Physics-Informed Surrogate Modeling for Lipid Nanoparticle Mixing and Scale-Up

George Weale
Columbia University
New York, NY, USA
george.weale@columbia.edu

***Abstract*—Lipid nanoparticle (LNP) formulation is governed by coupled hydrodynamics, solvent exchange, lipid association, nucleic-acid complexation, and downstream handling. A formulation that appears favorable in a small mixer can change its size distribution, encapsulation, or stability when geometry, throughput, raw materials, or purification conditions change. This paper develops a multiscale physics-informed surrogate framework that links mixer-scale transport descriptors to measured critical quality attributes (CQAs) while preserving uncertainty and out-of-domain warnings. The method combines computational fluid dynamics (CFD) or validated reduced-order transport simulations, a mechanistically constrained feature set, Gaussian-process or ensemble surrogate learning, and split-conformal uncertainty quantification. It treats scale-up as prediction under distribution shift rather than assuming that matching total flow rate is sufficient. A staged experimental design and cross-geometry validation strategy connect model qualification, physicochemical measurement, calibration, and human-reviewed experiment selection.**

***Index Terms*—lipid nanoparticles, microfluidics, surrogate modeling, scale-up, uncertainty quantification, design of experiments**

I. Process Modeling Framework

The paper contributes an auditable model-selection and validation workflow for LNP mixing and scale transfer. It connects transport simulation, formulation metadata, physicochemical measurement, uncertainty calibration, and decision gates so that each prediction can be traced to a declared model version and evidence boundary. The method seeks transferable process understanding rather than a universally optimal formulation.

LNPs are a useful test case for process modeling because quality depends on both composition and history. A small change in flow-rate ratio, total flow rate, aqueous pH, nucleic-acid concentration, lipid ratio, mixing geometry, or hold time can alter the local solvent-exchange trajectory. A purely black-box model can interpolate within its training data but may fail silently during scale-up. The framework encodes transport-relevant structure while retaining data-driven flexibility.

II. Transport-Informed Scale Transfer

The analysis tests whether a transport-informed surrogate predicts selected LNP CQAs across controlled changes in mixer scale and operating conditions more reliably, and with better-calibrated uncertainty, than composition-only or geometry-agnostic baselines. The primary CQAs are hydrodynamic diameter D_h, polydispersity index (PDI), encapsulation efficiency (EE), and a specified stability metric after a controlled hold or storage condition. Biological potency requires a separately approved assay design and is not used as a proxy for physicochemical process validation.

The framework addresses rapid solvent-shift mixing of an ethanolic lipid stream and an aqueous nucleic-acid stream. Mixer-scale features such as residence-time distribution and solvent-composition exposure serve as physically motivated predictors of observed CQAs, complementing rather than replacing molecular-scale characterization.

III. Related work and motivation

LNP delivery systems have made the manufacturing history of a formulation as important as its nominal composition. Reviews of LNP-enabled nucleic-acid delivery describe the central roles of ionizable lipids, composition, and particle structure [1], [2]. Microfluidic mixing has long been used to make nanoparticle formation more controllable by reducing and shaping the solvent-exchange time scale [5], [6]. In the specific LNP setting, microfluidic synthesis and associated internal morphology have been studied as linked to mixing conditions and formulation choices [3], [4]. Those studies motivate a scale-aware process model; they do not imply that a single flow setting or mixer geometry transfers across payloads.

Surrogate modeling is attractive because high-fidelity transport simulation and wet-lab characterization can both be expensive. Bayesian calibration and Gaussian processes offer a disciplined way to combine a computational model with observed discrepancy and predictive uncertainty [7], [8]. Bayesian optimization literature further shows how uncertain models can choose informative experiments rather than merely chase a noisy maximum [10], [11]. Conformal prediction is included here as a distinct uncertainty check because a model's posterior variance can be poorly calibrated when it is used outside its training support [9], [12].

The motivation for the framework is a common but avoidable scale-up mistake: treating total flow rate, or even flow-rate ratio, as a sufficient descriptor of mixing. Geometry, residence-time distribution, viscosity, dispersion, and solvent exposure can change together. The method therefore frames scale-up as a distribution-shift test and evaluates whether a surrogate conditioned on transport-derived features remains accurate and identifies conditions that require new measurements. Efficacy

and clinical performance require separate biological evidence beyond the physicochemical model.

IV. Cross-Geometry Validation Strategy

The multiscale, uncertainty-aware workflow tests feature transfer across mixer geometries rather than assuming it. It evaluates a declared set of physicochemical CQAs within a specified formulation family and preserves separate evidence requirements for molecular structure, biological potency, safety, efficacy, and regulatory suitability. The validation sequence qualifies one mixer geometry with tracer or imaging evidence, uses a randomized design with technical replicates, reserves an unseen geometry for scale-transfer testing, and reports calibration and abstention behavior alongside point-prediction accuracy. A model that satisfies these gates can rank a constrained experiment for qualified human review.

V. Mechanistic backbone

A. Flow and solvent-exchange model

For incompressible liquid flow, the CFD backbone solves

$$\nabla \cdot \boldsymbol{u} = 0, \tag{1}$$

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u}\right) = -\nabla p + \mu\nabla^2\boldsymbol{u}, \tag{2}$$

with mixer-specific boundary conditions and temperature-dependent viscosity where data support it. Let c_E denote local ethanol volume fraction. Its transport is approximated by

$$\frac{\partial c_E}{\partial t} + \boldsymbol{u}\cdot\nabla c_E = \nabla\cdot(D_{eff}\nabla c_E), \tag{3}$$

where D_{eff} is an effective diffusivity or dispersion parameter. If the flow model is reduced-order, its domain of validity must be demonstrated against at least a small set of CFD or tracer-mixing checks; it cannot simply be assumed to reproduce a different geometry.

The dimensionless groups used for cross-geometry comparison are

$$\mathrm{Re} = \frac{\rho U D_c}{\mu}, \qquad \mathrm{Pe} = \frac{U D_c}{D_{eff}}, \qquad \mathrm{Ca} = \frac{\mu U}{\sigma}, \tag{4}$$

where D_c is a documented characteristic length. Matching total flow rate without considering D_c, residence time, and flow partition can change these groups substantially; it is therefore not treated as a scale-up criterion.

B. Assembly proxy and feature construction

The model represents assembly through a set of process-history descriptors along Lagrangian tracer paths $\boldsymbol{X}_n(t)$:

$$\tau_{mix} = \inf\{t : \mathrm{Var}[c_E(\boldsymbol{X}_n(t))] \le \delta_{mix}\}, \tag{5}$$

$$\tau_{res} = \text{residence time to collection plane}, \tag{6}$$

$$\mathcal{E}_{solv} = \int_0^{\tau_{res}} c_E(\boldsymbol{X}_n(t))\,dt, \tag{7}$$

$$\mathcal{S}_{max} = \max_{t\in[0,\tau_{res}]}\left|\frac{dc_E(\boldsymbol{X}_n(t))}{dt}\right|. \tag{8}$$

The distribution, rather than only the mean, of these descriptors is retained. For example, broad residence-time tails can be relevant even when mean mixing time is unchanged. Inputs also include lipid molar ratios, total lipid concentration, nucleic-acid concentration, N/P ratio if appropriate to the chemistry, flow-rate ratio (FRR), total flow rate (TFR), pH, temperature, material lot identifiers, and purification/hold metadata.

VI. Surrogate model

A. Physics-informed mean function

For CQA y_j, input vector $\boldsymbol{x}$, and geometry label g, the surrogate is

$$y_j = m_j(\boldsymbol{\phi}(\boldsymbol{x}, g); \boldsymbol{\beta}_j) + f_j(\boldsymbol{x}, g) + \epsilon_j, \tag{9}$$

where $\boldsymbol{\phi}$ contains the transport descriptors in (8), m_j is an interpretable regularized mean function, f_j is a flexible residual process, and ϵ_j is measurement noise. A Gaussian process (GP) residual provides a posterior mean and variance:

$$f_j \sim \mathcal{GP}\left(0, k_j\left([\boldsymbol{x}, g], [\boldsymbol{x}', g']\right)\right). \tag{10}$$

The kernel may include automatic relevance determination and a geometry component, but geometry is not allowed to become a shortcut that masks transport mismatch. A multi-output model may share information across D_h, PDI, and EE only after checking that the learned covariance does not degrade calibration for any individual CQA.

The phrase "physics-informed" denotes the use of mass conservation, mixer-derived features, dimensionless groups, and feasibility constraints. A composition-only baseline and a geometry-identifier baseline isolate the contribution of those transport terms from flexible statistical fitting.

B. Constraints and uncertainty

Some output constraints are structural: $D_h > 0$, $0 \le \mathrm{PDI} \le 1$ under the chosen reporting convention, and $0 \le \mathrm{EE} \le 1$. Appropriate transformations or constrained likelihoods shall be used, with inverse transformations applied before reporting. Predictive intervals are evaluated by empirical coverage, width, and conditional coverage over TFR, FRR, and geometry.

Because GP variance can be overconfident under model misspecification, the final interval is split-conformal:

$$C_{1-\alpha}(\boldsymbol{x}) = \left[\hat{y}(\boldsymbol{x}) - q_{1-\alpha}, \hat{y}(\boldsymbol{x}) + q_{1-\alpha}\right], \tag{11}$$

where $q_{1-\alpha}$ is computed from a held-out calibration set using a prespecified nonconformity score. Coverage is reported separately for in-domain interpolation, held-out operating conditions, and held-out geometry. Conformal validity can weaken under distribution shift; this limitation must be stated rather than concealed.

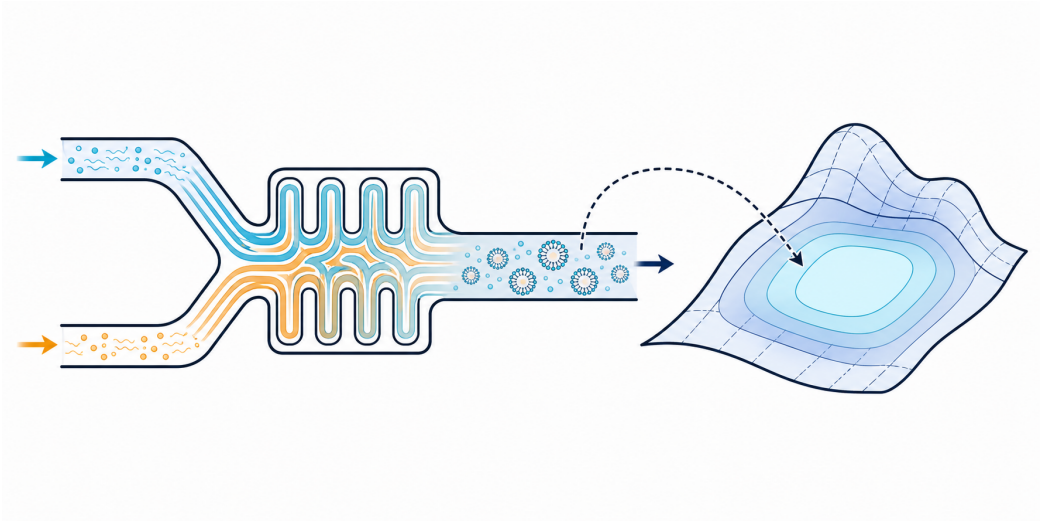

Fig. 1. Micromixer schematic linking solvent-exchange layers, particle formation, and an uncertainty-aware process-window representation.

VII. Experimental Design

A. Phased design

The experimental design is staged to prevent an expensive laboratory campaign from being driven by an unvalidated model.

1) **Phase 0: model qualification.** Define geometry, mesh or reduced-order discretization, property assumptions, tracer seeding, and mass-conservation checks. Compare a limited number of transport predictions with a suitable mixing or tracer measurement if available.
2) **Phase I: local formulation-operating map.** Use a space-filling, constrained design over composition, FRR, TFR, and pH in one geometry. Include randomized run order, technical replicates, and reference material where feasible.
3) **Phase II: controlled geometry shift.** Hold selected chemical conditions fixed while changing channel scale or mixer configuration. The predeclared aim is to test transport-feature transfer, not to maximize CQA values.
4) **Phase III: uncertainty-driven acquisition.** Select new points by expected improvement subject to an uncertainty threshold and explicit feasibility bounds. Reserve final test conditions before adaptive optimization begins.

B. Quality controls and measurement model

Each run should record the full preparation sequence, not merely the final CQA. A hierarchical measurement model can separate technical replicate variance from run-to-run variance:

$$\begin{aligned} y_{r\ell} &= \theta_r + \eta_{r\ell}, \\ \theta_r &\sim \mathcal{N}(\mu(\boldsymbol{x}_r), \sigma^2_{run}), \\ \eta_{r\ell} &\sim \mathcal{N}(0, \sigma^2_{assay}). \end{aligned} \tag{12}$$

The study should report outliers with a predeclared rule and retain failed formulations as censored or failure observations where appropriate. Deleting low-quality data after observing outcomes would bias both optimization and uncertainty estimates.

TABLE I
Experimental variables, evidence, and guardrails.

Factor or response	Evidence to record	Guardrail
TFR, FRR, geometry	Pump settings, flow verification, device ID	Do not infer equivalence from TFR alone.
Lipid and nucleic-acid inputs	Concentration assay, lot ID, preparation time	Separate lot effects from process effects.
D_h, PDI, EE	Instrument method, dilution, replicate values	Record assay uncertainty and failed reads.
Hold and purification	Buffer, temperature, timing, recovery	Treat as inputs, not unrecorded nuisance.

VIII. Scale-up hypothesis and test protocol

The protocol does not assume that a small-scale optimum transfers. It tests the hypothesis that the conditional distribution of CQAs is more stable across geometry when conditioning on ϕ, the transport descriptors, than when conditioning on TFR and FRR alone:

$$p(y \mid \mathrm{TFR}, \mathrm{FRR}, \mathrm{chemistry}, g) \text{ versus } p(y \mid \boldsymbol{\phi}, \mathrm{chemistry}, g). \tag{13}$$

The primary validation is leave-one-geometry-out or leave-one-scale-out testing. Metrics include mean absolute error, root mean squared error, negative log predictive density, 90% interval coverage, and worst-decile error. A surrogate is not called transferable merely because average accuracy is acceptable; it must also flag inputs whose transport features or chemistry lie outside the training support.

An out-of-distribution score can combine standardized distance in feature space with predictive disagreement among independently trained models. When the score exceeds a prespecified threshold, the system returns "experiment recom-

TABLE II
MODEL-USE GATES FOR SEQUENTIAL QUALIFICATION AND SCALE TRANSFER.

Gate	Required evidence before the next use
Transport	Mass balance, mesh or reduced-model qualification, and documented feature calculation.
Measurement	Replicate-aware CQA record, assay metadata, and visible treatment of failed reads.
Prediction	Held-out-condition error and interval calibration, not only training fit.
Transfer	Reserved geometry or scale test with an explicit out-of-domain policy.
Acquisition	Human-reviewed feasible set and a reason to sample beyond a high-scoring predicted point.

mended" rather than an optimization recommendation. This abstention behavior is a core safety feature, not a failure.

IX. Model diagnostics and decision gates

The study separates a good interpolation model from a useful scale-transfer model. A model can obtain low error within a densely sampled mixer setting while still failing on a new geometry, a different lot, or an altered purification sequence. The diagnostic record therefore includes feature-support plots, residuals by geometry and CQA, interval calibration by operating region, and a review of failed or censored formulations. The record is reviewed before any adaptive acquisition decision, not after an apparent optimum has been selected.

The decision rule is conservative. If the transport model has not been qualified, the surrogate is a composition-and-operating-condition model and must be described as such. If prediction intervals are miscalibrated, an acquisition score cannot be interpreted as risk aware. If the final held-out geometry is outside support, the model's correct action is to recommend measurement rather than extrapolate. These gates make the study falsifiable: a null finding that transport features do not improve held-out transfer is more useful than a visually appealing response surface with unknown validity.

X. Optimization as a constrained decision problem

If the model is adequately validated, the next experiment is chosen by

$$\boldsymbol{x}_{next} = \underset{\boldsymbol{x}\in\mathcal{X}_{safe}}{\arg\min}\, \mathbb{E}[\mathcal{L}(y(\boldsymbol{x})) \mid \mathcal{D}] - \kappa\, \mathcal{I}(\boldsymbol{x}; \text{model}), \qquad (14)$$

where $\mathcal{X}_{safe}$ imposes chemical, equipment, and quality constraints; $\mathcal{L}$ encodes a declared CQA target; and $\mathcal{I}$ values information gain. The objective must never be a hidden weighted average. Any trade-off among size, PDI, EE, material use, throughput, and stability should be shown through a Pareto set or stakeholder-approved utility function. Model outputs rank experiments for human review rather than initiating autonomous execution.

XI. Comparator models and ablation protocol

The physics-informed model must be compared with baselines that make the source of any improvement visible. The minimum baseline set is a composition-and-operating-condition model that excludes transport features, the same model with an opaque geometry identifier, and a transport-feature model with the geometry held out. If the first baseline performs as well as the physics-informed model, the correct conclusion is that the additional transport computation was not useful for the evaluated task. If the geometry-identifier model performs well only within observed devices, that result suggests memorization rather than a scale-transfer mechanism.

An ablation should also test the contribution of residence-time descriptors, solvent-exposure descriptors, material-lot metadata, and conformal calibration independently. Each ablation uses identical time or run-level splits and retains all failed formulations. The report should present a performance-by-regime table and a calibration plot before any aggregate score. This is necessary because an average error can hide a model that is accurate at common settings but unreliable near a process boundary.

A held-out sequence evaluates experiment-selection quality before operational use. The comparison measures whether an acquisition selects feasible, informative experiments with an uncertainty statement that remains calibrated after each observation, rather than whether it merely finds a high predicted score. This criterion grounds the workflow in experimental learning rather than an unverified optimization surface.

XII. Material and batch effects

Material provenance is treated as an experimental factor rather than background noise. Lipid lot, nucleic-acid preparation, buffer, storage interval, device batch, and assay operator can each change an observed CQA without indicating a transport effect. The data schema therefore retains stable identifiers and preparation timestamps. When replication permits it, a hierarchical model can separate within-run assay variation, run-to-run variation, and lot-associated variation. If replication does not permit that separation, the limitation should be reported rather than absorbed into a single unexplained noise term.

The scale-transfer test should also protect against confounding between geometry and calendar time. For example, all small-mixer runs should not be performed before all larger-mixer runs if material degradation or a new assay operator could coincide with the geometry change. Blocking, randomized run order within practical constraints, and repeated reference formulations provide a more credible comparison. These details prevent a false attribution of manufacturing drift to an inferred transport mechanism.

XIII. Threats to validity and limitations

The largest threat is unmeasured chemistry. Transport features can explain only the part of the response that is controlled by hydrodynamic and solvent-exchange history. Lipid chemistry, nucleic-acid construct, ionization behavior, assay

variation, and post-mixing handling may dominate in some regimes. CFD parameters can be uncertain, and continuum transport cannot establish molecular structure. Results from a model formulation may not generalize to another payload or indication. Safety, efficacy, manufacturability, and regulatory suitability require distinct evidence and review.

XIV. Reproducibility package

A reproducibility package releases or archives geometry files where permitted; mesh-convergence and mass-balance reports; property assumptions; simulation configurations; input and assay metadata schemas; prespecified data splits; feature-generation code; model seeds and hyperparameters; calibration diagnostics; and a manifest of all completed, failed, and excluded runs. Proprietary chemistry can be represented by stable identifiers, with its absence recorded as a limit on replication.

XV. Data governance and experiment ledger

The protocol requires an experiment ledger that joins a formulation identifier to raw-material lots, preparation timestamps, mixer configuration, pump settings, collection and purification steps, assay files, and all quality-control flags. The ledger should be append-only after a run is completed. Corrections are allowed, but they must create a new version with a reason and reviewer identity. This rule protects against an unintentional mismatch between the data used for modeling and the materials that were actually mixed.

Data splits, exclusions, and adaptive-acquisition choices should also be committed before response values for the reserved test set are examined. An experiment selected because it was operationally infeasible, an assay that failed, or a batch that was delayed is still informative about deployment conditions. The ledger should retain these events with a declared status instead of deleting them from the modeling table. When sensitive formulation details cannot be released, the record should expose stable pseudonymous identifiers, variable ranges, and an explanation of which conclusions depend on the withheld detail.

The final claim register links each statement about predictive accuracy, calibration, transfer, or an acquisition recommendation to a model version and held-out evaluation definition. A response surface becomes evidence only with the corresponding inputs, calibration set, and out-of-domain policy. This governance layer makes the evaluation auditable and distinguishes explanatory process-window figures from measured manufacturing results.

XVI. Conclusion

This paper develops a conservative method for studying LNP mixing and scale-up with a physics-informed surrogate. It links mixer transport histories, CQAs, and uncertainty-aware scale transfer through explicit model qualification, held-out geometry tests, calibrated intervals, and decision gates. Preregistered simulations and carefully documented experiments evaluate whether transport features improve prediction and support more informative experiment selection while preserving human review.

References

[1] X. Hou, T. Zaks, R. Langer, and Y. Dong, "Lipid nanoparticles for mRNA delivery," *Nat. Rev. Mater.*, vol. 6, pp. 1078–1094, 2021.

[2] J. A. Kulkarni, P. R. Cullis, and R. van der Meel, "Lipid nanoparticles enabling gene therapies: from concepts to clinical utility," *Nucleic Acid Ther.*, vol. 28, no. 3, pp. 146–157, 2018.

[3] N. M. Belliveau *et al.*, "Microfluidic synthesis of highly potent limit-size lipid nanoparticles for in vivo delivery of siRNA," *Mol. Ther. Nucleic Acids*, vol. 1, e37, 2012.

[4] A. K. K. Leung *et al.*, "Lipid nanoparticles containing siRNA synthesized by microfluidic mixing exhibit an electron-dense nanostructured core," *J. Phys. Chem. C*, vol. 116, no. 34, pp. 18440–18450, 2012.

[5] R. Karnik *et al.*, "Microfluidic platform for controlled synthesis of polymeric nanoparticles," *Nano Lett.*, vol. 8, no. 9, pp. 2906–2912, 2008.

[6] A. D. Stroock *et al.*, "Chaotic mixer for microchannels," *Science*, vol. 295, no. 5555, pp. 647–651, 2002.

[7] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J. Roy. Stat. Soc. B*, vol. 63, no. 3, pp. 425–464, 2001.

[8] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[9] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*. New York, NY, USA: Springer, 2005.

[10] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[11] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, 2016.

[12] R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani, "Predictive inference with the jackknife+," *Ann. Stat.*, vol. 49, no. 1, pp. 486–507, 2021.