Molecular Therapy
Methods & Clinical Development
Original Article

Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells

Tam N.T. Nguyen,[1] Sha Sha,[2] Moo Sun Hong,[1] Andrew J. Maloney,[1] Paul W. Barone,[2] Caleb Neufeld,[2] Jacqueline Wolfrum,[2] Stacy L. Springs,[2] Anthony J. Sinskey,[2,3] and Richard D. Braatz[1,2]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; [2]Center for Biomedical Innovation, Massachusetts Institute of Technology, Cambridge, MA, USA; [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

Manufacturing of recombinant adeno-associated virus (rAAV) viral vectors remains challenging, with low yields and low full:empty capsid ratios in the harvest. To elucidate the dynamics of recombinant viral production, we develop a mechanistic model for the synthesis of rAAV viral vectors by triple plasmid transfection based on the underlying biological processes derived from wild-type AAV. The model covers major steps starting from exogenous DNA delivery to the reaction cascade that forms viral proteins and DNA, which subsequently result in filled capsids, and the complex functions of the Rep protein as a regulator of the packaging plasmid gene expression and a catalyst for viral DNA packaging. We estimate kinetic parameters using dynamic data from literature and in-house triple transient transfection experiments. Model predictions of productivity changes as a result of the varied input plasmid ratio are benchmarked against transfection data from the literature. Sensitivity analysis suggests that (1) the poorly coordinated timeline of capsid synthesis and viral DNA replication results in a low ratio of full virions in harvest, and (2) repressive function of the Rep protein could be impeding capsid production at a later phase. The analyses from the mathematical model provide testable hypotheses for evaluation and reveal potential process bottlenecks that can be investigated.

INTRODUCTION

Recombinant adeno-associated virus (rAAV) is the replication-incompetent form of the wild-type parvovirus AAV and is employed for *in vivo* gene-therapy treatments.[1] Its reduced toxicity, robust and long-term transgene expression, and ability to transduce both dividing and non-dividing cells as well as target a wide range of tissues has made rAAV the most widely used viral vector.[1,2] Increasing demand for gene-therapy products, with 359 *in vivo* therapies in clinical trials in the first half of 2020 and a predicted 10 to 20 approved cell and gene-therapy treatments by 2025, has created a need for manufacturing large quantities of rAAVs.[3,4] However, current state-of-the-art viral vector manufacturing methods still fall short of meeting current and future demands.

rAAV viral vectors are manufactured by providing mammalian cells or insect cells with genetic elements that express components for rAAV production via infection, transfection, or genetic engineering. In general, the crude harvest of rAAV contains only a small fraction of capsids that contain a therapeutic element, from 5% to 30%,[5–7] contributing to low viral vector yield and high cost because of the extra purification steps to remove empty particles. Transient transfection of mammalian cells with multiple plasmids is the first and remains the most commonly used method to produce rAAVs.[8] The flexibility and speed of transient transfection are advantageous in early-stage clinical product development compared to other methods of viral vector manufacturing, such as using a stable expressing cell line.[9] In double transfection, the helper virus genes required for viral replication are either provided by a helper virus infection or cloned onto the same plasmid that contains the *rep/cap* genes, whereas in triple transfection, the helper genes are cloned into a third plasmid. Compared to double transfection, triple transfection avoids the potential for incorporation of unwanted viral genomes (vgs) in the product, thus ensuring product safety. However, triple transfection is characterized by low yield. Recent process development efforts on rAAV manufacturing have demonstrated transfection in suspension culture that can produce up to 1×10^{14} vg/L after purification.[10–13] At the current productivity, the total required production of the viral vector, in terms of liters of culture, to treat a population of patients for one disease can be on the order of hundreds of thousands. For instance, the prevalence of Duchenne muscular dystrophy is 19.8 per 100,000 live male births, and each whole-body therapy-treatment requirement of at least 1×10^{15} vg/patient.[14,15] It follows that the total production needed to treat the whole population for this disorder will be approximately 1×10^{5} L of culture per year, with 140 million births per year worldwide.[16] Such demands for viral vectors prompt the need to optimize the production process and increase productivity. A quantitative analysis of the system via a mechanistic, mathematical model will be

Received 11 December 2020; accepted 8 April 2021;
https://doi.org/10.1016/j.omtm.2021.04.006.

Correspondence: Richard D. Braatz, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Rm. E19-551, Cambridge, MA 02139, USA.
E-mail: braatz@mit.edu

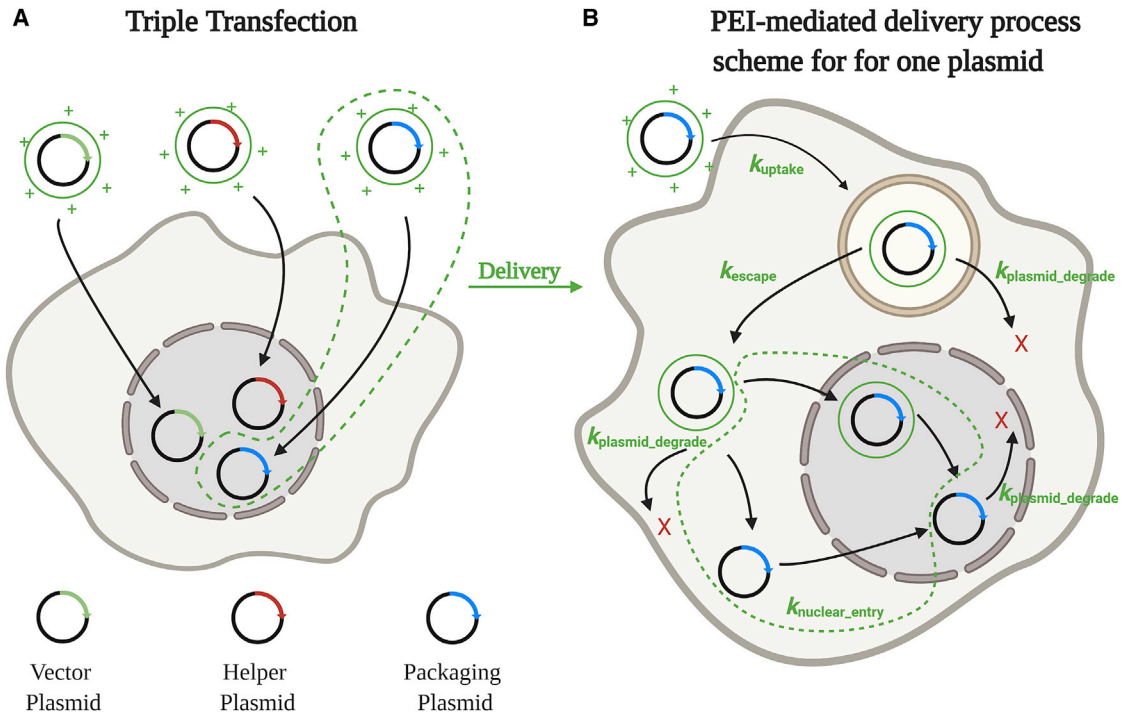

Figure 1. PEI-mediated plasmid-delivery schematic for triple transfection

(A) Triple transfection. (B) Model of plasmid delivery. Non-viral gene-delivery kinetics are assumed to be similar across all three different plasmids. The cellular uptake parameter (k_{uptake}) describes the combined process of the extracellular complex (*ExComplex*) binding to the cell membrane followed by its internalization into the cytosol in the form of endosomal complexes (*EndosComplex*). Endosomal escape (k_{escape}) then follows and releases the complexes into the cytosol (*CytoComplex*).PEI complexes in the cytosol can either migrate into the nucleus before unpacking to release plasmids, or they can unpack in the cytosol; all of the steps required to deliver the plasmids from inside of the cytosol complexes to inside of the nucleus are described by an effective kinetic event ($k_{nuclear\_entry}$) , and the final product is naked plasmid in the nucleus (*NuclearPlasmid*). In addition, the degradation rates of all plasmid-equivalent species in the cytosol are described by one single parameter ($k_{plasmid\_degrade}$).This framework was adapted from previous non-viral gene-delivery model[18,19] and simplified to ensure model identifiability/observability with available measurement data.

useful to organize and exploit information obtained from existing data, gain understanding of process dynamics, and predict responses to various process inputs. This approach to facilitate process development can avoid the high cost of experimentation compared to empirical optimization.[17]

To our knowledge, there is currently no mechanistic model that captures the triple-transfection process as a whole, starting from the delivery of plasmid to the synthesis and final formation of rAAV particles. To this end, we construct a mechanistic, single-cell, triple-transfection model based on AAV replication biology that includes gene delivery, virion replication, Rep protein and viral protein (VP) synthesis, capsid synthesis, and genome encapsidation (Figures 1 and 2). Kinetic parameters are identified using data from in-house transfection and literature experiments. The model predictions to varied plasmid ratio inputs are benchmarked against published literature results. Sensitivity analysis and dynamics of intermediate components reveal a possible explanation to the low ratio of full-to-empty capsid in crude harvest. Transfection strategies are discussed from an operational standpoint that have the potential to overcome the identified bottlenecks. Lastly, model applications in process development and future research for model improvements are presented.

RESULTS

The mechanistic model that describes the end-to-end viral vector production process contains two parts: non-viral gene delivery (Figure 1) and viral production network (Figure 2), with 21 species and 14 parameters in total. Parameter estimates are obtained from literature and fit to data via maximum likelihood (ML) estimation. Details on the model formulation and equations, as well as the parameter estimation strategy can be found in Mathematical modeling in Materials and methods.

Dynamics of polyethylenimine (PEI)-mediated delivery of plasmid DNA

Model simulation with optimized gene-delivery parameter values shows good agreement with the data used for fitting from Carpentier et al.[20] (Figures 3A and S1). Simulation results show total plasmid uptake peaked at 20 h post-transfection (hpt) (Figure 3A), and plasmid copies in the nucleus peaked later at 45 hpt (Figure S1). Only 5% of the initial plasmid added to the cell culture (total plasmid input) is transfected into the cells, and the amount of plasmid that is eventually trafficked inside the nucleus is much smaller—around 0.6% of the total plasmid input. To validate the dynamic trend of nuclear plasmid

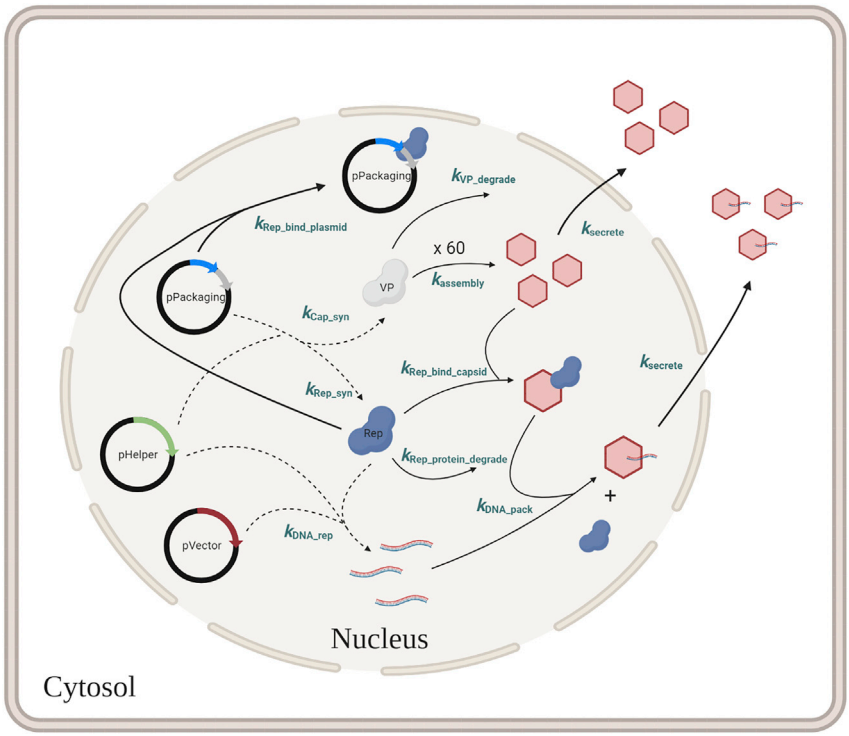

Figure 2. Viral vector synthesis network

Helper plasmid (*pHelper*) activates expression of the *rep/cap* gene on the packaging plasmid (*pPackaging*) and synthesis (k_{Cap_syn} and k_{Rep_syn}) of viral protein (*VP*) and Rep protein (*RepProtein*). Capsid proteins are assembled ($k_{assembly}$) into empty capsids in the nucleus (*EmptyCapNuc*), and each capsid particle consists of 60 protein subunits. With helper functions from the helper plasmid, the Rep protein replicates (k_{DNA_rep}) viral DNA (*vDNA*) from the transgene vector plasmid (*pVector*). Rep proteins dock on empty capsids ($k_{Rep\_bind\_capsid}$) to form intermediate complexes (*CapRepcomplex*), which then interact with viral DNA and encapsidate them inside capsids (k_{DNA_pack}) at a 1:1 ratio to form full virions inside the nucleus (*RepRCcomplex*). Regardless of their contents (empty or full), capsids can be secreted out of the nucleus ($k_{secrete}$) into the cytosol (*EmptyCapCyto* and *FullCapCyto*). Rep protein binding ($k_{Rep\_bind\_plasmid}$) forming bounded plasmids (*RepRCcomplex*) negatively regulates expression of the *rep/cap* gene on the packaging plasmid. Degradation of proteins ($k_{Rep\_protein\_degrade}$ and $k_{VP\_degrade}$) is possible during the process. Dotted lines imply that reactants do not get consumed in the reaction.

copies in PEI-mediated gene delivery, model simulation with the estimated parameters was generated and compared against the data from the study by Glover et al.[21] (Figure 3B). The model was able to describe the dynamic trend of the nuclear plasmid copies up to 24 hpt. The slight overprediction of the model can be attributed to the differences in cell lines used in the studies: (1) HEK293 cells have been shown to have better transfectability than HeLa cells at lower doses with calcium phosphate,[22] and (2) the study by Carpentier et al.[20] used HEK293-EBNA1 cells that express EBNA1 protein for episomal replication of the oriP-harboring plasmid, leading to overall higher plasmid copies.[23,24]

Parameter estimates are provided in Table 1. The uptake rate constant (k_{uptake}) is higher in the case estimated by the literature than the value estimated from the in-house experiment. The phenomenon is to be expected since Carpentier et al.[20] performed transfection in HEK293-EBNA1 cells that express EBNA1 proteins that enhances replication of the oriP-harboring plasmid.[23,24] The estimated degradation constant of plasmid-associated species ($k_{plasmid\_degrade}$) in the cytosol is 0.0195 h^{-1}; this rate is lower than the intrinsic instability constant of the uncomplexed plasmid (0.4621–0.8318 h^{-1}) found in other literature.[26] The estimated value is reasonable, as plasmids associated with complexes are expected to be more stable than uncomplexed plasmids. The smaller value of our plasmid nuclear entry constant ($k_{nuclear\_entry}$) compared to the range found in the literature is because several steps were combined into one parameter, effectively prolonging the timescale.

Dynamics of viral vector synthesis in triple transfection

Model simulation with estimated parameters provides a good fit with experimental data (Figure 4). Interestingly, the plasmid k_{uptake} estimated from in-house experimental data is significantly lower than that estimated from literature data. In-house experimental data show that viral DNA started replicating around 12 hpt and doubled every ~12 h (Figure 4A). The total amount of capsid produced per cell increased sharply at first, with ~80% of total capsids produced within the first 24 h and then plateaued. Total capsid particles and full virion production display a similar trend and saturate at a slightly later time point between 24 and 36 h; this trend is consistent with data from another transfection study, where the majority of full virions was produced in the first 24 h.[10] Given the assumption that newly synthesized capsid proteins are rapidly assembled, the plateau in capsid particle production would be caused by saturation in VP production regulated by total Rep protein production. The amount of full virion only accounts for a consistent 2%–3% of the total capsid produced (Figure 4B). Although most of viral DNA was encapsidated at 24 hpt, and DNA replication continued at an exponential rate, the rate of encapsidated DNA did not follow DNA replication and decreased significantly beyond that time point. By 48 hpt, only 26% of total viral DNA made it inside capsid particles (Figure 4A).

Model validation by varied input plasmid ratio

The model was validated by comparing its predictions using the estimated parameters for various plasmid ratio inputs against data published in literature by Grieger et al.[12] (Figure 5). Data from triple-transfection experiments using the same PEI:DNA mass ratio of 2:1

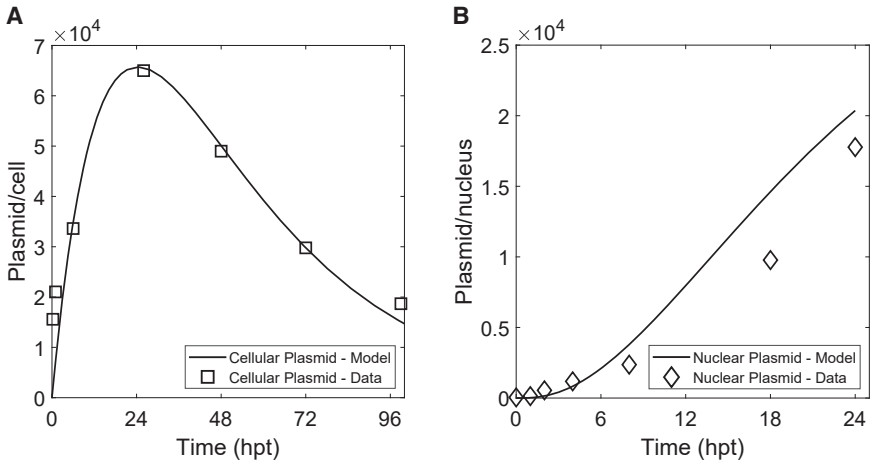

Figure 3. Dynamics of PEI-mediated gene delivery
(A) Plasmid copies per cell over time. The model was fitted to single-transfection data for HEK293-EBNA1 cells from Carpentier et al.[20] with initial plasmid input of 2.7×10^5 plasmids/cell. (B) Plasmid content in the nucleus over time. The model prediction was generated using estimated parameters from the Carpentier et al.[20] study and plotted against HeLa cell transfection data from Glover et al.[21] The initial plasmid input was 2.2×10^6 plasmids/cell.

as in-house experiments were used. Model outcomes for the four plasmid ratios experimentally tested in the aforementioned study were produced and plotted side by side. Data were normalized to the maximal value reported, and model predictions were normalized to the maximal value predicted among four cases. The model captured the relative change in full virion quantity as input plasmid ratio varied: case 2, where both helper and packaging plasmids ratios are increased, gives the highest amount of product, whereas case 4, in which only vector plasmid ratio is increased, gives the lowest yield of all. The observation suggests that optimal productivity requires a

Table 1. Model parameters

Parameter	Process	Value (95% confidence interval)	Unit	Evaluating method	Literature range
k_{uptake}	DNA uptake from medium into cell cytosol	2.78×10^{-2} (2.06×10^{-2}, 3.75×10^{-2})	h^{-1}	fit to literature measurement of total plasmid[20]	exponential decay of 9×10^4[19]
		1.19×10^{-3} (1.00×10^{-3}, 1.40×10^{-3})		fit to in-house measurement of vDNA	6.527×10^{-3} to 4.2×10^{-1}[18,25]
k_{escape}	endosomal escape	6.00×10^{-1}	h^{-1}	parameter for PEI 25K[18]	–
$k_{nuclear\_entry}$	nuclear entry	4.30×10^{-3} (2.4×10^{-3}, $7.7 \times 10^{-3)}$	h^{-1}	fit to literature measurement of nuclear plasmid[20]	8.75×10^{-2} to 3.54[18,25]
$k_{plasmid\_degrade}$	plasmid degradation	1.95×10^{-2} (1.09×10^{-2}, 3.48×10^{-2})	h^{-1}	fit to literature measurement of total plasmid[20]	0.4621 to 0.8318[26]
k_{Cap_syn}	capsid protein synthesis	6.50×10^4 (5.28×10^4, 8.00×10^4)	molecule/cell/h	fit to in-house measurement of vDNA, total capsid, and full virion	–
k_{Rep_syn}	Rep protein synthesis				
$k_{VP\_degrade}$	VP degradation	2.77×10^{-1}	h^{-1}	literature	0.23 to 0.35[27]
$k_{Rep\_protein\_degrade}$	Rep protein degradation	2.45×10^{-2}	h^{-1}	literature	0.01 to 6[28–30]
k_{DNA_rep}	transgene rescue and replication	3.10×10^{-7} (1.52×10^{-7}, 6.32×10^{-7})	$cell^2/molecule^2/h$	fit to in-house measurement of vDNA	–
$k_{assembly}$	capsid assembly	1.00×10^5	h^{-1}	fast step assumption[31]	–
$k_{secrete}$	capsid secretion from the nucleus to the cytosol	1.00×10^5	h^{-1}	fast step assumption	–
k_{DNA_pack}	viral DNA encapsidation	8.18×10^{-2} (1.1×10^{-9}, 6.08×10^6)	molecule/cell/h	fit to in-house measurement of full virion	–
$k_{Rep\_bind\_plasmid}$	binding of Rep protein on packaging plasmid	3.38×10^{-5}	molecule/cell/h	literature	1.0×10^4 $M^{-1}s^{-1}$[32,33]
$k_{Rep\_bind\_capsid}$	binding of Rep protein on empty capsid	5.70×10^{-3} (4.10×10^{-3}, 8.00×10^{-3})	molecule/cell/h	fit to in-house measurement of total capsid and full virion	–

vDNA, viral DNA.

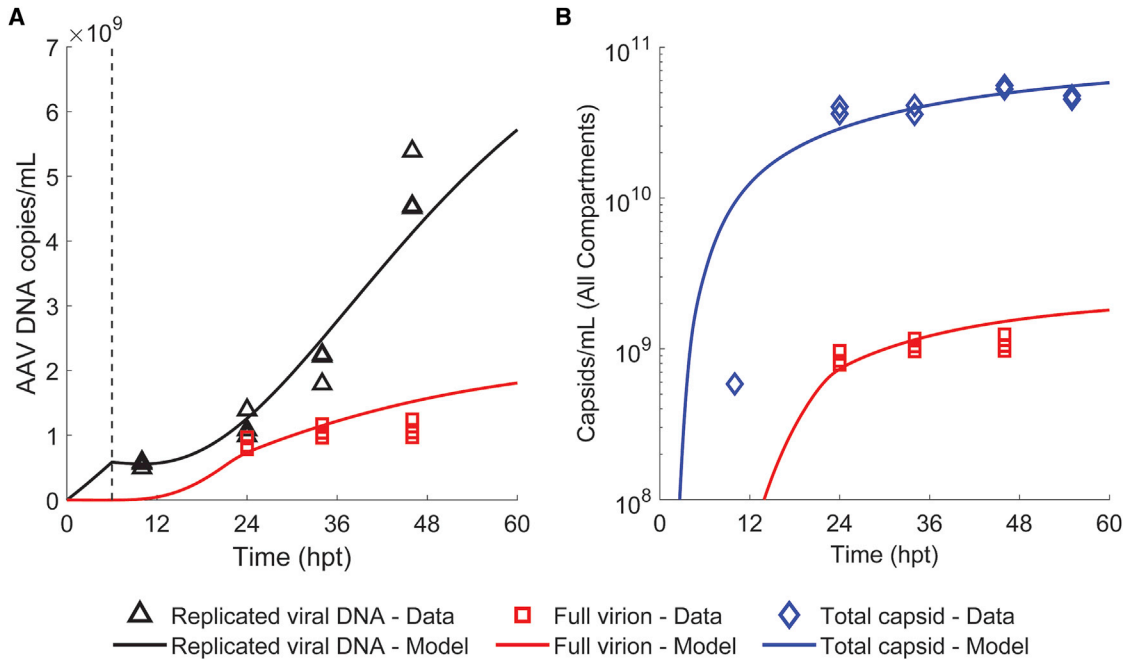

Figure 4. Dynamics of viral DNA and rAAV5 capsid synthesis in in-house triple-transfection experiments for HEK293 cells

(A) Dynamics of viral GFP DNA replication. (B) Dynamics of total and full rAAV capsid production. Viral production parameters were fitted to in-house triple-transfection experimental data. Measurements of total and full rAAV capsids are per-milliliter culture, harvested from both cell lysate and supernatant, normalized by the cell density at the sampling time. The dotted line indicates time point of media exchange. Initial plasmid input was 7.6×10^4/cell for each of the three plasmids.

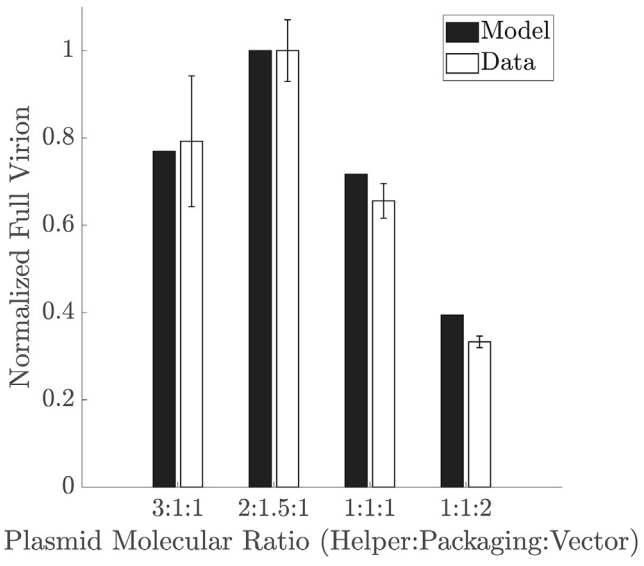

Figure 5. Quantity of full virions produced as a function of the molecular ratios of helper, packaging, and vector plasmids

Data/model simulation was normalized to the maximum value reported/predicted in the four cases. The experimental data with error bars were taken from a study by Grieger et al.[12]

balance between the packaging gene and helper gene expression and that the increase of the copies of therapeutic gene plasmid does not improve production. This result also verifies the capability of our model to capture the interactions between components during the viral production process.

Sensitivity analysis

The dynamic sensitivity profile of the optimized parameters was computed to evaluate their impacts on the final quantity of full virions as production progresses (Figure 6). Plasmid-delivery parameters (k_{uptake}, $k_{nuclear\_entry}$, k_{escape}) consistently display positive effects on the quantity of full virion. $k_{plasmid\_degrade}$ and degradation constant of Rep protein ($k_{Rep\_protein\_degrade}$) exhibit consistent inverse effects on full virion production. The assembly and secretion constants ($k_{assembly}$ and $k_{secrete}$) have little impact on full virion production, as expected from the fast step assumptions that we made in model formulation. Binding rate constant ($k_{Rep\_bind\_plasmid}$) has the largest negative impact on productivity. Interestingly, the sensitivity of full virion to the DNA replication rate constant (k_{DNA_rep}) and packaging constant (k_{DNA_pack}) switches from positive to negative at approximately 25 to 26 hpt. The packaging of DNA simultaneously releases the Rep protein to the free Rep protein pool as formulated in the model, and these Rep proteins can bind onto the packaging plasmid, limiting further *rep/cap* gene expression. On the other hand, the binding of Rep protein on empty capsid ($k_{Rep\_bind\_capsid}$) as the first step to the packing process has a positive impact on the full virion formation, and its sensitivity

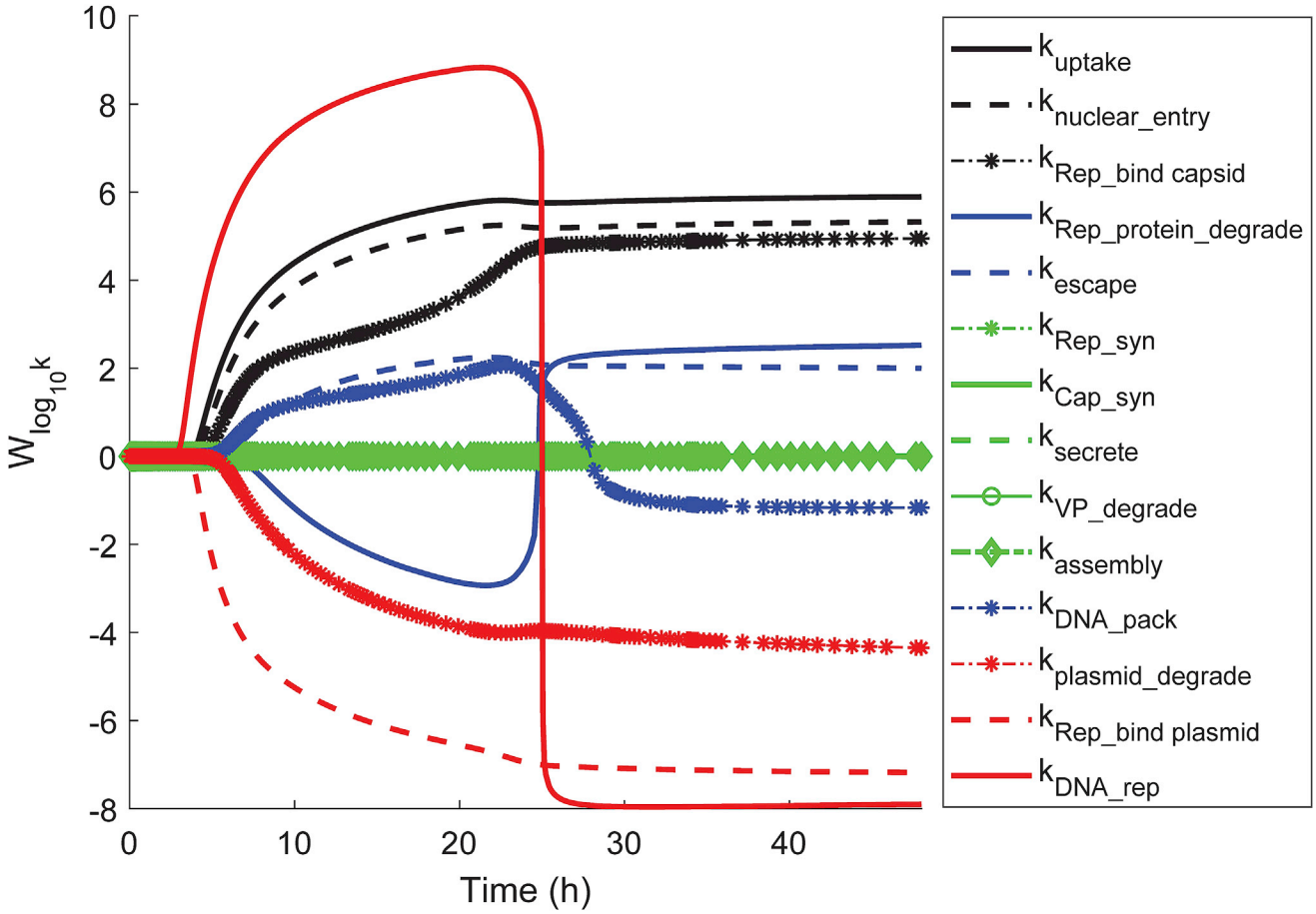

Figure 6. Dynamic profiles of the full virion sensitivities (Equation 17) for all of the model parameters.

dominates the sensitivity of the k_{DNA_pack}. The observation explains the wide confidence of the packing constant estimated. The analysis suggested factors that significantly impact the viral production process include the following: (1) plasmid trafficking, (2) Rep proteins function as a repressor to gene expression and catalyst to DNA packaging, and (3) DNA replication during first 24 hpt.

DISCUSSION

Bottlenecks in viral production

Gene-delivery efficiency

Transient transfection at its current state is not an efficient process. Although transfection efficiency could be as high as 60%, as seen in our experiment and in the literature,[34] plasmid copies in the nucleus after overcoming trafficking barriers are very low, as observed in most studies regarding gene delivery via non-viral transfection. Data from Carpentier et al.[20] and the in-house experiment show that only a maximal of 5% of initial plasmid input is taken up inside the cells, and less than 1% of initial plasmid input was trafficked inside of the nucleus. This result is also consistent with the conclusion from the study by Pollard et al.[35] Model sensitivity analysis indicates that plasmid-delivery parameters are limiting to full virion production; therefore, it is necessary to engineer a better transfection reagent or devise a delivery strategy that allows more plasmids to enter the nucleus. Varga et al.[18] performed a quantitative analysis on the performance of transfection reagents and found that PEI performs better than lipofectamine. Within the class of polymer transfection reagents, compared to PEI Max that we used, there are other transfection reagents such as PEI Pro and FectorVIR-AAV (Polyplus Transfection, New York, NY, USA) that boast increased production in functional viral titers in suspension cell culture. There has been evidence that an increase in PEI/DNA complex size over longer incubation time with cells had an adverse effect in uptake process efficiency;[36] thus another strategy to get more plasmids inside of the cells can be temporal dosing of cell cultures at lower dosage to prevent complexes from getting too large to enter the cells.

Saturation of capsid production and delayed DNA replication result in empty capsid

We visualized the reaction fluxes simulated by the model using in-house experimental settings (Figure 7B) to better understand the impact of their dynamics in full virion production. Rep protein (r_{RepSyn}) and VP production (r_{CapSyn}) begin shortly after plasmids successfully enter

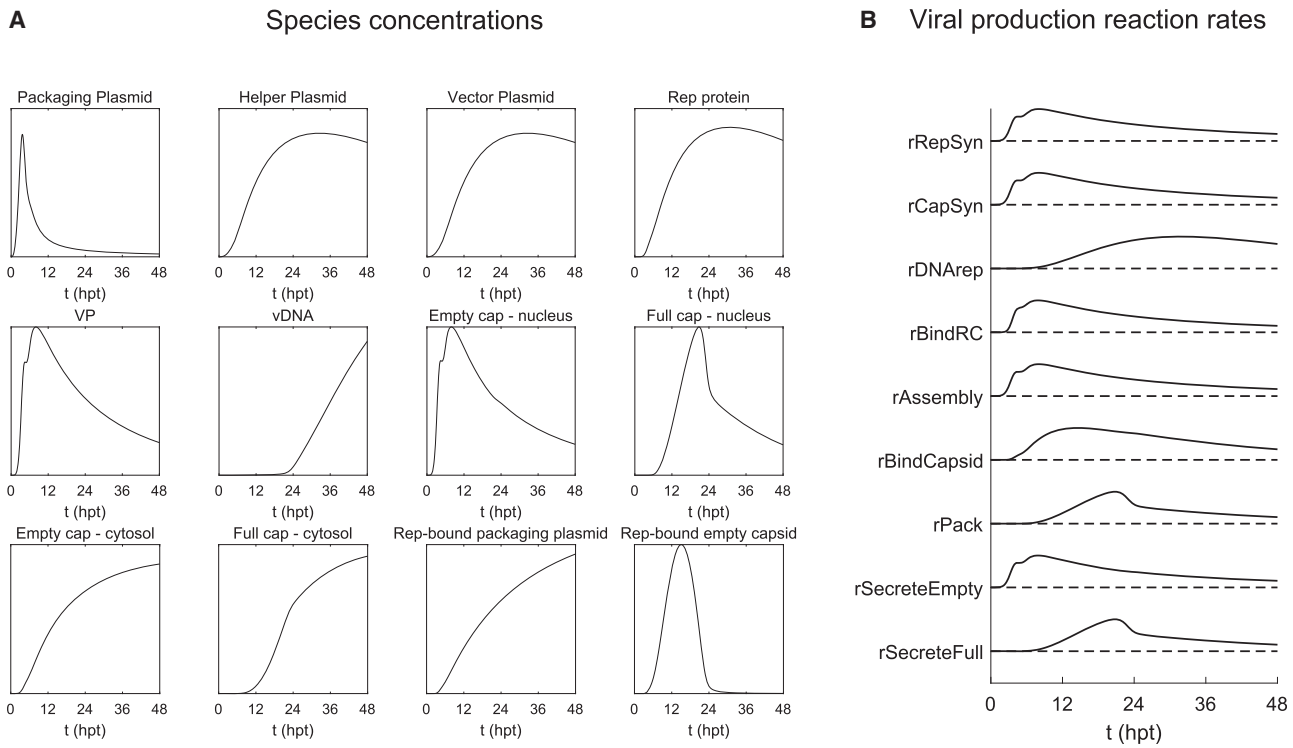

Figure 7. Model simulation of intermediate species.

Dynamic trends of the (A) species concentrations and (B) reaction fluxes. The values are qualitative. The simulation parameters were according to the in-house experiments described in Materials and methods.

the nucleus at around 6 hpt, peak around 12 hpt, and decline shortly after. Capsid particles are synthesized with the same timeline as the Rep protein and VP since the assembly is assumed to have fast kinetic; the majority of capsids are shown to assemble (r_{Assembly}) during the first 24 hpt. Beyond this point, unbound packaging plasmids that are active for protein synthesis are significantly reduced due to inhibitive binding of the Rep protein (Figure 7A). Analysis of reaction fluxes and intermediate species concentrations also reveals a possible explanation as to why the majority of capsids in harvest are usually empty. Viral DNA starts replicating later than capsid synthesis, around 12 hpt and peaks at ~24 hpt, and the trend remains steady until at least 48 hpt as simulated by the model (Figure 7B) and also observed in our data (Figure 4). As mentioned above, capsid production is significantly reduced by 24 hpt, and as postulated in the model formulation, replicated DNA in later phases cannot be packed due to the lack of new capsid production, as previously made empty capsid are already secreted out to the cytosol without DNA. This also explains why enhancing the viral DNA replication kinetics at an earlier time point will have a positive effect on full virion production, as indicated in our sensitivity analysis. DNA replication and capsid synthesis dynamics depend largely on Rep protein regulations, which simultaneously negatively and positively affect gene expression. Optimization of rAAV production will require balancing the *rep/cap* gene with helper gene expression as evident in Figure 5. An essential strategy to enable viral DNA replication before or in parallel with capsid production is early expression of Rep proteins, which is a required machinery. With the current method, the Rep protein and VP are linked on the same plasmid. The decoupling of the expression of the Rep protein and VP by expressing the Rep protein from a separate plasmid from the *cap* gene will give better control of gene expression levels. In fact, a study by Emmerling et al.[37] employed a four-plasmid transfection strategy by splitting the *rep/cap* gene on two different plasmids and achieved a higher rAAV yield that corresponded to an increase in VP production. Our analysis should also motivate a transfection strategy that alleviates protein production stress on cellular machinery as well as syncs the timelines of viral DNA and capsid production. From a process-operating standpoint, temporal dosing of plasmid at different time points is a potential solution, although it will be of concern whether the ability of the culture to uptake plasmids decreases over time.

Limitations and future improvements of the model

Currently, all three plasmids are assumed to have the same kinetics; however, the reality could be different: nonlinear effects of circular DNA length on plasmid transfection efficiency have been reported in literature.[38] Therefore, it is important to determine whether there is a significant difference in the kinetics across different plasmid uptake rates. Besides the assumption of similar uptake kinetics for all three different plasmids, several other simplifying assumptions were made to accommodate limited data availability: (1) transcription and translation steps were combined into one kinetic step; (2) capsid

assembly and secretion of capsid from the nucleus into the cytosol were assumed to have fast kinetics; (3) all $k_{plasmid\_degrade}$ species were described by one parameter; and (4) the DNA encapsidation rate was assumed to only depend on free replicated viral DNA and empty capsid in the nucleus. Experiments that elucidate these intermediate steps have potential for future work and improvements upon the model.

Although the developed model enables the prediction of changes in rAAV production in response to several process parameters such as cell growth rate and plasmid quantity, it has not yet captured the effects of many other process parameters that are commonly manipulated during process development, such as cell density, pH, dissolved oxygen, and media component concentrations on rAAV production. As an example, cell density at the time of transfection has also been reported to influence final product quantity through an undefined mechanism.[10,12] The enhancement of the media with sodium butyrate and certain peptones, which are known to increase recombinant protein expression, can have a positive effect on the recombinant viral vector productivity.[11,39–41] Therefore, the model can be improved by correlating the physiological state of the cell culture to its viral vector productivity. Extension of the model to include such correlations will allow predictions from more process production parameters for further support in process development.

Conclusion

At the current state of process development in viral vector manufacturing, evaluating the impact of system inputs on the outcome still relies on the empirical approach. Statistical and empirical approaches, such as design of experiment (DoE) done by Zhao et al.,[11] can be straightforward for optimizing processes but only within a defined parameter space, whereas mechanistic models can provide physical insights and understanding of the process, multivariable interactions and dynamics to guide novel production design, and extrapolated hypotheses that can be tested.[42,43] In this study, we develop a mechanistic model for the triple-transfection production platform to link transfection inputs to outputs and estimate the relevant parameters. Since the biological mechanism of rAAV synthesis will remain largely the same for other transfection methods and even productive infection or a stable production cell line, our model can be easily adapted to other production methods and serotypes with small modifications. For instance, a double transfection can be derived by removing one set of gene-delivery parameters and species and re-estimating protein synthesis constants. Transfection production data from Chahal et al.[10] showed that the quantities of genome-containing particles of different serotypes produced in the same platform are comparable; therefore, the proposed mechanistic model can be adapted to the production of various rAAV serotypes with slight parameter adjustments. This mechanistic model can provide valuable insights to experimental data, identify process bottlenecks, and guide future research. The leveraging of the model in a quality-by-design implementation has great potential to facilitate process development and advance gene-therapy manufacturing.

MATERIALS AND METHODS

Materials

FreeStyle 293-F cells and 293 Expression Medium were purchased from Thermo Fisher Scientific (Waltham, MA, USA). A set of plasmids, including pAAV-GFP (part no. AAV-400), pRC5 (part no. VPK-425), and pHelper (part no. 340202), was purchased from Cell Biolabs (San Diego, CA, USA) for AAV5 production with GFP as the transgene. Plasmid pRC5 is 7.3 kb long with endogenous promoter p5 and p19 for Rep2 and p40 and p81 for Cap5. Plasmid pHelper is 11.6 kb with endogenous sequences for E2A and E4. Plasmid pAAV-GFP is 5.4 kb long and constructed with a cytomegalovirus (CMV) promoter and GFP between two inverted terminal repeat sequences. These plasmids were amplified using One Shot Stbl3 *E. coli* from Thermo Fisher Scientific (Waltham, MA, USA) and purified by the Plasmid Mega Kit from QIAGEN (Hilden, Germany). The purified plasmids were further sterilized by 0.22 μm filtration and stored at −20°C. PEI transfection agent PEI Max was purchased from Polysciences (Warrington, PA, USA) and dissolved in MilliQ water to make a solution of 1 μg/μL (pH adjusted to 7.0) by adding 1 M HCl. The solution was sterilized by 0.22 μm filtration and stored at 4°C. The AAV5 ELISA kit was from Progen (Wayne, PA, USA). DNA primers were from Integrated DNA Technologies (Coralville, IA, USA). The Monarch Genomic DNA Purification Kit, the DNase buffer, DNase I, and proteinase K were from New England Biolabs (Ipswich, MA, USA). The iTaq Universal SYBR Green Supermix was from Bio-Rad Laboratories (Hercules, CA, USA).

Transfection method

Cell culture

A vial of FreeStyle 293-F cells was thawed and cultured in FreeStyle 293 expression medium. The culture was conducted at 37°C with 135 rpm agitation in a humidified incubator with 5% CO_2. The cells were cultured at 30 mL vol in 125 mL shake flasks (flat bottomed with vented caps), and subcultures were performed between 0.3 and 3 million cells/mL. At all times, cell viability was maintained above 90%. Samples of 50 μL cell-culture samples were taken daily from the shake flasks, mixed with 50 μL trypan blue stain, and analyzed for cell count and viability using the Countess II Automated Cell Counter (Thermo Fisher Scientific, Waltham, MA, USA).

Cell transfection for AAV5 production

Prior to transfection, an aliquot of cells at the exponential growth phase was transferred to new shake flasks and diluted to a cell density of 1 million cells/mL to a total volume of 24 mL by adding prewarmed medium. Each plasmid solution was quantified using a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The three plasmids were mixed at 1:1:1 molar ratio, and the total plasmid dosage was 2 μg plasmid per milliliter of culture. 1 mL of transfection cocktail was prepared by the sequential addition of the plasmid mix, culture medium, and the PEI Max solution. The amount of PEI Max added was calculated so that the PEI:DNA mass ratio came out to 2:1. The transfection cocktail was vortexed vigorously for 10 s, incubated at room temperature for 10 min, and then

added to the cell culture. The final volume of the culture in each shake flask was 25 mL. For the control culture, 25 mL culture was set up at 1 million cells/mL without adding transfection cocktail. All of the shake flasks were cultured as described above. Media exchange was performed 6 h after transfection: the cells were transferred to 50 mL centrifuge tubes and centrifuged at 300 × *g* for 5 min to remove the supernatant; then re-suspended in 25 mL fresh, warmed medium; and transferred back to the original shake flasks and cultured as described above. Culture samples of 2 mL were taken at the time points of 10, 24, 34, 46, and 55 h. A volume of 50 μL culture broth from each sample was used to measure the cell count and viability immediately. The rest of cell-culture samples were aliquoted into 500 μL in each Eppendorf tube and centrifuged at 1,000 rpm for 5 min. Supernatant and cells were separated into different tubes and stored at −80°C until the assays below were performed.

Analytical methods

Measurement of AAV capsids

The quantity of AAV capsids in the supernatant and cell pellet was measured. To quantify the capsids in the cells, one sample tube containing the cells from 500 μL culture was thawed on ice. A volume of 100 μL pre-chilled lysis buffer composed of 150 mM NaCl, 50 mM Tris-HCl (pH 8.5), was added to the cells. The cells were suspended, briefly vortexed, and then processed by a total of three cycles of freezing in an ethanol bath (a mixture of dry ice and 70% ethanol) for 10 min and thawing in a 37°C water bath for 10 min. The sample was clarified by centrifuging at 12,100 × *g*, 4°C, for 15 min, followed by transferring the supernatant to a new Eppendorf tube and discarding the precipitates. The samples were then analyzed by an AAV5 ELISA kit according to the instruction from the manufacturer (Progen, Wayne, PA, USA). To quantify the capsids in the supernatant, an aliquot of the supernatant saved from cell culture was thawed and further clarified by centrifuging at 12,100 × *g*, 4°C, for 15 min. The clarified supernatant was then analyzed by the AAV5 ELISA kit. The final absorbance from the ELISA reactions was read by a microplate reader (BioTek Instruments, Winooski, VT, USA). The concentration of capsids in the cell lysate solution was converted to the capsids' quantity per cell based on the cell number included in the 500-μL culture sample.

Measurement of intracellular plasmids and AAV genome

A sum of the quantity of pAAV-GFP and AAV-GFP genome replicated during AAV production was measured from cells. With the procedures described in the following, we lysed th_e cells without separating the cytosol and nucleus; therefore, the quantified DNA was a population from the entire cells. To begin with, one sample tube containing the cells from 500 μL culture was thawed on ice. A wash step was performed to cell pellets to remove extracellular plasmids: 1 mL cold 1× PBS was added to suspend cells; the solution was centrifuged at 1,000 rpm and 4°C for 10 min. The supernatant was discarded, and the cell pellets were collected at the bottom of the tubes. DNA was extracted from cell pellets by the Monarch Genomic DNA Purification Kit following the instruction from the manufacturer. The Monarch Genomic DNA Purification Kit uses the same principles for extracting and purifying DNA as the DNeasy Blood & Tissue Kit (QIAGEN), which has been used to extract DNA at the range of 5 kb size.[44] In addition, there is an equal chance for the positive (+) or negative (−) sense DNA strand to have been encapsidated in AAV capsids. Once released by proteinase K, these complementary strands can form double-stranded DNA.[45] Briefly, cell pellets were suspended in 100 μL cold PBS, mixed with 1 μL proteinase K and 100 μL lysis buffer, and incubated at 56°C for 5 min. Subsequently, the sample was mixed with 400 μL binding buffer and transferred to a spin column for centrifugation at 1,000 × *g*, room temperature for 3 min, followed by centrifugation at 13,000 × *g* for 1 min. The column was then placed on a clean tube and washed with 500 μL wash buffer and a centrifugation at 13,000 × *g* for 1 min, followed by one more wash using 500 μL wash buffer and a centrifugation at 18,000 × *g* for 1 min. The spin column was then transferred to a clean tube and mixed with 100 μL elution buffer that was pre-heated to 60°C and incubated for 1 min. The spin column was then centrifuged at 13,000 × *g* for 1 min. The eluent was collected and ready for qPCR analysis. To perform qPCR analysis, 5 μL of the extracted DNA was mixed with 2.5 μL of the 4 μM forward primer, 2.5 μL of the 4 μM reverse primer, and 10 μL of iTaq 2× Universal SYBR Green Supermix. The quantity of pAAV-GFP was measured in the samples by a LightCycler 480 System (Roche, Basel, Switzerland) with the following conditions: 95°C for 10 min for the initial denaturation and activation of enzyme, 40 cycles of 95°C for 15 s (for denaturation), 55°C for 15 s (for primer annealing), and 68°C for 1 min (for extension). The primers used targeted the GFP sequence. The forward primer was 5′-GCAAAGACCCCAACGAGAAG-3′, and the reverse primer was 5′-TCACGAACTCCAGCAGGACC-3′. The plasmid pAAV-GFP was diluted to a series concentration from 2×10^9 to 2×10^2 molecules/μL using DNase-free water and used as the standards. The DNase-free water was used as the non-template control.

Measurement of AAV encapsidated genome

To measure the copy number of the AAV genome that was encapsidated in AAV particles, one sample tube containing the cells from 500 μL culture was thawed and lysed by three cycles of freeze and thaw, following the same procedures described in the method for AAV capsid measurement. In the 100-μL lysate, 10 μL of 10× DNase buffer and 5 μL DNase (10 U) were added. The mixture was incubated at 37°C for 1 h, followed by a heating step at 75°C for 10 min to deactivate the DNase. The solution was then cooled down on ice. A 2-μL of proteinase K (1.6 U) was subsequently added to the solution and incubated at 50°C for 1 h. The solution was then heated at 95°C to deactivate the proteinase. Then, the encapsidated genome was quantified from the samples using the primers targeting GFP and the qPCR procedures as described in the method for intracellular plasmids and AAV genome measurement.

Mathematical modeling

Model formulation

The mechanistic model describes kinetic behavior of transient transfection at the subcellular level and is separated into two parts: the

trafficking of plasmid from the supernatant to the nucleus and the vg replication process. Schematics of the two parts of the model are shown in Figures 1 and 2. The model has a total of 21 species and 14 parameters.

The first part of the model describes the non-viral gene-delivery process. This part was adapted from previously developed non-viral, gene-delivery models.[18,19,25] Our model was simplified from the models in the literature to exclude some intracellular intermediate species (nuclear pore complex, plasmid associated with PEI versus plasmid-free plasmid, plasmid bound to transport protein) to utilize measurable states and parameters. Specifically, the model bypasses certain steps and combines species to measurable states to ensure model identifiability.[46]

The mass balances of the species in the first part of the model are the following:

$$\frac{d(ExComplex)}{dt} = -k_{uptake}(ExComplex), \quad \text{(Equation 1)}$$

$$\frac{d(EndosComplex)}{dt} = k_{uptake}(ExComplex) - \left(k_{escape} + k_{plasmid\_degrade} + \mu\right)(EndosComplex)[0,1], \quad \text{(Equation 2)}$$

$$\frac{d(CytoComplex)}{dt} = k_{escape}(EndosComplex) - \left(k_{nuclear\_entry} + k_{plasmid\_degrade} + \mu\right)(CytoComplex) \quad \text{(Equation 3)}$$

, and

$$\frac{d(NuclearPlasmid)}{dt} = k_{nuclear\_entry}(CytoComplex) - \left(k_{plasmid\_degrade} + \mu\right)(NuclearPlasmid). \quad \text{(Equation 4)}$$

These equations include four major steps, namely the uptake of PEI/plasmid complex from culture medium into the cytosol (k_{uptake}) in the form of endosomal vesicles, the escape of complexes from endosomal vesicles into the cytosol (k_{escape}), the entry of plasmid/complex from the cytosol to inside the nucleus ($k_{nuclear\_entry}$), and the overall $k_{plasmid\_degrade}$, which describes the combined effect of possible degradation of vesicles containing complexes and uncomplexed plasmids due to various factors such as nucleases in the cytosol.[26] In addition, intracellular species dilution due to cell division is described by a mutual cell growth parameter (μ).

In the second part of the model, the knowledge of rAAV production mechanism[47] was adapted to construct a viral vector synthesis network from the three plasmids once they are successfully delivered inside the cell nucleus. The main steps capturing viral production are described in Figure 2, including Rep protein and VP synthesis, capsid assembly, viral DNA replication, DNA packaging into the capsids, and capsid secretion from the nucleus into the cytosol. The model does not include the secretion of viral particles from the cytosol to the supernatant; this step can be added and is expected to vary for different serotypes. The mass balances of the species in the second part of the model are the following:

$$\frac{d(RepProtein)}{dt} = k_{Rep\_syn}(pPackaging)(pHelper) - k_{Rep\_bind\_plasmid}(RepProtein)(pPackaging) - k_{Rep\_bind\_capsid}(RepProtein)(EmptyCapNuc) + k_{DNA\_pack}(CapRepComplex)(vDNA) - \left(k_{Rep\_protein\_degrade} + \mu\right)(RepProtein), \quad \text{(Equation 5)}$$

$$\frac{d(VP)}{dt} = k_{Cap\_syn}(pPackaging)(pHelper) - \left(k_{VP\_degrade} + \mu\right)(VP) - 60k_{assembly}(VP), \quad \text{(Equation 6)}$$

$$\frac{d(vDNA)}{dt} = k_{DNA\_rep}(RepProtein)(pHelper)(pVector) - k_{DNA\_pack}(CapRepComplex)(vDNA) - \mu(vDNA), \quad \text{(Equation 7)}$$

$$\frac{d(EmptyCapNuc)}{dt} = k_{assembly}(VP) - k_{Rep\_bind\_capsid}(RepProtein)(EmptyCapNuc) - (k_{secrete} + \mu)(emptyCapNuc), \quad \text{(Equation 8)}$$

$$\frac{d(FullCapNuc)}{dt} = k_{Rep\_bind\_capsid}(RepProtein)(EmptyCapNuc) - (k_{secrete} + \mu)(FullCapNuc), \quad \text{(Equation 9)}$$

$$\frac{d(EmptyCapCyto)}{dt} = k_{secrete}(EmptyCapNuc) - \mu(EmptyCapCyto), \quad \text{(Equation 10)}$$

$$\frac{d(FullCapCyto)}{dt} = k_{secrete}(FullCapNuc) - \mu(FullCapCyto), \quad \text{(Equation 11)}$$

$$\frac{d(RepRCcomplex)}{dt} = k_{Rep\_bind\_plasmid}(RepProtein)(pPackaging), \quad \text{(Equation 12)}$$

$$\frac{d(CapRepcomplex)}{dt} = k_{\text{Rep\_bind\_capsid}}(RepProtein)(EmptyCapNuc) - k_{\text{DNA\_pack}}(CapRepComplex)(vDNA). \quad \text{(Equation 13)}$$

The cellular process corresponding to each kinetic constant is shown in Table 1. The model assumptions and simplifications of the actual events are the following:

(1) One kinetic rate is used to describe the combined kinetics of transcription and translation and does not consider the transportation of mRNA into the cytosol and protein into the nucleus after synthesis.
(2) The synthesis rates of Rep protein and VP are described by bilinear kinetics to capture the role of the helper plasmid, specifically the E2A gene, in activation of *rep/cap* gene expression[48] (Equations 5 and 6). Although the helper protein E1A is also crucial in transcription activation and repression relief,[49,50] this protein is stably expressed in HEK293 cells,[8,51,52] so its effects on capsid production are not explicitly modeled.
(3) Viral DNA requires both AAV helper functions of the E4 protein[53] and the Rep protein to replicate,[54] and the replication kinetics are formulated as trilinear reactions, including both helper plasmid and Rep protein as reactants (Equation 7).
(4) Capsid proteins are VPs of three different types (VP1, VP2, and VP3) but are denoted as one species and assumed to have the same relative synthesis rates.
(5) The Rep protein has been reported to negatively regulate the promoter p5 to the *rep* gene as well as a translational inhibitor of p40 mRNA in AAV2.[55,56] Although the plasmid used for in-house experiments has the cap gene from AAV5, it has the same promoter p40, and the Rep protein regulatory function remains the same in our case. The Rep protein regulation is captured via inhibitive binding to the packaging plasmid $k_{\text{Rep\_bind\_plasmid}}$, which effectively inactivates that plasmid for gene expression after binding. The binding of such Rep protein to the packaging plasmid does not reduce the quantity of plasmid present but rather simulates the effect of reduced transcription and translation rates; the bound plasmids are still present but no longer active in protein synthesis activities.
(6) Packaging of the vg into preformed capsids requires the helicase activity of Rep protein as well as the presence of Rep protein/capsid complex intermediates;[57–59] thus viral DNA encapsidation is formulated as a two-step process, in which Rep protein first docks on $k_{\text{Rep\_bind\_capsid}}$, and DNA is packaged via strand-displacement synthesis $k_{\text{DNA\_pack}}$. The Rep protein is postulated to detach from the capsid after packaging, as there has been evidence that Rep proteins accumulate mostly inside the nucleus.[58]

Parameter estimation strategy

Determination of gene-delivery parameters. Existing models for this process contain parameters that do not pertain to the cell line that was used in in-house transfection experiment[18,19] or were only validated for shorter time courses (24 h or less[18,19,25]) than the length of transfection considered (48 h or more). In addition, uptake kinetics can vary between laboratories depending on transfection conditions. Therefore, kinetic parameters were re-estimated to ensure accuracy of the plasmid-delivery dynamics. Although the quantitative dynamic measurement of intracellular plasmid content, including both cytosolic and nuclear plasmids, has been reported in the literature, the corresponding measurement of plasmid content in the nucleus alone is rarely reported considering the often laborious protocol of nuclei isolation. Since the viral vector production requires component synthesis from all three plasmids, quantification of the plasmid copy number in the nucleus relative to the intracellular amount is necessary to constrain plasmid-delivery kinetic constants, specifically $k_{\text{nuclear\_entry}}$. In this work, the data of a single transfection reported by Carpentier et al.[20] were leveraged to estimate the gene-delivery parameters because (1) they reported quantitative measurements of plasmids both in the cell and the nucleus, and (2) the cell line and transfection used in the experiment (single transfection of HEK293 with PEI40K) were the most similar to our system. Some data were extracted from the figures[20] using WebPlotDigitizer. Although the measurement of plasmids both in the cell and the nucleus cannot uniquely determine k_{escape} and $k_{\text{nuclear\_entry}}$, we assume prior knowledge on the k_{escape} using results from a previous study.[18]

Estimates of delivery parameters can vary between laboratories due to differences in transfection conditions. The plasmid k_{uptake}, which directly affects transfection efficiency (percentage of cells effectively transfected), can change depending on experimental conditions, including the chemical structure and molecular weight of the transfection agent, media composition, cell density, cell health, cell line, complex incubation times, and plasmid dosage. In particular, utilization of HEK293-EBNA1 cells that express the EBNA1 that enhances oriP plasmids will lead to higher plasmid copies overall and will affect the estimated value of k_{uptake}.[23,24] Therefore, the k_{uptake} was re-estimated to fit in-house experimental data. However, the feature of this cell line will not affect the nuclear entry barrier and the related $k_{\text{nuclear\_entry}}$ parameter. Loss of intracellular plasmid content (μ) also depends on the doubling time of cells in each experiment. Cell cultures in the literature mentioned above had a faster doubling time (18 h) compared to our in-house experiments (~30 h on average), suggesting that the plasmid dilution rate would be lower in our case. However, it is reasonable to assume that the rates of endosomal escape and nuclear entry are similar for the same cell line and for the same transfection reagent. Although PEI 40K was used in an in-house experiment instead of PEI 25K in the study by Carpentier et al.,[20] data provided from the manufacturer[60] suggest that the two reagents are close in transfection efficiency, as measured in level of protein expression, thus making it reasonable to apply the k_{escape} for PEI 25K previously reported[18] and the $k_{\text{nuclear\_entry}}$ estimated from the data by Carpentier et al.[20] to estimate model parameters with in-house experimental data.

Determination of viral production parameters. Due to the intricate viral synthesis network, limited availability of dynamic data, and

the lack of prior knowledge on some parameter values, steps are taken to ensure that all parameter values estimated were unique and make physical sense. We leveraged structural analysis and prior parameter knowledge to make assumptions based on observations from literature in order to reduce the number of degrees of freedom and estimate a unique set of parameters.

Protein synthesis constants. Myers and Carter[31] reported the dynamics of Rep protein and VP during wild-type AAV production by co-infection with adenovirus and showed that the synthesis rates of Rep protein and VP are comparable. Therefore, the same value was used for both Rep and VP synthesis constants (k_{Rep_syn} and k_{Cap_syn}, respectively) in the estimation.

Protein degradation constants. Although assembled capsids are stable and do not degrade easily, Grosse et al.[27] found that unassembled capsid proteins are prone to degradation and have a half-life of 2 to 3 h, in which the first-order kinetic constant equivalence is 0.23 to 0.35 h^{-1}. On the contrary, Redemann et al.[28] reported that Rep proteins are more stable and have a half-life of at least 15 h, which translates to a first-order degradation rate constant of 0.0462 h^{-1} or lower. The VP degradation constants ($k_{VP\_degrade}$) in mammalian cells including GFP protein in HEK293 cells, as reported in literature, fall within the 0.01 to 0.06 h^{-1} range.[29,30] The protein degradation constant can be assumed to be symmetrically and normally distributed in a logarithmic with base 10 ($\log_{10}$) scale, and its range contains 95% confidence intervals. Therefore, the degradation constants are fixed at 0.2773 h^{-1} for VP and 0.0245 h^{-1} for the Rep protein.

Rep protein-binding constant. To our knowledge, the association rate constants of the Rep protein to the respective promoter and capsid have not been reported. Transcription factor-binding constants *in vitro* have been reported to be on the order of 10^6 (insect) to 10^8 (bacteria) $M^{-1}s^{-1}$.[32] However, the binding constants observed in *in vivo* systems can be two orders of magnitude lower than *in vitro* as a result of intracellular diffusion and other proteins crowding around binding sites.[33,61] Because mammalian cells and insect cells are both eukaryotic, whereas bacteria cells are prokaryotic, the $k_{Rep\_bind\_plasmid}$ is fixed at the lower limit adjusted by a factor of 100, at 10^4 $M^{-1}s^{-1}$. Unit conversion to $(molecule/cell)^{-1}h^{-1}$ was performed using 15 μm as the average diameter of HEK nucleus.[62]

Capsid assembly constant. Myers and Carter's study[31] suggested that capsid proteins are assembled into viral capsids rapidly following synthesis. Therefore, fast kinetics is assumed based on this observation, and $k_{assembly}$ is fixed to be at least 1,000 times faster than the rate of VP synthesis.

Packing and secretion constants. The relationship between DNA packing rates and capsid secretion rates dictates full:empty capsid ratios in the final product, as viral DNA encapsidation into viral capsids only happens inside the nucleus in the case of parvovirus. The total full virions in in-house experimental data does not distinguish between capsids inside or outside the nucleus. The majority of harvested rAAV particles are empty, as reported in literature and observed in experimental data, suggesting that secretion of capsids from the nucleus might be much faster than DNA packing. Accordingly, secretion is assumed to have fast kinetics, and $k_{secrete}$ was set to be arbitrarily large while estimating k_{DNA_pack} and $k_{Rep\_bind\_capsid}$.

Computational methods

Variances of parameters predetermined by the literature ($k_{Rep\_protein\_degrade}$, $k_{VP\_degrade}$) or estimated from gene-delivery data ($k_{nuclear\_entry}$) will affect the determination of the remaining viral production parameters (k_{uptake}, k_{Rep_syn}, k_{Cap_syn}, k_{DNA_rep}, k_{DNA_pack}, $k_{Rep\_bind\_capsid}$). ML estimation, i.e.,

$$\min_{\boldsymbol{k}} \left(\boldsymbol{y}-\boldsymbol{u}\left(\boldsymbol{k},\boldsymbol{k}_0^*\right)\right)^{\mathrm{T}}\boldsymbol{V}_{\boldsymbol{y}}^{-1}(\boldsymbol{k})\left(\boldsymbol{y}-\boldsymbol{u}\left(\boldsymbol{k},\boldsymbol{k}_0^*\right)\right), \text{ and} \qquad \text{(Equation 14)}$$

$$\boldsymbol{V}_{\boldsymbol{y}}=\left.\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{k}_0}\right|_{\boldsymbol{k}_0^*}\boldsymbol{V}_{\boldsymbol{k}}\left.\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{k}_0}\right|_{\boldsymbol{k}_0^*}^{\mathrm{T}}+\boldsymbol{V}_{\varepsilon}, \qquad \text{(Equation 15)}$$

was used to capture the effect of predetermined parameter variances in the objective function, where k is the vector of parameter estimates, k_0 is the vector of predetermined parameters, y is the vector of experimental observations, u is the vector of model predictions, and V_y is the matrix of experimental data variance taking into account the predetermined parameter variance V_k and the normal noise V_ε. Detailed derivations can be found in the Materials and methods in a past work by our laboratory.[63]

Simulation and optimization were performed entirely in MATLAB (MathWorks, Waltham, MA, USA). The differential equations were solved using a stiff ordinary differential equation (ODE) solver (ode15s). The best-fit parameters were computed using a multi-start approach with an interior-point algorithm to look for the global minima by minimizing the objective function locally from 1,000 random starting points. Since the values of the parameters span a large range in order of magnitude, the optimization was performed over $\log_{10} k$ to search over a larger numerical space and speed convergence and ensure that the estimated confidence intervals are positive.

The sensitivities of model species with respect to the model parameters $\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{k}}$ were calculated by integrating the sensitivity equations along with the state equations.[64] Let u denote the vector of state variables, f denote the vector of kinetic equations, W denote the sensitivity matrix, and J_u and J_k denote the Jacobian matrices with respect to the states and parameters, respectively. Then

$$\boldsymbol{W}(t)=[\boldsymbol{W}_1(t)|\cdots|\ \boldsymbol{W}_{11}(t)] \text{ where } \boldsymbol{W}_i(t)=\frac{\partial \boldsymbol{u}}{\partial k_i}, \qquad \text{(Equation 16)}$$

$$\boldsymbol{W}'^{(t)}=\boldsymbol{J}_{\boldsymbol{u}}\boldsymbol{W}+\boldsymbol{J}_{\boldsymbol{k}}$$

The Jacobian matrices are defined $\boldsymbol{J}_{\boldsymbol{u}}=\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}$, and $\boldsymbol{J}_{\boldsymbol{k}}=\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{k}}$, and the partial derivatives can be calculated for each time step by using the auto

differentiator ADiGator.[65] Since we optimize over $\log_{10} k$, the sensitivities are adjusted correspondingly as follows:

$$\boldsymbol{W}_{\log_{10} k_i} = \frac{\partial \boldsymbol{u}}{\partial \log_{10} k_i} = \ln 10 \frac{\partial \boldsymbol{u}}{\partial \ln k_i} = (\ln 10) k_i \boldsymbol{W}_i, \qquad \text{(Equation 17)}$$

$$\boldsymbol{V}^{-1}_{\log_{10} \boldsymbol{k}} = \left((\ln 10) k_i \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{k}} \right)^{\mathrm{T}} \boldsymbol{V}_{\boldsymbol{y}}^{-1} \left((\ln 10) k_i \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{k}} \right). \qquad \text{(Equation 18)}$$

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2021.04.006.

ACKNOWLEDGMENTS

This work was done in Cambridge, MA, USA. The graphical abstract and Figures 1 and 2 were created with BioRender (https://biorender.com/). This work was supported by the US Food and Drug Administration (grant ID: 1R01FD006584-02, Continuous Viral Vector Manufacturing based on Mechanistic Modeling and Novel Process Analytics). A.J.M. was partially supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under grant number 1122374. T.N.T.N. was partially supported by a MathWorks Engineering Fellowship.

AUTHOR CONTRIBUTIONS

Conceptualization, T.N.T.N., S.S., A.J.M., and R.D.B.; methodology, T.N.T.N.; formal analysis, T.N.T.N. and M.S.H.; investigation, S.S.; writing – original draft, T.N.T.N. and S.S.; writing – review & editing, T.N.T.N., S.S., M.S.H., A.J.M., P.W.B., C.N., J.W., S.L.S., A.J.S., and R.D.B.; funding acquisition, J.W., S.L.S., A.J.S., and R.D.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nat. Rev. Drug Discov. *18*, 358–378.
2. Nonnenmacher, M., and Weber, T. (2012). Intracellular transport of recombinant adeno-associated virus vectors. Gene Ther. *19*, 649–658.
3. Alliance for Regenerative Medicine (2020). Advancing Innovation During COVID-19: ARM Global Regenerative Medicine & Advanced Therapy Sector Report, H1, https://alliancerm.org/sector-report/h1-2020-report/.
4. U.S. Food and Drug Administration (2019). Statement from FDA Commissioner Scott Gottlieb, M.D. and Peter Marks, M.D., Ph.D., Director of the Center for Biologics Evaluation and Research on new policies to advance development of safe and effective cell and gene therapies. January 15, 2019, https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-and-peter-marks-md-phd-director-center-biologics.
5. Adamson-Small, L., Potter, M., Falk, D.J., Cleaver, B., Byrne, B.J., and Clément, N. (2016). A scalable method for the production of high-titer and high-quality adeno-associated type 9 vectors using the HSV platform. Mol. Ther. Methods Clin. Dev. *3*, 16031.
6. Joshi, P.R.H., Cervera, L., Ahmed, I., Kondratov, O., Zolotukhin, S., Schrag, J., Chahal, P.S., and Kamen, A.A. (2019). Achieving high-yield production of functional AAV5 gene delivery vectors via fedbatch in an insect cell-One Baculovirus system. Mol. Ther. Methods Clin. Dev. *13*, 279–289.
7. Wang, Q., Dong, B., Pokiniewski, K.A., Firrman, J., Wu, Z., Chin, M.P., Chen, X., Liu, L., Xu, R., Diao, Y., and Xiao, W. (2016). Syngeneic AAV pseudo-particles potentiate gene transduction of AAV vectors. Mol. Ther. Methods Clin. Dev. *4*, 149–158.
8. Clément, N., and Grieger, J.C. (2016). Manufacturing of recombinant adeno-associated viral vectors for clinical trials. Mol. Ther. Methods Clin. Dev. *3*, 16002.
9. Wright, J.F. (2009). Transient transfection methods for clinical adeno-associated viral vector production. Hum. Gene Ther. *20*, 698–706.
10. Chahal, P.S., Schulze, E., Tran, R., Montes, J., and Kamen, A.A. (2014). Production of adeno-associated virus (AAV) serotypes by transient transfection of HEK293 cell suspension cultures for gene delivery. J. Virol. Methods *196*, 163–173.
11. 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.
12. 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.
13. Blessing, D., Vachey, G., Pythoud, C., Rey, M., Padrun, V., Wurm, F.M., Schneider, B.L., and Déglon, N. (2018). Scalable production of AAV vectors in orbitally shaken HEK293 cells. Mol. Ther. Methods Clin. Dev. *13*, 14–26.
14. Romitti, P.A., Zhu, Y., Puzhankara, S., James, K.A., Nabukera, S.K., Zamba, G.K.D., Ciafaloni, E., Cunniff, C., Druschel, C.M., Mathews, K.D., et al.; MD STARnet (2015). Prevalence of Duchenne and Becker muscular dystrophies in the United States. Pediatrics *135*, 513–521.
15. Flanigan, K. (2017). Gene transfer clinical trial to deliver rAAVrh74.MCK.GALGT2 for Duchenne muscular dystrophy. ClinicalTrials, gov: NCT03333590, https://clinicaltrials.gov/ct2/show/NCT03333590.
16. Department of Economic and Social Affairs (2019). World Population Prospects 2019. Annually interpolated demographic indicators: Estimates, 1950−2020 (United Nations), https://population.un.org/wpp/Download/SpecialAggregates/EconomicTrading/.
17. Kyriakopoulos, S., Ang, K.S., Lakshmanan, M., Huang, Z., Yoon, S., Gunawan, R., and Lee, D.Y. (2018). Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing. Biotechnol. J. *13*, e1700229.
18. Varga, C.M., Tedford, N.C., Thomas, M., Klibanov, A.M., Griffith, L.G., and Lauffenburger, D.A. (2005). Quantitative comparison of polyethylenimine formulations and adenoviral vectors in terms of intracellular gene delivery processes. Gene Ther. *12*, 1023–1032.
19. Varga, C.M., Hong, K., and Lauffenburger, D.A. (2001). Quantitative analysis of synthetic gene delivery vector design properties. Mol. Ther. *4*, 438–446.
20. Carpentier, E., Paris, S., Kamen, A.A., and Durocher, Y. (2007). Limiting factors governing protein expression following polyethylenimine-mediated gene transfer in HEK293-EBNA1 cells. J. Biotechnol. *128*, 268–280.
21. Glover, D.J., Leyton, D.L., Moseley, G.W., and Jans, D.A. (2010). The efficiency of nuclear plasmid DNA delivery is a critical determinant of transgene expression at the single cell level. J. Gene Med. *12*, 77–85.
22. Schmidt, V. Transfection of HeLa and HEK293-Cells with METAFECTENE (Biontex). Date accessed: March 1, 2021.
23. Yates, J., Warren, N., Reisman, D., and Sugden, B. (1984). A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc. Natl. Acad. Sci. USA *81*, 3806–3810.
24. Sun, X., Goh, P.E., Wong, K.T.K., Mori, T., and Yap, M.G.S. (2006). Enhancement of transient gene expression by fed-batch culture of HEK 293 EBNA1 cells in suspension. Biotechnol. Lett. *28*, 843–848.
25. Zhou, J., Yockman, J.W., Kim, S.W., and Kern, S.E. (2007). Intracellular kinetics of non-viral gene delivery using polyethylenimine carriers. Pharm. Res. *24*, 1079–1087.
26. Lechardeur, D., Sohn, K.J., Haardt, M., Joshi, P.B., Monck, M., Graham, R.W., Beatty, B., Squire, J., O'Brodovich, H., and Lukacs, G.L. (1999). Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. Gene Ther. *6*, 482–497.

27. Grosse, S., Penaud-Budloo, M., Herrmann, A.-K., Börner, K., Fakhiri, J., Laketa, V., Krämer, C., Wiedtke, E., Gunkel, M., Ménard, L., et al. (2017). Relevance of assembly-activating protein for adeno-associated virus vector production and capsid protein stability in mammalian and insect cells. J. Virol. *91*, e01198-17.

28. Redemann, B.E., Mendelson, E., and Carter, B.J. (1989). Adeno-associated virus rep protein synthesis during productive infection. J. Virol. *63*, 873–882.

29. Dahari, H., Ribeiro, R.M., Rice, C.M., and Perelson, A.S. (2007). Mathematical modeling of subgenomic hepatitis C virus replication in Huh-7 cells. J. Virol. *81*, 750–760.

30. Schwake, G., Youssef, S., Kuhr, J.T., Gude, S., David, M.P., Mendoza, E., Frey, E., and Rädler, J.O. (2010). Predictive modeling of non-viral gene transfer. Biotechnol. Bioeng. *105*, 805–813.

31. Myers, M.W., and Carter, B.J. (1980). Assembly of adeno-associated virus. Virology *102*, 71–82.

32. Young, S.M., Jr., McCarty, D.M., Degtyareva, N., and Samulski, R.J. (2000). Roles of adeno-associated virus Rep protein and human chromosome 19 in site-specific recombination. J. Virol. *74*, 3953–3966.

33. Tungtur, S., Schwingen, K.M., Riepe, J.J., Weeramange, C.J., and Swint-Kruse, L. (2019). Homolog comparisons further reconcile in vitro and in vivo correlations of protein activities by revealing over-looked physiological factors. Protein Sci. *28*, 1806–1818.

34. Cervera, L., González-Domínguez, I., Segura, M.M., and Gòdia, F. (2017). Intracellular characterization of Gag VLP production by transient transfection of HEK 293 cells. Biotechnol. Bioeng. *114*, 2507–2517.

35. Pollard, H., Remy, J.S., Loussouarn, G., Demolombe, S., Behr, J.P., and Escande, D. (1998). Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in mammalian cells. J. Biol. Chem. *273*, 7507–7511.

36. Xie, Q., Xinyong, G., Xianjin, C., and Yayu, W. (2013). PEI/DNA formation affects transient gene expression in suspension Chinese hamster ovary cells via a one-step transfection process. Cytotechnology *65*, 263–271.

37. Emmerling, V.V., Pegel, A., Milian, E.G., Venereo-Sanchez, A., Kunz, M., Wegele, J., Kamen, A.A., Kochanek, S., and Hoerer, M. (2016). Rational plasmid design and bioprocess optimization to enhance recombinant adeno-associated virus (AAV) productivity in mammalian cells. Biotechnol. J. *11*, 290–297.

38. Hornstein, B.D., Roman, D., Arévalo-Soliz, L.M., Engevik, M.A., and Zechiedrich, L. (2016). Effects of circular DNA length on transfection efficiency by electroporation into HeLa cells. PLoS ONE *11*, e0167537.

39. Pham, P.L., Perret, S., Cass, B., Carpentier, E., St-Laurent, G., Bisson, L., Kamen, A., and Durocher, Y. (2005). Transient gene expression in HEK293 cells: peptone addition posttransfection improves recombinant protein synthesis. Biotechnol. Bioeng. *90*, 332–344.

40. Grünberg, J., Knogler, K., Waibel, R., and Novak-Hofer, I. (2003). High-yield production of recombinant antibody fragments in HEK-293 cells using sodium butyrate. Biotechniques *34*, 968–972.

41. 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.

42. Gernaey, K.V., Lantz, A.E., Tufvesson, P., Woodley, J.M., and Sin, G. (2010). Application of mechanistic models to fermentation and biocatalysis for next-generation processes. Trends Biotechnol. *28*, 346–354.

43. Hong, M.S., Sun, W., Lu, A.E., and Braatz, R.D. (2020). Process analytical technology and digital biomanufacturing of monoclonal antibodies. Biopharm. Process. *23*, 122–125.

44. Fliedl, L., Kast, F., Grillari, J., Wieser, M., and Grillari-Voglauer, R. (2015). Optimization of a quantitative PCR based method for plasmid copy number determination in human cell lines. N. Biotechnol. *32*, 716–719.

45. Guerin, K., Rego, M., Bourges, D., Ersing, I., Haery, L., Harten DeMaio, K., Sanders, E., Tasissa, M., Kostman, M., Tillgren, M., et al. (2020). A novel next-generation sequencing and analysis platform to assess the identity of recombinant adeno-associated viral preparations from viral DNA extracts. Hum. Gene Ther. *31*, 664–678.

46. Guillaume, J.H.A., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., and Stigter, J.D. (2019). Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. Environ. Model. Softw. *119*, 418–432.

47. Maurer, A.C., and Weitzman, M.D. (2020). Adeno-associated virus genome interactions important for vector production and transduction. Hum. Gene Ther. *31*, 499–511.

48. Matsushita, T., Elliger, S., Elliger, C., Podsakoff, G., Villarreal, L., Kurtzman, G.J., Iwaki, Y., and Colosi, P. (1998). Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther. *5*, 938–945.

49. Shi, Y., Seto, E., Chang, L.-S., and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell *67*, 377–388.

50. Chang, L.S., Shi, Y., and Shenk, T. (1989). Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. J. Virol. *63*, 3479–3488.

51. Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. *36*, 59–74.

52. Louis, N., Evelegh, C., and Graham, F.L. (1997). Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. Virology *233*, 423–429.

53. Huang, M.M., and Hearing, P. (1989). Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. J. Virol. *63*, 2605–2615.

54. Li, J., Samulski, R.J., and Xiao, X. (1997). Role for highly regulated rep gene expression in adeno-associated virus vector production. J. Virol. *71*, 5236–5243.

55. Pereira, D.J., McCarty, D.M., and Muzyczka, N. (1997). The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J. Virol. *71*, 1079–1088.

56. Trempe, J.P., and Carter, B.J. (1988). Regulation of adeno-associated virus gene expression in 293 cells: control of mRNA abundance and translation. J. Virol. *62*, 68–74.

57. Dubielzig, R., King, J.A., Weger, S., Kern, A., and Kleinschmidt, J.A. (1999). Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. J. Virol. *73*, 8989–8998.

58. Wistuba, A., Weger, S., Kern, A., and Kleinschmidt, J.A. (1995). Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. J. Virol. *69*, 5311–5319.

59. King, J.A., Dubielzig, R., Grimm, D., and Kleinschmidt, J.A. (2001). DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. EMBO J. *20*, 3282–3291.

60. Fisher Scientific (2021). Polysciences, Inc. PEI MAX - Transfection grade linear polyethylenimine hydrochloride (MW 40,000), 1 g, https://www.fishersci.com/shop/products/polyethylenimine-hcl-max/NC1038561.

61. Li, G.W., Berg, O.G., and Elf, J. (2009). Effects of macromolecular crowding and DNA looping on gene regulation kinetics. Nat. Phys. *5*, 294–297.

62. Dietmair, S., Hodson, M.P., Quek, L.-E., Timmins, N.E., Gray, P., and Nielsen, L.K. (2012). A multi-omics analysis of recombinant protein production in Hek293 cells. PLoS ONE *7*, e43394.

63. Hong, M.S., Velez-Suberbie, M.L., Maloney, A.J., Biedermann, A., Love, K.R., Love, J.C., Mukhopadhyay, T.K., and Braatz, R.D. (2021). Macroscopic modeling of bioreactors for recombinant protein producing Pichia pastoris in defined medium. Biotechnol. Bioeng. *118*, 1199–1212.

64. Caracotsios, M., and Stewart, W.E. (1985). Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations. Comput. Chem. Eng. *9*, 359–365.

65. Weinstein, M.J., and Rao, A.V. (2017). Algorithm 984: ADiGator, a toolbox for the algorithmic differentiation of mathematical functions in MATLAB using source transformation via operator overloading. ACM Trans. Math. Softw. *44*, 21.