Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Mechanistic Digital Twin for Dual-AAV Bioprocessing and Multi-Gene Base-Editing Payload Delivery

FAIZAN AHMAD

1996faizanahmad@gmail.com

Jamia Hamdard https://orcid.org/0000-0002-7768-9411

Research Article

Keywords: adeno-associated virus, dual-AAV, digital twin, bioprocess modelling, base editing, identifiability, global sensitivity analysis, uncertainty quantification, capsid quality

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11747176/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Mechanistic Digital Twin for Dual-AAV Bioprocessing and Multi-Gene Base-Editing Payload Delivery

Faizan Ahmad*

*Department of Medical Elementology and Toxicology, Jamia Hamdard University, Delhi, India

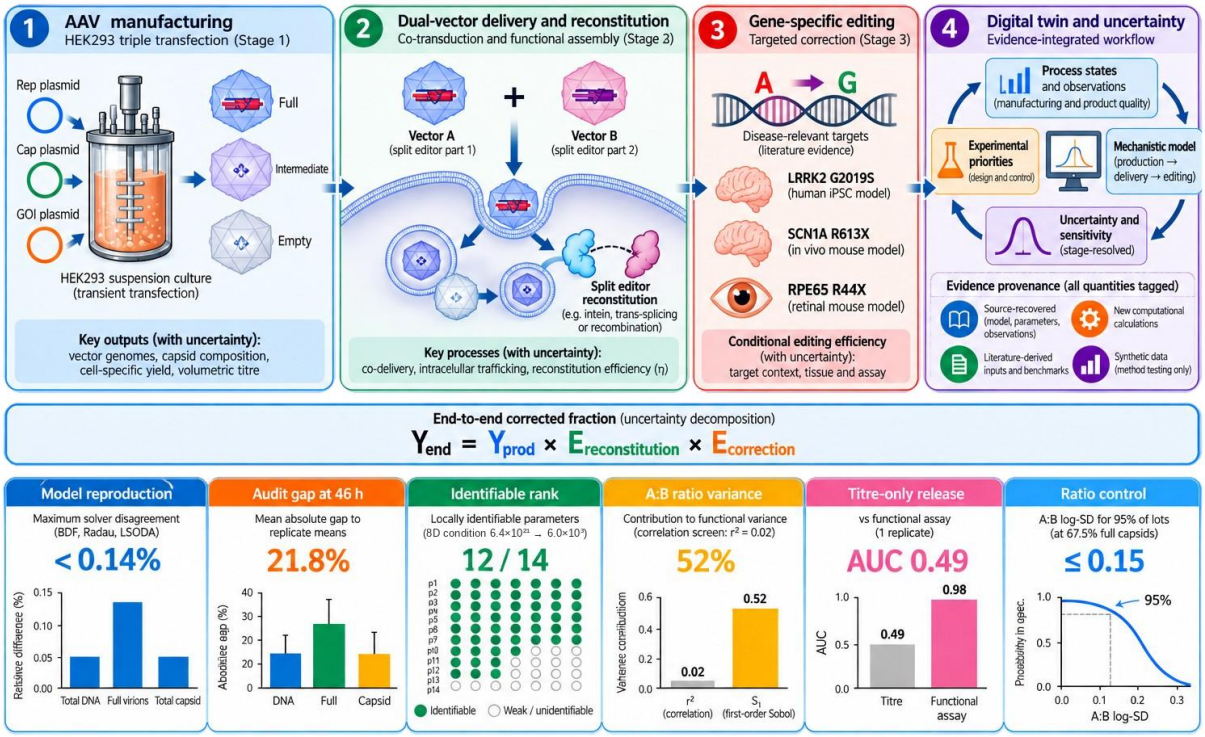
Correspondence
Faizan Ahmad
Department of Medical Elementology and Toxicology
Jamia Hamdard University, Delhi
India
Email: 1996faizanahmad@gmail.com; faizmedi16@gmail.com

Abstract

Dual-AAV delivery of oversized editors couples vector manufacturing, product quality, co-delivery and gene correction, yet process models rarely carry this coupling or state what their data can identify. We built a modular computational digital twin spanning manufacturing, dual-vector reconstitution and gene-specific correction. An independent reconstruction of a published 21-state production model agreed across three stiff solvers (< 0.14%) but retained a 21.8% mean absolute gap to source observations at 46 h; its 14-parameter space was rank-deficient (12/14; condition number 6.4×10^{21}), improving to 6.0×10^{3} after reduction. Fourteen scenario studies traced product quality to function: within the assumed surrogate equations and input distributions, variance-based sensitivity assigned 52% of functional variance to the A:B ratio (correlation screen: 2%), and titre-only release classified function at chance (AUC 0.49) versus 0.98 for a functional assay. All analyses are computational; no experimental validation is claimed.

Keywords: adeno-associated virus, dual-AAV, digital twin, bioprocess modelling, base editing, identifiability, global sensitivity analysis, uncertainty quantification, capsid quality

Graphical Abstract



Highlights

- Developed a modular digital twin linking AAV manufacturing, reconstitution and gene correction.
- Independent stiff-solvers reproduced the source model within 0.14%.
- Parameter reduction improved identifiability of the mechanistic production model.
- A vector stoichiometry explained 52% of simulated functional variance.
- Functional assays outperformed titre-based release, achieving an AUC of 0.98.

1. Introduction

AAV vector engineering has moved from a largely vector-centric problem toward a coupled systems problem involving payload architecture, capsid biology, manufacturing, analytics, delivery, and therapeutic mechanism. The approximately 4.7-kb AAV genome capacity is sufficiently restrictive that many therapeutic coding sequences and contemporary gene editors cannot be encoded as a single conventional vector once promoters, regulatory elements, and other sequence requirements are included. Dual-AAV strategies were developed to distribute oversized payloads between two particles and subsequently reconstitute functional expression through mechanisms that can include homologous recombination, trans-splicing, or protein trans-splicing. The approach is powerful, but it introduces an additional dependency that is often underrepresented in process models: the therapeutic unit is not one particle but a successful pair of particles arriving in the same cell and producing a functional molecular product [1-3].

Base editing makes this systems problem particularly explicit. Adenine and cytosine base editors can be substantially larger than the practical AAV payload limit. Levy et al. demonstrated dual-AAV delivery of split base editors using intein-mediated protein trans-splicing across multiple mouse tissues, providing a direct experimental precedent for treating vector pairing and protein reconstitution as part of the delivery problem. Conversely, Davis et al. demonstrated that compact, size-optimized ABEs can in some settings fit into a single AAV genome, illustrating that payload architecture itself can be an optimization variable. More recently, dual-AAV ABE studies have demonstrated disease-relevant editing in specific models, including SCN1A R613X and RPE65 R44X. Thus, a computational framework intended to support future editing platforms should be able to represent both manufacturing constraints and the downstream consequences of split versus compact payload architectures [2,3,14,15].

The upstream production problem is equally multivariable. Triple-plasmid transient transfection of HEK293 cells remains a major production strategy because it is comparatively flexible and can be adapted to suspension culture. Grieger et al. demonstrated suspension HEK293 production with continuous harvest and reported high vector genome productivity under optimized conditions. Subsequent work has shown that plasmid ratios, transfection reagent-to-DNA ratios, pH, viable cell density, complexation time, feeding strategy, and perfusion can materially influence vector genome yield and product quality [4,6-8]. Deng et al. further showed that high-cell-density fed-perfusion strategies can alter unit yield and upstream economics. These studies collectively indicate that process development is not adequately described by a single scalar titer endpoint.

The quality dimension is equally important. AAV preparations contain mixtures of genome-containing, partially filled, and empty capsids, and full-to-total capsid ratio is widely used as a quality attribute. Analytical methods themselves are imperfect: interlaboratory comparisons have spectrometry and other orthogonal approaches have been developed to improve characterization [9,10]. A digital twin that ignores this distinction can therefore create a false impression of process improvement when genome titer rises but product quality deteriorates. In the present framework, vector genomes and capsid states are retained separately wherever the underlying model permits.

Mechanistic mathematical models provide a natural substrate for digital twins because they explicitly connect latent biological states to measurable process outputs. Moser et al. described mechanistic models as a central component of bioprocess digital twins, emphasizing their utility for prediction, optimization, monitoring, and control. The value of a digital twin, however,

depends on more than numerical sophistication. It requires source provenance, parameter identifiability, numerical stability, explicit uncertainty representation, and a clear boundary between simulation and experimental evidence. A model with many states and parameters can be visually convincing while remaining poorly identifiable from sparse observations. Conversely, a lower-dimensional representation can be useful if it preserves the observable behavior that the available data can actually support [12].

The present study was designed around this evidence problem. Rather than treating every computational output as equally credible, we built an evidence hierarchy. Published mechanistic code and observations were treated as source material; independent reimplementation was treated as numerical reproduction; parameter reduction was treated as an identifiability exercise; Monte Carlo and sensitivity analyses were treated as uncertainty propagation under assumptions; and published disease-specific experiments were treated as external biological anchors. Synthetic Bayesian calibration and filtering were deliberately separated from biological validation because synthetic truth is known by construction but does not establish that the underlying biological model is correct.

Three disease-associated variants were selected to test how the framework behaves across different evidence contexts. LRRK2 G2019S provides a human iPSC-derived correction context in which ABE correction has been investigated, but it is not a direct dual-AAV production experiment. SCN1A R613X provides the strongest direct connection to the present architecture because recent work has used dual-AAV9 ABE delivery in a disease model. RPE65 R44X provides a retinal dual-AAV split-ABE context with functional rescue measurements in an established mouse model [13-15]. The objective was not to rank these diseases or to claim equal predictive validity, but to demonstrate that the same computational architecture can preserve differences in evidence provenance.

The study therefore addressed five questions: (i) can a published AAV mechanistic model be independently reconstructed with numerically stable behavior; (ii) which dimensions of its parameter space are locally identifiable from the available observations; (iii) how do process variables propagate into dual-AAV end-point uncertainty; (iv) can literature-derived downstream constraints be coupled to manufacturing uncertainty without disguising them as new experiments; and (v) can the same computational infrastructure expose its own inference, filtering, and control failures before those methods are transferred to experimental process development?

2. Materials and Methods

2.1 Computational evidence architecture

For a simplified end-to-end quantity, Y_prod is the normalized production-side contribution, and the latter two terms represent downstream efficiencies. This factorization is not intended as a complete biological model of dose-response, receptor kinetics, intracellular trafficking, or editing-window occupancy. It is an uncertainty-accounting device that makes the origin of uncertainty explicit.

The architecture also separates four classes of evidence: (1) source-recovered parameters and observations; (2) new computational calculations performed on those source materials; (3) external literature-derived quantities used as inputs or contextual benchmarks; and (4) synthetic

observations generated solely to test numerical inference and control algorithms. This classification is carried through the Results and Discussion so that a computationally generated value is never presented as if it were a newly measured biological result.

2.2 Simplified five-state manufacturing model

The first model was intentionally compact and was used for design-space exploration rather than source recovery. The state vector consisted of viable cell density X, glucose concentration G, free viral genomes V_gen, empty capsids C_emp, and full vectors V_full. The model equations were:

$dX/dt = [\mu(G) - k_{lysis} V_{full}]X$

$dG/dt = D(G_{feed} - G) - [\mu(G)/Y_{X/G}]X$

$dV_{gen}/dt = k_{transf} f(r)X - k_{pack} V_{gen} C_{emp}$

$dC_{emp}/dt = k_{cap} X - k_{pack} V_{gen} C_{emp}$

$dV_{full}/dt = k_{pack} V_{gen} C_{emp}$.

The growth term was $\mu(G)=\mu_{max}G/(1+G)$, and vector-balance was represented by $f(r)=1-2|r-0.5|$. The parameter set was $\mu_{max}=0.035$ h−1, $Y_{X/G}=0.55$, $k_{transf}=0.85$, $k_{cap}=1.02$, $k_{pack}=0.012$, $k_{lysis}=0.005$, and $G_{feed}=25$ mM. These values are explicitly computational starting values and were not claimed to be calibrated biological constants.

A two-dimensional computational design space was evaluated over dilution rate D from 0 to 0.12 h−1 and vector split ratio from 0.30 to 0.70. Twenty-five levels were used along each axis, generating 625 conditions. The purpose was to expose model structure, boundary behavior, and interactions rather than to identify an experimentally optimal operating point. Because the split-ratio penalty is explicitly symmetric around 0.5, a maximum at 50:50 is partly encoded by construction and should not be interpreted as an emergent biological discovery.

Reactor configuration. As written, dilution enters only the glucose balance; cells, genomes, capsids and vector are retained and nothing is harvested, i.e. a perfusion configuration with full retention and no bleed. The computational sweep assumed $X_0 = 1.0$, $G_0 = 25$ mM, $V_{gen,0} = C_{emp,0} = V_{full,0} = 0$, a simulation horizon $T = 96$ h, and $V_{full}(T)$ as the response. These initial conditions, horizon and response definition are assumptions used for the scenario analysis, not values recovered from the source paper; robustness was examined by varying the assumed settings.

Table 1. Five-state model definitions and Stage-1 computational summary

State	Symbol	Interpretation	Measurement role
Viable cells	X	Productive cell population	Potential online biomass proxy
Glucose	G	Carbon substrate	Potential online metabolite proxy
Free viral genomes	V_gen	Unpackaged vector genomes	Upstream intermediate
Empty capsids	C_emp	Capsid shells lacking packaged genome	Quality-related state

Full vectors	V_full	Genome-containing vectors	Primary production endpoint
Key Stage-1 result	625 conditions; D = 0-0.12 h−1; split = 0.30-0.70	Highest simulated response at D=0.12 h−1 and split=0.50	Boundary/scenario optimum; 50:50 partly encoded by model
Ablation / closure	Four nested model configurations	Growth→genome generation→packaging→lysis; packaging closure satisfied	Structural/implementation check, not experimental mass balance
Sobol	512 Saltelli base samples; 7,168 evaluations; ±30% ranges	Transfer/lysis-related terms among largest total-order drivers	Conditional variance allocation

Note: All values are computational. The Stage-1 constants are illustrative model inputs, not newly calibrated biological constants.

2.3 Source-faithful reconstruction of the Nguyen 21-state model

The central mechanistic component was reconstructed from Nguyen et al., whose published model describes recombinant AAV production following triple plasmid transfection of HEK293 cells [5]. The independent implementation retained the source model's 21-state structure, source parameter vector, observation definitions, initial conditions, and media-exchange event. The state system includes helper-, packaging-, and vector-associated uptake compartments; Rep protein; VP protein; viral DNA; nuclear and cytoplasmic empty and full capsids; and Rep/capsid intermediate complexes. The source implementation uses a 6-h media exchange, with initial plasmid inputs of 7.6×10^4 molecules per cell for each of the three plasmids.

The recovered source parameter vector included uptake, endosomal escape, nuclear entry, plasmid degradation, Rep synthesis, Cap/VP synthesis, Rep and VP degradation, DNA replication, Rep binding to packaging plasmid, capsid assembly, Rep binding to capsids, DNA packaging, and secretion terms. No refitting was performed for the source-faithful reconstruction. This distinction is critical: the numerical exercise asks whether the published model can be reproduced independently, not whether a new parameter set can be found that fits the source observations more closely.

Observation mapping was preserved exactly. Total viral DNA was defined from the source model's specified state combination; encapsidated DNA was derived from full-capsid states; and total capsid was constructed from the source's empty and full nuclear/cytoplasmic states. Altering this mapping could create an artificial fit improvement while changing the scientific meaning of the observable, so observation definitions were treated as part of the source model rather than as an implementation detail.

The independent implementation was integrated with BDF, Radau, and LSODA using relative tolerance 10−2 and absolute tolerance 10−3. The comparison was performed at source sampling times. MATLAB ODE15s identity was not claimed because the MATLAB runtime was

unavailable; instead, agreement among independent stiff solvers was used as a reproducibility check on the recovered dynamical system.

2.4 Forward time-point audit

In addition to source reconstruction, a forward consistency audit compared the source-fixed trajectory with the reported observations at later sampling times. This is not a temporal holdout because the original source parameter estimates were derived using the complete source dataset. It therefore measures forward consistency of a fixed published model rather than out-of-sample predictive performance. Discrepancies at 46 h were computed against the mean of the source replicates (total viral DNA −14.3%, full virions +44.7%, total capsid −6.3%; mean absolute discrepancy 21.8%).

This distinction is methodologically important. A source model may be numerically stable and still display meaningful deviations from the experimental data. Conversely, a refitted model may reduce residuals without being identifiable. The manuscript therefore reports numerical stability, observational agreement, and inferential identifiability as separate axes rather than compressing them into a single model-quality score.

2.5 Mechanistic ablation and mass-balance closure

Four nested model configurations were examined: growth only; growth plus genome-generation processes; genome generation plus packaging without lysis; and the complete model containing the lysis term. The purpose was to determine which model layers create qualitative changes in the flux, k_pack V_gen C_emp, and comparing the integrated flux with the final full-vector state. This is a mathematical conservation check on the implementation and is not equivalent to an experimental material balance because no measured mass-flow streams were generated in this study.

The ablation analysis also provides a guard against overinterpreting a complex model. If a downstream result is essentially unchanged after removal of a mechanistic term, that term may not be identifiable from the available observables even if it is biologically meaningful. Conversely, large changes after an ablation identify structural dependence of the model output on that mechanism.

2.6 Global sensitivity analysis

Global sensitivity analysis was conducted using Saltelli sampling for six Stage 1 kinetic inputs over ±30% computational ranges. A base sample size of 512 produced 7,168 model evaluations. First-order indices quantify the contribution of individual inputs without interaction terms, whereas total-order indices include the effect of interactions. The analysis was therefore interpreted as variance decomposition conditional on the selected parameter ranges, not as an empirical ranking of biological mechanisms.

A second variance-decomposition analysis was performed on the SCN1A end-to-end scenario. Production-side and downstream parameters were separated so that uncertainty could be attributed to manufacturing versus reconstitution/correction. This stage-wise decomposition is central to the digital-twin concept because it turns a final endpoint into a set of experimentally addressable uncertainty sources.

2.7 Dual-AAV reconstitution and gene-specific correction modules

Stage 2 was treated as a functional co-delivery/reconstitution term rather than a simple vector-dose multiplier. The literature supports this distinction: dual-AAV systems require both components to be present in the same target cell, and vector preparation quality can influence reconstitution. Ferreira et al. reported that higher-quality vector preparations produced substantially greater split-protein reconstitution than low-quality preparations in their tested systems, illustrating why full/empty composition can propagate into downstream function [11].

Stage 3 was represented as a conditional correction efficiency. The three disease models were intentionally assigned different evidentiary status. LRRK2 G2019S used the human iPSC correction study by Chang et al. as a biological anchor [13]. SCN1A R613X used the 2026 dual-AAV9 ABE study by Nelson et al., which directly addressed this recurrent variant in a disease model [14]. RPE65 R44X used the dual-AAV split-ABE rd12 study by Jo et al. [15]. Levy et al. provided the broader precedent for dual-AAV ABE/CBE delivery and intein-mediated reconstitution [2].

The LRRK2 case requires special caution. Chang et al. demonstrated ABE-mediated correction in patient-derived iPSCs and phenotypic rescue in derived dopaminergic neurons, but that study does not establish the same production, co-transduction, or in vivo delivery pathway represented in the present digital twin. The LRRK2 module is therefore a literature-anchored scenario, not a direct recreation of an AAV experiment.

The SCN1A case has a closer architecture match. Nelson et al. reported dual-AAV9 delivery of an optimized ABE targeting SCN1A R613X, with DNA and mRNA editing measured in neocortex and functional outcomes in a mouse model [14]. Those findings are not re-generated here; they provide external evidence context for comparison with the computational architecture. The RPE65 case similarly uses an external dual-AAV split-ABE context. Jo et al. reported correction of the Rpe65 R44X allele in rd12 mice and associated restoration of RPE65-related molecular and retinal functional readouts [15]. The present model does not simulate those experimental animals mechanistically; it uses the literature as a downstream evidence anchor.

2.8 Monte Carlo uncertainty propagation

Twenty thousand Monte Carlo draws were generated for each gene-specific scenario. The sampling distributions were constructed from the adopted computational ranges and literature-informed quantities. For each draw, production, reconstitution, and correction terms were propagated to the end-to-end output. The reported mean, central 95% interval, and threshold exceedance probability describe the adopted scenario distribution. They are not confidence intervals around measured parameters and should not be interpreted as probabilities of clinical success.

2.9 Identifiability, singular-value analysis, and constrained reparameterization

The source model contains 14 kinetic parameters. Identifiability was evaluated through a local sensitivity matrix: the Jacobian of the 31 log10 residuals defined below with respect to the log10 parameter vector. Singular-value decomposition was used to distinguish strongly informed parameter combinations from weakly informed directions. At a relative singular-value threshold of 10^{-12} the full model had numerical rank 12/14 (10/14 at 10^{-6}), with a condition number of

approximately 6.4×10^{21}. Two directions were numerically null (relative singular values below 10^{-15}): one is the Cap-protein degradation rate (kProteinDegCap), to which the outputs are insensitive because assembly is effectively instantaneous, and one is a combination dominated by kAssembly with kBindCapsid and kSecrete; a third near-null direction (5×10^{-10}) involves the same three parameters.

A source-constrained eight-dimensional formulation fixed source-supported parameters (kEscape, kProteinDegRep, kProteinDegCap, kAssembly, kSecrete) and tied Rep- and Cap-synthesis terms. The eight free coordinates were kUptake, kNuclearEntry, kPlasmidDeg, the common Rep/Cap synthesis term, kDNArep, kBindRCplasmid, kBindCapsid and kPack. Residuals were unweighted log10(model/observation) over all 31 source observations (12 total viral DNA, 10 total capsid, 9 full virion). Optimisation used scipy least_squares in log10-parameter space from the source values and from random starts (numpy seed 1, SD 0.8 log10 units); the best of three refined basins is reported (SSR 0.668084, log10 RMSE 0.146803), identical across BDF, Radau and LSODA. Other local minima were found (SSR 0.687, 0.704, and a stalled basin near 2.2). The fitted vector, solver settings and seed are archived (fit8d_FROZEN.json). The leading five singular values gave an approximate condition number of 17.0.

Weak-direction profiles were generated by perturbing selected singular coordinates while reoptimizing the remaining identifiable coordinates. These profiles were intentionally interpreted as local stability diagnostics rather than confidence intervals. The weak directions showed only small changes in SSR over the tested local range, confirming that sparse observations provide limited information about certain microscopic combinations.

Identifiability representation	**SSR**	**Condition number**	**Interpretation**
Full 14D source parameterization	3.1144	6.4×10^{21}	Locally ill-conditioned; rank 12/14
Source-constrained 8D	0.668084	6.00×10^{3}	Usable local reduced representation
Leading 5D singular core	0.668084	~17.0	Strongly conditioned local observable core

2.10 Synthetic Bayesian calibration

Bayesian calibration was tested using synthetic observations generated from known model parameters. The purpose was to test whether the inference implementation can recover known computational truth under a controlled setting. Synthetic observations combined dense X/G measurements with sparser vector-output observations to mimic the general structure of online and offline process measurements. Four MCMC chains were initialized from dispersed starting values. R-hat and integrated autocorrelation time were used as convergence diagnostics.

The analysis was deliberately classified as a method test. A successful posterior fit to synthetic data would establish that the software can recover parameters under the synthetic observation model; it would not establish that the biological model is correct. Conversely, non-convergence is

a useful warning that the inference pipeline requires additional sampling or reparameterization before application to real process data.

2.11 Synthetic state estimation and process control

A UKF was used to estimate hidden states from synthetic observations of X and G only, with a bootstrap particle filter as a non-Gaussian comparison. Filter performance was evaluated using RMSE, NIS, and NEES. A PI controller was then applied to a synthetic glucose-control problem. These calculations test whether the software architecture can connect observations to latent states and control inputs. They do not constitute online bioreactor validation.

2.12 Hydrodynamic scoping

A generic 50-L vessel geometry was used for first-order Reynolds-number, mixing, and kLa calculations across rotational speed. The calculation was designed as a scoping layer linking process kinetics to plausible engineering operating ranges. It was not intended to substitute for reactor-specific CFD, experimentally measured power input, oxygen-transfer testing, or shear characterization.

2.13 External literature benchmark

The upgraded external benchmark uses Grieger et al., who evaluated plasmid ratios 3:1:1, 2:1.5:1, 1:1:1, and 1:1:2 under PEI:DNA ratios of 2:1 and 4:1 in suspension HEK293 production [4]. Their study identified the 2:1.5:1 plasmid ratio with PEI:DNA 2:1 as the highest vector-genome-per-cell condition among the tested combinations. Because the accessible article reports these values graphically rather than as a machine-readable numerical table, the eight bar values used for the benchmark (Supplementary Table S2) are approximate digitizations. They are not raw experimental data and no digitization-derived error bars are presented.

2.14 AAV2 dual-AAV scenario extension

The extension reuses a frozen core: a central quality scenario of 67.5% full capsids, a 1:1 A:B ratio, a central normalised functional output of 5.50% ($\eta = 0.055/0.675 = 0.081481$), an anion-exchange recovery scenario of 0.703 and a vector-genome production rate of 66.33 VG per transfected cell per hour. The production anchor is an AAV2 HEK293 suspension process [21], the dual-AAV functional benchmark is a retinal dual-vector study [22], and the capsid-state framing follows reports that intermediate capsids can contribute to vector-genome titre without contributing to measured potency [19,20]. The extension adds no experimental observations; the capsid-quality factors, competition coefficients, assay CVs and K_D below are hypothesis parameters.

Stage definitions and dose normalisation. We distinguish five quantities at each stage: (i) particles (capsids: full, intermediate and empty); (ii) vector genomes (VG), which include genomes in intermediate capsids; (iii) successful cell entry; (iv) functional protein after reconstitution; and (v) editing. The surrogate function = $\eta \cdot F_full \cdot q_cap \cdot h(A:B)$ is defined per unit of full-capsid dose at matched A:B, so the full fraction enters once, through F_full. The total-VG index is a separate manufacturing quantity and is not multiplied by F_full in the function. When dose is varied (Study 4) it is defined in full-capsid units, so a change in full fraction is not applied a second time.

Manufacturing-to-delivery mass balance. If a harvest with full fraction F_in is purified with recoveries R_full and R_nonfull, the output full fraction is F_out = R_full·F_in / (R_full·F_in + R_nonfull·(1 − F_in)). The 3.09% full capsids of the 21-state model at 46 h and the 67.5% central scenario used here are not linked in this paper. If they were, F_out = 0.675 from F_in = 0.0309 would require R_nonfull/R_full = 0.01535, i.e. about 65-fold selective recovery of full over non-full capsids. A scalar anion-exchange recovery (0.703) does not specify this. We therefore treat the two values as separate literature-anchored scenarios and make no claim of a calibrated continuous trajectory.

Neutral co-delivery baseline. Under independent Poisson entry of the two vectors at total effective multiplicity m and ratio r = A:B, P(both) = (1 − exp(−m r/(1+r)))·(1 − exp(−m/(1+r))). In the low-dose limit the penalty relative to 1:1 is $4r/(1+r)^2$ (0.889 at 2:1; 0.640 at 4:1). The ratio penalty h used here (0.318 at 2:1; 0.135 at 4:1) is steeper and therefore encodes additional reconstitution or competition penalties; it is a named hypothesis and is not derived from co-delivery physics.

Yield per batch versus potency per dose. The scenarios assume a fixed administered dose. Under a fixed dose, titre (qVG × transfection × AEX recovery) changes the number of doses and the cost per dose, not the delivered function. The independent sampling of titre and function drivers therefore produces zero titre–function association by construction. A yield-limited dose would couple them and would need a process-supported dependence; none is modelled, so the Sobol and AUC results are conditional illustrations.

Study 1 represented F_full + F_intermediate + F_empty = 1 on a grid of 105 compositions (full 0.20–0.90, intermediate 0–0.60). Intermediate capsids retained a vector-genome contribution but no potency contribution.

Study 2 applied a multiplier q_cap (0.70–1.20, 51 levels) to the full-particle contribution at 50%, 67.5% and 85% full capsids.

Study 3 imposed representative coefficients of variation (VG_A/VG_B 8%, full_A/full_B 7%, functional readout 10%) and propagated 20,000 draws.

Study 4 crossed total dose (0.125–8 × central, seven levels) with A:B ratio (0.25–4, five levels); the matched-reference normalised endpoint was multiplied by a saturating exposure term D/(D + K_D) with K_D = 1 central-dose unit.

Study 5 set D_eff = D_full/(1 + αE + βI) over $\alpha, \beta \in [0, 2]$ (41 levels each) at the central composition.

Study 6 treated η as uncertain (prior mean 0.0815, SD 0.025) and updated it with synthetic observations at three full-capsid fractions, generated as y = eta_true x F + N(0, 0.003^2) with eta_true = 0.0815 and seed 20261026.

Study 7 drew 30,000 joint samples of qVG, transfection efficiency, full fraction, A:B ratio, η, AEX recovery and q_cap and reported Pearson and Spearman correlations with normalised function and a total vector-genome index.

Study 8 mapped regimes on a 141 × 141 grid reconstitution-limited; full ≥ 60% outside that window is stoichiometry-limited; full < 60% with a balanced ratio is quality-limited; and full < 60% with an imbalanced ratio is compound-limited.

2.15 In-depth studies

The in-depth studies use a surrogate reconstructed from the packaged result files: normalised function = $\eta \cdot F_full \cdot q_cap \cdot h(A:B)$ with $\eta = 0.0815$, and total vector-genome index = $136.69 \cdot qVG \cdot$transfection efficiency$\cdot$AEX recovery. The ratio penalty h ($h(1) = 1$, symmetric in log ratio) was read from the Study 8 grid at 90% full capsids; the surrogate reproduces all 30,000

Study 7 outputs to within 3.4×10^{-5}. The studies are computational scenarios, their assumptions are hypotheses, and none generates experimental data.

Study 9 estimated first- and total-order Sobol indices with Saltelli sampling (16,384 base samples; Saltelli first-order and Jansen total-order estimators; 300 bootstrap resamples) by resampling the empirical marginals of the seven Study 7 inputs independently.

Study 10a treated η as Normal (prior SD 31% of mean) and computed the posterior SD after n independent lots with relative observation error $\sqrt{(CV_fn^2 + CV_F^2)}$, for CV_fn of 5%, 10% and 20% and CV_F of 0 or 7%.

Study 10b computed the Fisher information for (log η, log K_D) from $\log y = \log \eta + \log D - \log(D + K_D)$ with $K_D = 1$, three replicates per dose and 10% CV, for five dose designs.

Study 11 scored the 30,000 Study 7 lots with five rules (titre with 8% CV; full fraction with 7% CV; their product; a functional assay with 10% CV and 1 or 3 replicates, using log-normal measurement error), accepted the same fraction of lots under each rule, and reported AUC and consumer risk against the specification function ≥ 3.30% (60% of central); replicate sweeps used 1, 2, 3, 6 and 12 replicates. Measurement noise was mean-preserving log-normal ($sigma^2 = \ln(1 + CV^2)$), replicates were averaged arithmetically, the accepted share equalled the share of lots meeting specification (54.6%), and reported values are means over 50 measurement-noise seeds (saved run: seed 20261027).

Study 12 drew log-normal A:B ratios (200,000 draws per setting; log-SD 0–0.5) at 15 full-capsid fractions (0.50–0.85) and computed the share meeting specification.

Study 13 computed rank correlation and conditional exceedance for titre and function in the Study 7 draws.

Study 14 computed the Fisher information for (α, β) from $\log D_eff = -\log(1 + \alpha E + \beta I)$ over the 105 Study 1 compositions, with three replicates per lot, 10% CV and true $\alpha = \beta = 1$, for four lot subsets; 95% joint regions used $\Delta\chi^2 = 6.18$.

2.16 Figures, statistics and reproducibility

All figures were generated from the packaged result files by the scripts in the accompanying repository (Python; Matplotlib). A verification script re-derives the numbers quoted in the text from the result files where those version-matched files are available. No inferential statistics beyond the stated Monte Carlo, Pearson and Spearman summaries were used, and no experimental replicates were generated.

3. Results

3.1 A framework that keeps four evidence classes separate

The framework links three stages (Fig. 1a-c): manufacture of two vector components by triple-plasmid transfection of HEK293 cells, co-delivery and reconstitution of the split editor, and gene-specific correction, wrapped in a digital-twin loop of process states, observations, a mechanistic model, uncertainty and experimental priorities (Fig. 1d). Every quantity is tagged by origin: source-recovered, newly computed, literature-derived or synthetic (Fig. 1e). The results below follow that logic. We first test whether a published production model can be reproduced and what its data can identify (Sections 3.3–3.7), then follow product quality through to a functional endpoint in separately sourced scenarios (the 3.09% full-capsid output of the 21-state model at 46 h and the 67.5% central AAV2 scenario are not linked by a calibrated purification step; if linked, the full-capsid fraction would have to rise from 3.09% to 67.5%, which requires about 65-fold selective recovery of full over non-full capsids) (Sections 3.8–3.11), then use that endpoint to ask decision-level questions about sensitivity, measurement design, lot release and process control (Sections 3.12–3.15), and finally allocate uncertainty and test the inference machinery itself (Sections 3.16–3.18) and close with a summary of the results (Section 3.19).

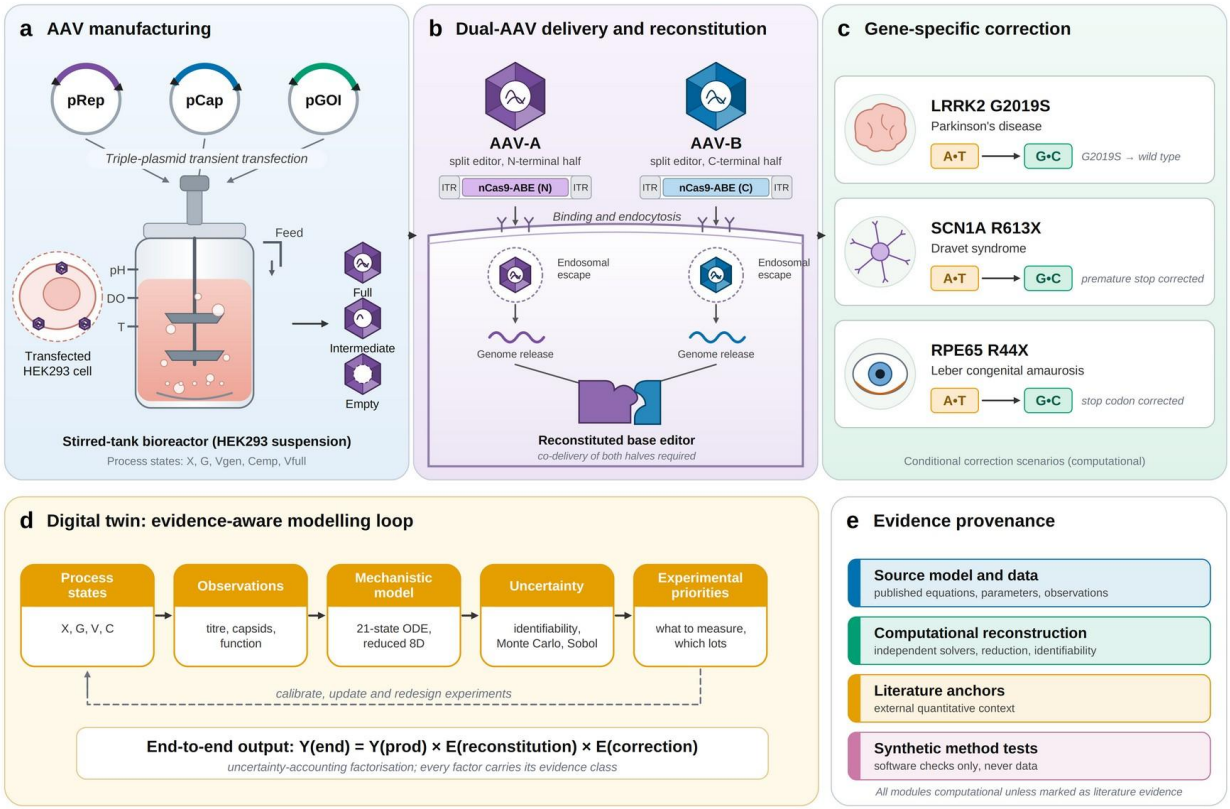


Figure 1: Mechanistic digital-twin architecture and evidence framework: *a, Triple-plasmid transient transfection of HEK293 suspension cultures produces full, intermediate and empty capsids (states X, G, V_gen, C_emp, V_full). b, Dual-AAV co-transduction and reconstitution of a split editor. c, Gene-specific correction scenarios (LRRK2 G2019S, SCN1A R613X, RPE65 R44X) with their evidence context. d, The digital-twin loop and the uncertainty-accounting factorisation Y_end = Y_prod × E_reconst × E_correct. e, Evidence classes carried through the study. All modules are computational unless marked as external literature evidence.*

3.2 Simplified production model: design space, ablation and sensitivity

Design-space sweep. The 625-condition Stage 1 sweep covered dilution rate and vector split ratio. As written (Section 2.2), dilution enters only the glucose balance, so the reactor is a perfusion with full retention of cells, genomes and capsids and no bleed or harvest term. With initial glucose equal to the feed concentration, glucose never became limiting (minimum 22.5 mM against a half-saturation constant of 1 mM), and the simulated response varied by less than 0.05% across the whole dilution range (initial conditions, horizon and response definition: Section 2.2). The maximum at D = 0.12 h-1 is therefore the upper edge of the sampled range of an almost flat surface, not evidence that higher dilution is beneficial. The 50:50 result is also structural because $f(r) = 1 - 2|r - 0.5|$ penalises imbalance by construction.

This boundary behaviour is a model-development requirement, not a result. The simplified model contains no mechanism that penalises or rewards dilution once glucose is replete (no washout of retained cells, no medium cost, no product loss to the permeate), so no interior optimum can exist. A full chemostat version, in which every state is diluted, places the optimum at D = 0 because of washout. A calibrated model would need explicit harvest, retention and bleed balances before an operating dilution rate is discussed.

Ablation and mass-balance. The nested ablation analysis showed that growth alone cannot reproduce the vector-production trajectory because viral genome generation and capsid packaging are the mechanisms that convert biomass into vector output. Adding genome production produced and capsid shells into full particles; the lysis term altered the late trajectory by reducing the viable-cell pool as full-vector burden accumulated. The analysis therefore confirmed that the simplified model's endpoint is mechanistically dependent on multiple sequential transformations rather than on biomass alone.

The packaging mass-balance calculation provided an internal implementation check: the integrated packaging flux matched the terminal full-vector state within the numerical closure reported in the computational audit. This supports algebraic consistency of the state equations. It should not be confused with a measured mass balance because no experimental input/output stream data were generated.

Global sensitivity. In the simplified production model, total-order Sobol analysis identified the cell-lysis and transfer-related terms as major contributors to simulated output variance under the adopted ±30% ranges. This result should be read as a statement about the mathematical model and its uncertainty ranges: a parameter can have a large Sobol index because the model output is structurally sensitive to it, because its uncertainty range is broad, or because it interacts with other uncertain inputs.

The end-to-end SCN1A analysis provided a second, more operational result. Approximately 70% of simulated output variance arose from the production side and approximately 30% from downstream reconstitution/correction terms. Thus, under the adopted scenario distributions, reducing uncertainty in upstream manufacturing would be expected to have a larger effect on endpoint uncertainty than refining downstream parameters alone. This is a hypothesis for experiment prioritization, not a claim about the biological hierarchy of mechanisms.

3.3 Independent reconstruction of the source 21-state model

The source-faithful reconstruction produced the following principal BDF outputs: total viral DNA of 5.5909×10^{8} at 10 h, 1.2610×10^{9} at 24 h, 2.4875×10^{9} at 34 h, and 4.1233×10^{9} at 46 h. Corresponding full-vector quantities were 1.557×10^{7}, 7.5326×10^{8}, 1.1888×10^{9}, and 1.5728×10^{9}. Total capsid quantities were 9.3210×10^{9}, 2.9514×10^{10}, 4.0549×10^{10}, and 5.0923×10^{10}.

Table 2. Source-faithful 21-state model trajectory

Time	Total viral DNA	Full virions	Total capsid	Full/total capsid
10 h	5.5909×10^{8}	1.557×10^{7}	9.3210×10^{9}	0.17%
24 h	1.2610×10^{9}	7.5326×10^{8}	2.9514×10^{10}	2.55%
34 h	2.4875×10^{9}	1.1888×10^{9}	4.0549×10^{10}	2.93%
46 h	4.1233×10^{9}	1.5728×10^{9}	5.0923×10^{10}	3.09%

The full-to-total-capsid fraction rose from about 0.2% at 10 h to about 3.1% at 46 h across the reconstructed trajectory. This behavior is consistent with the qualitative feature emphasized in the original mechanistic study: genome replication and capsid synthesis are not perfectly synchronized, producing a large population of capsids that are not genome-filled [5]. Importantly, the present calculation does not claim that these percentages represent a universal AAV manufacturing property. They are outputs of this specific source model, source parameterization, source observation definitions, and initial conditions.

The reconstruction is shown against the source replicates in **Fig. 2a.** Capsid filling, the quality attribute that matters most for the dual-AAV analyses that follow, is low and rises slowly (Fig. 2b): 0.17% at 10 h, 2.55% at 24 h, 2.93% at 34 h and 3.09% at 46 h. Observed filling, computed as mean genome copies over mean capsid count at the same time points, was 2.25%, 2.75% and 2.00% at 24, 34 and 46 h, so the model overestimates filling at the final time point. Throughout this paper the upstream reference is the source-parameter trajectory (3.09% full capsids at 46 h). The 8D refit gives 3.70%, but it is a descriptive reduction with strongly displaced parameters and is used only for identifiability analysis, not as a manufacturing prediction. Either value is far below the 67.5% central scenario used in the downstream studies, which is an independent AAV2 scenario and not a calibrated purification outcome.

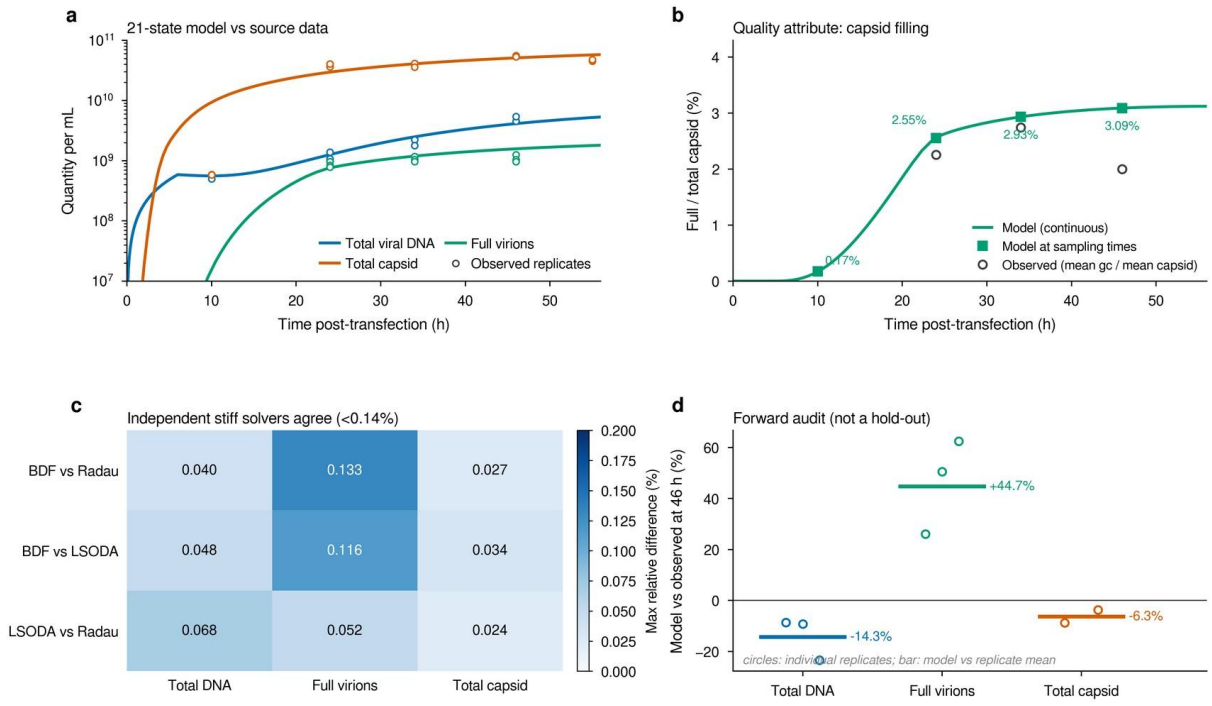


Figure 2. Independent reconstruction and numerical audit of the 21-state AAV production Model : *a, Model trajectories (lines) for total viral DNA, total capsid and full virions per mL, with source replicates (open circles). b, Full-to-total capsid fraction: continuous model (line), model at sampling times (squares) and observed ratio of mean genome copies to mean capsid count (open circles). c, Maximum relative differences between BDF, Radau and LSODA integrations (all <0.14%). d, Model-to-observation discrepancy at 46 h for each replicate (circles) and against the replicate mean (bars); this is a forward-consistency audit, not a hold-out test.*

3.4 Solver cross-check establishes numerical stability but not biological correctness

Independent BDF, Radau, and LSODA integrations agreed closely. BDF versus Radau differed by 0.0403% for total DNA, 0.13295% for full virions, and 0.0273% for total capsid. BDF versus LSODA differed by 0.04845%, 0.11577%, and 0.03365%, respectively. LSODA versus Radau differed by 0.06849%, 0.05215%, and 0.02384%. The largest discrepancy across the comparisons was therefore below 0.14%.

This result is strong evidence for numerical reproducibility of the independently implemented equations under the selected tolerances. It does not prove identity with the original MATLAB ODE15s execution, because the original runtime was not available. It also does not establish agreement with experimental measurements. Numerical stability and biological validity are separate dimensions and are reported separately here.

Table 3. Numerical reproducibility and forward consistency

Solver comparison	Total DNA max relative difference	Full virion max relative difference	Total capsid max relative difference
BDF vs Radau	0.0403%	0.13295%	0.0273%
BDF vs LSODA	0.04845%	0.11577%	0.03365%
LSODA vs Radau	0.06849%	0.05215%	0.02384%

Forward 46-h audit	Total DNA −14.3%; full virion +44.7%; total capsid −6.3% (vs replicate means)	Mean absolute discrepancy 21.8%	Forward consistency of source-fixed model; not temporal holdout

Note: Solver agreement establishes numerical stability under tested tolerances. The forward audit is not out-of-sample validation because source parameters were estimated using the complete source dataset.

3.5 Forward consistency audit exposes a nonzero observational gap

With the source parameter set held fixed, the model was propagated to the 46-h observation (Fig. 2d). Relative to the mean of the source replicates, total viral DNA was 14.3% below the observation, total capsid 6.3% below, and full virions 44.7% above; the mean absolute discrepancy was 21.8%. Individual replicates scatter widely around these means, and the discrepancy for full virions ranges from +26% to +62% across the three replicates, so the sign of the gap is robust but its size is not. (Comparing against the first replicate alone, as in an earlier draft of this work, gives −8.6%, +26.1% and −8.7%; we use replicate means throughout.)

The source model also overshoots total capsid at 10 h by 15.9-fold (9.32×10^{9} versus 5.85×10^{8} per mL); these two observations contribute 2.89 of the sum of squared log10 residuals (3.11) at the source parameters. The 8D refit reduces the overshoot to 2.1-fold.

These deviations do not make the model unusable; they show that an independently reproduced source implementation retains residual disagreement with the published observations, concentrated in the full-virion measurement. A computational twin intended for experimental use must reproduce both the governing equations and the measurement layer. Residual differences can arise from parameter uncertainty, measurement noise, observation mapping, missing mechanisms or implementation differences, and the present study, which generated no new measurements, cannot say which dominates.

3.6 Identifiability gate: the full 14-parameter model is not equally learnable

The local singular-value analysis identified 12 numerically informed directions out of 14 at a relative threshold of 10^{-12} (10 of 14 at 10^{-6}). The singular spectrum decreased across many orders of magnitude, ending in two effectively null directions. The resulting condition number of approximately 6.4×10^{21} demonstrates extreme sensitivity of individual parameter estimates to small perturbations when the full parameterization is considered.

The practical implication is not that the mechanistic model is biologically wrong. Rather, the available observations do not provide enough independent information to estimate every microscopic rate simultaneously. Several parameter combinations can generate nearly indistinguishable trajectories. This is a classic inverse-problem limitation: mechanistic detail increases the number of latent degrees of freedom faster than sparse measurements increase information.

The source-constrained eight-dimensional fit substantially improved conditioning to approximately 6.00×10^{3} while achieving SSR = 0.668084 and log10 RMSE = 0.146803. The leading five-dimensional singular core had an approximate condition number of 17.0. The five

coordinates should be interpreted as local observable directions in parameter space rather than as five directly measured biological rates. This distinction is essential because singular vectors are mathematical combinations of original parameters.

The singular spectrum, correlation structure and singular-vector loadings are shown in Fig. 3. The 14-parameter space contains many strongly correlated pairs (Fig. 3c) and two numerically null directions (Fig. 3 a,b). In the source-constrained eight-dimensional representation, five directions dominate and three are weak (Fig. 3a,d); the weak directions v6-v8 load mainly on p2–p6 and p8 (v6: kPlasmidDeg and kBindRCplasmid; v7: kRepCapSyn, kDNArep and kNuclearEntry; v8: kPack).

Table 4. Identifiability and model-reduction comparison

Representation	**Free dimensions**	**SSR**	**log10 RMSE**	**Condition number**	**Local rank / interpretation**
Full source parameterization	14	3.1144	0.3170	6.4×10^{21}	12/14 at relative threshold 10^{-12} (10/14 at 10^{-6})
Source-constrained 8D (refit)	8	0.668084	0.146803	6.00×10^{3}	Reduced, locally usable representation
Leading SVD core	5	0.668084	—	17.0	Same fitted vector; SSR equal by construction

Note: *Identifiability is local and data-dependent. The 5D coordinates are singular-vector combinations, not directly measured biological rates. The 5D row reports the conditioning of the leading singular directions at the same fitted vector, so its SSR equals the 8D value by construction.*

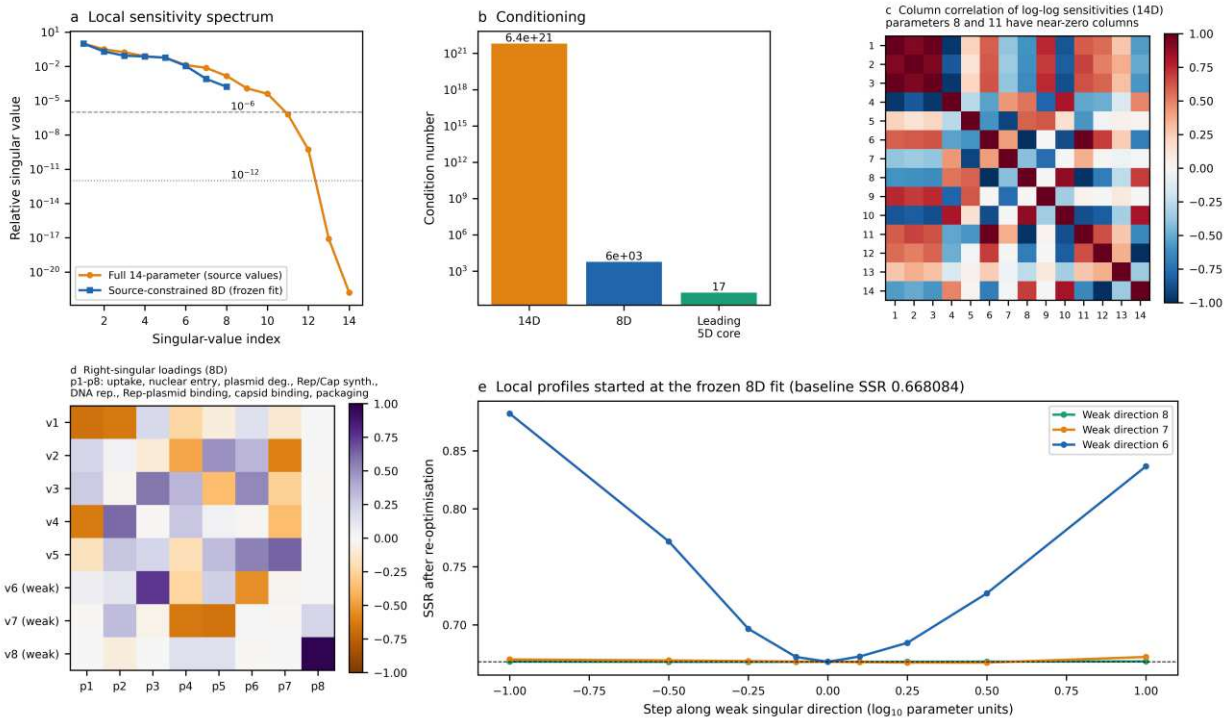


Figure 3: Local identifiability of the 21-state model. *(a) Relative singular values of the log-log sensitivity matrix (31 residuals) for the full 14-parameter model at source values and for the 8D representation at the frozen fit; dashed lines mark relative thresholds 10^{-6} and 10^{-12}. (b) Condition numbers: 6.4×10^{21} (14D), 6.00×10^{3} (8D) and 17.0 (leading five directions). (c) Correlation between columns of the 14D sensitivity matrix (parameters 8 and 11 have near-zero columns). (d) Right-singular-vector loadings of the 8D representation (p1–p8: kUptake, kNuclearEntry, kPlasmidDeg, common Rep/Cap synthesis, kDNArep, kBindRCplasmid, kBindCapsid, kPack). (e) Local profiles along weak directions 6–8 after re-optimisation of the remaining coordinates, started at the frozen 8D fit (SSR 0.668084). This fit replaces an earlier fit (SSR 0.731223) whose parameter vector was not archived; the earlier profiles could not be reproduced and are not shown. Profiles are local diagnostics, not confidence intervals.*

3.7 Reduced representations and model parsimony

The full 14-dimensional representation, evaluated at the source parameters, had SSR 3.1144 and a condition number of 6.4×10^{21}. The constrained eight-dimensional representation, which was refitted, reached SSR 0.6681 with condition number 6.00×10^{3}, and its leading five-dimensional core had a condition number of 17.0 (Table 4). Because the 14D value is the unrefitted source model and the 8D value follows refitting, the lower SSR reflects refitting rather than parsimony alone; the robust conclusion is the improvement in conditioning by about eighteen orders of magnitude without loss of fitted observable behaviour. Profiles along the weak directions were shallow (Fig. 3e): in the regenerated profiles SSR changed by less than 0.001 along direction 8, by 0.004 along direction 7 and by 0.005 along direction 6 over a step of 0.1 log10 units (direction 6 rises to 0.21 at one log10 unit). Profiles were regenerated from the frozen 8D fit (baseline SSR 0.668084; Fig. 3e), so profile and fit refer to the same parameter vector. Earlier quoted SSR values (0.7208–0.7312, 0.731223) came from runs whose parameter vectors were not saved and are superseded. The profiles are local, the observation set is fixed and nonlinear correlations outside the explored neighbourhood were not mapped, so global identifiability is not established.

The refitted vector departs strongly from source values along weakly identified directions (kPack ×1500, kRepCapSyn ×0.05, kBindRCplasmid ×0.01), so the 8D fit is a descriptive reduction, not a biological parameter estimate. Under it the full-to-total capsid fraction at 46 h is 3.70% (3.09% at source parameters) and full virions at 46 h exceed the replicate mean by about 52%. The upstream reference used throughout is the source-parameter trajectory (3.09%).

This supports an architecture in which the mechanistic model remains available for forward simulation while a lower-dimensional latent representation is used for calibration and online inference. A future experimental campaign should target measurements that interrogate the weak mechanisms directly, rather than collect more samples of the same aggregate endpoint.

3.8 Capsid quality links manufacturing to function

Manufacturing outputs matter for dual-AAV delivery only through the particles that can empty capsids that sum to one (Study 1; Fig. 4a). In the base model intermediate particles contribute vector genomes but no potency, so normalised function depends only on the full fraction and rises linearly from 1.63% at 20% full capsids to 7.33% at 90% (5.50% at the central 67.5% scenario). The ternary contours in Fig. 4a are therefore parallel to the full-capsid axis by construction; the study is a structural statement of an assumption, not a discovery.

Study 2 added a dimensionless capsid-quality multiplier q_cap (0.70–1.20) on the full-particle contribution. Across that range function at the central composition moved from 3.85% to 6.60%, and across 50%, 67.5% and 85% full capsids the span was 2.85–4.89%, 3.85–6.60% and 4.85–8.31% (Fig. 4b). All values collapse onto a single relationship, function = $\eta \cdot F_{full} \cdot q_{cap}$, with η = 0.055/0.675 = 0.0815, which reproduces the tabulated values to machine precision (Fig. 4c). A 20% change in capsid quality therefore has the same effect on function as a 20% change in full fraction, which makes explicit that a shift in either quantity can mask the other in an endpoint assay.

3.9 Dose, stoichiometry and impurity competition

Dual-AAV function depends on the pair, not on either vector alone. Study 4 combined total dose (0.125–8 × central) with the A:B ratio (0.25–4). At the highest-quality scenario the matched-reference endpoint was 5.50% at 1:1, 1.75% at 2:1 or 1:2, and 0.74% at 4:1 or 1:4, a sharp penalty for imbalance. These penalties (relative factors 0.318 and 0.135 against 1:1) are assumed, not derived: an independent-Poisson co-delivery null gives 0.889 and 0.640 in the low-dose limit (4r/(1+r)^2), so the extra penalty is a hypothesis to be calibrated against measured co-delivery and reconstitution. Multiplying by a saturating exposure term D/(D + K_D), with K_D = 1 central-dose unit, gave an absolute functional index that rose from 0.61% to 4.89% at 1:1 across the 64-fold dose range, but only from 0.19% to 1.56% at 2:1 and from 0.08% to 0.66% at 4:1 (Fig. 4d,e). Doubling the dose of an imbalanced preparation therefore recovers far less function than correcting the ratio. K_D is a hypothesis parameter and must be estimated experimentally.

Study 5 asked whether non-functional particles could suppress effective exposure, using $D_{eff} = D_{full}/(1 + \alpha E + \beta I)$ with competition coefficients α (empty) and β (intermediate) between 0 and 2. At the central composition (32.5% non-full capsids) equal coefficients of 1 reduced function from 5.50% to 4.15% (a 25% loss) and coefficients of 2 reduced it to 3.33% (−39%; Fig. 4f). The functional form and coefficients are sensitivity assumptions, not established AAV2 biology, but

they show that impurity competition would have a quantitatively large effect, equivalent, because function scales linearly with full fraction, to a 25–39% reduction in full-capsid fraction.

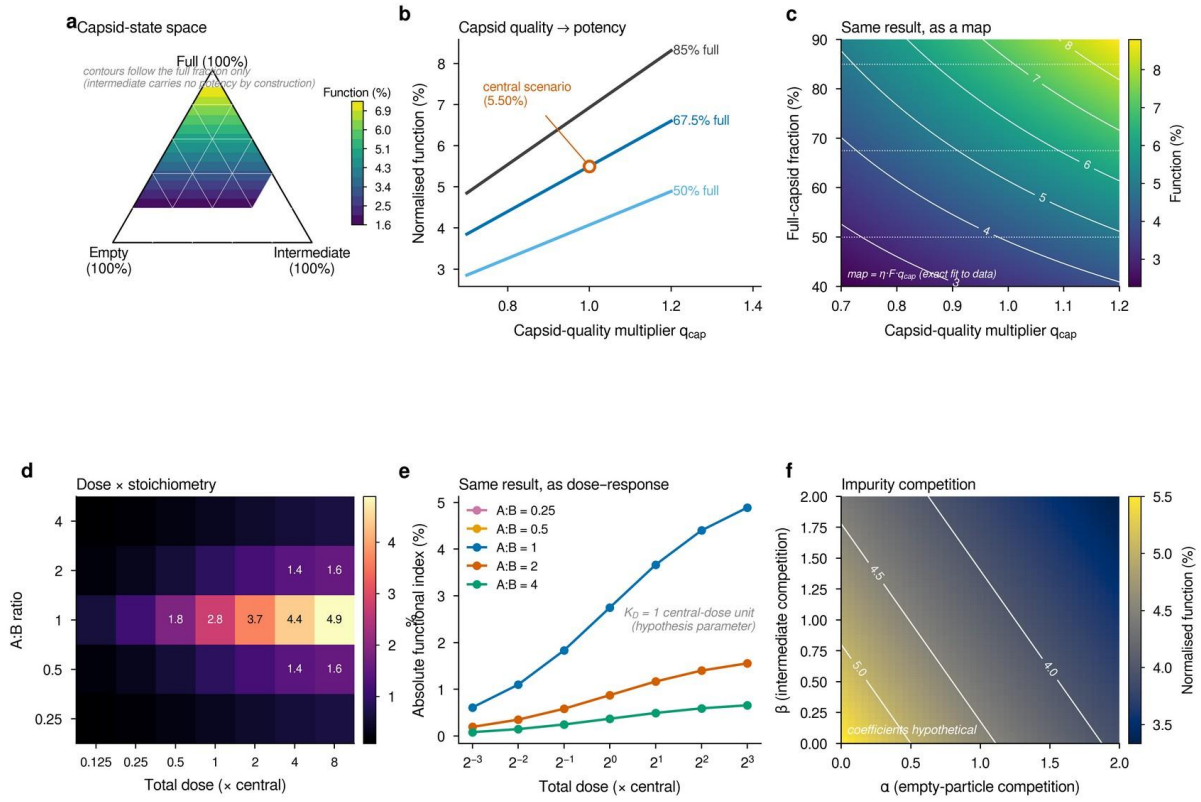


Figure 4: Capsid quality, dose, stoichiometry and impurity competition *a, Normalised function across full, intermediate and empty capsid compositions (Study 1; contours follow the full fraction by construction). b, Capsid-quality multiplier q_cap versus function at three full-capsid fractions, with the central scenario (5.50%) marked (Study 2). c, The same relationship as a continuous map, function = η·F_full·q_cap (exact fit to the tabulated values). d, Absolute functional index over total dose and A:B ratio (Study 4). e, The same data as dose–response curves, K_D = 1 central-dose unit (hypothesis parameter). f, Effect of empty-particle (α) and intermediate-particle (β) competition coefficients on function (Study 5; hypothetical coefficients). All studies are computational scenarios.*

3.10 Measurement uncertainty and Bayesian

Product-quality attributes are themselves measured with error. **Study 3** imposed representative assay coefficients of variation (8% on vector-genome ratio, 7% on full-fraction ratio, 10% on the functional readout) and propagated 20,000 draws. Model-propagated function had a median of 4.83% (2.5–97.5th percentile 3.55–5.71%); adding synthetic assay noise widened the interval to 3.17–6.25% (median 4.79%), with a residual standard deviation of 0.55 percentage points (Fig. 5a,b). Because function scales linearly with full fraction, the 95% interval with assay noise (about ±32% of the median) is equivalent to a ±32% shift in full-capsid fraction. The median sits below the 5.50% central value, most likely because noise in the measured A:B ratio can only move function away from the 1:1 optimum. The CVs are scenario values, not assay-validation claims.

Study 6 demonstrates how future validation data would update the one structural parameter that links quality to function. SYNTHETIC DEMONSTRATION: observations were generated as y = eta_true x F + N(0, 0.003^2) with eta_true = 0.0815, F = 0.50, 0.675 and 0.85 and random seed

20261026, giving y = 0.0424, 0.0564 and 0.0716. With the prior eta ~ N(0.0815, 0.025^2) the conjugate Gaussian posterior is 0.0841 (SD 0.0025), a ten-fold contraction (Fig. 5c); a grid integration gave the same values. These data are synthetic by design, were generated by the script in the repository (aav_repair_suite.py) and are not evidence about AAV2 or about the biological model; the study tests the updating workflow only. Observations stored in earlier drafts (0.080, 0.118, 0.137) are about twice the model prediction under y = eta x F and are not used.

3.11 Which process variables drive function, and where the design fails

Study 7 propagated 30,000 joint draws of seven process and product variables to two endpoints. Variables that raise total vector genomes did not move function: vector-genome production rate (qVG) correlated with total vector genomes at r = 0.77 (Spearman 0.76) but with function at r ≈ 0.00, and transfection efficiency at r = 0.55 against r = −0.01. Conversely, full fraction (r = 0.48), η (r = 0.35) and capsid-quality factor (r = 0.25) drove function and not titre, and AEX recovery affected titre (r = 0.29) only (Fig. 5d,e). Pearson and Spearman coefficients agreed within 0.02 except for the A:B ratio, whose Pearson −0.15 versus Spearman −0.01 reflects its symmetric, non-monotonic penalty. Because the endpoint definitions encode this separation, the screen shows decoupling inherent in the model structure rather than discovering it; it is a correlation screen, not causal proof or a variance decomposition. Titre and function inputs were sampled independently, so zero association is built in; the scenario assumes a fixed administered dose, under which titre changes the number of doses and cost rather than delivered function. A yield-limited dose would couple them and is not modelled.

Study 8 partitioned the design space into four operational regimes using a 60% full-capsid threshold and a 0.75–1.33 A:B window (Fig. 5f). Of the 141 × 141 grid (log-spaced in A:B) 7.7% was reconstitution-limited, 10.1% quality-limited, 35.6% stoichiometry-limited and 46.7% compound-limited. The reconstitution-limited region, where both criteria are met, contained the highest function values (3.10–7.33%), whereas the maxima in the other regimes were 4.85% (quality-limited), 4.48% (stoichiometry-limited) and 2.96% (compound-limited). Because the grid is log-spaced the percentages describe design-grid area, not probability. The labels are operational definitions for quality-by-design visualisation, not demonstrated mechanisms. The eight studies are summarised in Table 5.

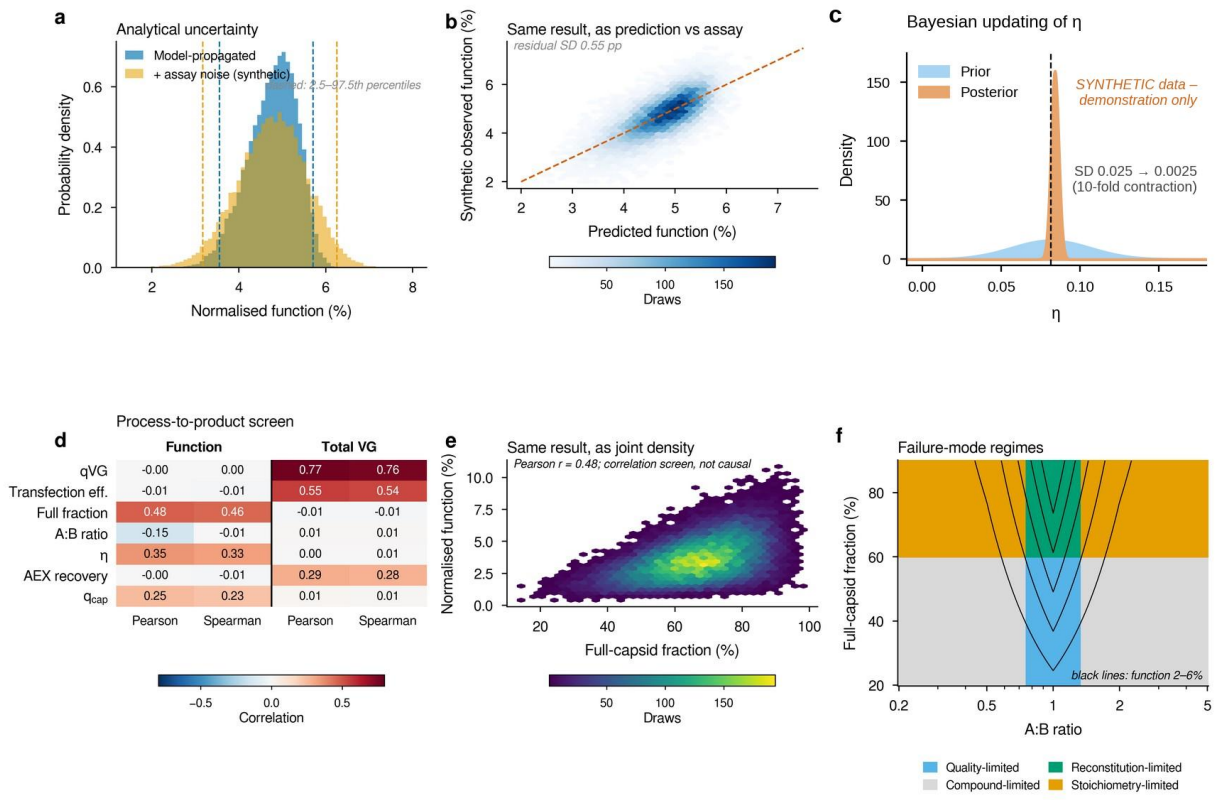


Figure 5: Measurement uncertainty, Bayesian updating, driver screening and failure-mode regimes (Studies 3, 6, 7 and 8). *a, Distribution of normalised function from model propagation and with synthetic assay noise (20,000 draws; dashed lines, 2.5–97.5th percentiles; Study 3). b, The same draws as predicted versus synthetic observed function (hexagonal bins; dashed line, identity). c, Prior and posterior for η using synthetic observations (Study 6; y = η_true·F + N(0, 0.003²), seed 20261026). d, Pearson and Spearman correlations of seven drivers with normalised function and with total vector genomes (Study 7; 30,000 draws). e, Joint density of full-capsid fraction and normalised function. f, Failure-mode regimes over full-capsid fraction and A:B ratio, with function contours (Study 8). Synthetic observations are demonstrations only; correlations are screening results and do not establish causation.*

The results of Studies **4, 5, 8 and 12** are shown together as response landscapes in **Fig. 6.** The dose-stoichiometry surface is a ridge along 1:1 that rises with dose but collapses on either side **(Fig. 6a).** The competition surface is close to planar, falling from 5.50% at α = β = 0 to 3.33% at α = β = 2 (Fig. 6b). In the failure-mode landscape **(Fig. 6c),** function peaks in the reconstitution-limited corner; the regime boundaries cut across a smooth function surface, so a regime change is a classification boundary and not a cliff in function. The tolerance surface from **Study 12 (Fig. 6d)** shows a plateau of near-complete compliance that falls away as ratio variability grows, more steeply at low full-capsid fractions.

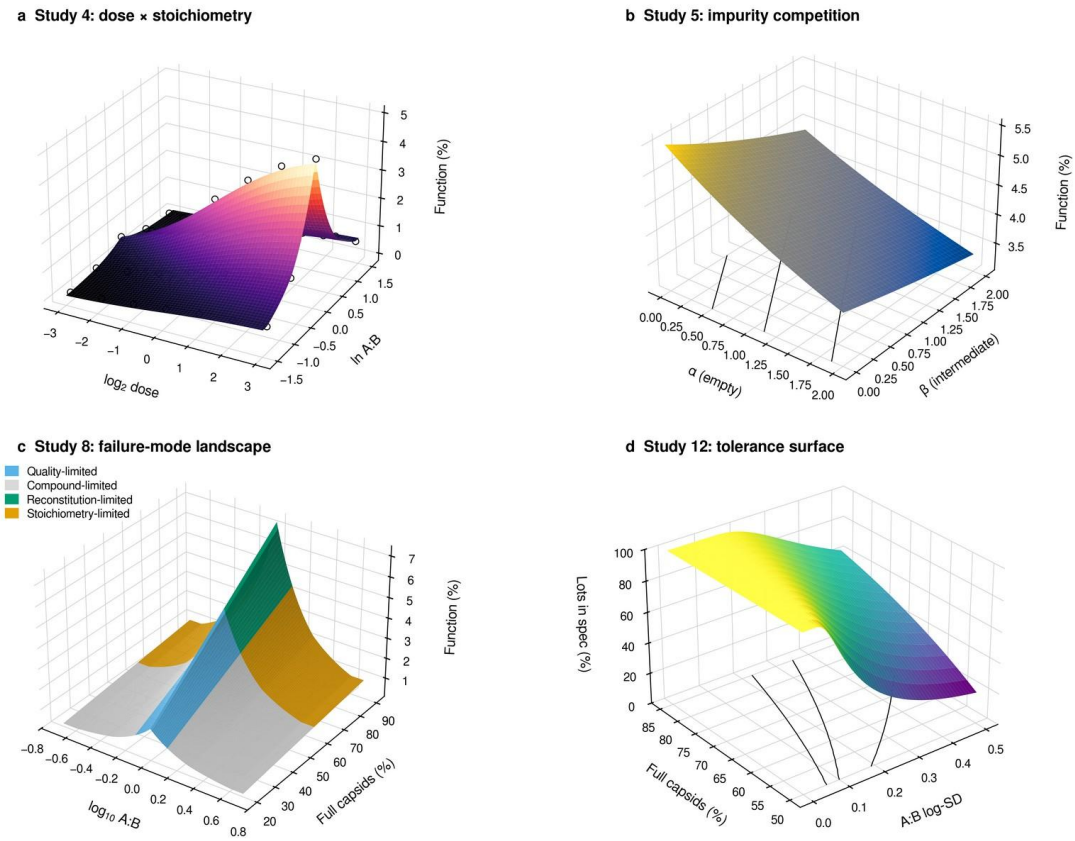


Figure 6: Response landscapes: *a, Absolute functional index over log₂ dose and ln A:B ratio (surface, surrogate; points, tabulated Study 4 values). b, Normalised function over the competition coefficients α and β, with contour projections at 4, 4.5 and 5%. c, Normalised function over full-capsid fraction and A:B ratio, coloured by operational regime. d, Share of lots meeting specification over A:B dosing-ratio variability and full-capsid fraction, with contour projections at 50, 80 and 95%.*

Table 5. Eight-study AAV2 dual-AAV computational extension

AAV2 dual-AAV extension study	Model element	Scenario / result	Evidence status
Study 1	Full + intermediate + empty quality-state model	F_full + F_intermediate + F_empty = 1; intermediate contributes VG but not potency in base model	Structural computational extension
Study 2	Capsid quality multiplier q_cap	Range 0.70–1.20	Sensitivity scenario

Study 3	Analytical uncertainty	VG ratio CV 8%; full ratio CV 7%; functional readout CV 10%	Representative uncertainty scenario
Study 4	Dose × A:B ratio	Matched-reference endpoint plus saturating D/(D+K_D), K_D=1 central-dose unit	Hypothesis/scenario parameter
Study 5	Impurity competition	D_eff = D_full/(1+αE+βI)	Sensitivity functional form
Study 6	Bayesian updating	η = 0.055/0.675 = 0.081481; synthetic observations for updating workflow	Synthetic demonstration
Study 7	Process-to-product propagation	qVG, transfection, full fraction, A:B ratio, η, AEX recovery, q_cap → endpoints	Correlation screen; not causal proof
Study 8	Failure-mode map	Four model-defined regimes based on full fraction and A:B ratio	Computational regime labels

Note: All eight studies are computational extensions and introduce no new experimental observations.

3.12 Variance-based sensitivity overturns the correlation screen (Study 9)

The correlation screen of Section 3.11 cannot register non-monotonic effects. We therefore repeated the propagation as a variance decomposition (Saltelli sampling, 16,384 base samples and 262,144 model evaluations per output, 300 bootstrap resamples; Fig. 7a,b). The analysis uses a surrogate, function = η·F_full·q_cap·h(A:B), that reproduces the 30,000 Study 7 outputs to within 3.4×10^{-5} (h is the ratio penalty tabulated in Studies 4 and 8), and resamples the empirical marginals of the seven inputs, which are treated as independent (largest pairwise correlation in the sampled inputs 0.014).

For functional output the A:B ratio carried the largest share of variance (first-order index S_1 = 0.52, 95% bootstrap interval 0.47–0.56; total-order S_T = 0.57), followed by full fraction (S_1 = 0.22), η (0.13) and q_cap (0.065); first-order indices summed to 0.93, so about 7% of variance arises from interactions. The correlation screen had scored the A:B ratio at $r^2 = 0.02$ ($r = -0.15$), fourth of the four functional drivers, because its penalty is symmetric about 1:1 and therefore invisible to linear and rank correlation. For total vector genomes, qVG (S_T = 0.60), transfection efficiency (0.31) and AEX recovery (0.09) accounted for all variance. No input had a total-order

index above 0.01 for both outputs (Fig. 7b), so titre and function respond to disjoint sets of process variables, and the variable that matters most for function is the one a correlation-based screen would have discarded. The result is conditional on the assumed ratio penalty and input distributions.

3.13 What to measure and how often

Calibration needs experiments that can actually separate the parameters. Study 10a asked how many independent lots must be assayed to learn η, the parameter linking full-capsid fraction to function. Starting from the prior (SD 31% of its mean), a functional assay with 10% CV and a 7% full-fraction measurement error per lot reduces the posterior SD to 10% of the mean with 2 lots and to 5% with 6 lots; with a 20% assay CV the numbers rise to 5 and 18 lots **(Fig. 7c).** These are lower bounds that assume independent lots, unbiased assays and a correct model form.

Study 10b asked whether a dose series can separate η from the half-saturation constant K_D. With three replicates per dose and 10% CV, a three-dose design spanning K_D (0.125, 1 and 8 central units) achieved SDs of 6.2% for η and 10.5% for K_D, and seven doses 4.5% and 7.9%. Designs that missed the transition did badly: high doses only gave 14% and 91%, low doses only 29% and 37%, and a design bracketing K_D tightly (0.5–2) gave 13% and 25% (Fig. 7d). Every design left η and K_D positively correlated (0.84–0.99), so the dose range, not the number of replicates, determines identifiability.

Study 14 applied the same Fisher-information logic to the competition coefficients α and β. A lot set containing only empty-plus-full particles estimates α (SD 0.048) but carries no information on β, and the reverse holds for intermediate-plus-full lots (SD of β 0.066). Eighteen lots in which empty and intermediate fractions co-vary ($|E - I| \leq 0.05$) gave SDs of 0.68 for both coefficients with a correlation of −0.99 (condition number 184), whereas the full 105-lot composition grid gave SDs of 0.032 and 0.036 with a correlation of −0.44 (condition number 2.6; Fig. 7e,f). Separating impurity effects therefore requires lots that vary empty and intermediate fractions independently. These figures assume the competition form of Study 5, true coefficients of 1, three replicates per lot and 10% CV; they bound what a real campaign could achieve.

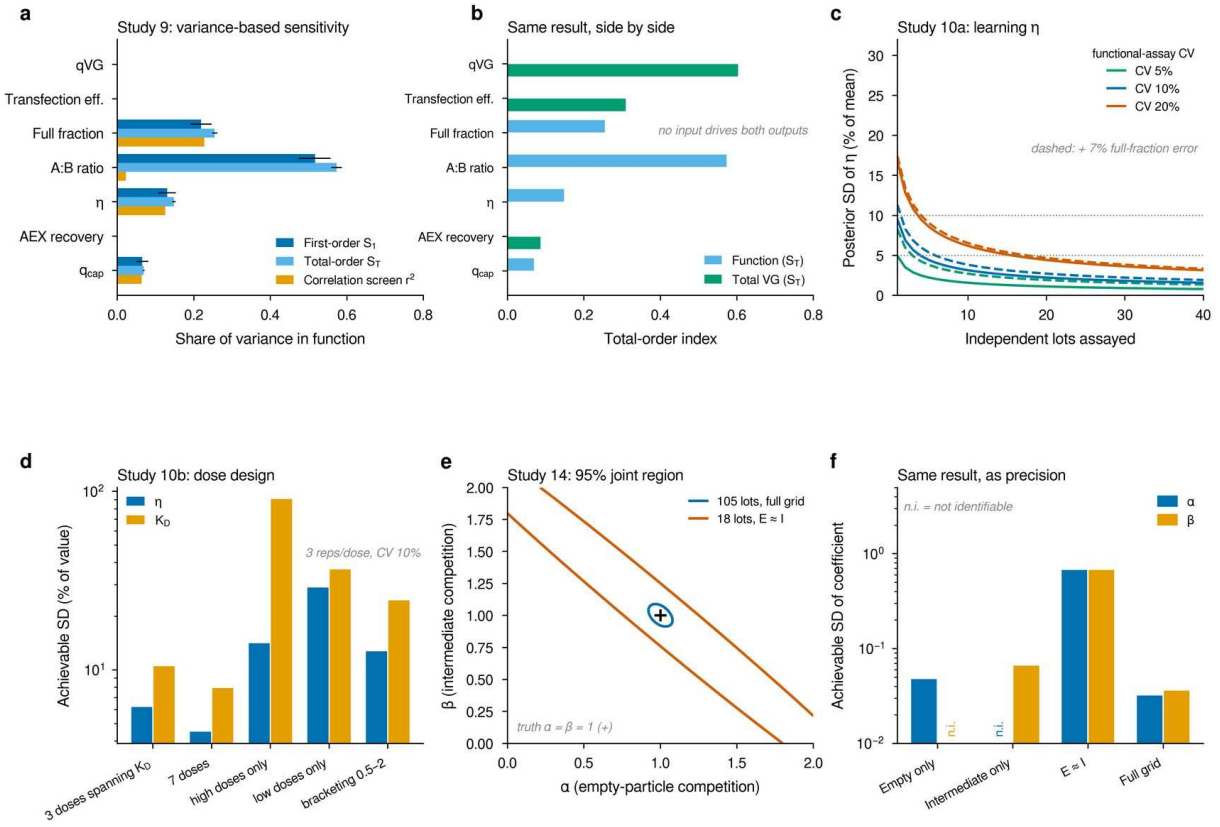


Figure 7: Variance-based sensitivity and measurement design *a, First-order (S_I) and total-order (S_T) Sobol indices for normalised function (95% bootstrap intervals) with the squared correlation of the earlier screen. b, Total-order indices for function and for total vector genomes (VG) by input. c, Posterior SD of η versus independent lots assayed at three functional-assay CVs (dashed lines add a 7% full-fraction error). d, Achievable SD of η and K_D for five dose designs (3 replicates per dose, 10% CV). e, 95% joint regions for the competition coefficients α and β from 105 lots across the composition grid and from 18 lots with co-varying empty and intermediate fractions (true α = β = 1). f, Achievable SD of α and β by lot design (n.i., not identifiable). All are model-based design bounds.*

3.14 Lot release: titre cannot stand in for function

We next asked which measurable quantity could support a release decision. A lot was defined as good if its function was at least 60% of the central scenario (3.30%), which 54.6% of the 30,000 simulated lots met. Each rule scored every lot with measurement noise, and we accepted the same share of lots under every rule (54.6%, the share of lots meeting specification; 16,372 of 30,000) so that consumer risk, the share of accepted lots below specification, can be compared directly (Fig. 8a,b). Titre alone (8% CV) performed no better than chance (AUC 0.49; consumer risk 45.8%). Full fraction alone (7% CV) reached AUC 0.70 (32.2%), and multiplying titre by full fraction was worse than full fraction alone (AUC 0.65; 35.9%) because titre adds noise unrelated to function. A functional assay with 10% CV reached AUC 0.98 (6.4%) with one replicate and 0.99 (3.7%) with three. Bringing consumer risk below 5% required at least two replicates (6.4%, 4.6%, 3.7%, 2.7% and 1.9% for 1, 2, 3, 6 and 12 replicates; means over 50 noise seeds, SD 0.06–0.12 percentage points; Fig. 8d).

Study 13 shows the same decoupling at the lot level (Fig. 8c). Titre and function were uncorrelated (Spearman $\rho = -0.008$). Among lots in the top titre quartile, 45.8% fell below specification, against 45.4% overall, and 25.5% sat in the bottom function quartile, as expected by chance. A high-titre lot is thus no safer than any other lot. Both studies are upper bounds in one respect: the functional assay here measures the model endpoint exactly apart from its 10% noise, whereas a real potency assay adds bias and its own model error.

3.15 Stoichiometry control requirements

Because the ratio penalty carries most of the functional variance **(Section 3.12),** we asked how tightly the A:B ratio must be controlled. With log-normal ratio variability of a given log-SD, a lot meets specification when its function is at least 3.30%. At 67.5% full capsids and central η and q_cap, at least 95% of lots meet specification only if the log-SD of the dosing ratio is no more than 0.15 (about 15%), and at least 80% if it is no more than 0.24 **(Fig. 8e,f).** Higher-quality material buys tolerance: the 95% limit is 0.06 at 50% full capsids and 0.22 at 85%, a 3.7-fold relaxation. The window in which the ratio penalty stays at or above 0.8 is 0.87–1.14. At a log-SD of 0.10 the mean function at the central composition is 11.9% below its 1:1 value. All limits follow from the tabulated penalty, which is a hypothesis **(Study 4),** and would shift with a measured curve; the study quantifies the consequence of the assumption, not a validated control limit.

Table 6: summarises the six in-depth studies and their evidence status.

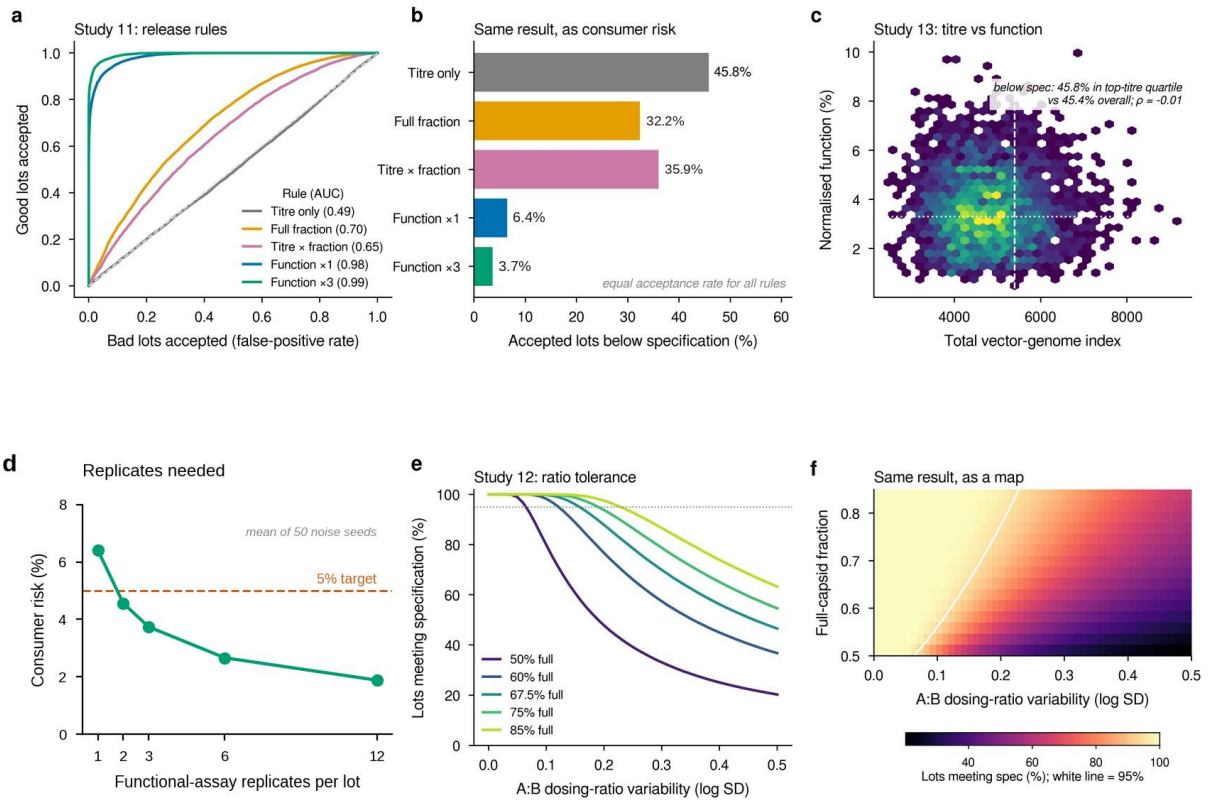


Figure 8: Lot-release rules and stoichiometry tolerance *a, Receiver operating characteristic curves for five release rules (AUC in legend). b, Share of accepted lots below specification at equal acceptance (54.6% of lots, the share meeting specification); mean of 50 measurement-noise seeds. c, Total vector-genome index versus function (hexagonal bins; dashed lines, top-titre quartile and specification). d, Consumer risk versus functional-assay replicates (mean of 50 measurement-noise seeds). e, Share of lots meeting specification versus A:B dosing-ratio variability at five full-capsid fractions. f, The same result as a map over full fraction and ratio variability (white line, 95%). Specification: function ≥ 60% of the central scenario (3.30%).*

Figure 9 integrates the decision-level results. A Sankey diagram of total-order indices shows the two disjoint driver sets for function and titre (**Fig. 9a**). A slope chart shows the A:B ratio moving from fourth place under the correlation screen to first under variance decomposition (**Fig. 9b**). A radar chart compares the five release rules on four performance measures (**Fig. 9c**). Ridgeline densities show how the distribution of function shifts and narrows as the A:B ratio departs from 1:1, with the median in specification only for ratios between 0.73 and 1.48 (Fig. 9d).

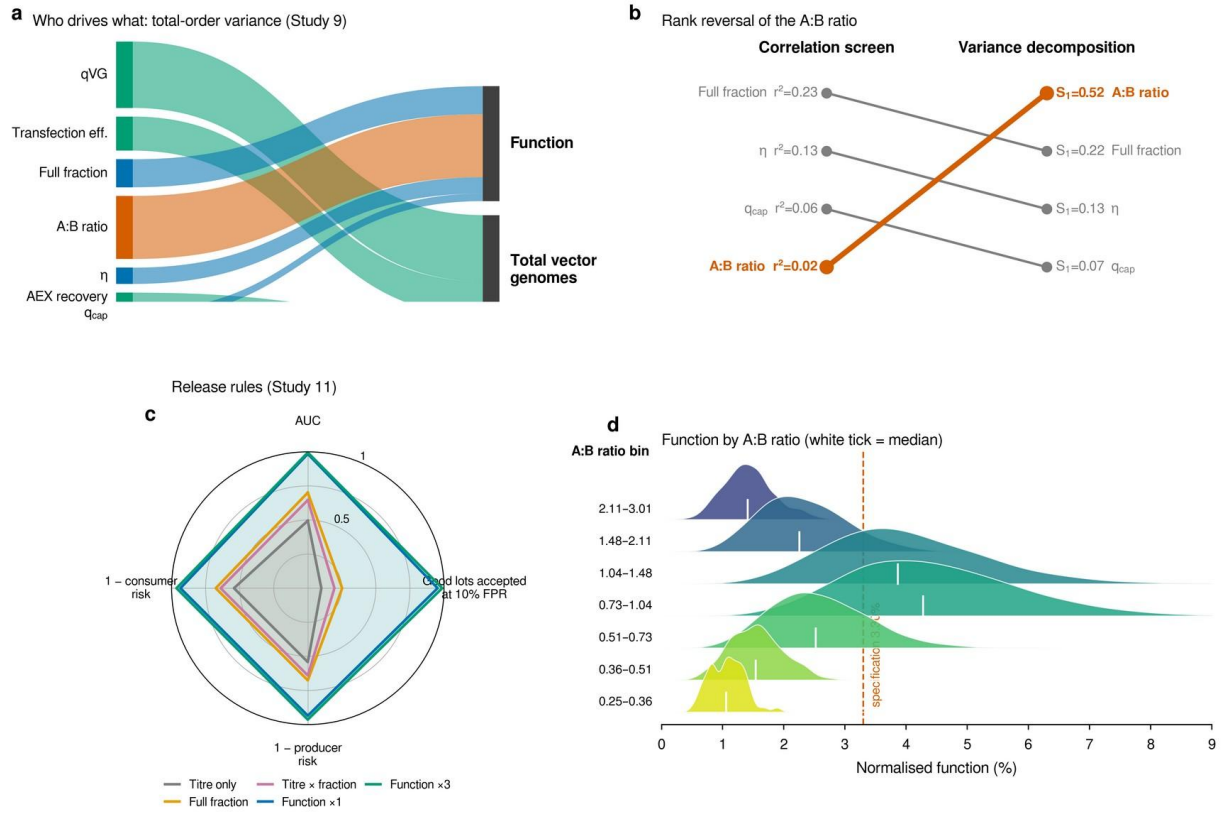


Figure 9: Drivers and decisions at a glance: *a, Sankey diagram of total-order Sobol indices from seven inputs to function and to total vector genomes (bands below 0.01 omitted). b, Rank of each functional driver under the correlation screen (r^2) and under first-order variance share (S_1); the A:B ratio is highlighted. c, Radar chart of five release rules on AUC, good lots accepted at a 10% false-positive rate, 1 − producer risk and 1 − consumer risk. d, Ridgeline densities of function for lots grouped by A:B ratio (white tick, median; dashed line, specification 3.30%).*

Table 6: In-depth studies 9-14 on the verified surrogate

Study	Question	Key result	Evidence status
9	Which inputs carry variance?	A:B ratio S_1 0.52 (r^2 0.02); full fraction 0.22; η 0.13; q_cap 0.065; titre and function inputs disjoint	Computational; ratio penalty hypothesised
10a	Lots needed to learn η	2 lots for ≤10%, 6 lots for ≤5% (10% assay CV, 7% F error)	Idealised lower bound
10b	Can a dose series separate η and K_D?	Spanning design 6.2%/10.5%; high-dose-only 14%/91%	Fisher bound; K_D hypothetical
11	Which release rule works?	AUC 0.49 titre, 0.70 full fraction, 0.98 functional assay	Upper bound; assay is exact endpoint plus noise
12	How tight must A:B control be?	log-SD ≤0.15 for 95% of lots at 67.5% full	Conditional on ratio penalty
13	Does high titre mean good function?	ρ = −0.008; 45.8% vs 45.4% below spec	Structural decoupling
14	Can α and β be separated?	Full grid SD 0.03; E ≈ I lots SD 0.68, r = −0.99	Fisher bound; hypothetical form

Note: All studies are computational scenarios built on the surrogate function = η·F_full·q_cap·h(A:B); no experimental data were generated.

3.16 Gene-specific scenarios and allocation of uncertainty

Chaining production, reconstitution and correction (Y_end = Y_prod × E_reconstitution × E_correction) for three disease contexts gave end-to-end Monte Carlo means of 6.7% (central 95% interval 5.5–8.1%) for LRRK2 G2019S, 59.9% (50.2–70.3%) for SCN1A R613X and 10.7% (5.9–15.5%) for RPE65 R44X, with threshold exceedance of 0%, 54.2% and 17.2% (Fig. 10a,b; Table 7). These scenarios should not be interpreted as a ranking of biological efficacy among diseases. LRRK2 was lowest because the model had to bridge from an iPSC correction study (24.5% corrected clones [13]) to an AAV delivery context, and the quantities are not equivalent; the SCN1A value lies near the 59% bulk-neocortex DNA editing and 97% mRNA editing reported for

dual-AAV9 delivery [14], but the experimental endpoint is tissue-specific and the correspondence is contextual, not a replication. The RPE65 scenario is an external anchor to functional rescue in rd12 mice [15], and is not a re-analysis of that dataset.

Variance decomposition of the SCN1A scenario attributed about 70% of output variance to production-side terms and about 30% to downstream reconstitution and correction (**Fig. 10c**). Under the adopted ranges, reducing upstream uncertainty would therefore narrow endpoint uncertainty more than refining downstream parameters alone. This is a prioritisation hypothesis, conditional on assumed distributions, not a claim about the biological hierarchy of mechanisms.

Table 7: Gene-specific Monte Carlo uncertainty propagation

Gene / scenario	Modeled mean corrected fraction	Central 95% computational interval	Threshold exceedance	External evidence context
LRRK2 G2019S	6.7%	5.5–8.1%	0%	Human iPSC correction study; not direct dual-AAV production context
SCN1A R613X	59.9%	50.2–70.3%	54.2%	Recent dual-AAV9 ABE disease-model context
RPE65 R44X	10.7%	5.9–15.5%	17.2%	Dual-AAV split-ABE retinal disease- model context

Note: Scenario distributions are conditional on adopted modeling/literature-informed inputs and are not empirical efficacy probabilities or clinical-success probabilities.

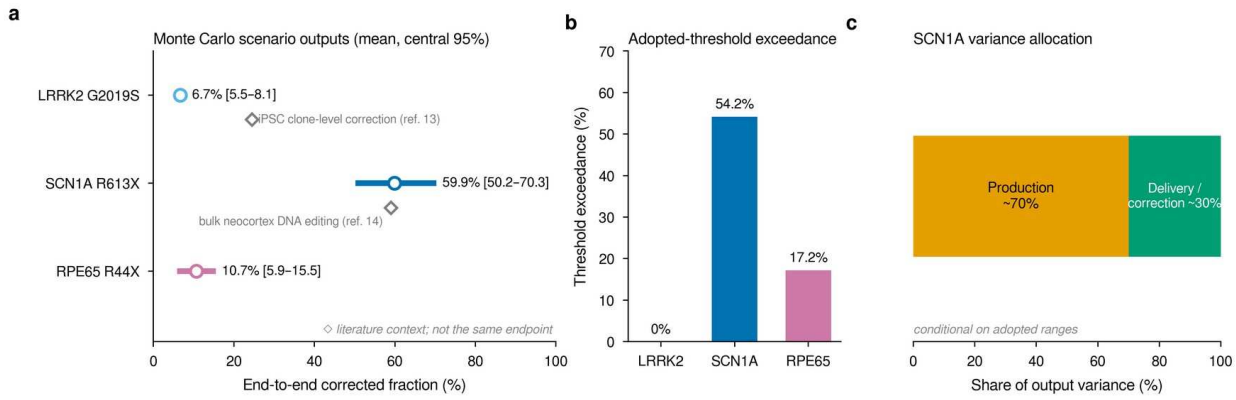


Figure 10: End-to-end scenario uncertainty and variance allocation. *a, Mean and central 95% computational interval of 20,000 Monte Carlo draws per scenario; grey diamonds show external literature context that is not the same endpoint (iPSC clone-level correction, ref. 13; bulk neocortex DNA editing, ref. 14). b, Fraction of draws exceeding the adopted threshold. c, Allocation of SCN1A output variance to production (~70%) and downstream (~30%) terms.*

Intervals are conditional on adopted distributions and are not efficacy or clinical-success probabilities.

3.17 Synthetic method tests act as gates

The synthetic Bayesian analysis produced R-hat values of approximately 1.21–1.58 and integrated autocorrelation times of approximately 26–33 iterations under the tested sampling budget. The chains therefore failed the intended convergence criterion. This failure was retained rather than removed from the manuscript because it identifies a genuine computational limitation. Increasing chain length, improving initialization, using the reduced parameterization, and adopting more informative observation models are potential remedies for future work.

The result also demonstrates why synthetic-data testing should precede biological calibration. In synthetic experiments, the true parameter vector is known. If the inference algorithm cannot reliably recover that vector, applying it to real AAV data would make it difficult to distinguish biological non-identifiability from algorithmic failure. The negative result therefore functions as a quality-control gate for future calibration.

UKF state estimates based on synthetic X and G observations produced RMSE values equivalent to approximately 4.8–5.3% of the peak synthetic truth. The particle-filter comparison produced RMSE values of 1.68 versus 2.31 for the tested alternatives, depending on the state/error aggregation used in the computational audit. NIS and NEES diagnostics indicated uncertainty miscalibration: approximately 38.3% of NIS values and 63.3% of NEES values fell within the nominal 95% consistency bounds, below the expected 95% coverage target.

The PI control simulation reduced the modeled glucose-variability amplitude by approximately 16.1%. This is a software-level demonstration that a state-estimation/control loop can be connected to the process model. It is not evidence that the same controller would stabilize a physical AAV bioreactor. Real deployment would require sensor latency, actuator limits, process disturbances, controller tuning, and experimental safety constraints that are not represented here.

Table 8. Synthetic Bayesian, state-estimation, and control diagnostics

Method test	Input / configuration	Quantitative result	Status
Synthetic Bayesian calibration	4 MCMC chains; synthetic observations	R-hat ≈1.21–1.58; IACT ≈26–33	Non-converged computational diagnostic retained
UKF synthetic state estimation	Synthetic X/G observations	RMSE ≈4.8–5.3% of peak synthetic truth	Method test only
Particle-filter comparison	Synthetic non-Gaussian state estimation	RMSE 1.68 vs 2.31 for tested alternatives	Method test only
NIS consistency	Synthetic filtering	38.3% inside nominal 95% bounds	Uncertainty miscalibration
NEES consistency	Synthetic filtering	63.3% inside nominal 95% bounds	Uncertainty miscalibration

PI control	Synthetic glucose control	~16.1% reduction in modeled glucose-variability amplitude	Software-level demonstration only

Note: Synthetic truth is known by construction. These analyses test computational machinery and do not validate the biological model or a physical bioreactor.

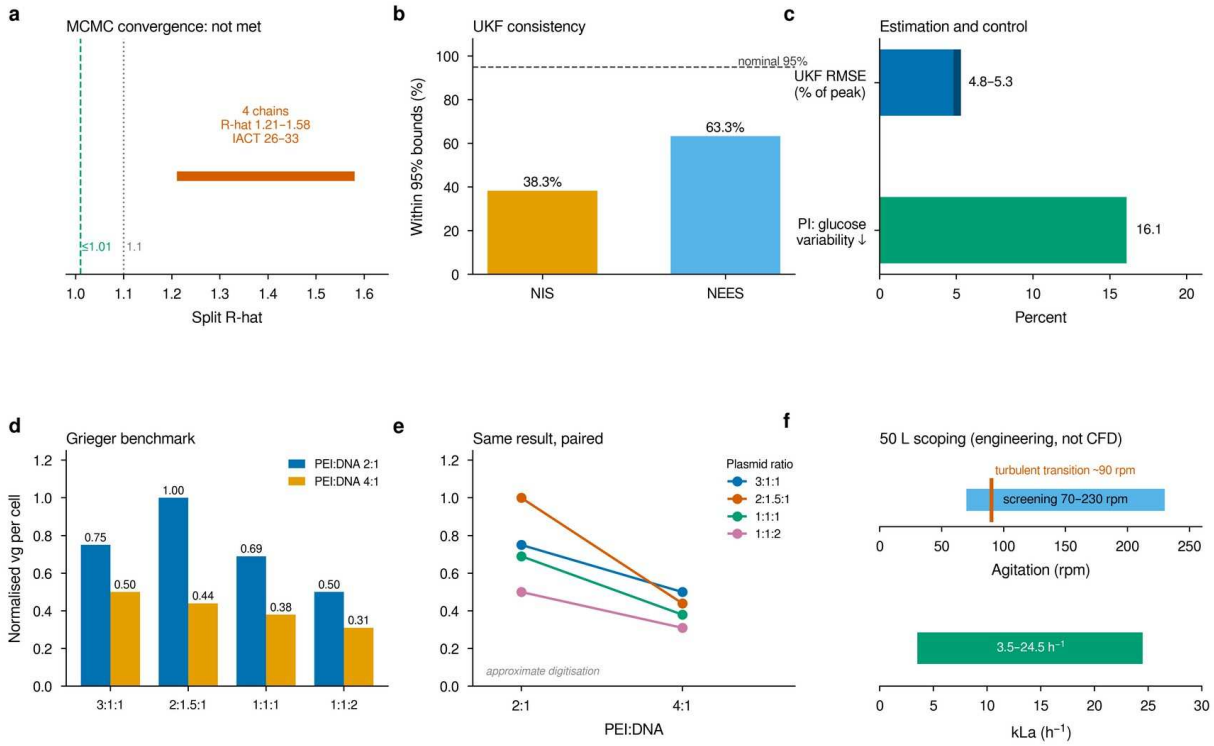


Figure 11: Synthetic method gates, external benchmark and engineering scope. *a, Split R-hat of four MCMC chains on synthetic observations (1.21–1.58; integrated autocorrelation times 26-33) against thresholds of 1.01 and 1.1. b, Fraction of NIS and NEES values within nominal 95% bounds for the unscented Kalman filter. c, UKF RMSE as a percentage of peak synthetic truth and simulated glucose-variability reduction under PI control. d, Approximate digitised benchmark values (vector genomes per cell, normalised to 80,000) for four plasmid ratios at PEI:DNA 2:1 and 4:1 (Grieger et al., ref. 4). e, The same values as paired change. f, 50 L agitation screening window with turbulent-transition estimate (top) and estimated kLa range (bottom); engineering scoping, not CFD. Panels a–c are software-level tests on synthetic data; d is a graphical digitisation, not raw data.*

Table 9. External benchmark, weak-direction, and engineering-scope checks

Audit / scenario	Key quantitative result	Manuscript interpretation
Forward 46-h consistency audit	Total DNA −14.3%; full virion +44.7%; total capsid −6.3% (vs replicate means); mean absolute discrepancy 21.8%	Numerically reproduced source model retains a nonzero observational gap
Weak direction profiles	Selected local weak directions show shallow SSR changes; regenerated profiles: SSR change <0.001 (direction 8), 0.004 (direction 7), 0.005 (direction 6) at a step of 0.1 log10 units, baseline 0.668084 (profiles regenerated from the frozen 8D fit)	Sparse aggregate observations weakly constrain selected microscopic combinations
Grieger external benchmark	Approximate digitization preserves published 2:1.5:1 > other tested ratios at PEI:DNA 2:1	Literature plausibility benchmark, not validation
Hydrodynamic scoping	Turbulent transition ~90 RPM; kLa ~3.5–24.5 h−1; screening window ~70– 230 RPM	Engineering scoping range under generic 50-L assumptions

Note: The Grieger values are approximate graphical digitizations. Hydrodynamic values are generic engineering scoping calculations, not reactor specifications.

3.19 Results at a glance

Figure 12 condenses the study into six headline results. The independent reconstruction agreed across three stiff solvers to better than 0.14%, but a 21.8% mean absolute gap to the source observations remained at 46 h. Only 12 of 14 parameter directions were locally identifiable (relative singular-value threshold 10^{-12}), and constrained reduction lowered the condition number from 6.4×10^{21} to 6.0×10^{3}. In the scenario studies the A:B ratio carried 52% of functional variance, which the correlation screen had scored at 2%. A titre-only release rule performed at chance (AUC 0.49) against 0.98 for a one-replicate functional assay, and 95% of lots stay in specification only if the dosing-ratio log-SD is at most 0.15 at 67.5% full capsids. Together they support one design conclusion: titre and function are governed by different variables, so a dual-AAV twin must measure capsid state, pair stoichiometry and a functional readout.



Figure 12: Results at a glance. *(A) Maximum relative difference among BDF, Radau and LSODA solver trajectories. (B) Mean absolute discrepancy between the source-fixed model trajectory and the source replicate means at 46 h. (C) Locally identifiable parameter directions (12 of 14 at a relative singular-value threshold of 10^{-12}; 10 of 14 at 10^{-6}) and condition numbers (6.4×10^{21} for the full model, 6.0×10^{3} for the 8D representation). (D) Variance-based first-order Sobol index for the A:B ratio compared with its squared correlation from the screening analysis. (E) Area under the curve of titre-only versus one-replicate functional-assay release rules. (F) Share of lots in specification versus A:B dosing-ratio log-SD at 67.5% full capsids (95% at log-SD 0.15). All panels are computational results; the solver comparison is a numerical reproducibility check and is not predictive validation.*

4. Discussion

4.1 Principal contribution: uncertainty provenance rather than a single predictive score

The main contribution of this study is the explicit separation of uncertainty across production, vector pairing/reconstitution, and gene-specific correction. In a conventional endpoint-only analysis, a low editing fraction can be interpreted as poor vector production, inadequate co-transduction, inefficient molecular reconstitution, weak editing chemistry, unfavorable target sequence context, or combinations of these factors. A digital twin becomes useful when it can preserve these causal alternatives and quantify how strongly each contributes under a stated model.

The approximately 70:30 production-to-downstream variance split in the SCN1A scenario illustrates the value of this decomposition. Under the adopted ranges, production uncertainty dominated. If that pattern persisted after experimental calibration, process-development resources could be directed toward improving upstream measurement and control before investing disproportionate effort in downstream model refinement. However, because the current result is conditional on assumed distributions, it should be treated as a prioritization hypothesis rather than a biological conclusion.

4.2 Titre and function separate through product quality

The eight-study extension adds a practical message to the identifiability result. In the model, process levers that raise vector-genome titre (qVG, transfection efficiency, AEX recovery) are uncorrelated with the functional endpoint, while the levers that move function (full fraction, η and capsid quality) do not change titre (Section 3.11). Because the endpoint definitions build this separation in, we present it as a design principle rather than a finding: a digital twin that reports only vector genomes would rank process changes incorrectly for dual-AAV products. Downstream

steps matter in the same way, since anion-exchange recovery changes the quantity of product but not its quality in this model, in line with analyses of downstream process economics [18]. Studies 4 and 5 add that stoichiometric imbalance and non-functional particles can erase the benefit of higher dose, so that doubling dose is a poor substitute for correcting the ratio. Together these results define what a calibrated twin must measure: capsid state, pair stoichiometry and an orthogonal functional readout, in addition to titre.

4.3 From scenario analysis to decisions

The in-depth studies show where the scenario framework has direct consequences for practice. First, screening choices matter: a correlation screen ranked the A:B ratio last among functional drivers ($r^2 = 0.02$), whereas a variance decomposition ranked it first ($S_1 = 0.52$), so non-monotonic factors such as stoichiometry need variance-based or design-of-experiments methods. Second, release criteria should be built on function. Titre-based rules classified functional quality at chance level and multiplying titre by full fraction degraded the better of the two, while two replicates of a functional assay brought consumer risk below 5%. Third, calibration should be designed for identifiability: dose series must straddle the transition to separate potency from saturation, and impurity effects can be separated only with lots that vary empty and intermediate fractions independently, so that replicating similar lots adds little. Fourth, the dosing ratio must be controlled more tightly than most assays resolve (log-SD ≤ 0.15 for 95% of lots at 67.5% full capsids), although higher-quality material relaxes this limit. These conclusions are conditional on the assumed ratio penalty, competition form and noise levels, and each is a hypothesis to test experimentally, not a validated specification.

4.4 Relationship to the AAV manufacturing literature

The framework is consistent with an increasingly multivariable view of AAV manufacturing. Grieger et al. demonstrated that suspension-cell production and continuous harvest can support scalable vector manufacture [4]. Park et al. showed that plasmid-ratio optimization can simultaneously affect genome titer and full-capsid fraction [6]. Coplan et al. used multivariate DoE to identify transfection pH, production pH, complexation time, viable cell density, and transfection reagent-to-DNA ratio as influential variables [7]. Deng et al. extended process intensification through high-cell-density fed-perfusion and systematic optimization of transfection variables [8]. Johari et al. demonstrated that molecular control of Rep/Cap and helper expression can alter both vector titer and product quality [17], while Srinivasan et al. used mechanistic analysis to link Rep expression dynamics to filled-capsid enrichment [16].

These studies suggest that future digital twins should not treat the simplified k_transf, k_cap, and k_pack terms as universal biological constants. Instead, they should eventually become functions of measurable process variables such as cell density, transfection reagent ratio, plasmid composition, pH, complexation time, perfusion state, and Rep/Cap expression. The current model is therefore best understood as a structural scaffold onto which richer empirical dependencies can be attached.

4.5 Full/empty capsids should be modeled as coupled quantity-quality states

A single vector-genome titer cannot fully describe an AAV harvest. The source model itself illustrates why: a large capsid population can coexist with a relatively small full-particle

population. Analytical studies reinforce the importance of measuring full/empty composition with orthogonal methods and of accounting for interlaboratory variability [9,10]. In a future digital twin, the observation model should explicitly represent assay-specific uncertainty for qPCR/dPCR genome measurements, capsid assays, analytical ultracentrifugation, chromatography, mass spectrometry, or other methods. This would allow the twin to infer whether an apparent change is more likely biological or analytical.

4.6 Identifiability is a design constraint for future experiments

The 14D condition number of approximately 6.4×10^{21} is one of the most important findings because it places a hard limit on what can be learned from the current aggregate measurements. Adding more observations of the same observable at nearly identical temporal resolution may improve numerical precision without resolving the underlying parameter correlations. The more informative strategy is to add orthogonal observables that map preferentially to currently weak directions.

Candidate measurements include Rep abundance, VP abundance, intracellular plasmid states, viral DNA replication intermediates, nuclear versus cytoplasmic capsid pools, packaging intermediates, and higher-frequency sampling around the media-exchange transition. Such measurements would alter the sensitivity matrix itself rather than merely adding repeated points to an existing trajectory. The digital twin can therefore be used prospectively to choose measurements that are expected to collapse the weakest singular directions.

4.7 Why the external Grieger benchmark is useful but limited

The Grieger benchmark adds an independent experimental context without pretending to be a validation experiment. The direction of the plasmid-ratio effect is consistent with the broader literature showing that transfection composition is a major determinant of AAV yield. The benchmark also demonstrates a practical issue in computational reproducibility: published figures often contain information that cannot be recovered as exact machine-readable raw data. Approximate digitization can be useful for contextual analysis, but it must retain its uncertainty and provenance. For this reason, the present manuscript does not attach experimental error bars to the digitized values and does not label the benchmark as model validation.

4.8 Dual-AAV payload architecture creates a coupled manufacturing-delivery problem

Dual-AAV delivery can overcome the size constraint but makes manufacturing quality relevant to downstream reconstitution. Ferreira et al. demonstrated that vector preparation quality affected split-protein reconstitution and that more efficient split-inteins could improve reconstitution at lower vector doses [11]. This creates a direct conceptual link between the upstream full/empty state and downstream editing performance. A digital twin that reports only vector genomes would miss this coupling.

At the same time, compact single-AAV editors can alter the optimization problem. Davis et al. showed that size-optimized compact ABEs can achieve efficient in vivo editing in a single vector in selected settings [3]. Thus, the digital twin should eventually support architecture switching: single-AAV, dual-AAV split-intein, or other delivery schemes should be represented as alternative downstream constraints on a common production engine rather than as unrelated models.

4.9 Disease-specific evidence should remain heterogeneous

The three gene scenarios should not be homogenized. LRRK2 has human-cell correction evidence but not direct equivalence to an in vivo dual-AAV process. SCN1A has the closest direct dual-AAV9 ABE context. RPE65 has a dual-AAV retinal disease-model context with molecular and functional rescue. Treating all three as equally validated would create false symmetry. The manuscript instead uses evidence provenance as part of the model specification.

4.10 Negative computational results strengthen reproducibility

The Bayesian convergence failure and filter uncertainty miscalibration are scientifically useful because they identify parts of the computational pipeline that are not ready for biological calibration. In many computational studies, only successful outputs are reported, which makes it difficult for later users to distinguish a robust method from a method that simply was not stress-tested. Here, failed diagnostics are retained as explicit gates. The same principle should apply to future model versions: if a calibration or control layer fails synthetic validation, it should not be promoted to experimental use merely because the forward model appears plausible.

4.11 Implications for a future experimentally calibrated digital twin

The present framework suggests a sequential calibration strategy. First, production should be calibrated using time-resolved cell density, glucose, viral DNA, capsid, and full/empty measurements. Second, dual-vector co-delivery should be characterized using single-cell or cell-level co-transduction measurements. Third, split-editor reconstitution should be quantified independently of target correction. Fourth, gene-specific correction should be measured across target sequence contexts and editor doses. Finally, the calibrated modules should be integrated and tested prospectively on conditions not used for fitting.

This staged strategy is preferable to fitting an end-to-end endpoint in one step because it preserves parameter meaning and makes failures diagnosable. It also creates a natural path toward model predictive control: process measurements update latent manufacturing states, the state estimator updates the digital twin, and the controller proposes process adjustments subject to experimentally defined constraints.

4.12 Experimental priorities generated by the computational analysis

The sensitivity and identifiability results suggest several high-value experimental measurements. For production, candidate perturbations include plasmid ratios, DNA concentration, transfection reagent-to-DNA ratio, cell density at transfection, pH, media composition, perfusion rate, and harvest timing. For vector quality, orthogonal full/empty measurements should be collected because assay noise can otherwise be confounded with biological variation. For the dual-vector stage, paired-vector co-transduction should be quantified at the single-cell level rather than inferred solely from bulk vector dose. For editing, both molecular correction and functional phenotype should be measured where possible.

The most informative experiments are not necessarily those with the largest expected effect on titer. A perturbation that changes a weakly identifiable direction may provide more information about the model than a perturbation that produces a large but redundant response. The digital twin can therefore become an experimental-design tool, not merely a prediction engine.

Comparison with the supplied Biotechnology and Bioengineering reference article. Qin et al. (2026) evaluate a hybrid mechanistic model using ten 7-L mammalian-cell culture runs, five training runs for the lactate back-metabolism scenario, three external runs for predictive evaluation, and one lactate run-away run; they report a mean R^2 of 0.88 across viable cell density, mAb, glucose, and lactate, and describe real-time fine-tuning that reduced glucose RMSE by about 50% and lactate RMSE by about 30%. Their design therefore includes an explicit training/evaluation split and quantified prediction metrics against process measurements. The present AAV study is different in evidence maturity: it independently reconstructs a published model and audits source-model discrepancy, then uses sensitivity and scenario analysis to explore downstream functional consequences. The <0.14% solver difference is a numerical reproducibility check, not a predictive-accuracy metric; the 21.8% mean absolute discrepancy at 46 h is the relevant source-observation audit. Because independent experimental AAV runs are not available here, the manuscript does not claim external predictive validation comparable to Qin et al. Instead, the comparison identifies a clear next validation step: predefine training and holdout AAV production runs, calibrate using training data only, report per-output RMSE and R^2 with uncertainty intervals, and evaluate unchanged parameters on held-out runs before using the digital twin for process recommendations.

4.13 Limitations of the study

First, this is an entirely computational study. No new cell-culture, bioreactor, animal, or patient experiments were performed. Second, the simplified Stage 1 constants are illustrative computational inputs rather than newly estimated biological parameters. Third, the source-faithful model was reconstructed from published implementation and observations, and exact MATLAB ODE15s identity could not be certified. Fourth, the identifiability analysis is local and data-dependent; it does not establish global identifiability. Fifth, the five-dimensional SVD core represents mathematical parameter combinations rather than five directly measured biological rates.

Sixth, Monte Carlo intervals and threshold probabilities are conditional on assumed distributions. They are not clinical probabilities. Seventh, literature-derived gene-specific values are heterogeneous and cannot be interpreted as direct cross-gene comparisons. Eighth, the Grieger benchmark is based on approximate digitization rather than raw numerical data. Ninth, Bayesian, UKF, particle-filter, and PI-control analyses use synthetic data and therefore test software methodology rather than biological model validity. Tenth, hydrodynamic estimates use generic geometry and first-order correlations rather than reactor-specific CFD or experiments.

Finally, the current framework does not explicitly represent all clinically relevant factors, including capsid immunogenicity, pre-existing antibodies, tissue distribution, receptor abundance, intracellular trafficking, innate immune activation, off-target editing, bystander editing, or dose-dependent toxicity. These omissions are intentional at this stage because the objective was to establish a reproducible manufacturing-to-editing computational architecture before expanding into higher-dimensional biological detail.

5. Conclusion

We developed a detailed computational digital twin framework that links AAV manufacturing, dual-vector reconstitution, and gene-specific base-editing outcomes while explicitly preserving

uncertainty provenance and evidence boundaries. The independent reconstruction of a published 21-state mechanistic AAV production model demonstrated strong numerical stability across three stiff solvers, but the same reconstruction revealed that the complete 14-parameter space is poorly conditioned and cannot be treated as uniformly identifiable. Source-constrained reparameterization reduced the local condition number by many orders of magnitude, and a leading five-dimensional singular core retained the observable fit while providing a substantially better-conditioned representation. The broader computational analyses demonstrate how the framework can be used to interrogate design spaces, perform mechanistic ablation, allocate uncertainty, propagate literature-informed scenarios, test inference algorithms, and identify candidate experimental priorities. The approximately 70:30 production-to-downstream variance split in the SCN1A scenario illustrates how an end-to-end endpoint can be decomposed into experimentally actionable uncertainty sources. The upgraded external literature benchmark further demonstrates how independent published process data can be incorporated without being mislabeled as validation. The appropriate interpretation of this work is therefore as a reproducible computational scaffold and hypothesis-generation platform. It is not an experimentally calibrated digital twin, a clinical efficacy predictor, or a substitute for bioreactor and biological validation. Its principal value is to establish the architecture, diagnostics, and evidence discipline required for those future stages. The 14-study AAV2 scenario extension showed, within the stated assumptions, that titre drivers and functional drivers separate, that a stoichiometric imbalance or non-functional particles can offset a dose increase, and that representative assay error is equivalent to roughly a ±30% uncertainty in full-capsid fraction. Fourteen scenario studies further showed, conditional on the assumed surrogate equations and an assumed A:B penalty, that variance-based sensitivity ranks the A:B ratio first for function while correlation ranks it last, that titre-based release is uninformative about function, and that competition and saturation parameters need composition- and dose-diverse calibration lots.

Abbreviations

AAV, adeno-associated virus; AAV2, adeno-associated virus serotype 2; ABE, adenine base editor; AEX, anion-exchange chromatography; BDF, backward differentiation formula; CBE, cytosine base editor; CN, condition number; CV, coefficient of variation; DNA, deoxyribonucleic acid; dPCR, digital polymerase chain reaction; DoE, design of experiments; HEK293, human embryonic kidney 293 cells; IACT, integrated autocorrelation time; iPSC, induced pluripotent stem cell; K_D, half-saturation constant (in central-dose units); kLa, volumetric oxygen-transfer coefficient; LSODA, Livermore Solver for Ordinary Differential Equations with Automatic method switching; MCMC, Markov chain Monte Carlo; NEES, normalized estimation error squared; NIS, normalized innovation squared; ODE, ordinary differential equation; PEI, polyethyleneimine; PI, proportional-integral; qPCR, quantitative polymerase chain reaction; qVG, vector-genome production rate per transfected cell; RPM, revolutions per minute; RMSE, root-mean-square error; SCN1A, sodium voltage-gated channel alpha subunit 1; SSR, sum of squared residuals; SVD, singular-value decomposition; UKF, unscented Kalman filter; VG, vector genome.

Data Availability Statement

The synthetic cohort tables, parameter dictionaries, available provenance records, and computational analysis outputs are supplied in the accompanying private review archive. The archive has not yet been deposited in a public repository and therefore has no public repository URL or DOI. A public repository URL and persistent DOI must be added before submission. Analysis scripts and figure-generation code are included where available; the repository documents the computational workflow, data provenance, fixed-seed configurations, and current reproducibility limitations in docs/REPRODUCIBILITY.md. Full regeneration of every manuscript figure and table should not be assumed unless confirmed by the corresponding version-matched source files and execution records. The source model of Nguyen et al. (Mol Ther Methods Clin Dev 2021;21:642–655) and its data are available at github.com/tamntnguyen/AAV-Triple-Transfection-Mechanistic-Model (commit 4e7fb16c59ef2ca526618feca7fc8b25de79c65d; licence CC BY-ND 4.0) and are fetched, not redistributed, by the repository scripts; SHA-256 values of the input files are recorded. The adapted Python reimplementation is not included in the public repository pending written permission under the upstream licence; a clean-room replacement is not claimed as completed. Study 6 observations are synthetic. The Grieger et al. benchmark values are approximate graphical digitizations rather than raw numerical data. Earlier drafts quoted a refit SSR of 0.731223, an 8D condition number of 4.09×10^{4}, a 14D SSR of 1.7434 and a weak-direction profile baseline of 0.819 (range 0.7208–0.7312); the parameter vectors that produced them were not archived and the values could not be reproduced, so they are superseded by the regenerated fit archived with the source inputs.

Acknowledgments

The author would like to thank Mr Aamish Ahmad for his continued support and guidance throughout this work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The author declares no conflicts of interest.

Ethics Statement

This study was computational in nature and did not involve human participants, animal experiments, or identifiable clinical data. Therefore, ethical approval was not required.

References

1. McClements ME, MacLaren RE. Adeno-associated virus (AAV) dual vector strategies for gene therapy encoding large transgenes. Yale J Biol Med. 2017;90(4):611-623. PMID: 29259525.

2. Levy JM, Yeh WH, Pendse N, Davis JR, Hennessey E, Butcher R, et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. Nat Biomed Eng. 2020;4(1):97-110. doi:10.1038/s41551-019-0501-5. PMID: 31937940.

3. Davis JR, Wang X, Witte IP, et al. Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. Nat Biomed Eng. 2022;6(11):1272-1283. doi:10.1038/s41551-022-00911-4. PMID: 35902773.

4. Grieger JC, Soltys SM, Samulski RJ. 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):287-297. doi:10.1038/mt.2015.187. PMID: 26437810.

5. Nguyen TNT, Sha S, Hong MS, et al. Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. Mol Ther Methods Clin Dev. 2021;21:642-655. doi:10.1016/j.omtm.2021.04.006. PMID: 34095346.

6. Park S, Shin S, Lee H, et al. Enhancing the production of adeno-associated virus (AAV)2 and AAV9 with high full capsid ratio in HEK293 cells through design-of-experiment optimization of triple plasmid ratio. Biotechnol J. 2024;19(3):e2300667. doi:10.1002/biot.202300667. PMID: 38479987.

7. Coplan L, Zhang Z, Ragone N, et al. High-yield recombinant adeno-associated viral vector production by multivariate optimization of bioprocess and transfection conditions. Biotechnol Prog. 2024;40(3):e3445. doi:10.1002/btpr.3445. PMID: 38450973.

8. Deng Z, Lv YL, Wang XT, et al. Production of recombinant adeno-associated virus through high-cell-density transfection of HEK293 cells based on fed-perfusion culture. Hum Gene Ther. 2025;36(3-4):116-127. doi:10.1089/hum.2024.160. PMID: 39761134.

9. Kontogiannis T, Braybrook J, McElroy C, et al. Characterization of AAV vectors: a review of analytical techniques and critical quality attributes. Mol Ther Methods Clin Dev. 2024;32(3):101309. doi:10.1016/j.omtm.2024.101309. PMID: 39234444.

10. Lehman SE, Vreeland WN, Blaszczyk AJ, et al. Interlaboratory measurement of adeno-associated virus: comparative quantification of full and empty capsids. Hum Gene Ther. 2024;36(1-2):36-44. doi:10.1089/hum.2024.124. PMID: 39723438.

11. Ferreira MV, Fernandes S, Almeida AI, et al. Extending AAV packaging cargo through dual co-transduction: efficient protein trans-splicing at low vector doses. Int J Mol Sci. 2023;24(13):10524. doi:10.3390/ijms241310524. PMID: 37445701.

12. Moser A, Appl C, Brüning S, et al. Mechanistic mathematical models as a basis for digital twins. Adv Biochem Eng Biotechnol. 2021;176:133-180. doi:10.1007/10_2020_152. PMID: 33205260.

13. Chang KH, Huang CY, Ou-Yang CH, et al. In vitro genome editing rescues parkinsonism phenotypes in induced pluripotent stem cells-derived dopaminergic neurons carrying LRRK2 p.G2019S mutation. Stem Cell Res Ther. 2021;12(1):508. doi:10.1186/s13287-021-02585-2. PMID: 34551822.

14. Nelson AT, Hill SF, Simon M, et al. In vivo adenine base editing ameliorates Dravet syndrome phenotypes in a mouse model. Sci Transl Med. 2026;18(849):eadx5999.

15. Jo DH, Jang HK, Cho CS, et al. Visual function restoration in a mouse model of Leber congenital amaurosis via therapeutic base editing. Mol Ther Nucleic Acids. 2023;31:16-27. doi:10.1016/j.omtn.2022.11.021. PMID: 36589710.

16. Srinivasan P, Canova CT, Sha S, et al. Multidose transient transfection of human embryonic kidney 293 cells modulates recombinant adeno-associated virus2/5 Rep protein expression and influences the enrichment fraction of filled capsids. Biotechnol Bioeng. 2024;121(12):3694-3714. doi:10.1002/bit.28828. PMID: 39176568.

17. Johari YB, Pohle TH, Whitehead J, et al. Molecular design of controllable recombinant adeno-associated virus (AAV) expression systems for enhanced vector production. Biotechnol J. 2024;19(6):e2300685. doi:10.1002/biot.202300685. PMID: 38900035.

18. Lyle A, Stamatis C, Linke T, et al. Process economics evaluation and optimization of adeno-associated virus downstream processing. Biotechnol Bioeng. 2024;121(8):2435-2448. doi:10.1002/bit.28402. PMID: 37086200.

19. McColl-Carboni A, Dollive S, Laughlin S, Lushi R, MacArthur M, Zhou S, Gagnon J, Smith CA, Burnham B, Horton R, Lata D, Uga B, Natu K, Michel E, Slater C, DaSilva E, Bruccoleri R, Kelly T, McGivney JB 4th. Analytical characterization of full, intermediate, and empty AAV capsids. Gene Ther. 2024 May;31(5-6):285-294. doi: 10.1038/s41434-024-00444-2. Epub 2024 Feb 19. PMID: 38374348; PMCID: PMC11090809.

20. Frenkel R, Tribby D, Boumajny B, Larson N, Sampson M, Barney C, Bergelson S, Sosic Z, Yeung B. ACUVRA: Anion-Exchange Chromatography UV-Ratio Analysis-A QC-Friendly Method for Monitoring Adeno-Associated Virus Empty Capsid Content To Support Process Development and GMP Release Testing. AAPS J. 2022 Nov 22;25(1):3. doi: 10.1208/s12248-022-00768-0. PMID: 36414847.

21. Chahal PS, Schulze E, Tran R, Montes J, Kamen AA. Production of adeno-associated virus (AAV) serotypes by transient transfection of HEK293 cell suspension cultures for gene delivery. J Virol Methods. 2014 Feb;196:163-73. doi: 10.1016/j.jviromet.2013.10.038. Epub 2013 Nov 13. PMID: 24239634; PMCID: PMC7113661.

22. Trapani I, Colella P, Sommella A, Iodice C, Cesi G, de Simone S, Marrocco E, Rossi S, Giunti M, Palfi A, Farrar GJ, Polishchuk R, Auricchio A. Effective delivery of large genes to the retina by dual AAV vectors. EMBO Mol Med. 2014 Feb;6(2):194-211. doi: 10.1002/emmm.201302948. Epub 2013 Dec 15. PMID: 24150896; PMCID: PMC3927955.