BioProcessing JOURNAL
TRENDS & DEVELOPMENTS IN BIOPROCESS TECHNOLOGY

Published Online September 8, 2025

Characterization and Optimization of Biologics Manufacturing Using Space-Filling Designs and Machine Learning

By Dogan Ornek, Nida Sayed, Matthew Sousa, Connor Crue, Ravi Chari, Stephanie Delzell, Adalia Cardoso, Roy Lin, and Philip J. Ramsey

Abstract

This study assessed a novel statistical approach using space-filling designs (SFDs) and self-validating ensemble modeling (SVEM) machine learning to efficiently identify key process factors using recombinant adeno-associated virus type 9 (rAAV9) gene therapy manufacturing as a case study. Based on risk assessment of parameters that may impact rAAV9 production, we have evaluated six process parameters using 24-run SFDs generated by the JMP statistical software. SFDs are a new class of design of experiment (DoE) created with the objective of covering the entire design space as completely as possible; this in turn allows more accurate modeling of complex response surface behavior typically found in bioprocesses. SFDs are very flexible in structure, and typically require fewer runs for larger numbers of factors than the commonly used response surface designs. The commonly used classical and optimal response surface designs primarily distribute the design points near the boundary of the experimental region such that little or no data is collected in the interior. As a result, the experimenters have no direct knowledge of the kinetic behavior inside the design region. Without data describing the interior kinetic behavior, it is not possible to build accurate response surface models. In this study, we outlined the significance of SFDs and provided a framework for establishing key process parameters and their ideal operating ranges to achieve consistent and high-yield rAAV9 production in suspended HEK293 cells cultured in the Ambr® 15 Cell Culture microbioreactor system.

INTRODUCTION

Traditionally, process characterization proceeds through a two-step approach involving a screening study to evaluate main effects followed by response surface modeling to estimate main effects, two-factor interactions, and second-order effects. It was indicated that this approach requires 86 runs to characterize one unit operation for six factors[1], which is a lengthy, labor-intense, and costly study. Also, traditional design of experiments (DoEs) primarily distribute the design points near the boundary of the experimental region such that little or no data is collected in the interior. As a result, the experimenters have no direct knowledge of the kinetic behavior inside the design region. Without data describing the interior kinetic behavior, it is not possible to build accurate response surface models.

Space-filling designs (SFDs) are a new class of DoE created with the objective of covering the entire design space as completely as possible; this in turn allows more accurate modeling of complex response surface behavior typically found in bioprocesses.[2] SFDs are very flexible in structure, do not require any prior assumptions about response surface shape, and typically require fewer runs for larger numbers of factors than the commonly used response surface designs. These designs work well with neural network models, which are machine learning approaches to experimentation.[2] A neural network requires small design sizes and can be used in conjunction with SFDs to fit a model with minimal interpolation over the entire design space. Self-validating ensemble modeling (SVEM) is a technique based on modern machine learning methods that can be applied to small datasets.[3] It has been shown to build predictive models with a very high degree of accuracy, quickly and cost-effectively by combining the power of machine learning and DoEs. SVEM is predicated on a method of validation, referred to as *auto-validation* or *self-validation*.[2–3] Models are fit using a fractional weighted bootstrap (FWB) method. In statistics, *bootstrapping* refers to a statistical procedure that resamples a single dataset to create many simulated samples.[4]

This study evaluated SFD and SVEM methodologies using a recombinant adeno-associated virus type 9 (rAAV9) gene therapy manufacturing process as a case study. HEK293 cells were transiently transfected with AAV cis/trans plasmids and an adenovirus helper plasmid to optimize bioreactor parameters for maximizing rAAV9 titer and DNA encapsulation efficiency using these new DoEs. To identify those parameters and their ranges to be evaluated in this study, a comprehensive list of bioreactor input parameters was compiled by a thorough examination of the bioreactor production process flow diagram and a review of historical process performance and literature data review.[5–7] Based on the assessment of the associated risk levels, the following six parameters were identified to be potentially critical parameters: (1) total DNA (mg/10^6cells/mL); (2) transfection reagent (FectoVIR):DNA ratio; (3) gene of interest (GOI):packaging plasmid (pRep/Cap and pHelper) molar ratio; (4) transfection complex volume; (5) incubation time of transfection complex; and (6) pH of rAAV production process. JMP statistical software was used to design a 24-run, space-filling set of experiments to determine the effects of these six parameters on rAAV titer and DNA encapsulation, and to understand the critical process parameters that directly impact product quality.[8] The SVEM machine learning technique was applied to the SFD data to build predictive models for four responses: (1) viral genome (VG) titer; (2) total AAV capsid particle (CP) titer; (3) % full CP; and (4) gene transfection efficiency.

MATERIALS AND METHODS

Cell Line

HEK293 cells (suspension-adapted Expi293F™ by Gibco) were purchased from Thermo Fisher Scientific. Cells were thawed and passaged at 37°C, 8% CO_2, and 75% relative humidity in a shaker incubator (Forma Scientific) at 150 RPM speed of agitation. Cells were maintained at exponential growth phase by splitting every 3–4 days. Cells at passage number 4 (post-thawing) were used for transfection experiments. The basal maintenance and production medium was Expi293™ Expression Medium (by Gibco, Thermo Fisher Scientific).

Plasmids

The plasmids containing the required genes and DNA sequences to produce AAV9 via triple transfection were: (1) pAAV-GOI (with hrGFP as GOI) (Agilent); (2) pALD-X80 (Aldevron LLC); and (3) AAV9 Rep-Cap (GeneMedi). Propagation of pAAV-GOI and AAV9 Rep-Cap were performed in-house by transforming *E. coli* with plasmids. The plasmid DNA was then subsequently replicated upon culturing the transformed bacteria. By using a Qiagen plasmid purification kit, the plasmid DNA was retrieved from the harvested material.

Seed

Vial thaw was carried out in 125 mL PETG shake flask (Cell Treat) with working volume of 30 mL. Cells were cultured at 37°C, 8% CO_2, and 75% relative humidity in a shaker incubator (Forma Scientific) at 150 RPM speed of agitation. Cells were scaled up to 500 mL in a PETG shake flask (Cell Treat) with a working volume of 125 mL. To achieve the desired pre-seeding cell density (1.2×10^6 cells/mL) in the production bioreactor, and to prevent any detrimental effects from spent media, the cells were concentrated and resuspended with fresh media. Cells were then inoculated into fresh media 24 hours prior to triple transfection. This ensured that the cells would be in exponential growth phase at transfection. Triple transfection and AAV production were performed in pre-irradiated and single-use Ambr® 15 Cell Culture microbioreactors (Sartorius) with a working volume of 12 mL.

Production

A high-throughput bioreactor, process analytical technology (PAT) platform was used to carry out the AAV production process. The cell culture platform consisted of 24-way Ambr 15 microbioreactors placed in a biosafety cabinet integrated with a NOVA Flex2 multi-analyte analyzer. The Ambr 15 is comprised of two culture stations with 12 microbioreactors in each, a liquid handler, and deck positions for plates, tip boxes, and lids used for liquid additions. The hardware is integrated with software, which enables users to write a protocol/recipe to control and monitor online process parameters such as pH, dissolved oxygen (% DO), temperature, and agitation along with data from the Flex2 analyzer. Each Ambr 15 bioreactor is equipped with a vessel cap, in-line filter on gas supply, sparge tube, pitched blade impeller, pH sensor, and DO sensor. A clamp plate is fastened onto a culture station with 12 vessels, which connects them to the agitator motor and allows for gas exchange.

Pre-seeded cells (1.2×10^6 cells/mL) were cultured with agitation of 600 RPM, temperature at 36.8°C, and pH at 7.10 ± 0.05 (maintained with 0.5 M sodium bicarbonate and CO_2). Aeration was set at total 0.1 volume of air per unit volume of culture (vessel volume/minute, vvm) and O_2 flow was controlled by a one-way, proportional-integral-derivative (PID) control loop with DO set at 50% O_2 saturation. At 24 hours post-seeding, the bioreactor pH setpoint for each vessel was set per DoE design range of 6.8–7.4. CO_2 and base targets were defined at ±0.05 of pH setpoint, and the base trigger was defined at ±0.05 of base target. The culture volume inside each vessel was adjusted to accommodate varying transfection volumes, as defined by the DoE. Production was completed when the pre-seeded cells reached the desired density of 2.2×10^6 cells/mL. The process was monitored daily using the Ambr 15 control software.

Triple Transient Transfection

Recombinant AAV9 particles containing the green fluorescent protein (GFP) gene were produced via triple transient transfection of HEK293 cells. Five parameters of the transfection process were defined by the DoE design, as shown in **Table 1**.

TABLE 1. Transfection process of HEK293 cells.

Parameter	Description
1	DNA amount per 10^6 cells: 0.5–1.5 mg per 10^6 cells/mL
2	FectoVIR®-AAV (Sartorius) to total DNA ratio: 0.5–2
3	GOI plasmid to packaging plasmid molar ratio: 1–3. Packing plasmid includes pRep/Cap and pHelper plasmids maintained in a 1:1 ratio with each other.
4	Transfection complex volume: 1–10% of production volume
5	Complexation incubation time: 15–60 minutes

Per DoE, the calculated volume of each plasmid was combined into a 1.5 mL sterile Eppendorf tube containing a specific volume of fresh media. The calculated volume of transfection reagent was then added to the plasmid mixture, mixed, and incubated at rest for the pre-defined periods of time listed in **Table 2**. Once the incubation time was complete, the liquid handler of the Ambr 15 system added the complexation mixture to each vessel. Agitation during complex addition into the vessels was controlled by the system to limit shear stress.

Harvest of Viral Vectors

CPs were harvested from the cell culture broth 72 hours post-transfection. Cell lysis was carried out using a chemical lysis method. The lysis buffer comprised of 0.5% Triton X-100, 2 mM $MgCl_2$, and 0.01% Pluronic F-68 (Sigma-Aldrich); and 25 unit/mL Benzonase (Santa Cruz Biotech) in media was prepared fresh. One mL of lysis buffer was added to each vessel and incubated for 60 minutes at 37°C with gentle agitation. Lysate was clarified via centrifugation at 5,000×*g* for 40 minutes at 4°C followed by filtration through a 0.22 µm filter assembly. Clarified lysate samples were collected and stored at -80°C for analytics.

Daily Monitoring of Cell Culture

The Ambr 15 system, integrated with a NOVA Flex2 multi-analyte analyzer, enabled daily auto-sampling from each vessel. Culture samples were collected by the liquid handler and processed by the analyzer pre-transfection and every 24 hours post-transfection. The analyzer was used to determine concentrations of glucose, lactate, glutamine, glutamate, ammonia, sodium, potassium, and calcium. In addition, pH, pCO_2, pO_2, total cell density, viable cell density, viability, cell diameter, and osmolality were recorded.

Determination of Transfection Efficiency

A Countstar Rigel S6 Fluorescence Cell Analyzer (ALIT Life Science) was used to evaluate GFP transfection efficiency. Twenty µL of cell samples were loaded onto cell analyzer chamber slides. The GFP efficiency assay was used to first establish the baseline transfection value using pre-transfected cells followed by samples 24 hours post-transfection.

AAV Production Quantification

VGs in clarified lysate were quantified with a qPCR run on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Reactions used TaqMan Universal Master Mix (Thermo Fisher Scientific) and the following primer/probe set: 5′-ACAGCGGCAAGTTCTACAG-3′ (forward primer) and 5′-GTGCTGGATGAAGTG GTACTC-3′ (reverse primer), and 5′-6-FAM/TGATGAAGA/ZEN/GCAAGGGCGT GGT/IABkFQ/-3′ (TaqMan probe, where 6-FAM is the fluorescent reporter, ZEN is the internal ZEN quencher, and IABkFQ FQ quencher is Iowa Black). All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (IDT). The thermocycling temperature conditions used in the qPCR assays included an initial incubation at 50°C for two minutes followed by 95°C for ten minutes for denaturation. For amplification of DNA, cycles of 95°C for 15 seconds, then 60°C for one minute were repeated 40 times. Standard curves were generated using circular pAAV-GOI. All data were analyzed using the SDS 2.4 software (Computers and Structures, Inc. [CSI]) supplied with the PCR instrument. Results are reported as vg/mL.

Titer of intact viral CPs in the cell lysates were quantified by enzyme-linked immunosorbent assay (ELISA) using an AAV9 ELISA Titration Kit (Progen). The assay was performed according to the manufacturer's protocol, and the data was analyzed using GraphPad Prism 9 (Dotmatics). Results are reported as CPs/mL.

The percentage of full CPs (% CP) was calculated using the following equation, based on the assumption that all measured VGs were encapsulated:

$$\%\,full\ CPs = (vg/mL\ by\ qPCR)/(cp/mL\ by\ ELISA) \times 100$$

Design of Experiment and Model Selection

JMP 17 (SAS Institute Inc.) statistical software was used for 24-run SFD and subsequent analyses. There are seven types of SFDs.[9] Among them, *Fast Flexible Filling* designs were used because of their good space-filling properties and flexibility in allowing constraints and nominal factors. The SVEM method was used to build predictive models using SFD data and the details on how these models were constructed are discussed in the Results section. The SVEM analyses of the experimental data were performed using the Predictum SVEM and JMP software programs. To ensure model reliability, eight out of the 24 runs (33% of the data) were reserved as a validation set, independent of the training process. This split allowed for rigorous assessment of predictive accuracy, avoiding overfitting and ensuring generalizability to unseen conditions.

RESULTS

Space-Filling Designs Input and Output Data

The design matrix for the 24 runs and resulting AAV9 titer for the SFDs are provided in **Table 2**. A matrix of scatterplots for the six parameters is given in **Figure 1** and shows that SFDs distribute design points over the entire interior region. It illustrates that total VG concentration varied from 7.13×10^{8}–5.07×10^{10} (vg/mL), CP varied from 4×10^{9}–2.11×10^{12} (cp/mL), % full CP varied from 0.04–20.95%, and transfection efficiency varied from 11.61–62.4% for the complete experimental range.

Analysis of Experimental Data and Model Selection

The SVEM models were fit to the four responses using Predictum and by selecting SVEM > Neural in JMP. In the Neural window, 50 models were chosen to fit using a fractional weighted bootstrapping technique.[4] The TanH activation functional shape was chosen to fit a model, as combining multiple S-shaped functions predict nonlinear behavior of process very accurately. Model selection started with three TanH activation functions. Its neural network model diagram of the process with three models is shown in **Figure 2**. It shows that there is one layer of activation function. The layer is considered *hidden nodes* (or models) because their values are not known. SVEM algorithm fitted three different models and combined them together to give a prediction model for each of the responses. The algorithm uses a combination of models (*i.e.*, an *ensemble*) instead of a single model, which may result in instability. The model fit was then assessed for the number of models selected, based on root average square error (RASE). It is targeted to minimize RASE value, or the standard deviation of the difference between what is observed and what was predicted, by changing the number of models. All models that fit using SVEM were based on the forward selection method. The predictive modeling results showed that six combined models for VG, seven combined models for CP, six combined models

TABLE 2. Input and output parameters and outputs for SFDs.

Run #	Parameters						Responses			
	DNA (µg/mL)	FextoVIR: DNA	GOI: Packaging Plasmid	Transfection Volume (%)	Incubation Time (min)	pH	VG (vg/mL)	CP (cp/mL)	(%) Full Capsid	Transfection Efficiency (%)
1	1.40	0.67	1.00	1.14	54.61	7.39	2.92×10^{9}	7.45×10^{11}	0.39	11.61
2	0.95	0.91	1.56	3.36	42.75	7.37	8.38×10^{8}	1.91×10^{12}	0.04	39.43
3	1.04	1.91	2.70	2.64	58.08	7.34	1.05×10^{10}	8.94×10^{11}	1.17	34.53
4	1.48	1.52	2.63	9.25	40.13	7.36	5.06×10^{10}	2.27×10^{12}	2.23	45.01
5	0.54	0.89	1.38	7.42	16.15	7.19	9.71×10^{9}	7.97×10^{10}	12.18	55.39
6	0.56	1.88	2.17	5.43	20.52	7.39	2.04×10^{10}	1.26×10^{12}	1.62	62.40
7	1.07	1.06	1.13	8.42	31.34	7.28	5.07×10^{10}	1.06×10^{12}	4.78	57.78
8	0.50	1.76	2.99	4.56	54.24	7.11	1.44×10^{10}	6.89×10^{11}	2.09	30.53
9	0.63	0.72	2.95	9.81	48.97	7.23	4.06×10^{10}	4.25×10^{11}	9.55	37.97
10	1.33	1.52	1.06	2.16	17.65	7.06	8.38×10^{8}	4.00×10^{9}	20.95	49.23
11	0.68	1.97	1.62	1.63	34.06	7.12	7.13×10^{8}	4.00×10^{9}	17.83	48.94
12	1.11	0.51	1.68	4.42	23.76	7.20	1.32×10^{10}	9.89×10^{11}	1.33	33.67
13	1.49	0.78	2.73	1.09	29.10	6.83	5.59×10^{9}	6.69×10^{11}	0.84	15.94
14	0.81	1.41	2.92	1.45	26.49	7.31	5.54×10^{9}	7.52×10^{11}	0.74	25.30
15	0.91	1.05	2.50	3.81	22.24	6.91	3.22×10^{10}	2.11×10^{12}	1.53	46.88
16	1.26	1.75	2.28	9.50	28.82	7.15	3.28×10^{10}	6.98×10^{11}	4.70	51.32
17	0.75	0.53	1.02	9.74	19.34	6.84	1.91×10^{10}	1.61×10^{11}	11.86	36.77
18	1.47	1.84	1.54	6.22	25.21	6.87	3.99×10^{10}	7.66×10^{11}	5.21	49.01
19	1.34	1.13	1.42	8.86	58.90	6.84	3.01×10^{10}	1.23×10^{12}	2.45	34.52
20	1.15	1.25	2.78	7.53	43.86	7.09	3.25×10^{10}	1.49×10^{12}	2.18	45.69
21	0.64	1.61	1.25	6.85	43.44	6.80	4.54×10^{10}	6.13×10^{11}	7.41	33.47
22	0.80	0.60	2.41	6.45	36.91	7.05	3.85×10^{10}	1.26×10^{12}	3.06	39.62
23	0.75	0.75	2.02	2.47	59.98	7.00	6.91×10^{10}	1.32×10^{12}	0.52	17.59
24	1.28	1.40	1.72	5.34	47.91	6.95	2.25×10^{10}	1.04×10^{12}	2.16	29.30

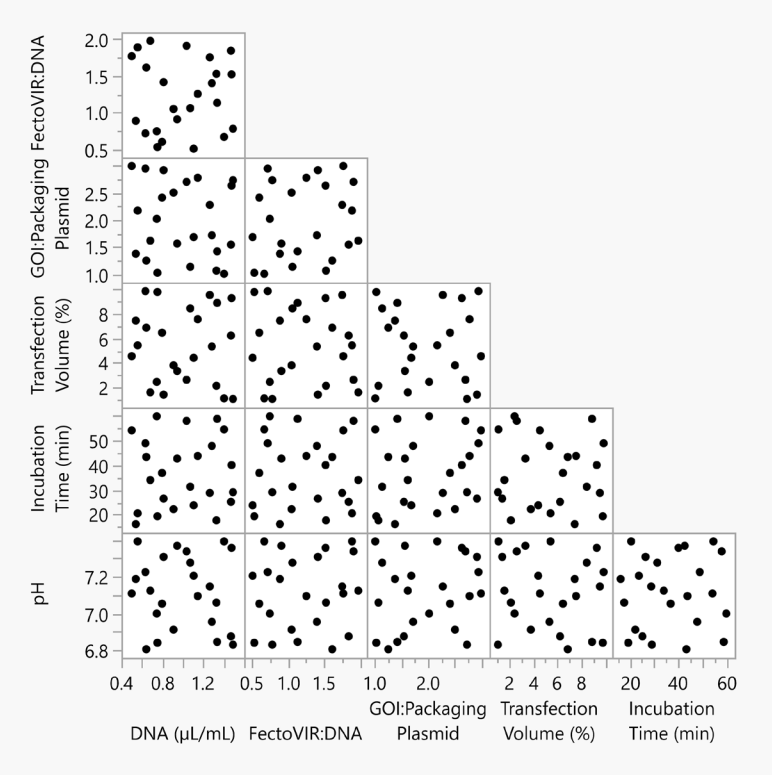

FIGURE 1. A matrix of scatterplots for the six parameters.

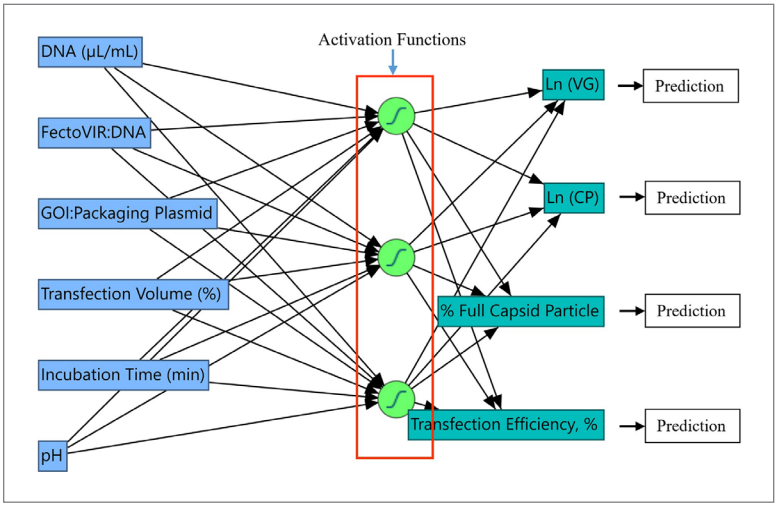

FIGURE 2. Neural network diagram.

for % full CP, and six combined models for transfection efficiency provided the lowest RASE score. The prediction plots in this setting suggest well-fitting models.

Prediction Profiler and Assessing Process Capability

The JMP prediction profiler (**Figure 3**) shows that increased transfection volume, GOI:packaging plasmid ratio, incubation time, and DNA concentration give a higher CP and VG titer but negatively affect the % full CPs. Low levels of culture pH yielded high levels of CP, VG, and % full CP. Transfection efficiency has a strong positive correlation with transfection volume, FectoVIR:DNA ratio, and pH, and has a strong negative correlation with incubation time.

Increasing the total plasmid DNA and GOI:packaging plasmid ratio did not significantly improve VG, and negatively impacted % full CP, which corroborates previous findings.[5] The depletion of some key components from the medium might have resulted in low % full CP at higher GOI:packaging plasmid and DNA concentration. These results suggest that there is potential for rAAV titer improvement by optimizing the medium and feeding strategies.

The % of full capsids and VG are the most important factors for an efficient separation of full capsids in the downstream purification process. At least 10–15% of full CP is needed in the harvest to enable the rAAV purification process. Under optimized conditions, **Figure 3** suggests that the production bioreactor can produce % full CP of 15.82%, VG of 2.19×10^{10} vg/mL, and CP of 9.10×10^{10} cp/mL, which is comparable to the Grieger *et al.* Wave bioreactor results.[10]

Using the predictive model and the simulator in the JMP prediction profiler, the sensitivity of the four output parameters' response to 20% random variation within the optimized parameters was examined. Using the results of the simulation, a 95–95% tolerance interval was estimated using the distribution platform. The tolerance interval indicates, with 95% confidence, that at least 95% of the batches will have % full CP in the range of 11.30–17.98%, VG in the range of 6.89×10^{9}–6.12×10^{10} vg/mL, CP in the range of 2.05×10^{10}–4.25×10^{11} cp/mL, and transfection efficiency of 37.4–67.3% (**Figure 4**). This material meets the purification process requirements for % full CP.

Variable Importance

Variable importance is a concept used in machine learning to assess the significance of process parameters in a model. It quantifies how critical a parameter is for making accurate predictions. It helps identify key drivers of model predictions, simplifies models by removing less influential variables, and improves interpretability and decision-making for machine learning. Quality by design (QbD) also requires assessing the importance of the process parameters to determine which factors to tightly control. Based on JMP variable importance analysis, the overall impact of transfection volume and GOI:packaging, and incubation time explains, respectively, 39.9% and 26.5%, and 20.4% of total variation across all four responses (**Table 3**). These three parameters are important variables that must be tightly controlled to produce high, full CP titer. The colorized prediction profiler is shown in **Figure 3**. Dark red indicates a parameter with high variable importance value.

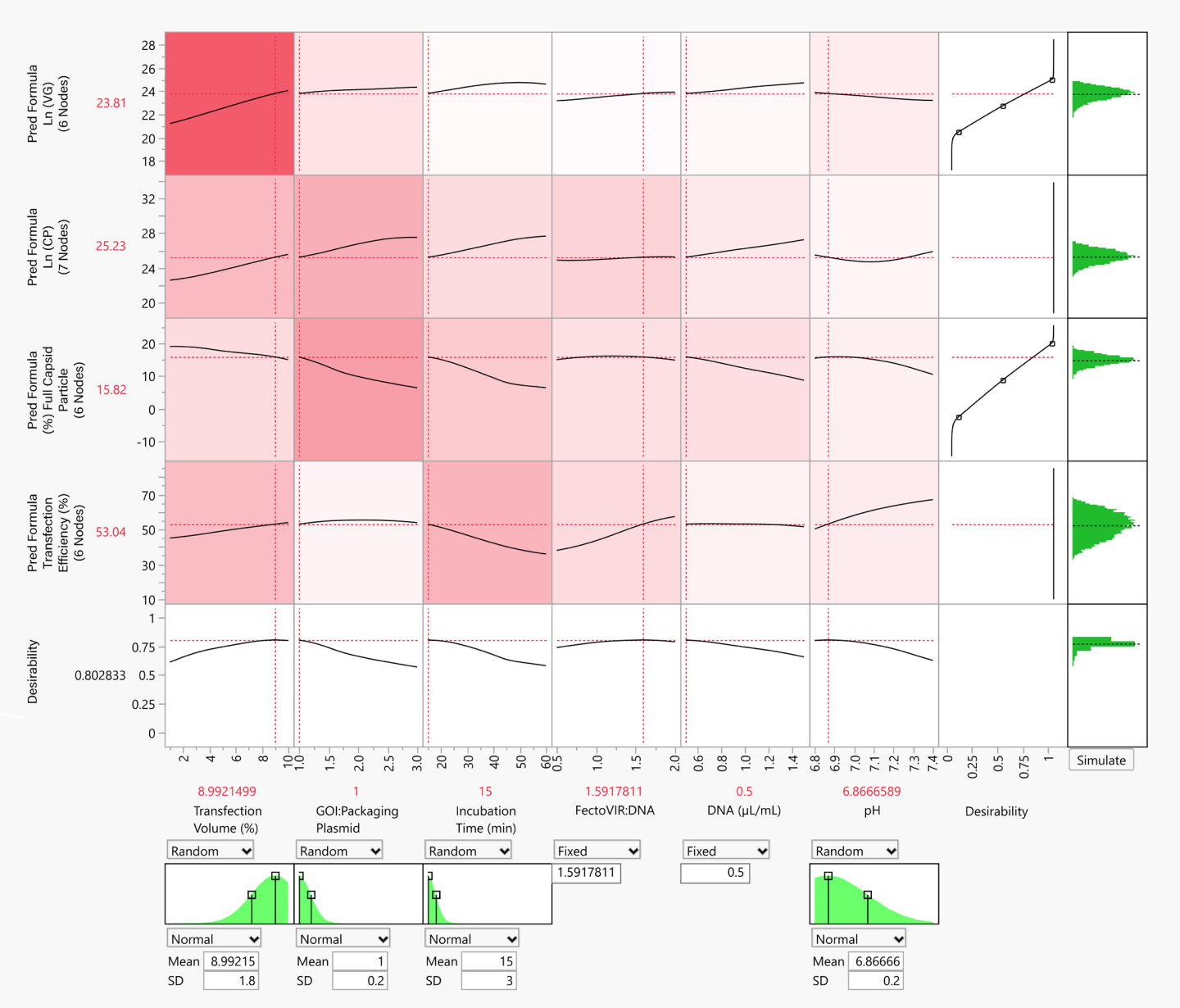

FIGURE 3. Colorized prediction profiler and simulation at optimal setting for combined four output parameters (% transfection efficiency, % full capsid particles, natural log of CP, and natural log of VG). Dark red means parameters have high variable importance value.

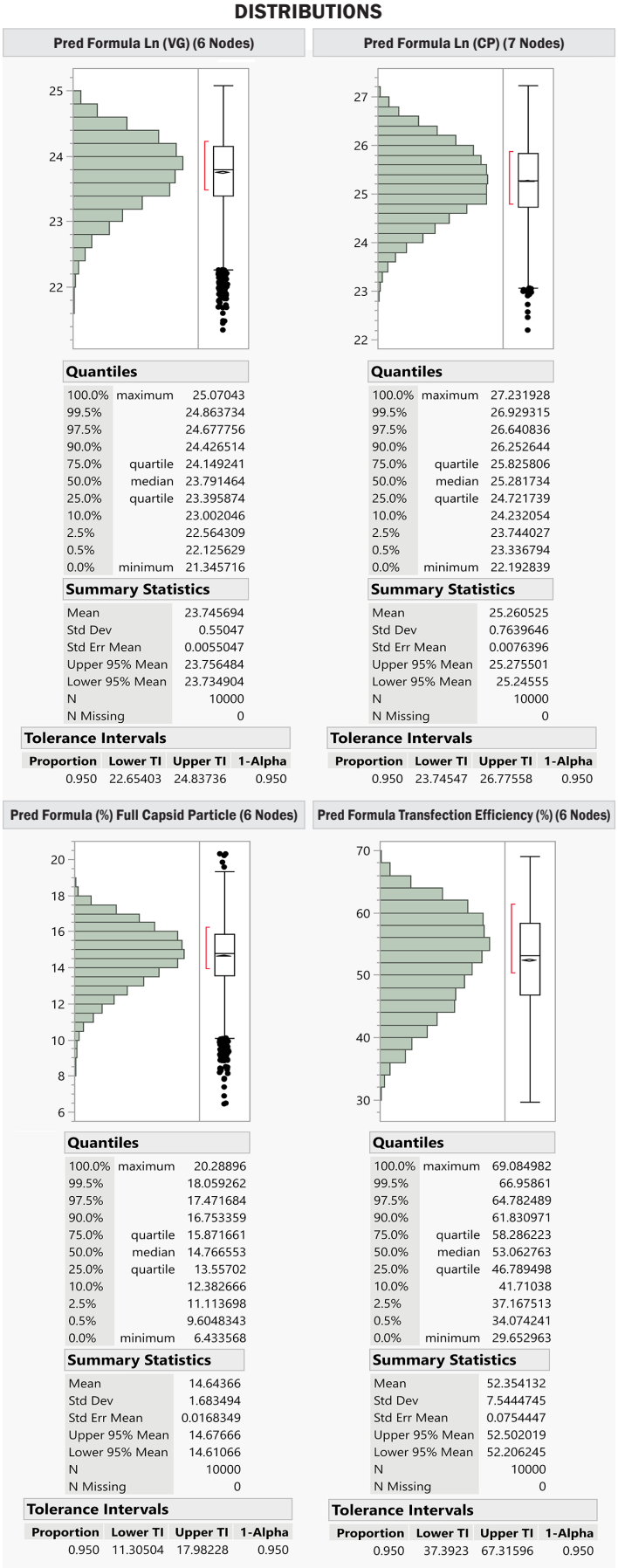

FIGURE 4. Tolerance intervals for VG, CP, % full CP, and transfection efficiency.

TABLE 3. Assessment of experimental parameter importance.

Response	Parameters	Main Effect	Weights (%)
Overall	Transfection Volume (%)	0.313	39.9
	GOI:Packaging Plasmid	0.177	26.5
	Incubation Time (min)	0.158	20.2
	FectoVIR:DNA	0.094	13.2
	DNA (µL/mL)	0.058	8.8
	pH	0.047	8.2
Prediction Formula VG (6 Nodes)	Transfection Volume (%)	0.726	78.9
	GOI:Packaging Plasmid	0.086	13.5
	pH	0.071	9.6
	DNA (µL/mL)	0.015	3.2
	Incubation Time (min)	0.009	2.6
	FectoVIR:DNA	0.005	1.3
Prediction Formula CP (7 Nodes)	GOI:Packaging Plasmid	0.275	39.3
	Transfection Volume (%)	0.201	31.0
	FectoVIR:DNA	0.142	19.5
	Incubation Time (min)	0.120	16.4
	DNA (µL/mL)	0.069	10.5
	pH	0.028	5.3
Prediction Formula % Full Capsid Particle (6 Nodes)	GOI:Packaging Plasmid	0.334	48.4
	Incubation Time (min)	0.175	24.6
	Transfection Volume (%)	0.032	16.1
	DNA (µL/mL)	0.120	15.3
	FectoVIR:DNA	0.080	13.9
	pH	0.037	8.1
Prediction Formula Transfection Efficiency (6 Nodes)	Incubation Time (min)	0.328	37.2
	Transfection Volume (%)	0.294	33.5
	FectoVIR:DNA	0.146	18.2
	pH	0.051	9.8
	DNA (µL/mL)	0.029	6.3
	GOI:Packaging Plasmid	0.013	4.9

CONCLUSIONS

The SVEM machine learning technique was applied to the 24-run SFD data to efficiently build predictive models and then to establish important process parameters (variable importance). Using the predictive model the operating ranges of process parameters were established for optimum and robust AAV9 production in suspended HEK293 cells in a bioreactor operating setting. Through this study, it was shown that all six factors that were identified during the risk assessment can potentially impact % full capsid and VG production under the conditions tested, with transfection volume, GOI:packaging plasmid, and incubation time having the strongest influence. This study demonstrates that the 24-run Ambr 15 Cell Culture system, in combination with the SFD/SVEM machine learning technique, enables a systematic investigation of critical process parameters and rapid, high-throughput process improvement and optimization in one set of experimentation. This new statistical technique provides a significant cost and time-saving opportunity to develop, optimize, and characterize more meaningful processes than a traditional two-stage DoE that requires 86 runs.[1]

Although not all studies are the same, our 24-run SFD/SVEM findings match with the historical, traditional rAAV DoE results on important process parameters. In one study, it was shown that transfection volume, initial cell density, and total plasmid DNA played important roles in rAAV production in shake flasks using a one-factor-at-a-time (OFAT) experimental method.[10] Another study found that rAAV production was optimized using traditional DoE methodology, and data suggested that total DNA, GOI:packaging plasmid, and cell density were shown to be critical parameters for optimal full-capsid rAAV production.[5] In another study, a traditional DoE approach was used to develop models to optimize DNA complexation and full capsid formation in transient rAAV production.[11] Incubation time, cocktail volume (transfection volume), and polyethylenimine (PEI, transfection reagent):DNA ratio and total DNA/cell were identified to have strong correlation with product quality attributes.

Author Disclosures: No competing financial interests exist.

References

[1] Erler, A.S.; de Mas, N.; Ramsey, P.; Henderson, G. Efficient biological process characterization by definitive-screening designs: the formaldehyde treatment of a therapeutic protein as a case study. *Biotechnol Lett* **2013**, 35(3): 323–9. https://doi.org/10.1007/s10529-012-1089-y

[2] Ramsey, P. Accelerating innovation with space filling mixture designs, neural networks and SVEM. Discovery Summit Americas Presentation, **2021**. https://community.jmp.com/t5/Discovery-Summit-Americas-2021/Accelerating-Innovation-with-Space-Filling-Mixture-Designs/ta-p/398669

[3] Lemkus, T.; Ramsey, P.; Gotwalt, C.; Weese, M. Self-validated ensemble models for design of experiments. *Chemom Intell Lab Syst* **2021**, 219: 104439. https://doi.org/10.1016/j.chemolab.2021.104439

[4] Gotwalt, C.; Xu, L.; Hong, Y.; Meeker, W.Q. Applications of the fractional-random-weight bootstrap, **2018**. https://arxiv.org/abs/1808.08199.

[5] Zhao, H.; Lee, K.J.; Daris, M.; Lin, Y.; Wolfe, T.; Sheng, J.; Plewa, C.; Wang, S.; Meisen, W.H. Creation of a high-yield AAV vector production platform in suspension cells using a design-of-experiment approach. *Mol Ther Methods Clin Dev* **2020**, 18: 312-20. https://doi.org/10.1016/j.omtm.2020.06.004

[6] Moço, P.D.; Xu, X.; Silva, C.A.T.; Kamen, A.A. Production of adeno-associated viral vector serotype 6 by triple transfection of suspension HEK293 cells at higher cell densities. *Biotechnol J* **2023**, 18(9):2300051. https://doi.org/10.1002/biot.202300051

[7] Guan, J.S.; Chen, K.; Si, Y.; Kim. T.; Zhou, Z.; Kim, S.; Zhou, L.; Liu, X.M. Process improvement of adeno-associated virus (AAV) production. *Front Chem Eng* **2022**, 4:830421. https://doi.org/10.3389/fceng.2022.830421

[8] Kontogiannis, T.; Braybrook, J.; McElroy, C.; Foy, C.; Whale, A.S.; Quaglia, M.; Smales, CM. Characterization of AAV vectors: a review of analytical techniques and critical quality attributes. *Mol Ther Methods Clin Dev* **2024**, 32(3): 101309. https://doi.org/10.1016/j.omtm.2024.101309

[9] SAS Institute Inc. JMP® Pro 18 Space-Filling Design Methods; SAS Institute Inc.: Cary, NC, **2025**.

[10] Grieger, J.C.; Soltys, S.M.; Samulski, R.J. 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* **2016**, 24(2):287-97. https://doi.org/10.1038/mt.2015.187

[11] Fu, Q.; Lee, Y.S.; Green, E.A.; Wang, Y.; Park, S.Y.; Polanco, A.; Lee, K.H.; Betenbaugh, M.; McNally, D.; Yoon, S. Design space determination to optimize DNA complexation and full capsid formation in transient rAAV manufacturing. *Biotechnol Bioeng* **2023**, 120(11): 3148–62. https://doi.org/10.1002/bit.28508

About the Authors

Dogan Ornek[1]*, Nida Sayed[1], Matthew Sousa[1], Connor Crue[1], Ravi Chari[1], Stephanie Delzell[1], Adalia Cardoso[1], Roy Lin[1], and Philip J. Ramsey[2]

1. Process Science and Development, Boston Institute of Biotechnology, LLC, 225 Turnpike Road, Southborough, Massachusetts 01772 USA

2. Mathematics and Statistics, University of New Hampshire, Durham, New Hampshire 03824 USA

Corresponding Author:

Dogan Ornek, PhD, Vice President of Process Science and Development, Boston Institute of Technology, LLC
Phone: 1-508-333-2586 | Email: don.ornek@bostonbib.com