio employed in the transfection process, and the total amount of transfection reagent.[6–10]

The use of design of experiments (DOE) to study, characterize, and optimize a process is well established in the industry and widespread in multiple and different fields, such as food manufacturing, agriculture, material design, and even e-commerce and big data analysis, among others. This systematic and efficient approach allows researchers to study the effects and interactions of multiple variables simultaneously, minimizing experimental runs and providing statistically robust models for prediction of optimal solutions.[11–14] However, applying DOE to tackle biological problems can be challenging, due to the high variability of biological systems and the complicated nature of interactions of biological components.[2,15] In the past, optimization of rAAV production was mainly based on one-factor-at-a-time (OFAT) methods, which are easier to approach but more time consuming.[16–19] Recently, DOE approaches started gaining attention as a tool for rAAV production optimization. By systematically varying the different parameters, different DOE approaches lead to an efficient exploration of different experimental spaces.[11] Traditionally, response surface methodologies (RSMs) have been employed for modeling and optimizing responses in biological systems. Central composite design (CCD) and Box-Behnken design (BBD) are the most-used approaches to optimize biological processes.[20–26] Regarding rAAV optimization,

Received 20 May 2024; accepted 23 August 2024;
https://doi.org/10.1016/j.omtm.2024.101329.

[3]These authors contributed equally

**Correspondence:** Jesús Lavado-García, The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.
**E-mail:** jlavgar@dtu.dk

most of these studies also rely on RSM, and the factors used in the designs are sometimes not independent, making it challenging to analyze the generated models.[10,27,28] These models employ second-order polynomials to identify optimal combinations of the studied variables.[29] While a CCD combines factorial points at the extremes of the studied limits of the variables, BBD tests combinations of factors at the midpoints of the limits of the experimental space and lacks points at the edges, ensuring that all design points remain within the safe operating zone, avoiding extreme factor settings simultaneously.[30] One of the advantages of using these RSM methods is that they can present rotatability, meaning that all tested points in the experimental space have equal distance from the design center, ensuring a consistent variance prediction in all runs and enhancing the design's efficiency. A CCD model can always be made rotatable (RCCD), adjusting the limits of some of the runs, while a BBD is fully rotatable only for specific designs.[30] Moreover, depending on the biological response, we might need to test both the edges and the midpoints of the experimental space. In this case, a CCD can also be made to test both, adjusting its axial points to the center, in a face-centered CCD (FCCD). This configuration lacks rotatability and, thus, does not allow for assessment of potential curvature within the system.[30] These designs can be generated and executed in multiple orthogonal blocks, allowing for the independent estimation of model terms and uncontrolled random effects. This approach minimizes variation in regression coefficients, contributing to a more robust and accurate analysis of the experimental factors.[31,32]

Apart from RSM designs, mixture design (MD) is a specialized experimental approach used to study the effects and interactions of different factors when varying their proportions in a mixture, always adding up to a fixed quantity. This dependency on a constant total is a key distinguishing feature of MD compared to RSM, where each factor varies independently. MDs are particularly valuable in fields like chemistry, pharmaceuticals, and food science, where products often consist of a blend of various ingredients, and understanding the optimal ratio combination is crucial.[33,34] Park et al. showed that an MD approach can be used to study rAAV production and plasmid ratios in the triple-transfection protocol.[35]

However, these studies differ in the parameters selected as factors and do not compare the outcome and potential of each methodology. In this study, we conducted a comparative analysis of four distinct DOE methodologies, a complete fractional RCCD, FCCD, BBD, and MD, with the aim to evaluate their differences when studying and optimizing rAAV production (Figure 1). We intend not only to find optimal parameters for the rAAV triple-transfection method with each model but also to analyze advantages and disadvantages for the use of each type of model, facilitating the identification of the most suitable DOE approach tailored to individual cases.

## RESULTS

### Model diagnostics

The first step before analyzing the constructed models is to check if the obtained data meet the requirements for the assumptions of the models to be valid. The first requirement is that the residuals (the difference between predicted and observed values in each run) are normally distributed. This was routinely checked for all models using the normal probability plot (normal probability vs. the externally studentized residuals) and the Box-Cox plot. To achieve a normal distribution of the residuals, a log transformation was applied to all exponential data from volumetric productivity (Vp). Log(Vp) data fulfilled the requirements for the models to be valid. Likewise, the presence of outliers was routinely checked in all models for both responses—log(Vp) and viability—by the representation of residuals vs. predicted (Figures 2D, 3D, 4D, and 5D for log(Vp) and Figures 2H, 3H, 4H, and 5H for viability) and residuals vs. row (Figures 2E, 3E, 4E, and 5E for log(Vp) and Figures 2I, 3I, 4I, and 5I for viability). No outliers were detected in any of the models, emphasizing the validity of the conclusion that arose from the model analysis.

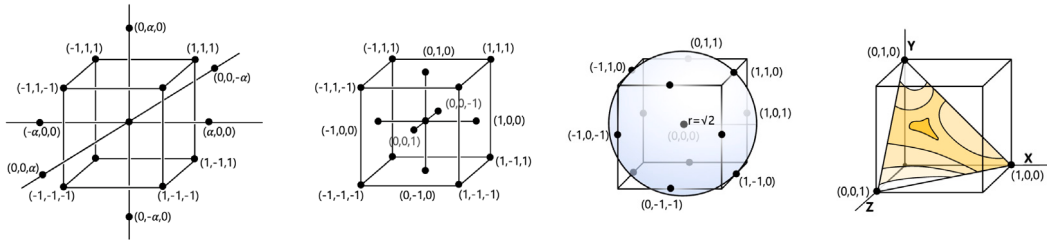
### Blocking is required when uncontrolled factors are introduced

To study whether the strategy of dividing the runs into three orthogonal blocks was successful, we analyzed the corresponding half-normal plots of each model, with each block treated as a fixed factor (Figures 2B–2E, 3B–3E, and 4B–4E). In these plots, the y axis represents the normal estimate (orthog t), while the x axis presents the normal quantile for each RSM (RCCD, FCCD, or BBD), each response (log(Vp) or viability), and each time point (48 or 72 hours post-transfection [hpt]). The blue line intersects the origin with a slope equal to Lenth's estimate of σ, and any significant effect will not conform to this line. The primary objective was to identify significant effects in each scenario, with a particular focus on whether the blocking strategy was necessary for our system.

For instance, for log(Vp) at 72 hpt in the RCCD model (Figure 2B), block 1 and block 2 are significant, and therefore, this source of variability must be considered. Since the blocks were assigned based on the qPCR plate used, it was henceforth treated as a random effect. Every qPCR plate used was in nature different, and the noise introduced into the system by each plate was random in nature. After defining blocking as a random effect, we repeated the model analysis using the restricted maximum likelihood method (REML). We then compared the $p$ value, the root mean square error (RMSE) and R-square, and the percentage of variability attributed to blocking for the three RSM models at both 48 and 72 hpt with and without blocking as a random effect (Table S9).

When analyzing viability as a response in each of our models, blocking was not required, as the samples were all quantified at the same time and did not suffer from the external source of uncontrolled noise coming from the qPCR plate. This was supported by two facts: (1) the highest variability observed in viability due to blocking was 5.66% for FCCD at 72 hpt (Table S4), while in BBD and RCCD, the blocking accounted for only 0.87% and 0.00% of the variability, respectively (Tables S6 and S2), and (2) the highest $p$ value for all models concerning viability was 0.0022 (Table S9).

Regarding the response of log(Vp), blocking had a different impact on each model. For RCCD, blocking improved the model fitness,




| | RCCD 4 factors | | | | | FCCD 4 factors | | | Box-Behnken 4 factors | | | Mixture Design[a] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Limit | -2 | -1 | 0 | 1 | 2 | -1 | 0 | 1 | -1 | 0 | 1 | 0.1 | 0.6 |
| Plasmid (μg/mL) | 0.1 | 0.375 | 0.65 | 0.925 | 1.2 | 0.375 | 0.65 | 0.925 | 0.3 | 0.65 | 1 | 0.2 | 1.2 |
| FectoVIR (μg/mL) | 1.5 | 2 | 2.5 | 3 | 3.5 | 2 | 2.5 | 3 | 2 | 2.5 | 3 | | |
| Number of runs | | | 32 | | | | 32 | | | 32 | | 27 | |
| Highest Log(Vp) | | | 10.84 | | | | 11.02 | | | 9.78 | | 12.77 | |
| Viability at highest Log(Vp) | | | 53.3 | | | | 81.3 | | | 83.9 | | 78.8 | |
| Best harvest timepoint for maximizing Log(Vp) (hpt) | | | 72 | | | | 48 | | | 48 | | 72 | |
| Best harvest timepoint for maximizing Viability (hpt) | | | 48 | | | | 48 | | | 48 | | 48 | |
| Response at 72hpt | **Log(Vp)** | | | **Viability** | | **Log(Vp)** | | **Viability** | **Log(Vp)** | | **Viability** | **Log(Vp)** | **Viability** |
| R-square | 0.819 | | | 0.868 | | 0.590 | | 0.886 | 0.678 | | 0.951 | 0.952 | 0.968 |
| Adjusted R-square | 0.649 | | | 0.745 | | 0.207 | | 0.779 | 0.377 | | 0.905 | 0.918 | 0.945 |
| p-value | 0.246 | | | **0.0003** | | 0.288 | | **0.0003** | 0.427 | | **<.0001** | **0.0002** | **<.0001** |

**Figure 1. Comparison of models**

Model comparison between RCCD, FCCD, Box-Behnken design (BBD), and mixture design (MD) with the selected limits, characteristics, and statistical parameters for both log(Vp) and viability responses for each model. [a]Runs for MD are calculated by adding MD (13 runs) and FCCD (12 runs). Two additional runs are added to every model, accounting for the two required controls for qPCR.

increasing the R-square value from 0.57 to 0.72 at 48 hpt and from 0.29 to 0.81 for 72 hpt, with a subsequent decrease in the *p* value from 0.2526 to 0.1158 and 0.9396 to 0.246, respectively. Furthermore, the blocking accounted for 64.67% of the total variability at 72 hpt (Table S2).

In contrast, in the FCCD, blocking worsened the model fitness at 48 hpt, raising the *p* value from 0.0597 to 0.2877, with a decrease in the R-square from 0.68 to 0.589. At 72 hpt, the obtained model fitness parameters were not relevantly different, with an R-square value of 0.59 when blocking was applied vs. 0.58. As expected, only 0.89% of the variability originated from the blocking at 72 hpt.

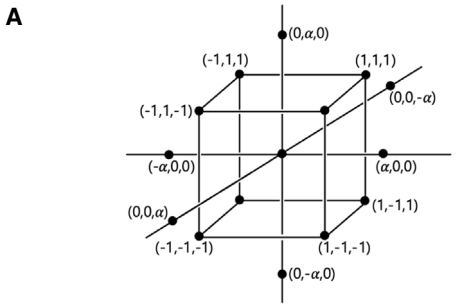
Finally, for the BBD, the introduction of blocking at 48 hpt was not largely different from the parameters obtained after blocking. For instance, the *p* value and R-square with blocking were 0.0684 and 0.6516, whereas the same parameters without blocking were 0.0474 and 0.70. However, at 72 hpt, blocking was crucial to decreasing the *p* value from 0.7885 to 0.4273 and to increasing the R-square from 0.38 to 0.67757. The variability stemming from the blocking was 34.39% at 72 hpt. For both RCCD and BBD, blocking accounted for a substantial proportion of the observed variability, 65% and 35%, respectively (Table S9), considering that blocking improved both *p* value and R-square fitness parameters.

The large disparities between the three models further affirmed the random nature of the uncontrolled factor coming from the different qPCR plates, underscoring the need for blocking in our experimental setup. Hence, blocking must not only be encouraged but considered a crucial factor when not all samples can be accommodated in a single qPCR run or any other equipment generating high intrinsic variability and uncontrolled noise.

### Different models work differently for each response

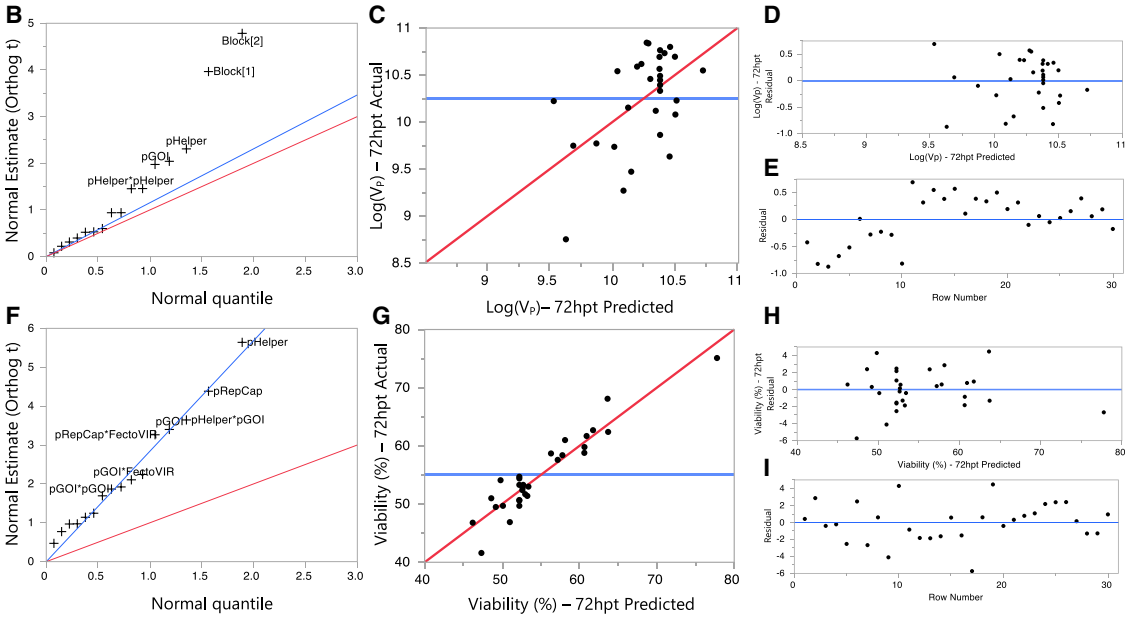
DOE approaches serve as invaluable tools in optimizing bioprocess conditions for multiple responses simultaneously. Our study highlights the variability in model efficiency among different DOE approaches when optimizing viability and Vp during rAAV production, as well as the variability in the identification of factors significantly affecting the studied responses.

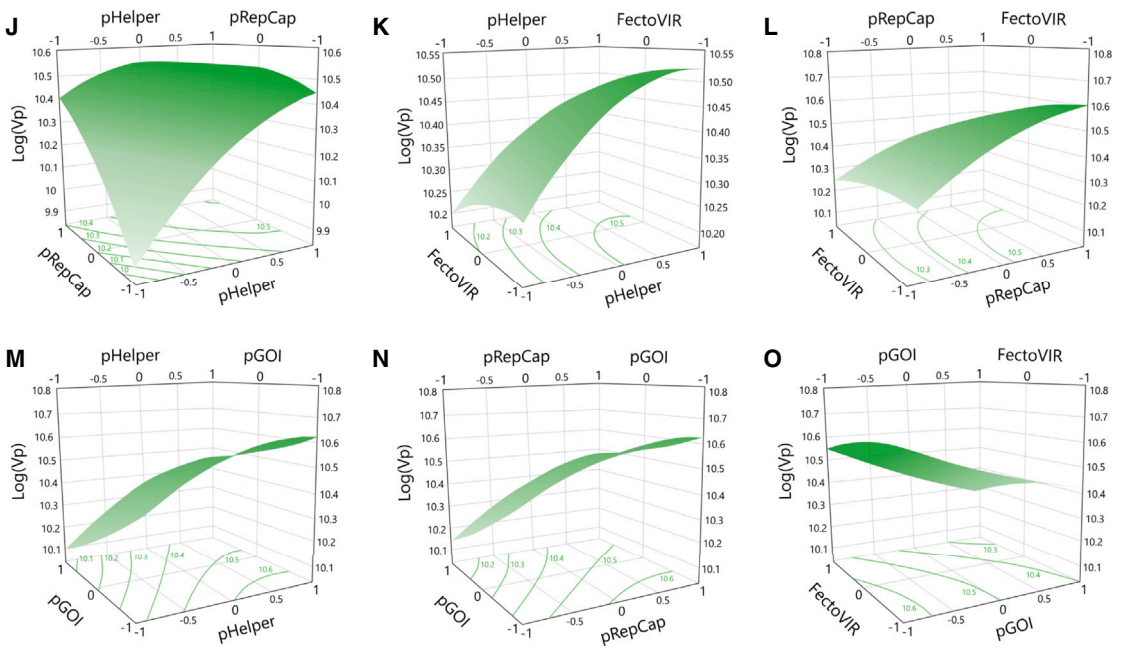
Regarding the response of log(Vp), we discovered significant variability in the ability of the models to identify significant factors. The only and most significant factor affecting response variability




**Rotatable Central Composite Design (RCCD)**

| Limit | -2 | -1 | 0 | 1 | 2 |
|---|---|---|---|---|---|
| Plasmid (μg/mL) | 0.1 | 0.375 | 0.65 | 0.925 | 1.2 |
| FectoVIR (μg/mL) | 1.5 | 2 | 2.5 | 3 | 3.5 |




*(legend on next page)*

at 72 hpt in RCCD appears to be pHelper ($p = 0.04$) (Table S2), underscoring the importance of helper functions in rAAV productivity. In contrast, when using FCCD, the model could identify only the transfection reagent as significant ($p = 0.01$) (Table S4). BBD failed to identify any significant factors affecting log(Vp) (Table S6), while employing a MD followed by a two-factor FCCD provided insight into previously unstudied interactions, highlighting all three plasmids, the transfection reagent, and the total DNA amount as significant factors (Table S8).

Moreover, we observed variability in the model fitting for log(Vp) among the studied approaches. MD followed by a two-factor FCCD yielded the highest R-square (0.952) (Table S8) and was the only significant fitting ($p = 0.0002$) for this response. The R-square value for RCCD was the second highest, with a value of 0.82 (Table S2), followed by BBD and FCCD (0.68 and 0.59, respectively) (Tables S4 and S6). In addition, all RSM models analyzed by REML exhibited high $p$ values (>0.25) (Tables S2, S4, and S6), compared to MD and two-factor FCCD for optimizing rAAV Vp.

Similar to log(Vp), we observed variability between the studied models regarding the response of viability. In all cases, pGOI and two-factor interactions between the transfection reagent and at least one of the plasmids were identified as significant. pHelper and pRepCap were significant for viability in all models at 72 hpt, except BBD (Tables S2, S4, S6, and S8). In the case of MD, all three plasmids were identified as significant ($p < 0.002$) (Table S8), while in the coupled two-factor FCCD, the transfection reagent, the total DNA amount, and the interactions between these two were all significant ($p < 0.01$) (Table S4). However, lower variability was observed in the model fitting when studying viability, indicating that all models could be efficiently used for optimizing this response. In this case, all models presented significant $p$ values (<0.0003) (Tables S2, S4, S6, and S8) and high R-square values. MD-two-factor FCCD had the highest R-square value (0.97) (Table S8), followed by BBD (R-square = 0.95) (Table S6), FCCD (R-square = 0.89) (Table S4), and last, RCCD (R-square = 0.87) (Table S2).

Here, we show that minor differences in the design of the experimental space can greatly affect the ability of the model to identify significant interactions between the factors and their effects in the studied responses. Interestingly, we observe that among RSM approaches, RCCD would be the most efficient option for optimizing rAAV productivity. However, it would be the least preferred option for viability optimization, as BBD showed the best performance among the three RSM approaches, highlighting that selecting and using an RSM method to analyze and optimize both responses at the same time can be challenging. Regarding only viability, BBD and MD followed by a two-factor FCCD emerged as preferred options. However, for simultaneous optimization of both viability and log(Vp), MD followed by a two-factor FCCD outperformed the RSM approaches. These findings underscore the importance of selecting the appropriate DOE approach tailored to the specific response of interest in bioprocess optimization studies.

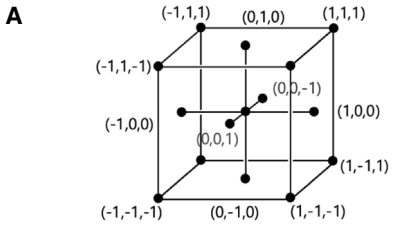
### Significant factors for harvesting-time-point selection

The selection of an appropriate harvest time point is crucial, as it significantly impacts both the quantity and the quality of the final product. Balancing our productivity metric with viability considerations is essential, as low viability levels may trigger protease release, potentially compromising product quality.[36] Our results indicate that different parameters should be considered when selecting the harvesting time point. For the four-factor RCCD model, the highest log(Vp) values are observed at 72 hpt, with high viability achievable at both 72 and 48 hpt (Figure 6). When harvesting at 72 hpt, implementing a blocking method is crucial, as it significantly affects Vp (Figure 2B). However, this trend is not observed at 48 hpt, where pGOI and pHelper play a more significant role in response variability (Figure S1B). Plasmid interactions with the transfection reagent were identified as significant factors for viability at both 48 and 72 hpt.

For the four-factor FCCD model, the highest log(Vp) and viability values among the runs were both achieved at 48 hpt (Table S3), indicating that 48 hpt would be the optimal harvest time point when using this model to analyze our system. The amount of the transfection reagent emerged as the most significant factor for Vp at both time points (Tables S3 and S4). The transfection reagent was significant only for viability at 48 hpt, with pGOI and pHelper affecting culture viability more than other factors (Tables S3 and S4). When optimizing log(Vp) using BBD, we observed the highest values at 48 hpt, reaching 9.88, followed by high viability levels above 75% (Table S5). At this time point, the transfection reagent appeared to be the most significant factor for both log(Vp) and viability (Figures S3A and S3D). At 72 hpt, similar to RCCD, blocking was observed to be crucial for optimizing log(Vp) (Figure 2B; Table S9). However, viability was still observed to be mostly affected by the transfection reagent (Figure 4F). Last, the highest log(Vp) values, achieved with MD followed by a two-factor FCCD, were observed at 72 hpt, with viability levels exceeding 78% (Figure 1; Table S8), highlighting the efficacy of this model in finding optimal values for log(Vp) while maintaining favorable viability levels. Notably, this model yielded the highest log(Vp) values among all studied

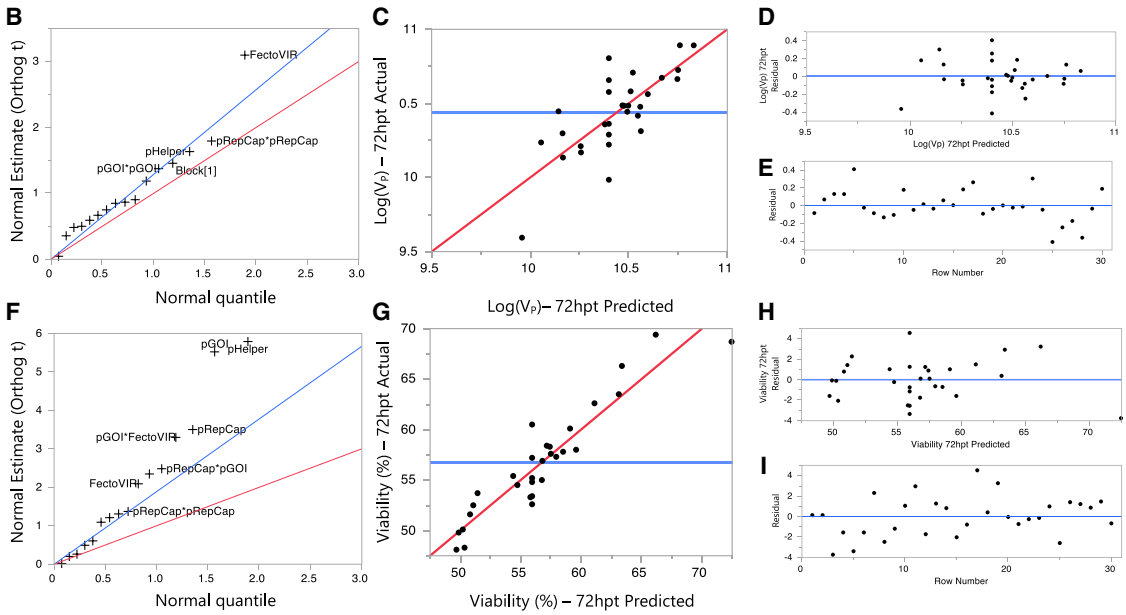
**Figure 2. RCCD 72 hpt**

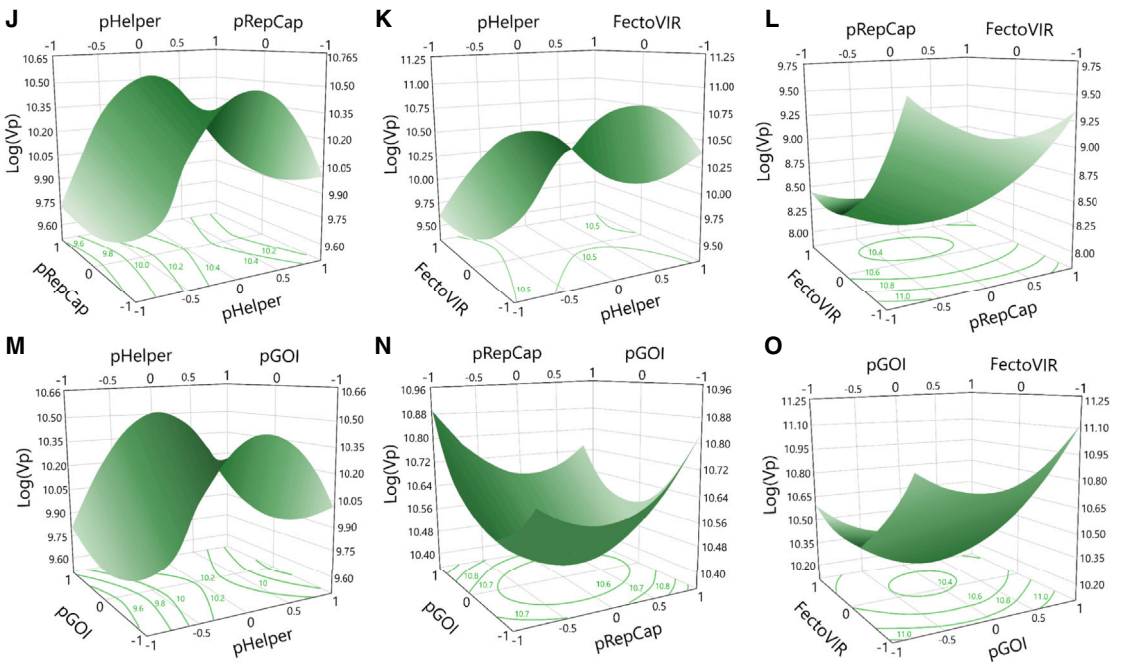
(A) Graphic representation of the experimental space of a four-factor RCCD (left) and limits used for the model (right). (B) Normal estimate (orthog t) against normal quantile at 72 hpt showing the absolute value of the effects to identify parameters that are deviating from normality. The red line has a slope of 1, whereas the blue line passes through the origin with a slope of Lenth's estimate of σ. (C) Comparison between actual and predicted log(Vp) at 72 hpt with the line of fit in red and the mean value in blue. (D) Residual of log(Vp) vs. predicted log(Vp) at 72 hpt to check if the values are randomly scattered around 0 (blue). (E) Residual vs. row number visualization at 72 hpt with a line at 0 (blue). (F), (G), (H), and (I) are equivalent to (B), (C), (D), and (E) for viability. (J–O) Surface plots for log(Vp) and (J) pHelper and pRepCap, (K) pHelper and FectoVIR, (L) pRepCap and FectoVIR, (M) pHelper and pGOI, (N) pGOI and pRepCap, and (O) pGOI and FectoVIR. Contour lines with log(Vp) values are at the bottom for all plots.




**Face Centered Composite Design (FCCD)**

| Limit | -1 | 0 | 1 |
|---|---|---|---|
| Plasmid (μg/mL) | 0.375 | 0.65 | 0.925 |
| FectoVIR (μg/mL) | 2 | 2.5 | 3 |




*(legend on next page)*

approaches, reaching 12.7 (Figures 1A, 2, 3, 4, and 5). Combining these approaches provided insight into new interactions, and we were able to identify the transfection reagent, the total amount of DNA, and their interactions as the most significant factors for both responses at 72 hpt (Figures 5A and 5E). At 48 hpt, Vp and viability were mainly driven by the three plasmids involved in the triple transfection (Figures S2A and S2D).

Our data revealed that each DOE model offered a unique perspective on the optimal harvest time point for rAAV production. These findings underscore the complexity of harvest-time-point selection in rAAV production, emphasizing the need to properly select the design approach to avoid misleading conclusions regarding both productivity and viability metrics.

### Different optimal ratios were obtained using each model

We derived the optimal plasmid ratio and FectoVIR-AAV (FV) concentration from each strategy using the generated models to maximize log(Vp) at 72 hpt. For some models, visualizing the optimal regions of the experimental space was clear (Figures 2K, 2L, and S3I–S3K).

However, when considering both responses at the same time and for some models like FCCD and BBD, the presence of saddle points (Figures 3J–3L) required the use of the desirability function (Equations 4 and 5) to determine the optimal ratio for maximum productivity. We prioritized log(Vp) with a desirability of 1 and viability with a desirability of 0.3.

Ultimately, the suggested optimal concentration of each plasmid and FV derived from each model was different. For instance, the optimal DNA concentration ratio for the four-factor RCCD at 72 hpt was 0.72:0.93:0.38 (pHelper:pRepCap:pGOI) with an FV concentration of 2 μg/mL, whereas for the four-factor FCCD it was 0.56:0.38:0.38 with a total of 2 μg/mL FV. Interestingly, when validating both optimal solutions, both RCCD and FCCD optimal values were non-significantly different. For the BBD, the optimal ratio was 0.66:0.64:0.38 coupled with 2.5 μg/mL FV. The only value shared by the three RSM models was the 0.38 μg/mL pGOI concentration. However, whereas FCCD required the same concentration of pRepCap, both BBD and RCCD demanded a higher concentration of the remaining plasmids, with pRepCap being the one with the highest concentration in RCCD (0.99 μg/mL) and pHelper being the highest concentration in BBD (0.66 μg/mL). Moreover, the optimal total amount of DNA in each model ranged from 2.09 μg/mL in RCCD to 1.31 μg/mL for FCCD, with BBD falling in the middle at 1.67 μg/mL. The calculated ratio between DNA and FV was 1.19 for RCCD, and despite the different plasmid ratios, BBD and FCCD had similar DNA:FV ratios of 0.67 and 0.66, respectively. Regarding the MD, the optimal concentration ratio was 0.80:0.70:0.50 with a total amount of DNA of 2 μg/mL and 2 μg/mL FV. From the optimal plasmid ratio, the subsequent FCCD found the optimal ratio DNA:FV at 2.37.

All optimal values were successfully validated (Figure 6), showing that the selection of suboptimal DOE models can provide misleading optimal values. For instance, the optimal ratio and consequently highest log(Vp) achieved with BBD was on average 10.4, whereas the maximum log(Vp) predicted in MD followed by FCCD was 12.77 (Figures 1 and 6), proving how the optimal value obtained from BBD was indeed suboptimal. The combination of MD and two-factor FCCD achieved a process improvement of 109-fold compared to the non-optimized condition (Figure 6).

### Coupling MD with FCCD outperformed RSM approaches

Among the various DOE approaches studied, MD followed by a two-factor FCCD emerged as the most efficient in enhancing both viral productivity and viability. This approach yielded the highest log(Vp) values and required only 27 experimental runs in total, compared to the four-factor RSM models that consisted of 32 runs. This indicates high efficiency in experimental resource utilization (Figure 5A) and reduced variability due to the absence of the need for blocking. Using a D-optimal MD, we determined that the optimal plasmid ratio in the mixture was 0.4:0.35:0.25 for pHelper:pGOI:pRepCap. After fixing this optimal ratio, we explored whether a two-factor RCCD or FCCD could further optimize the total DNA amount and the transfection reagent volume. The FCCD emerged as the preferred option, demonstrating better model fitting (Figures S5 and S6) and complementing the MD to identify transfection conditions that further improved log(Vp) and viability values to 12.7 and 78.8%, respectively (Figure 6). Moreover, the two-factor FCCD model after MD exhibited satisfactory R-square and adjusted R-square model fitting values at 72 hpt (Table S8). Importantly, this was the only approach to yield significant $p$ values for both responses (0.0002 for log(Vp) and <0.0001 for viability), unlike the RSM designs. Validation runs further confirmed the model's fitting and accuracy in predicting viability and log(Vp) values at 48 and 72 hpt (Figure 6).

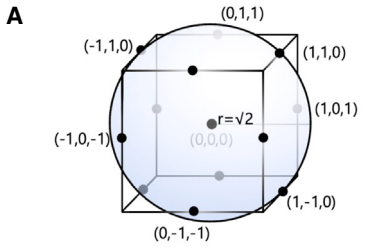
## DISCUSSION

### Full vs. reduced models

A common strategy for better model fit and more accurate predictions is reducing the model to where only the significant interactions

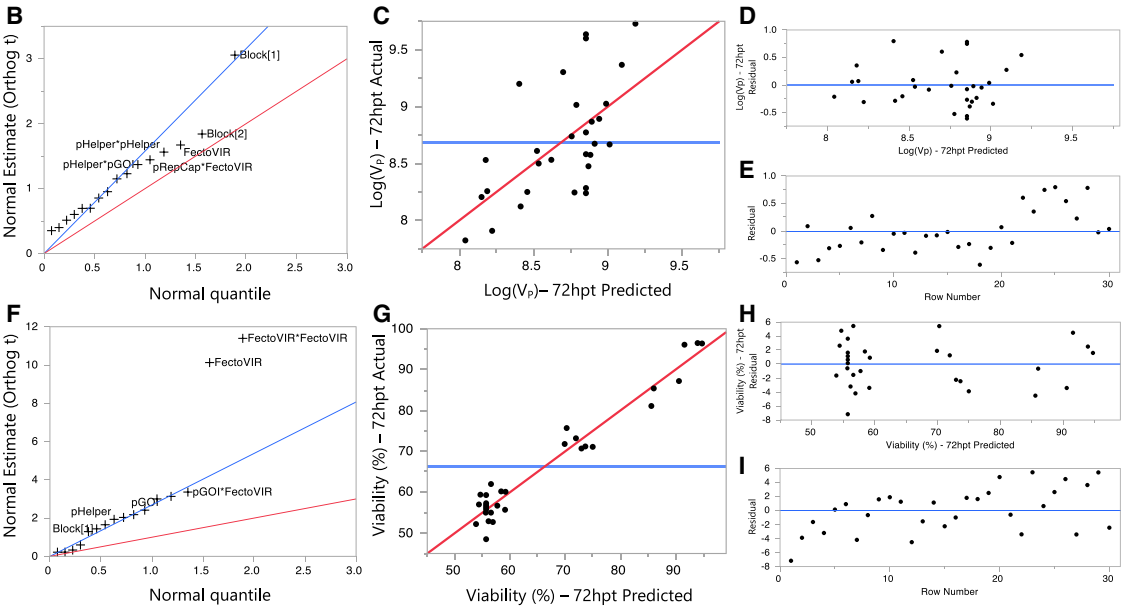
**Figure 3. FCCD 72 hpt**

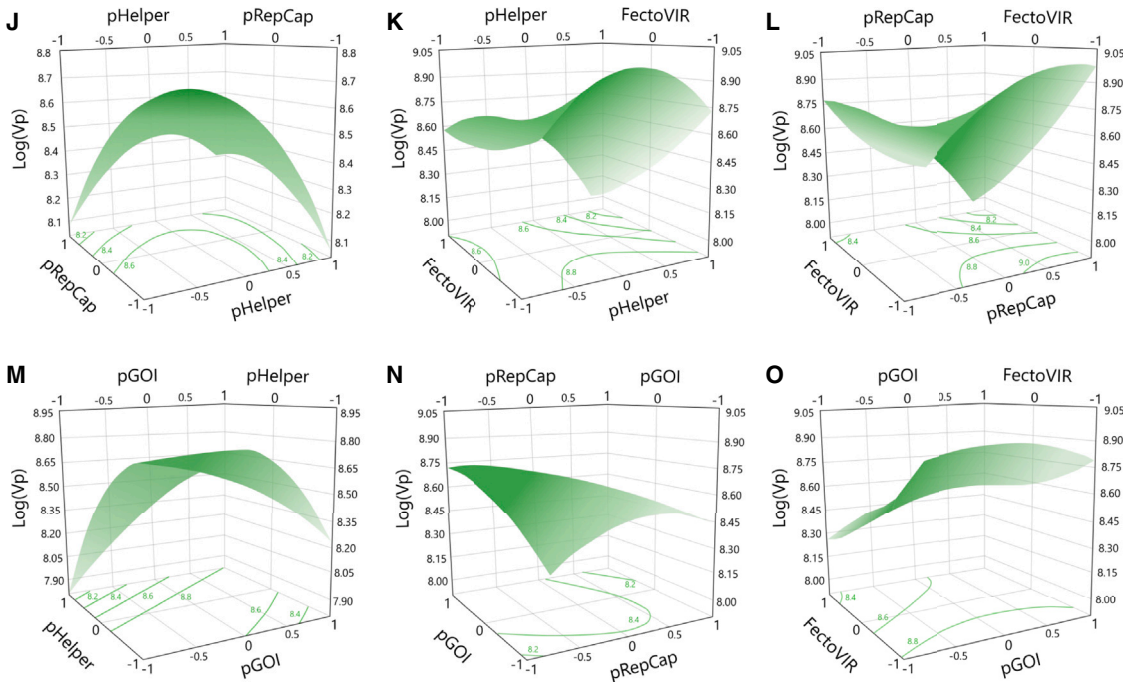
(A) Graphic representation of the experimental space of a four-factor FCCD (left) and limits used for the model (right). (B) Normal estimate (orthog t) against normal quantile at 72 hpt showing the absolute value of the effects to identify parameters that are deviating from normality. The red line has a slope of 1, whereas the blue line passes through the origin with a slope of Lenth's estimate of σ. (C) Comparison between actual and predicted log(Vp) at 72 hpt with the line of fit in red and the mean value in blue. (D) Residual of log(Vp) vs. predicted log(Vp) at 72 hpt to check if the values are randomly scattered around 0 (blue). (E) Residual vs. row number visualization at 72 hpt with a line at 0 (blue). (F), (G), (H), and (I) are equivalent to (B), (C), (D), and (E) for viability. (J–O) Surface plots for log(Vp) and (J) pHelper and pRepCap, (K) pHelper and FectoVIR, (L) pRepCap and FectoVIR, (M) pHelper and pGOI, (N) pGOI and pRepCap, and (O) pGOI and FectoVIR. Contour lines with log(Vp) values are at the bottom for all plots.




**Box-Behnken Design (BBD)**

| | Limit | -1 | 0 | 1 |
|---|---|---|---|---|
| | Plasmid (μg/mL) | 0.3 | 0.65 | 1 |
| | FectoVIR (μg/mL) | 2 | 2.5 | 3 |




*(legend on next page)*

are considered. In this way, higher R-square values and lower model *p* values can be achieved, improving predictions. However, in our study, the reduced models had lower performance than their respective full ones in almost all cases (Tables S1–S10). The optimization of the triple transfection for rAAV system is a particular biological challenge, as we know beforehand that all studied factors influence the final titer and are relevant for our final responses, both Vp and viability. All components of the triple transfection are essential for rAAV production, and their interactions can have profound effects on the overall yield and quality of the viral vectors. We know we cannot neglect any of them for successful rAAV optimization, and theoretically, the model able to identify these four factors as significant will explain better our system and most probably lead to better predictions. The synergistic effects and interactions between the plasmids involved in triple transfection can be overlooked in some models; this is why the analysis of the full model also helps compare the ability to identify these interactions in the different tested designs. Regarding the two-factor interactions, to have a working model for both Vp and viability, we need to consider the significant terms of both responses when reducing the model. Therefore, the models needed to be reduced based on the response with the highest number of significant factors. These reduced models did not present better-fitting parameters. This underscores the complexity of comprehensive modeling in biological processes where multiple factors interact with one another and even more in the production of rAAVs.

### Blocking

We have demonstrated the need for blocking when uncontrolled factors are introduced, such as when using qPCR for rAAV quantification. It is crucial to emphasize that the failure to apply blocking when experimentally necessary could lead to misleading conclusions. This became evident when we individually assessed the effect of blocking on each RSM approach. Interestingly, FCCD appeared not to require blocking, as evidenced by the lack of relevant changes in the *p* value with and without blocking and the low variability introduced due to blocking (Table S9). Conversely, RCCD showed the opposite effect—the introduction of blocking increased the model fitness and was able to explain a high percentage of the data variability. In addition, BBD demanded blocking only at 72 hpt. We hypothesize that the different impacts of blocking on each RSM are attributable to their unpredictable random effect. Therefore, to prevent incorrect conclusions and suboptimal solutions, it is necessary to first statistically test if blocking is influencing the studied response when uncontrolled factors are involved. Only in the absence of such factors would blocking not be recommended.

### Optimal designs

Optimal designs are a different type of DOE approach where the runs are placed in the experimental space in such a way that the resulting model minimizes the variance of either the regression coefficients (A optimal and D optimal) or the prediction variance in the design space (I optimal), among others. These designs are especially useful when the experimental region is irregular, the model is non-standard, the sample size is limited by the researcher, or the factors are constrained.[30] In this case, only MD was constrained by definition, since no plasmid concentration could reach 0; otherwise, no rAAVs would be produced. Consequently, an optimal design was appropriate to select the runs in the constrained experimental space. Among all optimal designs, the most widely used is D optimal. D-optimal designs maximize the determinant of the information matrix, thereby minimizing the determinant of the variance-covariance matrix of the regression coefficients. For the rest of the RSM approaches, there was no need to select an optimal design, as the experimental space was not constrained, and the standard design could be applied.

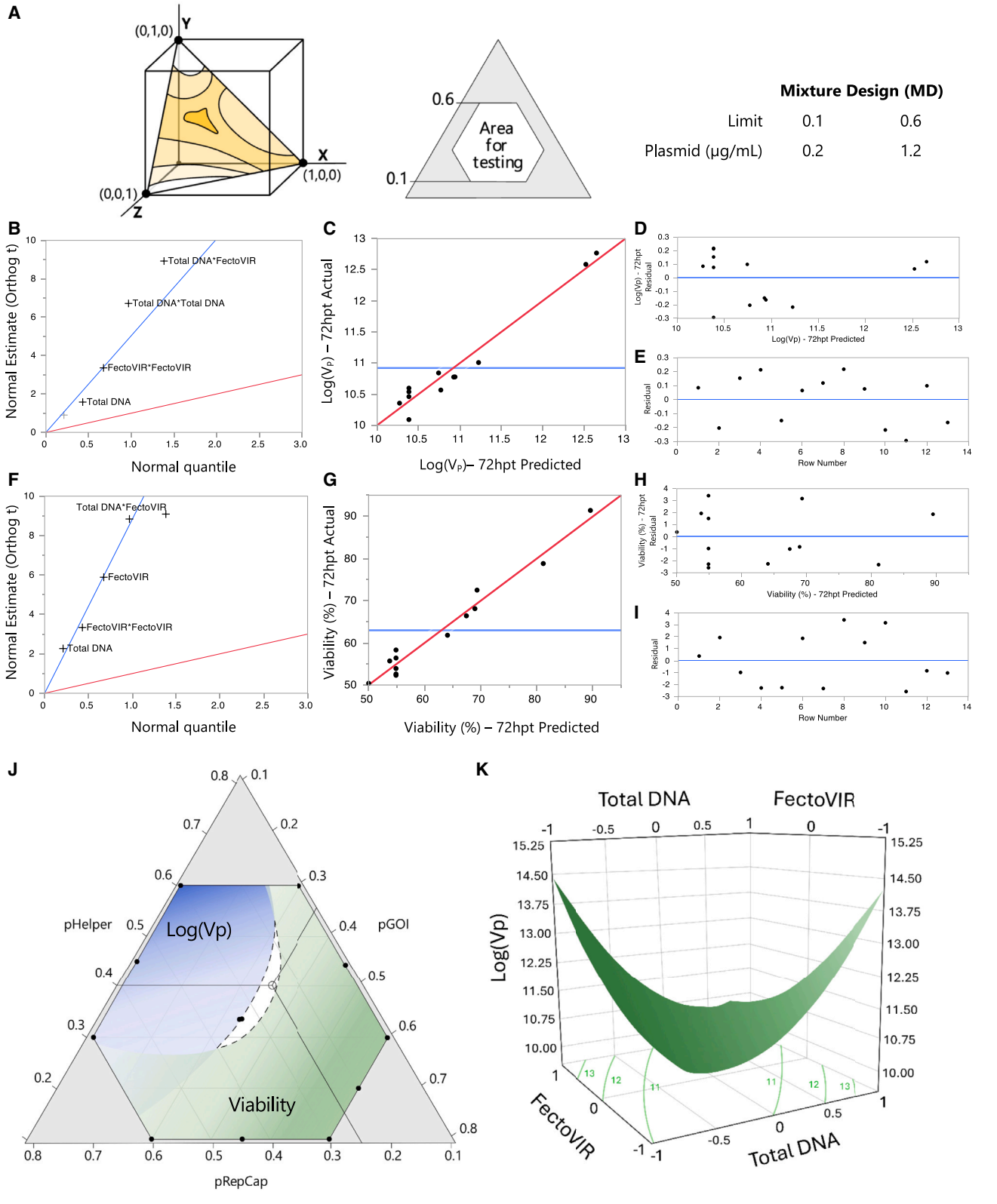
### The biological understanding of the system leads to MD

The primary objective of this work was to provide necessary tools to address challenges encountered in optimizing current rAAV manufacturing processes. When selecting an appropriate model, a thorough understanding of the biological context is crucial. In the triple-transfection process, there are many factors that affect the final outcome, and they should be appropriately analyzed to avoid masking and confounding effects in the models. The cell density is crucial for transfection-based processes. However, the biological reason behind the effects of the cell density in the process is substantially different from the biological reason to define a certain plasmid ratio, for instance. The cell density effect (CDE) acts as the limiting factor for cell density, imposing a threshold beyond which transfection efficiency and cell-specific productivity decrease. This is a common phenomenon shared across different cell-line platforms and expression methods, unrelated to the expression of AAV elements.[4,37] Therefore, to thoroughly study the interactions between AAV elements and reduce variables to optimize the experimental design, the cell density can be fixed at a value that is not yet affected by the CDE in our system.

Subsequently, the three plasmids and the transfection reagent can be defined as independent factors and were treated as such in all RSM designs (RCCD, FCCD, and BBD), resulting in two different ratios: pGOI:pHelper:pRepCap and DNA:transfection reagent. However, the total DNA amount added to the cell can lead to independent effects, driven solely by the DNA total concentration. These effects are

**Figure 4. BBD 72 hpt**

(A) Graphic representation of the experimental space of a four-factor BBD (left) and limits used for the model (right). (B) Normal estimate (orthog t) against normal quantile at 72 hpt showing the absolute value of the effects to identify parameters that are deviating from normality. The red line has a slope of 1, whereas the blue line passes through the origin with a slope of Lenth's estimate of σ. (C) Comparison between actual and predicted log(Vp) at 72 hpt with the line of fit in red and the mean value in blue. (D) Residual of log(Vp) vs. predicted log(Vp) at 72 hpt to check if the values are randomly scattered around 0 (blue). (E) Residual vs. row number visualization at 72 hpt with a line at 0 (blue). (F), (G), (H), and (I) are equivalent to (B), (C), (D), and (E) for viability. (J–O) Surface plots for log(Vp) and (J) pHelper and pRepCap, (K) pHelper and FectoVIR, (L) pRepCap and FectoVIR, (M) pHelper and pGOI, (N) pGOI and pRepCap, and (O) pGOI and FectoVIR. Contour lines with log(Vp) values are at the bottom for all plots.




*(legend on next page)*

overlooked when using RSM approaches. When optimizing ratios using an RSM approach, the total DNA concentration should be fixed. Thus, a relationship between all plasmid concentrations will emerge in the form of a ratio, and they can no longer be treated as independent variables, as an increase in one would cause a proportional decrease in the others. This could lead to suboptimal solutions or misleading estimate factors when analyzing these models. Consequently, the MD is necessary to successfully study the system. The different possible DNA mixtures lead to different expression balances of the AAV and adenoviral proteins, affecting cellular processes and homeostasis. On the other hand, FV affects the cell through the transfection process. The transfection process itself affects viability and cellular homeostasis independently,[38] and the factors driving this homeostasis disruption are DNA amount—regardless of the content of the plasmids—and transfection reagent. To study both phenomena independently, the MD approach separates the analysis of the plasmid interactions and their effects on the studied responses, treating the three plasmids as parts of a total mixture. The biological effects of expressing the different AAV elements and their interactions with cellular homeostasis are separate from the effects of just a high transfected DNA amount, regardless of the coded proteins. The first step of the MD studies the expressed proteins and the significance of their effects on the studied responses and the process. This is the main biological reasoning behind the optimal plasmid ratio. Subsequently, this optimal ratio is used to study the separate effects of the total amount of DNA and transfection reagent concentration. Although still related to the coded proteins, these effects have a separate biological effect, shared by any transfection-based process. Using four-factor RSM approaches, these two biological scenarios are mixed, leading to confounded effects. The coupling of MD-RSM enabled us to better analyze biological phenomena by separating variables according to their biological relevance. This approach provides insight into the role of the total DNA amount, something that is impossible when using RSM methodologies, as described above.

One of the most crucial variables in the process is the gene of interest (GOI), whose effects and interactions in the cell will substantially vary for each rAAV-based product. Unless engineered not to be expressed in the host cell line,[39] the GOI will have an impact on the cell metabolism and physiology, with potential interactions with the rest of the AAV elements that the DOE model should be able to identify and analyze. Moreover, the limits of the experimental space to study new GOIs will be constrained to their effect on the cell. Since the behavior of the GOI may differ depending on the final product, we anticipate that testing other GOIs will yield different responses. Our results showed that a D-optimal quadratic MD approach is the most suitable design to study these effects, as it is able to focus on plasmid interactions better than RSM approaches.

### Two-factor FCCD showed relevant relationship between DNA and the transfection reagent

The two-factor FCCD after MD allowed for the specific study of the effects of total DNA in the system. Here, we observed two areas for optimal log(Vp) at 72 hpt. The first area corresponded to the lowest concentration of DNA (0.93 μg/mL) and the highest concentration of FV (2.7 μg/mL), while the second area occurred with the highest DNA concentration (3.06 μg/mL) and the lowest FV concentration (1.29 μg/mL). To understand the appearance of these opposing solutions, we investigated the dynamics of the transfection system.

There are two different strategies to increase the number of plasmid copies in the cell: increase the plasmid concentration in the culture or increase the transfection reagent. However, high concentrations of transfection reagent can be cytotoxic and trigger cellular death.[38] Therefore, it is crucial to find the ideal concentrations to maximize endogenous protein synthesis pathways while not compromising viability.

For instance, when DNA amount and FV are at their highest, it could potentially result in the highest amount of DNA in the cell, the highest cytotoxic protein synthesis, and FV toxicity. Therefore viability and log(Vp) drastically decreased (Figure 5; Table S8). This is supported by the two opposing areas for optimal log(Vp) in the two-factor FCCD. In addition, the cytotoxic effect of FV can be observed, as increasing concentrations lead to a steady decrease in viability. The obtained optimal DNA concentration of 1 μg of DNA per milliliter of culture agrees with the manufacturer's recommendations for optimal expression. However, the observed ratios to maximize log(Vp) were 3.06:1.29 and 0.93:2.70 DNA:FV, diverging from the generally recommended 1:1. Using an MD-FCCD approach allows the analysis of the DNA-transfection reagent interaction showing that when DNA concentration is close to 1 μg/mL, a high concentration of FV will be necessary to ensure high productivity. However, when DNA concentration is high, an FV concentration close to 1.29 μg/mL should suffice to achieve maximum productivity.

Alternatively, in the two-factor RCCD after the MD, FV being toxic at high concentrations could also be observed (Tables S8 and S10). However, on the opposite end, with a low concentration of FV, log(Vp) did not suffer any relevant change, regardless of the quantity of DNA. This response proved that FCCD is more suitable to better study all biological phenomena and provide better-defined ratios.

**Figure 5. MD 72 hpt**

(A) Graphic representation of the experimental space of a three-factor MD (left) and limits used for the model (right). (B) Normal estimate (orthog t) against normal quantile at 72 hpt showing the absolute value of the effects to identify parameters that are deviating from normality. The red line has a slope of 1, whereas the blue line passes through the origin with a slope of Lenth's estimate of σ. (C) Comparison between actual and predicted log(Vp) at 72 hpt with the line of fit in red and the mean value in blue. (D) Residual of log(Vp) vs. predicted log(Vp) at 72 hpt to check if the values are randomly scattered around 0 (blue). (E) Residual vs. row number visualization at 72 hpt with a line at 0 (blue). (F), (G), (H), and (I) are equivalent to (B), (C), (D), and (E) for viability. (J) Experimental space showing how both log(Vp) in blue and viability in green change at different ratios. The threshold for log(Vp) is set at 10.4 and for viability at 53%. (K) Surface plot for log(Vp), total DNA, and FectoVIR. Contour lines with log(Vp) values are at the bottom.

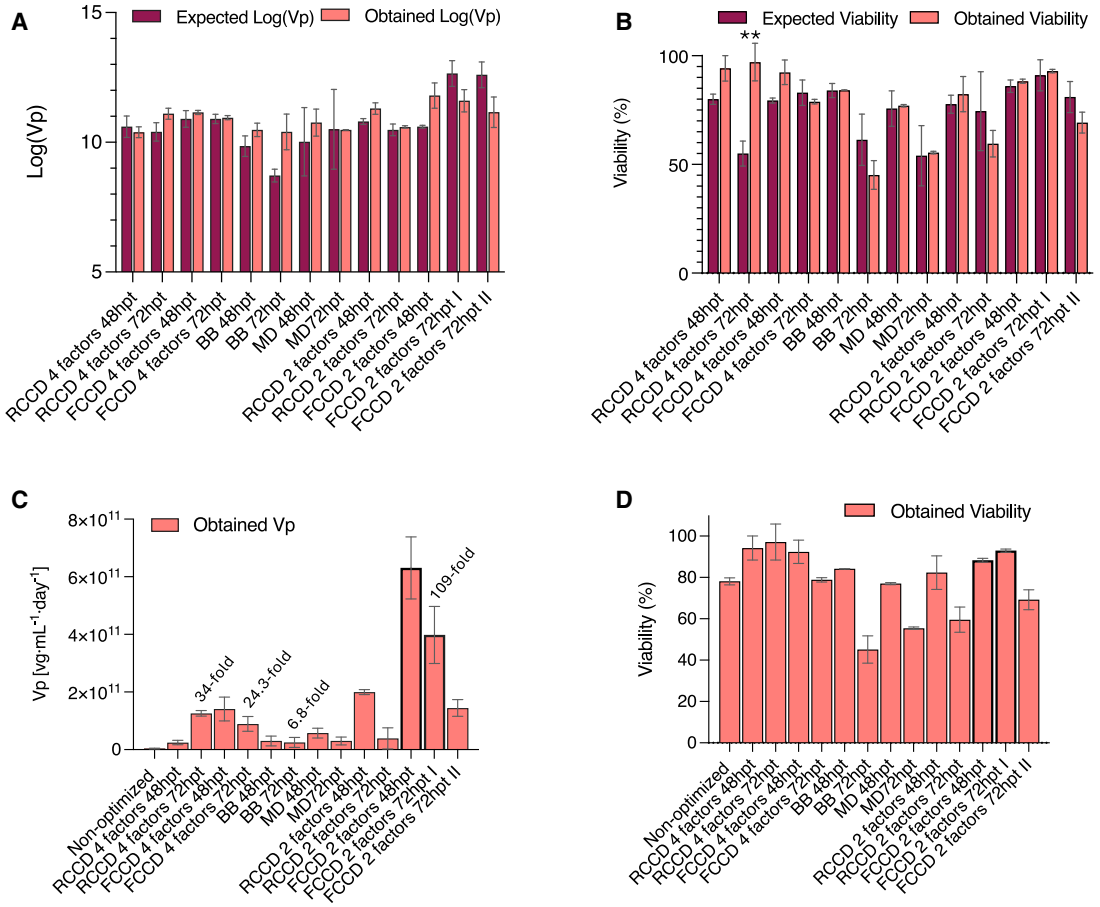



**Figure 6. Validations and process improvement**

(A and B) Predicted and actual experimental values for all validated conditions at 48 and 72 hpt for all models regarding (A) log(Vp) and (B) viability. Only viability for RCCD at 72 hpt presented a significant ($p = 0.01$) difference between predicted and obtained values using multiple comparison two-way ANOVA. The rest were non-significant, validating the models. (C and D) (C) Obtained volumetric (Vp) productivity and (D) obtained viability for all validated optimal points in each model compared to the non-optimized condition prior to any DOE optimization. Conditions framed in black are those showing the highest process improvement and the selected approach: MD + FCCD. Fold improvements are shown for the four studied DOE approaches at 72 hpt. Error bars represent standard error of the model for the predicted values and standard error, with $N = 3$ for most experimental values. MD 72 hpt had $N = 7$ and the non-optimized condition $N = 9$.

### Recommendations for rAAV optimization

Achieving standardized guidelines or specific optimal transfection conditions that work for every rAAV production system is not feasible due to the significant variability introduced by factors, including the GOI, producer cell line, medium, and AAV serotype, among others. In this work, we offer a robust methodological framework that researchers can follow to tailor their optimizations to their specific projects and rAAV systems. The first recommendation is to always consider blocking in the experimental design. It should be one of the first steps in the designing phase, as it can be critical for accounting for potential confounding of the factors to study in the model. The second recommendation is to couple MD with an FCCD as the best approach to systematically optimize plasmid ratios and the amount of transfection reagent and total DNA. This approach has proven significantly better than conventional RSM methods. The third recommendation is to understand and consider the biological context of the system to optimize. The biological implications of each factor to be included in the DOE design is crucial when selecting the appropriate design. Moreover, our last recommendation when determining optimal conditions is to consider all relevant responses together when analyzing the different generated models to ensure that predictions, beyond being statistically sound, will indeed be applicable in the biological system.

## MATERIALS AND METHODS

### Cell culture

HEK293SF-3F6 cells from the National Research Council of Canada (NRCC) were cultivated in disposable polycarbonate 125-mL baffled shake flasks equipped with vented caps (Corning Life Sciences, USA) in 20 mL of HyCell TransFx-H (Cytiva Life Sciences, USA) supplemented with 4 mM GlutaMAX (Gibco, Life Technologies, Thermo

Fisher, USA) and 0.1% Pluronic F-68 non-ionic surfactant (Gibco, Life Technologies). Cultures were maintained at 37°C and 5% $CO_2$ in a Hera Cell 150 incubator (Thermo Fisher, USA) with agitation at 130 rpm provided by Celltron (Infors HT, Switzerland) to ensure optimal growth conditions.

Samples were collected for cell count and viability assessment at 24, 48, and 72 hpt using NucleoCounter-250 (Chemometec, Denmark) according to the manufacturer's instructions.

#### Triple transient transfection in HEK293SF-3F6 cells

HEK293SF-3F6 cells were subjected to triple transfection using the following plasmids: PF1451 (pGOI), pXR2 (pRepCap), and pXX6-80 (pHelper). PF1451 (Plasmid Factory, Germany) carries the GOI GFP flanked by inverted terminal repeats (ITRs), pXR2 (National Gene Vector Biorepository, USA) contains the Rep and Cap genes from AAV2, and pXX6-80 (National Gene Vector Biorepository, USA) provides the helper genes E2 and E4. HEK293SF-3F6 cells were seeded at $1 \times 10^6$ cells/mL the day prior to transfection to ensure that the cultures were at $2 \times 10^6$ cells/mL when the transfection was performed. The transfection mix corresponded to 5% of the total working volume. Briefly, DNA was added to the corresponding volume of culture medium and vortexed for 10 s, followed by the addition of the necessary amount of the transfection reagent FV (Polyplus, France). The DNA and FV mix was vortexed three times for 3 s and incubated in accordance with the manufacturer's instructions. The ratios between each plasmid and between the DNA and the transfection reagent were determined by the corresponding experimental design.

#### DOE and models used

In this study, we analyzed the effect of viability, measured in percentage of live cells, on the production of rAAV, and we maximized the Vp, expressed in $vg \times mL^{-1} \times day^{-1}$. Vp data were transformed to log(Vp) as presented in Equation 1 to achieve normal distribution of residuals. Normal probability and Box-Cox plots were routinely used to validate that the log transformation was in fact required and successful. With a normal distribution of residuals, four different designs were used: an RCCD, a BBD, an FCCD, and an MD:

$$Log\ (V_P) = \log_{10} \frac{titer\ \left(\frac{vg}{mL}\right)}{time\ (day)}. \quad \text{(Equation 1)}$$

The first three designs (RCCD, BBD, and FCCD) treat all four factors as independent variables (pHelper, pGOI, pRepCap, and FV) to predict optimal plasmid and DNA:FV ratios. MD includes only pHelper, pGOI, and pRepCap, presenting the mathematical constraint of the sum of all plasmids being constant, to study plasmid ratios. To study the DNA:FV ratio, fixing the optimal plasmid ratio from MD, a two-factor FCCD and RCCD were fit. Considering the nature of the triple-transfection system, all designs were constrained to have a minimum quantity of each plasmid and FV to ensure successful production of rAAV.

The experimental design and statistical analysis were conducted using JMP 16 Pro (SAS Institute, USA) for all designs.

#### RSMs

All RSM models (RCCD, BBD, and FCCD) fit the experimental space to a second-order polynomial equation (Equation 2) where y is the response, either log (Vp) or viability (%); $\beta_0$ is the offset term; $\beta_i$ is the linear coefficient; $\beta_{ii}$ is the quadratic coefficient; $\beta_{ij}$ is the interaction coefficient with $x_i$ and $x_j$ as the independent variable; and $\epsilon$ is the noise observed in the response:

$$y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ii} x_i^2 + \sum \sum \beta_{ij} x_i x_j + \varepsilon. \quad \text{(Equation 2)}$$

In terms of rotatability, the difference between the three designs lies in the variance of the predicted response at points of interest. If the variance is the same for all points that are at the same distance from the center, the design will be rotatable. To ensure rotatability, the distance from the axial points to the center of the experimental space, $\alpha$, must be equal to the fourth root of the number of points in the factorial portion of the design ($n_F$) (Equation 3). For instance, in the four-factor RCCD presented in this work, $n_F$ equals 16, meaning that to have rotatability, $\alpha$ equals 2, while the FCCD has, by definition, $\alpha$ equal to 1, and therefore, it is not rotatable:

$$\alpha = (n_F)^{1/4}. \quad \text{(Equation 3)}$$

Regarding BBD, all experimental runs lie on a sphere of radius $\sqrt{2}$ from the center of the experimental space, making BBD rotatable for our four-factor design.

All RSM designs screened the four factors at three levels: a low level coded as −1, an intermediate level coded as 0, and a high level coded as +1 (Figure 1).

Due to the high complexity of some models, we used the desirability function (Equation 4) to identify the optimal values.[40] In the desirability function, $T$ is the target, $L$ is the lower limit, $y$ is the response to be maximized, and $r$ is a number between 0 and 1 to show the relevancy of the response. To analyze more than one response simultaneously, such as viability and log(Vp) in our case, we merged both desirabilities in Equation 5, where $k$ is the response:

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y - L}{T - L}\right)^r & L \le y \le T \\ 1 & y > T \end{cases}, \quad \text{(Equation 4)}$$

$$D = d_1^{1/k}\ d_2^{1/k} \ldots d_i^{1/k}. \quad \text{(Equation 5)}$$

Finally, the contribution of the quantification method to the observed variability was studied as both fixed and random effects for RCCD, BBD, and FCCD, as they required more than 24 runs,

and quantification by qPCR was limited to batches of 24 samples simultaneously. To do this, all RSM designs were split into three blocks of 10 runs each, with a total number of 6 central points (2 per block) to ensure the robustness of the design. To analyze the blocks as fixed effects, the significance was calculated by the analysis of variance (ANOVA). When the blocks were treated as random effects, the significance of each model was determined by the REML.[41]

### MD

The MD was defined as a mixture of the three required plasmids for the successful production of rAAV with a minimum of 10% and a maximum of 60% of each plasmid in the final mixture.

Compared to a cube (RCCD and FCCD) or a sphere (BBD) for the experimental space, in the case of MD we selected an optimal design to place the runs in the experimental space so they minimize the variance of the estimators. This is achieved by minimizing the covariance matrix. This D-optimal criterion is one of the most popular designs and it is available in JMP, Minitab, Design-Expert, and other software packages.

All factors were fitted to the quadratic formula presented in Equation 6, where β corresponds to the expected response to the pure blend $x_i = 1$ and $x_j = 0$ when $j \neq i$:

$$E(y) = \sum \beta_i x_i + \sum \sum \beta_{ij} x_i x_j. \quad \text{(Equation 6)}$$

No blocking was required, as the number of runs was only 12 and, consequently, the significance of the model was determined by an ANOVA.

### qPCR-based rAAV quantification

Cell culture was harvested at 48 and 72 hpt and diluted 1:1 with lysis buffer (Tris HCl 400 mM, Triton X-100 1%, and $MgCl_2$ 20 mM, adjusted to pH 7.5). After 1 h incubation at 37°C and agitation at 130 rpm, the lysate was centrifuged for 20 min at 4,000 × *g* and 4°C. The supernatants were stored at −70°C for long-term storage.

For quantification, samples were thawed in a controlled manner at room temperature. Then, 5 μL of sample was mixed with 2 μL of 1 U/μL DNase I (Thermo Fisher Scientific, USA) and 13 μL of 10× DNase I reaction buffer with $MgCl_2$ (Thermo Fisher Scientific). The mix was incubated for 16 h at 37°C. To inhibit the activity of DNase I, 4 μL of 50 mM EDTA (Thermo Fisher Scientific) was added to the mix, followed by a 30-min incubation at 70°C. Last, 5 μL of proteinase K (Thermo Fisher Scientific) was added to the sample and incubated for 2 h at 55°C. The proteinase K was inactivated for 15 min at 95°C. Two microliters of the freshly treated sample was mixed with 0.5 μL of 10 μM forward primer (5′-ACGTCAATGGGTGGAGTATTT-3′) and reverse primer (5′-AGGTCATGTACTGGGCATAAT-3′) binding to the GFP sequence with 5 μL of Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent Technologies, USA) and 0.15 μL of 2 μM reference dye.

Amplification was executed in a QuantStudio 5 real-time PCR system (Thermo Fisher Scientific) with the following conditions: 50°C for 2 min, 95°C for 10 min; 40× 95°C for 15 s, 60°C for 1 min, 95°C for 15 s, 60°C for 1 min, and 95°C for 1 s. Plates run in the QuantStudio 5 real-time PCR system were analyzed via ThermoCloud (Thermo Fisher Scientific).

To ensure the reliability of the generated data, two controls were added: saturate lysate control (SLC) and transfected lysate control (TLC). The SLC was prepared from a non-transfected cell culture that underwent the lysis protocol in parallel with the transfected samples. Then, 5 μL of cell lysate was mixed with 40 ng of pGOI plasmid, 9 μL of 10× DNase I reaction buffer with $MgCl_2$, and 2 μL of 1 U/μL DNase I. The SLC aimed to verify that the DNase step was successful. For the generation of the TLC, a single plasmid transfection was performed with pGOI following the triple transfection, lysis, and the quantification protocol. The TLC aimed to guarantee that the effect of DNase I remained independent of whether the plasmid had been transfected or spiked afterward.

## DATA AND CODE AVAILABILITY

Raw data from viability and titer measurements for each run can be found in the supplemental information.

## ACKNOWLEDGMENTS

The authors would like to acknowledge generous support from Novo Nordisk Foundation. This work was supported by NNF20CC0035580 and NNF20SA0066621. L.K.N. is supported by NNF14OC0009473. J.L.-G. is supported by NNF22OC0078741 and Marie Skłodowska-Curie Actions (MSCA) postdoctoral fellowship 101105465. K.T. is supported by a DTUQ Alliance PhD grant.

## AUTHOR CONTRIBUTIONS

Investigation, K.T. and D.C.-T.; visualization, K.T., D.C.-T., and J.L.-G.; writing – original draft, K.T., D.C.-T., and J.L.-G.; writing – review & editing, K.T., D.C.-T., and J.L.-G.; supervision and review, L.K.N.; conceptualization, J.L.-G.; supervision, J.L.-G.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2024.101329.

## REFERENCES

1. Ma, C.-C., Wang, Z.-L., Xu, T., He, Z.-Y., and Wei, Y.-Q. (2020). The approved gene therapy drugs worldwide: from 1998 to 2019. Biotechnol. Adv. *40*, 107502. https://doi.org/10.1016/j.biotechadv.2019.107502.

2. Chancellor, D., Barrett, D., Nguyen-Jatkoe, L., Millington, S., and Eckhardt, F. (2023). The state of cell and gene therapy in 2023. Mol. Ther. *31*, 3376–3388. https://doi.org/10.1016/j.ymthe.2023.11.001.

3. Sung, Y.K., and Kim, S.W. (2019). Recent advances in the development of gene delivery systems. Biomater. Res. *23*, 8. https://doi.org/10.1186/s40824-019-0156-z.

4. Lavado-García, J., Pérez-Rubio, P., Cervera, L., and Gòdia, F. (2022). The cell density effect in animal cell-based bioprocessing: Questions, insights and perspectives. Biotechnol. Adv. *60*, 108017. https://doi.org/10.1016/j.biotechadv.2022.108017.

5. Gutiérrez-Granados, S., Cervera, L., Kamen, A.A., and Gòdia, F. (2018). Advancements in mammalian cell transient gene expression (TGE) technology for accelerated production of biologics. Crit. Rev. Biotechnol. *38*, 918–940. https://doi.org/10.1080/07388551.2017.1419459.

6. Ayuso, E., Mingozzi, F., and Bosch, F. (2010). Production, Purification and Characterization of Adeno-Associated Vectors. Curr. Gene Ther. *10*, 423–436. https://doi.org/10.2174/156652310793797685.

7. Zhao, H., Lee, K.-J., Daris, M., Lin, Y., Wolfe, T., Sheng, J., Plewa, C., Wang, S., and Meisen, W.H. (2020). Creation of a High-Yield AAV Vector Production Platform in Suspension Cells Using a Design-of-Experiment Approach. Mol. Ther. Methods Clin. Dev. *18*, 312–320. https://doi.org/10.1016/j.omtm.2020.06.004.

8. Bilal, A.S., Parker, S.N., Murray, V.B., MacDonnell, L.F., Thuerauf, D.J., Glembotski, C.C., and Blackwood, E.A. (2023). Optimization of Large-Scale Adeno-Associated Virus (AAV) Production. Curr. Protoc. *3*, e757. https://doi.org/10.1002/cpz1.757.

9. Bosma, B., du Plessis, F., Ehlert, E., Nijmeijer, B., de Haan, M., Petry, H., and Lubelski, J. (2018). Optimization of viral protein ratios for production of rAAV serotype 5 in the baculovirus system. Gene Ther. *25*, 415–424. https://doi.org/10.1038/s41434-018-0034-7.

10. Fu, Q., Lee, Y.S., Green, E.A., Wang, Y., Park, S.Y., Polanco, A., Lee, K.H., Betenbaugh, M., McNally, D., and Yoon, S. (2023). Design space determination to optimize DNA complexation and full capsid formation in transient rAAV manufacturing. Biotechnol. Bioeng. *120*, 3148–3162. https://doi.org/10.1002/BIT.28508.

11. N Politis, S., Colombo, P., Colombo, G., and M Rekkas, D. (2017). Design of experiments (DoE) in pharmaceutical development. Drug Dev. Ind. Pharm. *43*, 889–901. https://doi.org/10.1080/03639045.2017.1291672.

12. Ilzarbe, L., Álvarez, M.J., Viles, E., and Tanco, M. (2008). Practical applications of design of experiments in the field of engineering: a bibliographical review. Qual. Reliab. Eng. Int. *24*, 417–428. https://doi.org/10.1002/qre.909.

13. Yu, P., Low, M.Y., and Zhou, W. (2018). Design of experiments and regression modelling in food flavour and sensory analysis: A review. Trends Food Sci. Technol. *71*, 202–215. https://doi.org/10.1016/j.tifs.2017.11.013.

14. Drovandi, C.C., Holmes, C., McGree, J.M., Mengersen, K., Richardson, S., and Ryan, E.G. (2017). Principles of Experimental Design for Big Data Analysis. Stat. Sci. *32*, 385–404. https://doi.org/10.1214/16-STS604.

15. Catalán-Tatjer, D., Tzimou, K., Nielsen, L.K., and Lavado-García, J. (2024). Unravelling the essential elements for recombinant adeno-associated virus (rAAV) production in animal cell-based platforms. Biotechnol. Adv. *73*, 108370. https://doi.org/10.1016/j.biotechadv.2024.108370.

16. Durocher, Y., Pham, P.L., St-Laurent, G., Jacob, D., Cass, B., Chahal, P., Lau, C.J., Nalbantoglu, J., and Kamen, A. (2007). Scalable serum-free production of recombinant adeno-associated virus type 2 by transfection of 293 suspension cells. J. Virol. Methods *144*, 32–40. https://doi.org/10.1016/j.jviromet.2007.03.014.

17. Park, J.Y., Lim, B.P., Lee, K., Kim, Y.G., and Jo, E.C. (2006). Scalable production of adeno-associated virus type 2 vectors via suspension transfection. Biotechnol. Bioeng. *94*, 416–430. https://doi.org/10.1002/bit.20776.

18. Hildinger, M., Baldi, L., Stettler, M., and Wurm, F.M. (2007). High-titer, serum-free production of adeno-associated virus vectors by polyethyleneimine-mediated plasmid transfection in mammalian suspension cells. Biotechnol. Lett. *29*, 1713–1721. https://doi.org/10.1007/s10529-007-9441-3.

19. Grieger, J.C., Soltys, S.M., and Samulski, R.J. (2016). Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. Mol. Ther. *24*, 287–297. https://doi.org/10.1038/mt.2015.187.

20. Balakrishnan, R., Mohan, N., and Sivaprakasam, S. (2022). Application of design of experiments in bioprocessing: process analysis, optimization, and reliability. In Current Developments in Biotechnology and Bioengineering (Elsevier), pp. 289–319. https://doi.org/10.1016/B978-0-323-91167-2.00013-7.

21. Saleh, A.K., El-Gendi, H., Soliman, N.A., El-Zawawy, W.K., and Abdel-Fattah, Y.R. (2022). Bioprocess development for bacterial cellulose biosynthesis by novel Lactiplantibacillus plantarum isolate along with characterization and antimicrobial assessment of fabricated membrane. Sci. Rep. *12*, 2181. https://doi.org/10.1038/s41598-022-06117-7.

22. Limkar, M.B., Pawar, S.V., and Rathod, V.K. (2019). Statistical optimization of xylanase and alkaline protease co-production by Bacillus spp using Box-Behnken Design under submerged fermentation using wheat bran as a substrate. Biocatal. Agric. Biotechnol. *17*, 455–464. https://doi.org/10.1016/j.bcab.2018.12.008.

23. Lavado-García, J., Cervera, L., and Gòdia, F. (2020). An Alternative Perfusion Approach for the Intensification of Virus-Like Particle Production in HEK293 Cultures. Front. Bioeng. Biotechnol. *8*, 617. https://doi.org/10.3389/fbioe.2020.00617.

24. El-Naggar, N.E.-A., and El-Shweihy, N.M. (2020). Bioprocess development for L-asparaginase production by Streptomyces rochei, purification and in-vitro efficacy against various human carcinoma cell lines. Sci. Rep. *10*, 7942. https://doi.org/10.1038/s41598-020-64052-x.

25. Liu, W., Xiang, H., Zhang, T., Pang, X., Su, J., Liu, H., Ma, B., and Yu, L. (2021). Development of a New Bioprocess for Clean Diosgenin Production through Submerged Fermentation of an Endophytic Fungus. ACS Omega *6*, 9537–9548. https://doi.org/10.1021/acsomega.1c00010.

26. Ahleboot, Z., Khorshidtalab, M., Motahari, P., Mahboudi, R., Arjmand, R., Mokarizadeh, A., and Maleknia, S. (2021). Designing a Strategy for pH Control to Improve CHO Cell Productivity in Bioreactor. Avicenna J. Med. Biotechnol. (AJMB) *13*, 123–130. https://doi.org/10.18502/ajmb.v13i3.6365.

27. Coplan, L., Zhang, Z., Ragone, N., Reeves, J., Rodriguez, A., Shevade, A., Bak, H., and Tustian, A.D. (2024). High-yield recombinant adeno-associated viral vector production by multivariate optimization of bioprocess and transfection conditions. Biotechnol. Prog. *40*, e3445. https://doi.org/10.1002/btpr.3445.

28. Meade, O., Clark, J., McCutchen, M., and Kerwin, J. (2021). Exploring the design space of AAV transient-transfection in suspension cell lines. Methods Enzymol. *660*, 341–360. https://doi.org/10.1016/bs.mie.2021.08.003.

29. Dean, A., Voss, D., and Draguljić, D. (2017). Response Surface Methodology (Design and analysis of experiments), pp. 565–614. https://doi.org/10.1007/978-3-319-52250-0_16.

30. Montgomery, D.C. (2019). Response Surface Methods and Designs. In Design and Analysis of Experiments (Wiley), pp. 408–472.

31. Beg, S. (2021). Response Surface Designs and Their Applications in Pharmaceutical Development. In Design of Experiments for Pharmaceutical Product Development (Springer Singapore), pp. 27–41. https://doi.org/10.1007/978-981-33-4717-5_3.

32. Bhattacharya, S. (2021). Central Composite Design for Response Surface Methodology and Its Application in Pharmacy. In Response Surface Methodology in Engineering Science (IntechOpen). https://doi.org/10.5772/intechopen.95835.

33. Beg, S. (2021). Mixture Designs and Their Applications in Pharmaceutical Product Development. In Design of Experiments for Pharmaceutical Product Development (Springer Singapore), pp. 87–96. https://doi.org/10.1007/978-981-33-4717-5_8.

34. Buruk Sahin, Y., Aktar Demirtaş, E., and Burnak, N. (2016). Mixture design: A review of recent applications in the food industry. Pamukkale J. Eng. Sci. *22*, 297–304. https://doi.org/10.5505/pajes.2015.98598.

35. Park, S., Shin, S., Lee, H., Jang, J.H., and Lee, G.M. (2024). Enhancing the production of adeno-associated virus (AAV)2 and AAV9 with high full capsid ratio in HEK293 cells through design-of-experiment optimization of triple plasmid ratio. Biotechnol. J. *19*, 2300667. https://doi.org/10.1002/BIOT.202300667.

36. Snyder, R.O., Audit, M., and Francis, J.D. (2011). rAAV Vector Product Characterization and Stability Studies. Methods Mol. Biol. *807*, 405–428. https://doi.org/10.1007/978-1-61779-370-7_17.

37. Pérez-Rubio, P., Lavado-García, J., Bosch-Molist, L., Romero, E.L., Cervera, L., and Gòdia, F. (2024). Extracellular vesicle depletion and UGCG overexpression mitigate the cell density effect in HEK293 cell culture transfection. Mol. Ther. Methods Clin. Dev. *32*, 101190. https://doi.org/10.1016/j.omtm.2024.101190.

38. Lavado-García, J., Jorge, I., Cervera, L., Vázquez, J., and Gòdia, F. (2020). Multiplexed Quantitative Proteomic Analysis of HEK293 Provides Insights into Molecular

Changes Associated with the Cell Density Effect, Transient Transfection, and Virus-Like Particle Production. J. Proteome Res. *19*, 1085–1099. https://doi.org/10.1021/acs.jproteome.9b00601.

39. Lee, Z., Lu, M., Irfanullah, E., Soukup, M., and Hu, W.S. (2022). Construction of an rAAV Producer Cell Line through Synthetic Biology. ACS Synth. Biol. *11*, 3285–3295. https://doi.org/10.1021/acssynbio.2c00207.

40. Derringer, G., and Suich, R. (1980). Simultaneous Optimization of Several Response Variables. J. Qual. Technol. *12*, 214–219. https://doi.org/10.1080/00224065.1980.11980968.

41. Corbeil, R.R., and Searle, S.R. (1976). Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. Technometrics *18*, 31. https://doi.org/10.2307/1267913.